

# Harmonization of internal quality tasks in analytical laboratories case studies : water analysis methods using polarographic and voltammetric techniques

**2008**

**Njabulo Joyfull Gumede**

**(Reg. No: 20204180)**

## **DECLARATION**

This dissertation is submitted in partial compliance with the requirements for the Masters Degree in Technology: Chemistry, in the Faculty of Applied Sciences at the Durban University of Technology

I, Njabulo Gumede, do declare that this dissertation is representative of my own work, both in conception and execution.

.....  
Signature of the Student

.....  
Date of Signature

## **APPROVED FOR FINAL SUBMISSION**

.....  
Supervisor: Prof. K. Bisetty

.....  
HOD: Prof. G.G. Redhi

## **Acknowledgments**

First and foremost I would like to thank Prof K Bisetty for his help and guidance during the early stages of my research and up until the final stages of my work. I would like to also express my gratitude to Prof Salvado Sagrado who has mentored me in some advanced applications of chemometrics and Qualimetrics. I would like to thank him and the entire Valencia group for a warm welcome during my stay in Valencia.

I would also like to thank the Centre for Research Management and Development (CRMD) at DUT and NRF for financial support. I would like to thank the Chemistry Department at DUT for offering me the facilities to do my project and more especially the Laboratory Technicians for their support i.e. Mrs M. Xhakaza, Mrs P. Masuku, Mrs I Naicker, and Miss T Singh.

Not forgetting to mention Jimmy Chetty for organizing waste water samples for me. Dr Brent Newman from CSIR for providing marine water samples obtained from Durban harbour. I would like also to thank Myalowenkosi Sabela (In-service trainee student) for working with me in the nitrate project. Last but not the least I would like to thank my family, especially my mother for being supportive as a single mother and for the encouragement through these trying times, which is gratefully acknowledged for being patient while I was realizing my true potentials.

## Abstract

In this work, a holistic approach to validate analytical methods was assessed by virtue of Monte Carlo simulations. This approach involves a statement of the methods's scope (i.e. analytes, matrices and concentration levels) and requisites (internal or external); selection of the method's (fit-for-purpose) features; pre-validation and validation of the intermediate accuracy and its assessment by means of Monte Carlo simulations. Validation of the other method's features and a validity statement in terms of a "fit-for-purpose" decision making, harmonized validation-control-uncertainty statistics and short-term routine work with the aim of proposing virtually "ready-to-use" methods. The protocol could be transferred to other methods. The main aim is to harmonize the work to be done by research teams and routine laboratories assuming that different aims, strategies and practical viewpoints exist. As a result, the recommended protocol should be seen as a starting point. It is necessary to propose definitive (harmonized) protocols that must be established by international normalisation/accreditation entities. The Quality Assurance (Method verification and Internal Quality Control, IQC) limits, as well as sample uncertainty were estimated consistently with the validated accuracy statistics i.e.  $E \pm U$  ( $E$ ) and  $RSDi + U$  ( $RSDi$ ). Two case studies were used to assess Monte Carlo simulation as a tool for method validation in analytical laboratories, the first involves an indirect polarographic method for determining nitrate in waste water and the second involves a direct determination of heavy metals in sea water by differential pulse anodic stripping voltammetry, as an example of the application of the protocol. In this sense the uncertainty obtained could be used for decision making purposes as it is very tempting to use uncertainty as a commercial argument and in this work it has been shown that the smaller the uncertainty, the better the measurement of the instrument or the laboratory's reputation.

## **Aims**

The main thrust of this work is to explore the global method development and validation for the determination and speciation of nitrate, cadmium and lead, and also to estimate the accuracy profile from the uncertainty measurement obtained from polarography and voltammetry of the analytical assay from the validation data by means of Monte Carlo (MC) simulation. The main purpose of MC is in the use of simulation to perform approximately 100 000 experiments from results obtained using voltammetric and polarographic methods with given features and requisites. To observe and predict the method's behavior by virtue of a theoretical simulation of pseudo-experiments before method validation using a method's internal and external requisites, in order to predict how the method will behave before validation begins.

## **Objectives**

- To develop a step-by-step approach/protocol for harmonizing quality tasks in testing laboratories, for method development and also validating methods that are fit for their given purpose.
- To use MC simulated results to estimate uncertainty intervals for some key method validation parameters such as *E* (Relative error or bias under trueness assessment) and *RSD<sub>i</sub>* (Relative standard deviation under intermediate precision conditions, inter-laboratory studies) and simulate the uncertainty of the analytical results aimed at improving the quality of results (chemometrics and qualimetrics) and also, to be implemented for decision making purposes for routine laboratories and research laboratories willing to publish new methods.
- To use the experimental design for validation and progressive routine work to illustrate the calculation of bias, and intermediate precision.
- To use the experimental design to explore a wider range of experimental variables with a reduced number of experiments, and to use the validated method to monitor the levels of Cd, Pb and nitrates in sea water and waste water respectively.

### **Publications**

**Published manuscript:** (i) K. Bisetty, N. J. Gumede, L. Escuder-Gilabert and S. Sagrado, Determination of lead and cadmium in marine water by differential pulse anodic stripping voltammetry. fit-for-purpose partial validation and internal Quality aspects. Anal Bioanal Chem (2008) 392:277-286.

**Manuscript under review:** (ii) K. Bisetty, N.J. Gumede and S. Sagrado, Some strategies for harmonizing quality tasks related to analytical methods. A case study: Indirect differential-pulse polarographic determination of nitrate in water. Talanta, submitted in June 2008 (under review).

## TABLE OF CONTENTS

<b>Declaration .....</b>	<b>i</b>
<b>Acknowledgments .....</b>	<b>ii</b>
<b>Abstract .....</b>	<b>iii</b>
<b>Aims .....</b>	<b>iv</b>
<b>Objectives .....</b>	<b>iv</b>
<b>Publications.....</b>	<b>v</b>
<b>TABLE OF CONTENTS.....</b>	<b>1</b>
<b>CHAPTER 1 .....</b>	<b>7</b>
<b>INTRODUCTION.....</b>	<b>7</b>
<b>1.1 Quality of analytical results .....</b>	<b>7</b>
<b>1.2 The selection of the method.....</b>	<b>9</b>
<b>1.3 Deciding on the extent of validation .....</b>	<b>10</b>
<b>1.4. Objectives of an analytical procedure.....</b>	<b>11</b>
<b>1.5 The development of the method.....</b>	<b>11</b>
<b>1.6 What is method validation? .....</b>	<b>13</b>
<b>1.7 Objectives of the validation.....</b>	<b>14</b>
<b>1.8 The role of method validation.....</b>	<b>15</b>
<b>1.9 Stages of validation .....</b>	<b>16</b>
<b>1.10 Toxicity of heavy metals under investigation [9]. .....</b>	<b>18</b>
<b>1.11 The toxicity of nitrates in the environment .....</b>	<b>19</b>
<b>CHAPTER 2 .....</b>	<b>22</b>
<b>THEORETICAL PRINCIPLES RELATED TO THE EXPERIMENTAL DESIGN.....</b>	<b>22</b>
<b>2.1 Applicability, fitness for purpose and acceptability limits.....</b>	<b>22</b>

2.2. Intermediate precision and trueness studies .....	23
2.2.1 Precision .....	28
2.3 Harmonization of a method .....	29
2.3.1. Internal Quality protocol.....	29
2.3.2 Protocol for harmonizing the quality tasks (quantitative methods).....	30
2.4 Sensitivity, detection limit and quantification limit.....	40
2.4.1 Sensitivity .....	40
2.4.2 Limit of detection .....	40
2.4.3 Limit of quantification .....	41
2.5 Matrix effects on sample response.....	42
2.6 The study of accuracy .....	45
2.7 Measurement uncertainty .....	47
<b>CHAPTER 3.....</b>	<b>55</b>
<b>THEORETICAL PRINCIPLES OF VOLTAMMETRY.....</b>	<b>55</b>
3.1 Differential pulse voltammetry/polarography .....	55
3.1.1 Voltammetry.....	57
3.2 Anodic Stripping Voltammetry .....	59
3.3 Standard Addition .....	60
3.4 Rotating disc electrode (RDE) .....	61
3.4.1 Glassy-Carbon Electrodes .....	62
3.4.2 Metal Electrodes .....	63
3.4.3 Mercury Film Electrode .....	64
3.4.4 Chemically modified Electrodes.....	64
<b>CHAPTER 4.....</b>	<b>66</b>
<b>GENERAL EXPERIMENTAL WORK AND DESIGN.....</b>	<b>66</b>
4.1 Laboratory Experimental Work (Instrumental) .....	66
4.1.1 Apparatus (Case study 1) .....	66



4.2 Experimental Design .....	67
4.2.1 Pre-validation of the method and standardization of the signal.....	67
4.2.2 Experimental design: Monte Carlo Simulation.....	73
CHAPTER 5.....	75
CASE STUDY 1 .....	75
5.1 Purpose statement.....	75
5.2 Mixture Reference solution.....	75
5.2.1 Samples.....	76
5.2.2 Sample analysis Procedure .....	76
5.2.3 Accuracy validation.....	76
5.2.4 Verification .....	76
5.2.5 Internal Quality Control (IQC) .....	77
5.2.6 Limit of detection (LOD) estimation .....	77
5.3 Results and discussion .....	77
5.3.1 External and internal method requisites and method's features.....	78
5.3.2 Accuracy pre-validation .....	78
5.3.3 Accuracy validation using a CRM solution .....	79
5.3.4 Accuracy validation assessment by means of Monte Carlo simulation..	81
5.3.5 QA and uncertainty statistics related to accuracy validation estimates.	82
5.3.6 Initial routine work .....	83
5.3.7 Application of the protocol to a case study .....	83
5.4 Conclusion.....	95
CHAPTER 6.....	97
CASE STUDY 2 .....	97
6.1 Purpose Statement:.....	97
6.2 Preparation of the electrodes .....	97

6.2.1 Cell pre-treatment procedure .....	98
6.2.2 Sampling.....	98
6.2.3 Sample analysis procedure .....	98
6.2.4 Mixture reference solution.....	99
6.2.5 Accuracy validation and IQC studies.....	99
6.3 Conclusion.....	111
CHAPTER 7.....	113
SUMMARY OF THE TWO CASE STUDIES.....	113
7.1 Summary of the two case studies demonstrating Monte Carlo simulation's applicability in Quality control of Analytical Laboratory analysis .....	113
7.1.1 Summary of Case study 1 .....	113
7.1.2 Summary for Case study 2 .....	115
7.2 CONCLUSIONS.....	117
7.3 Recommendations .....	119
7.2.1. Miss-directions related to method validation .....	119
7.2.2 Areas to be explored.....	119
REFERENCES .....	120
APPENDICES.....	126
Appendix 1: Definitions.....	126
Appendix 2 Symbols for the variables used in the text.....	131
Appendix 3 – Results for pre-validation (case study 1) .....	134
Appendix 4 – Results for a validation protocol of 50ppm nitrate EU parametric value .....	147
Appendix 5 Sample analysis verification and quality control results .....	173
Appendix 6: Results for a pre-validation study of trace metal determination in marine water.....	181
Appendix 7: Results for 10ppb validation study for trace metal determination.....	189

<b>Appendix 8: Marine water and QC synthetic sample results for case study 2..</b>	<b>214</b>
--	------------

## **Table of figures**

<b>Figure 1: Life cycle of an analytical procedure.....</b>	<b>9</b>
<b>Figure 2: illustrates the life-cycle of the method .....</b>	<b>12</b>
<b>Figure 3: Types of biases that are usually obtained.....</b>	<b>27</b>
<b>Figure 4: Intermediate accuracy plot.....</b>	<b>34</b>
<b>Figure 5: Short-term routine work .....</b>	<b>38</b>
<b>Figure 6: The Horwitz function .....</b>	<b>46</b>
<b>Figure 7: Illustration of accuracy, precision and measurement uncertainty .....</b>	<b>48</b>
<b>Figure 8: The relationship between the cause-and-effect diagram .....</b>	<b>49</b>
<b>Figure 9: A typical DPV measurement process .....</b>	<b>56</b>
<b>Figure 10: Typical current-potential curve for a DP measurement.....</b>	<b>57</b>
<b>Figure 11: Typical standard addition curve.....</b>	<b>60</b>
<b>Figure 12: Schematic diagram of a glassy carbon electrode.....</b>	<b>63</b>
<b>Figure 13: The 3-point validation graph.....</b>	<b>71</b>
<b>Figure 14: The regression graph obtained for 3-point calibration curve .....</b>	<b>69</b>
<b>Figure 15: Estimated concentration obtained in 12 runs using a 1-point approach .....</b>	<b>71</b>
<b>Figure 16: Regression graph for a 1-point calibration plot.....</b>	<b>72</b>
<b>Figure 17: Summary of polarograms obtained for Case Study 1.....</b>	<b>87</b>
<b>Figure 18: Accuracy assessment plot .....</b>	<b>90</b>
<b>Figure 19: Illustrate the routine work stage.....</b>	<b>94</b>
<b>Figure 20: Harmonized protocol for internal quality aspects .....</b>	<b>100</b>
<b>Figure 21 Summary of voltammograms obtained for the two analytes for Case Study 2.....</b>	<b>104</b>

Figure 22: Accuracy assessment plot for the validated data.....	105
Figure 23: Mean control chart for the progressive routine work .....	110
<b>List of Tables</b>	
Table 1: %RSD from the Horwitz function .....	25
Table 2: Limit of detection and limit of quantitation .....	41
Table 3: Acceptable recovery percentages depending on the analyte concentration level .....	43
Table 4: Calculation of recoveries using the fortified as well as unfortified samples .....	44
Table 5: Requisite-feature relationships.....	84
Table 6: Results for tap water related to the $X_{3 \times 12}$ experimental designs for a pre-validation study .....	86
Table 7: $X_{6 \times 12}$ data matrix for nitrate .....	88
Table 8: Harmonized statistics from the $X_{6 \times 12}$ data matrix (Table 11), after eliminating those values with $ z  > 2$ .....	89
Table 9: Initial routine work step.....	92
Table 10: $X_{6 \times 12}$ data matrix (experimental Lead concentration in $\mu\text{g.L}^{-1}$ ) corresponding to the analysis of a mixture CRM solution containing $10 \mu\text{g.L}^{-1}$ Pb.....	103
Table 11: $X_{6 \times 12}$ data matrix (experimental Cadmium concentration in $\mu\text{g.L}^{-1}$ ) corresponding to the analysis of a mixture CRM solution containing $10 \mu\text{g.L}^{-1}$ Cd. ....	103
Table 12: Harmonized statistics from the $X_{6 \times 12}$ data matrix for a validation study, after eliminating those values with $ z  > 2$ (see appendix 2) .....	107
Table 13: Results for 14 harbour water samples and independent mixture CRM solutions ( $10 \mu\text{g.L}^{-1}$ Pb and Cd level) used as IQC .....	108

# CHAPTER 1

## INTRODUCTION

### 1.1 Quality of analytical results

The implementation of quality control systems in analytical laboratories has now, in general, been achieved globally. While this requirement has significantly modified the way that the laboratories are run, it has also improved the quality of the results. The key idea is to use analytical procedures which produce results that fulfill the user's needs and also help when making decisions [1, 2]. The role of the analytical chemist has not changed since the time analysts discovered that naturally occurring products were composite materials. For example, when it was discovered that carrots helped prevent 'night blindness' it was an analytical chemist who separated out the various components of carrot, characterized the compounds and identified the active component as  $\beta$ -carotene [3]. It was reported [1, 3] that innumerable types of analytical methods exist in the fields of analytical and bio analytical chemistry (e.g. Plasma protein interactions of drugs and their metabolites), biochemistry (e.g. determination of Quinine in tonic water), biology (e.g. determination of iron in blood plasma), clinical biology (e.g. development of new drugs for curing of cancer and tuberculosis), and related application domains, such as forensic, toxicological, environmental, agricultural and food analyses. Laboratories must be able to produce reliable data regardless of the type of method the scope and application [1] when performing analytical tests for a client or regulatory purposes. The social and economic impact of an analyst reporting an inaccurate result and as a result the customer reaching a false conclusion can have huge consequences [3], such as;

- In forensic analysis, it could lead to a wrongful conviction or the guilty party going freely.
- In trade, it could lead to the supply of sub-standard goods and the high cost of replacement following the loss of customers.
- In environmental monitoring, mistakes could lead to hazards being undetected or to the identification of unreal hazards.

- In the supply of drinking water, it could lead to harmful contaminants being undetected and also selling tap water as mineral water.
- In healthcare, the incorrect medication or the incorrect content of active ingredient in a tablet can be catastrophic for the patient [3].

Quality of analytical measurement data encompasses two essential criteria-utility and reliability. Utility means that analytical results must allow reliable decision making. A key aspect of reliability or validity of results is that they are comparable, whatever their origin. Comparability between results in the strict sense is provided by traceability to appropriate standards. Traceability to common reference standards underlies the possibility of making a comparison i.e. a distinction-between different results. If the results are also to be compared in terms of their quantities or levels of analyte, additional information on the analytical results is needed [1]. It underlines the importance of method validation; at present a recent approach based on the accuracy profile to illustrate how validation must be fully integrated into the basic design of the method has been published [4]. Quality of results reflects adequacy or (inadequacy) of a method in terms of the extent to which the method fulfils its requirements or fit for its intended analytical purpose. Quality is always a related notion, referring to the requirements fixed before hand on the basis of national or international regulations or customer needs [1]. The quality of measurements produced by analytical laboratories during the last decade has significantly improved, and laboratories have been requested to control their procedures and organisations more effectively [1]. The quality of an analytical data can be defined at two main levels:

- ***Metrological requirements.*** The end-users of chemical data expect that measurements are close to the true value for the sample submitted to an analysis. This is known as “accuracy”. The accuracy depends on the precision and the bias of the method and hence, there are several factors which significantly influence the uncertainty of a measurement. An important goal is to evaluate this uncertainty and express it in such a way that end-users are able to adequately use results produced by the laboratory.

- **Socio-economic requirements.** While metrological requirements are implicit and cannot be directly controlled by the end-user, the costs involved in a measurement can be controlled to some degree. There is always a strong pressure on laboratories globally to reduce costs and propose cost-effective services.

## 1.2 The selection of the method

The purpose of an analytical method is to deliver a qualitative and or quantitative result with an acceptable uncertainty level, with the term ‘validation’ gives rise to ‘measuring uncertainty’ [5]. When faced with a particular analytical problem, ideally the laboratory should first agree with the customer that an analytical requirement, which defines the performance requirements that a method must be suitable for solving the analytical problem. Additionally, the laboratory can evaluate existing methods for suitability and if necessary develop a new method. This iterative approach (see figure 1) of developing and evaluation continues until the method is deemed capable of meeting the requirements, further development is unnecessary and the analytical work can proceed [1].

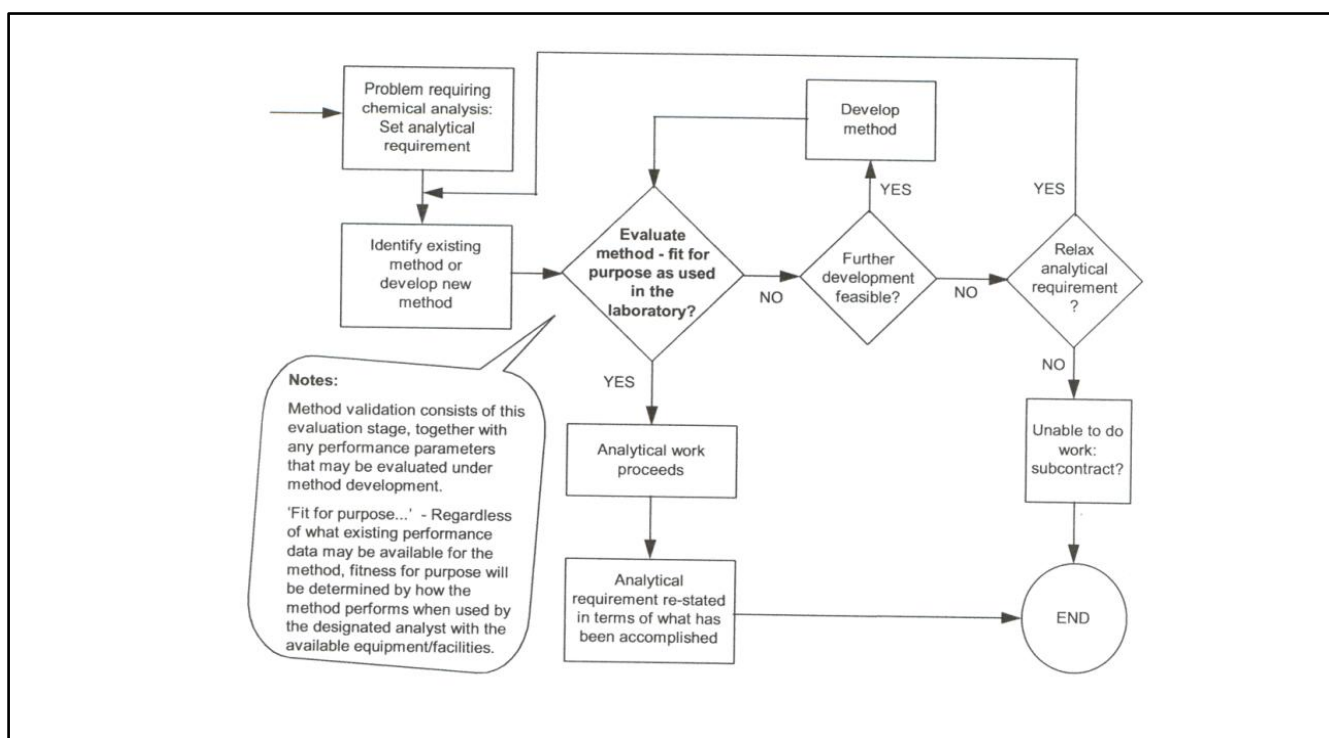


Figure 1: Life cycle of an analytical procedure

According to ISO 17025 on method selection, it is stated that: The laboratory shall use test or calibration methods, including methods for sampling, which meets the needs of the client and which are appropriate for the tests and / or calibration it undertakes preferably those published as international, regional or national standards. This statement agrees with the reasoning of ISO 9000:2000 standard. Once the method is selected, it is often necessary to perform several experiments in order to either adapt to the laboratory conditions or fully develop the method [4].

### **1.3 Deciding on the extent of validation**

The extent of validation and the choice of performance parameters to be evaluated depend on the status and experience of the analytical method. The validation plan is determined by the analytical requirements, as defined on the basis of customer needs or as laid down in regulations. When the method has been previously fully validated according to an international protocol, the laboratory does not need to conduct extensive in-house validation studies. It must verify only that it can achieve the same performance characteristics as outlined in the collaborative study [5]. The laboratory has to decide which method performance parameters need to be characterized in order to validate the method. Characterization of method performance is an extensive process and may inevitably be constrained by time and cost considerations. The implications of the constraints discussed above are particularly critical where the method is not going to be used on a routine basis. Validations of methods which are going to be used on a routine basis are relatively straight forward processes [2]. Validation requirements may be specified in guidelines within a particular sector of measurement relevant to the method and it is recommended that where these are available they are followed. Different International Conference on Harmonization of Technical Requirements for registration of Pharmaceuticals for human use United States Pharmacopeia (ICH/USP) guidelines [4] is set up for:

- Identification tests;
- Impurity tests;
- Assay tests.



An identification test ensures the identity of an analyte in a sample, by comparing it to a known reference material (RM). An impurity test is intended to confirm the identity of (limit impurity test) or accurately quantify (quantitative impurity test) an impurity, defined as an entity ‘which may normally not be present’. An assay test finally applies to the major component or active ingredient in a sample and quantifies the drug substance as such, as a whole, or the drug substance in a drug product [2, 5, and 9].

#### 1.4. Objectives of an analytical procedure

The objective of an analytical procedure is to demonstrate that the response varies linearly as a function of the concentration, that the bias and precision are less than (**X**%) or rather to quantify as accurately as possible each unknown quantity. These interrogations seem to be the question of interest for analytical chemists. The objective of a “good” analytical procedure is to be able to quantify as accurately as possible each unknown quantity that the laboratory has to determine [6]. What the analyst actually needs is the difference between the “measured value” ( $x$ ) and the true “value” ( $\mu_T$ ), which will always remain unknown, is as low as possible as or at least lower than an acceptable limit. This can be expressed as:

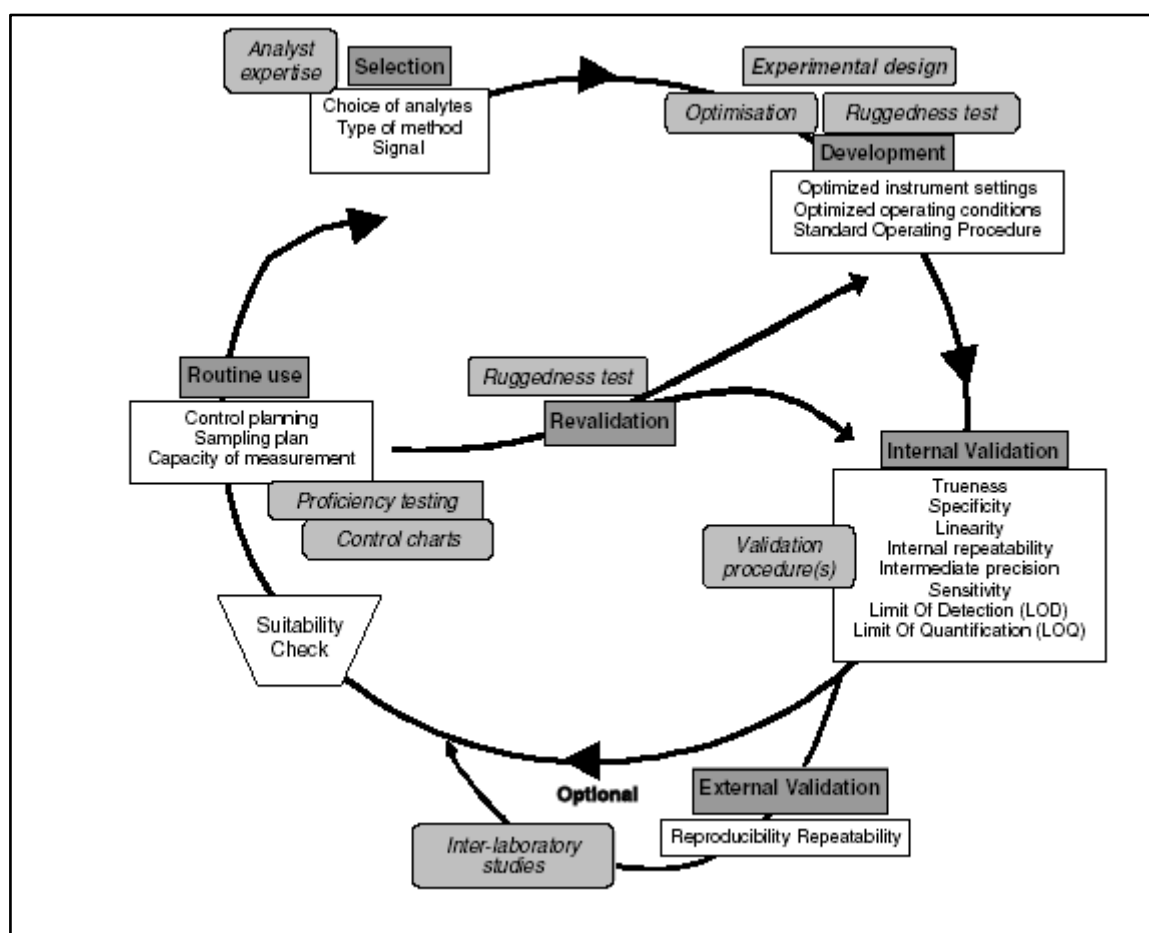
$$-\lambda < x - \mu_T < \lambda \Leftrightarrow |x - \mu_T| < \lambda \quad (1)$$

Where,  $\lambda$  is the acceptable limit which can be different depending on the requirements of the analyst or the objective of the analytical procedure. Obviously, the acceptance limit can vary according to the intended use of the analytical method for example 1-2% for the determination of a bulk pharmaceutical compound, 5% for the determination of the active ingredients in dosage forms, 15% in bio analysis, etc.). It can be noted that, not only acceptance limits for the performance of an analytical method but also the responsibility that the analyst has to take in the decision of accepting the performance of the method with respect to its intended use [6].

#### 1.5 The development of the method

Method development can be expressed in various ways (see figure 2). At one extreme, it may involve adapting an existing method or making minor changes so that it is suitable for a new application. For example, in one of our case studies a method required to determine nitrate in water might be adapted from an established method for

nitrite in water. The matrix is the same, and the two analytes have broadly similar properties. It is likely that the same principles of isolation (redox), identification, and quantitation that are applied to nitrites be applied to nitrates. On the other hand, the analytical chemist may start out by making few sketchy ideas and apply expertise and experience to devise a suitable method. It might also involve significant innovation based on novel exploitation of known properties of the analyte or measurand. This clearly involves a great deal of more work, and initially at least a degree of doubt as to whether the final method will be successful [2].



**Figure 2:** illustrates the life-cycle of the method which involves the analyst's expertise, selection of the method, experimental design, method development, robustness test, internal and external validation, suitability test, and routine work strategy which are the essential steps that have to be followed in developing a method [4].

As method development proceeds, regular reviews should be carried out to verify that the needs of the client are still being fulfilled and that the method is still fit for its purpose. Changing requirements to the development plan should be approved and authorised. When the development phase is finished, the draft of the standard operating procedure (SOP) can be written by a laboratory manager according to ISO 17025 accreditation system and Laboratory information management system (LIMS) [4].

### **1.6 What is method validation?**

Validating a method is investigating whether the analytical purpose of the method is achieved, which is to obtain analytical results with an acceptable uncertainty level. Analytical quality assurance is the complete set of measures a laboratory must undertake to ensure that it can always achieve high quality data. The ISO definition of validation is ‘confirmation by examination and provision of objective evidence that the particular requirements of a specified intended use are fulfilled’ [2, 5]. This is consistent with the Eurachem guide [2] definition where by validation is interpreted as being the process of defining an analytical requirement, and confirming that the method under consideration has performance capabilities consistent with what the application requires [2].

It is inadequate to try to merge development and validation in the same step. It is now acceptable to make a distinction between intra-laboratory (in-house) and inter-laboratory (or collaborative) validation. The first is universal and compulsory; the second is mainly applicable to methods that will be used by several laboratories or where the results can be used for economic decisions. For example, in the pharmaceutical industry, it is useless or impossible to perform a collaborative study for a new molecule under development. On the other hand, all methods used for environmental safety control must be inter-laboratory validated. Verification may occur at the end of the validation procedure, as proposed in the definition [4]. Method validation is needed to ‘confirm the fitness for purpose of a particular analytical method’, i.e. to demonstrate that ‘a defined method protocol, applicable to a specified type of test material and to a defined concentration rate of the analyte’-the whole is called the “analytical system” ‘is fit for a particular analytical purpose’. This analytical purpose reflects the achievement of analytical results with an acceptable standard of

accuracy. An analytical result must always be accompanied by an uncertainty statement, which determines the interpretation of the results. In simple terms, the interpretation and the use of any measurement fully depend on the uncertainty (stated at a level of uncertainty). Validation is thus a tool used to demonstrate that a specific analytical method actually measures what it is intended to measure, and thus is suitable for its intended purpose [5].

### **1.7 Objectives of the validation**

For a method to be validated it has to comply with the following requirements: Analytical measurements should satisfy an agreed requirement (a defined objective). Analytical measurements should be made using methods and equipment that have been tested to ensure that they are fit for purpose. Staff making analytical measurements should be qualified for and competent at undertaking the task (and should demonstrate that they can perform the analysis properly). There should be a regular independent assessment of the technical performance of a laboratory. Analytical measurements made in one location should be consistent with those made elsewhere. Organizations making analytical measurements should have well defined quality control and quality assurance procedures. Validation is the process of establishing the performance characteristics and limitations of a method, as well as identifying the influences that may change them (and to what extent). Validation is also the process of verifying that a method is suitable for its intended purpose (to solve a particular analytical problem)

[2, 4]. Knowing the characteristics of “true bias” and of “true precision” are parameters that will always remain unknown but that will be estimated by the measurements obtained in the validation phase, what is the objective of validation. In actual fact, it may seem reasonable to claim that the objective of validation is to give to the laboratories as well as to regulatory bodies “guarantees” that every single measure that will be later performed in routine analysis will be “close enough” to the unknown “true value” of the sample to be analyzed or at least that the difference will be lower than an acceptable limit taking into account the intended use of a method.

The goals of the validation are thus to minimize the consumer risk as well as the producer risk, as a result; the objectives of the validation cannot be simple limited to obtaining estimates of bias and variance but must be focused on the evaluation of the

risk even if these estimates are needed to evaluate the risk. Using the above objective, two basic notions have to be explained: “Close enough”, meaning; for example, that the realized measure in routine will be less than  $x\%$  ( $x$  retrieves itself to the acceptable limit  $\lambda$ ) of his “true unknown value”. ‘Guarantees’ meaning that it is very likely that whatever the measure is, the results will be “close enough” from the “unknown true value”. The objectives of validation are not simply to obtain estimates of trueness or bias and precision but also to evaluate those risks that can be expressed by the measurement uncertainty associated with the result. The final goal of the validation of an analytical method is to ensure that every future measurement in routine analysis will be close enough to the unknown true value for the content of the analyte in the sample [7, 8].

### **1.8 The role of method validation**

Validation is required for any method. As the definition says, validation always concerns a particular ‘analytical system’. This implies that, for a particular type of material and a particular operating range of concentrations, the method must be able to solve a particular analytical problem. As a result, ‘revalidation’ is needed whenever any component of the analytical system is changed or if there are any indications that the established method does not perform adequately any more. Method validation is closely related to method development. When a new method is being developed, some parameters are already being evaluated during the ‘development stage’ while in fact this forms part of the validation stage. An analysis is a complex multistage investigation of the values of the properties of materials, i.e. the identity and the concentration of a specific component in a specific sample material. The chemical analysis is a cyclic process in which the final objective is the generation of chemical information. Before any method validation is started, the scope of validation must be fixed, comprising of both the “analytical system” and analytical requirement”. A description of the analytical system includes the purpose and the type of method, the types of material or matrices for which the method is applied and a method protocol. This integrated process starts with defining the basic analytical problem (specifying the analytical requirement) and ends with evaluating and reporting the analytical results. In fact, the last step provides an answer to the initial problem, as stated by a client or based on regulatory requirements. Without a statement of uncertainty, a result must be

expressed with its expanded uncertainty, which, in general, represents a 95% confidence interval around the result. The probability that the mean measurement value is included in the expanded uncertainty is 95%, provided that it is an unbiased value that is made traceable to an internationally recognized reference or standard. In this way, the establishment of traceability and the calculation of method uncertainty are linked to each other. Before measurement uncertainty is estimated, it must be demonstrated that the result is traceable to a reference or standard which is assumed to represent the truth [1, 5].

### **1.9 Stages of validation**

The validation process involves four stages [5, 7] ‘concept’, ‘performance’, ‘operational’ and possible cross ‘validation’. Validation is a permanent process that starts from the very beginning of the life of the method until its retirement. In the concept or development phase, the analyst must identify and evaluate the impact of potential sources of variability that could later alter the global quality of the results. The objective today in development is no more to find a method that ‘works’, nor to elaborate smartly on an analytical method whose quality will have to be evaluated in a later stage; the objective becomes to build results of quality by means of an analytical method. As an application example the Societe Françoise des Sciences et Techniques Pharmaceutiques (SFSTP) guide indirectly addresses the issue of the development since, as clearly stated, preliminary knowledge on the performance of a method must be available before properly starting the characterisation stage. This formal validation stage must be seen as a set of experiments that will confirm to the regulatory agencies and the analyst himself that the method can indeed be used for its intended purpose. The validation phase can absolutely not be envisaged as a mean to estimate the performance of a method. If nothing or very little is known about the bias, the precision, the range or the limits or quantification before starting the validation itself, it is almost impossible or too expensive to specify the experiments to perform- i.e. selecting the levels of concentration, the number of runs, the number of replicates per run, the extreme concentration levels-while being able to give reasonable chance of success to end with a valid analytical method. Such an approach is counter-productive and unacceptable in an industrial perspective. If a developer proceeds to the validation stage with a sample size that is too small-with respect to the unknown performance, he

takes the risk of increasing significantly the costs for his laboratory either by accepting as valid a truly non valid method (high non-productive cost in routine) or rejecting a truly valid method (non-productive cost in development and potential delays in a project). The SFSTP guide warmly recommends starting with a ‘pre-validation’ phase whose objectives are:

- To identify a model to use for the calibration curve;
- Evaluate mainly the limits of quantitation;
- Provide good estimates of precision of a method for optimally sizing the ‘validation’ phase.

Discussions could arise around the word ‘pre-validation’ that could be understood by some analysts as steps to perform but not to report. This ‘pre-validation’ must however be viewed as a real validation first phase and documented accordingly. During the ‘pre-validation’, the model to be used as calibration curve will be identified and the quality of fit will be assessed only at this stage. The experiments proposed are designed to consistently evaluate the validity of a model. In the second phase, called ‘validation’, the objective is to mimic the routine practice that was envisaged. The model will be used as is and the parameters will of course be estimated based on the new data and no more investigation will be based on the quality of fit will be conducted, the same way it should be carried out during routine. In this second step, the experiments are designed to focus on the estimation of the bias and precision of the method, not on the calibration curve. If the model identified in the pre-validation is not adequate, then the bias and precision in the validation are impaired. The same reasoning applies to the limit of quantitation (LOQ); the range cannot be shortened after the validation results without impairing the global quality, i.e. the model identified in the previous phase has been demonstrated as adequate over the whole range and could induce bias if the range of application is charged. As was stated earlier, the knowledge of the bias, precision and limits of quantitation is mandatory for initiating a formal validation phase, but the proposed ‘pre-validation’ phase could be skipped if consistent estimates of the performance criteria are available to the analysts. That happens when methods are developed following a well structured strategy, such

as applying experimental design approaches. Unfortunately, the ‘trial and error’ approach is still widely used for developing new methods and so very little is known at the end of such a development process. The validation becomes unfortunately the very first opportunity to estimate the performance of the method. The validation of a method continues even after having successfully met all the requirements and documented the results [7].

### **1.10 Toxicity of heavy metals under investigation [9].**

Heavy metals are recognized as highly toxic and dangerous pollutants. They are placed second to pesticides in environmental importance. They may be the most harmful pollutants because, unlike many other pollutants, they are not biodegradable and are retained in the ecosystem indefinitely. One of the most important problems in oceanography and water resources sciences is the effect of the concentration and concentration changes of trace metals on the water system. Toxic metals enter the marine food chain, ultimately reaching man. Therefore, accurate determination of trace metal levels in various water systems is becoming increasingly important. As a result of their non-biodegradable nature, heavy metals accumulate in organs of man and exert progressively growing toxic actions. The most hazardous and important heavy metals are lead, mercury, cadmium, arsenic, thallium and selenium. These metals have history as occupational hazards, and some of them have been linked to cancer and heart diseases. Accordingly, they belong to the class of first-order priority in ecotoxicology [10]. Some metals that are essential nutrients e.g. copper, zinc and tin, also can exert toxic action, depending on their concentration levels. For these metals a narrow ‘concentration window’ exists between the toxic and the essential levels. As a result of this narrow concentration difference, reliable knowledge of the metal content in various matrices is mandatory [10]. US EPA has set the limits for heavy metals in sea water usually called Criteria Continuous Concentration (CCC) [11]. For Lead the CCC should be 8.1 ppb and 8.8 ppb for Cadmium. In this work a step-by-step protocol to evaluate and harmonize internal quality aspects of a method based on different joint strategies that has been employed successfully for water analysis was proposed [12, 13]. As an application example, data from an adapted method for determining Lead and Cadmium in marine water samples by anodic stripping voltammetry is evaluated [14].



The objective is to provide an almost 'ready-to-use' method that could be easily adapted (and validated) by testing laboratories subjected to ISO 17025 Quality System.

### **1.11 The toxicity of nitrates in the environment**

Nitrates and nitrites are nitrogen-oxygen units which combine with various organic and inorganic compounds. Nitrates have a higher distribution in nature. Nitrates are widely used in explosives, fertilizers, drugs, and many industrial products. Agricultural and sewage effluents as well as many other environmental samples and pollutants may also contain appreciable concentrations of nitrates and related ions [9, 15]. In food, nitrates are used for preservation of meat and in agriculture it is used as a fertilizer [16]. Once taken into the body, nitrates are converted to nitrites. Microbial action in soil or water decomposes water containing organic nitrogen first into ammonia, which is then oxidized to nitrates. Because nitrite is easily oxidized to nitrate, nitrate is the compound predominantly found in ground water and surface waters. Contamination of nitrogen containing fertilizers, including anhydrous ammonia as well as human natural organic wastes, can raise the concentration of nitrates in water [9, 17].

In 1974, a congress passed the safe drinking water act. This law requires US EPA to determine the safe levels of chemicals in drinking water, which do or may cause health problems. These non-enforceable levels, based solely on possible health risks and exposure, are called maximum contamination level goals (MCLG). The MCLG's for nitrates have been set at 10ppm, and nitrites at 1ppm, because EPA believes this level of protection would not cause any of the potential health risks. Based on these MCLG's, EPA has set an enforceable standard called maximum contamination level (MCL). MCL's are set as close as possible to MCLG's, considering the ability of public water systems to detect and remove contaminants using suitable technologies. The European Union (EU) has set their limit at 50ppm. In South Africa the maximum contamination level is set at 4.4ppm and 6ppm by South African bureau of Standards (SABS). In human, ingested nitrate is rapidly absorbed from the intestine and the proximal small amount is distributed throughout the body. Nitrates then enters the large bowel from the blood, where it is rapidly converted to highly reactive nitrite, in part by forming micro-organisms. The formed nitrite is reabsorbed into the blood, where it reacts with the ferrous  $\text{Fe}^{2+}$  ion of

deoxyhaemoglobin, forming met haemoglobin in the ferric  $\text{Fe}^{3+}$  valence state. Ferric ion is unable to transport oxygen. Nitrates are excreted rapidly in the urine. Nitrate is a precursor in the formation of N-nitroso compounds (NOC), a class of genotoxic compounds most of which are animal carcinogens.

In human body, nitrate is a stable, inert compound that cannot be metabolized by human enzymes. However, the nitrate-reducing activity of commensal bacteria may convert nitrate into nitrite and other bioactive nitrogen compounds that affect physiological processes and human health. Under acidic conditions in the stomach, nitrite is protonated to nitrous acid ( $\text{HNO}_2$ ), which in turn spontaneously yields dinitrogen trioxide ( $\text{N}_2\text{O}_3$ ), nitric oxide ( $\text{NO}$ ), and nitrogen dioxide ( $\text{NO}_2$ ).  $\text{NO}$  is a bioactive compound known to play a role in vasodilatation and in defense against periodontal bacteria and other pathogens.  $\text{N}_2\text{O}_3$ , on the other hand, is a powerful nitrosating agent capable of donating  $\text{NO}^+$  to secondary and tertiary amines to form potentially carcinogenic N-nitrosamines. Alternatively,  $\text{HNO}_2$  can be protonated to  $\text{H}_2\text{NO}_2$ , which reacts with amides to form N-nitrosamides. At neutral pH, nitrite can be reduced by bacterial activity to form  $\text{NO}$ , which can react with molecular oxygen to form the nitrosating compounds  $\text{N}_2\text{O}_3$  and dinitrogen tetroxide  $\text{N}_2\text{O}_4$ . In addition to the acid-catalyzed and bacterial-catalyzed formation of nitrosating agents, inducible  $\text{NO}$  synthase activity of inflammatory cells can also produce  $\text{NO}$ . Other sources of human exposure include preformed NOC found in preserved meats and fish, beer, certain occupational exposures, and tobacco products. Excessive levels of nitrates in drinking water have caused serious illness and sometimes death. The serious illness in infants is due to the conversion of nitrate to nitrite by the body, which can interfere with the oxygen-carrying capacity of the child's blood. The major source of nitrates in infants is drinking water used to dilute the baby formula [18, 19, and 20].

Methods in common use for the determination of nitrates are either direct or indirect. The indirect methods includes reduction of nitrate ion into ammonium, nitrite or nitric oxide gas followed by gasometry, or potentiometry, polarography, amperometry, spectrophotometry, fluorimetry, and atomic absorption spectroscopy. Electrochemical techniques represent an alternative, fast measuring and inexpensive group methods useful for monitoring environmental toxicity. Different types of

voltammetric methods are particularly suitable. The development of sensitive methods for measuring nitrates in the field is therefore of great interest and importance [21]. Electrochemical detection of contaminants is often employed as electrochemistry provides, clean, versatile, and cheap sensing methods with low limit of detections and high sensitivity [20]. The use of ion chromatography and potentiometric ion selective electrode with chemical sensors, on the other hand, provides an attractive direct approach for determining nitrates. Advantages offered by using potentiometric sensors are the ease of fabrication, simple monitoring instrumentation, fast response, application to turbid or coloured solutions, high sensitivity, wide linear dynamic ranges, low cost and possible interfacing with flow injection techniques [9]. As an application example, data from an adapted method for determining nitrate in water samples is evaluated. The determinations of nitrate have recently been a significant environmental subject of research [22]. Nitrate is a well-known contaminant largely encountered in different types of water samples. Due to its toxicity (mainly related to infants), health institutions established a maximum concentration in drinking water (e.g. European Community (EC) establish the parametric level at  $50 \text{ mg.L}^{-1}$  [22]. Several methods have been employed for nitrate determination in water, food and other matrices. Analytical techniques, such as spectrophotometry [23, 24, and 25], potentiometry [26], ion chromatography [27], high performance liquid chromatography [28, 29] and capillary electrophoresis [30] have been used; including, more recently, the use of biosensors [31]. The selected methodology, incorporating differential-pulse polarography, for the indirect determination of nitrate in water using o-nitro phenol [32] was evaluated in order to prepare a 'ready-to-use' format, for possible implementation by most testing laboratories. In this study Cd and Pb were determined in Harbor water and Nitrate was determined in industrial treated waste water following the methods published by Metrohm as application bulletins [14, 32].

## CHAPTER 2

### THEORETICAL PRINCIPLES RELATED TO THE EXPERIMENTAL DESIGN

#### 2.1 Applicability, fitness for purpose and acceptability limits

The method applicability is a set of features that cover, a part from the performance specifications, information about the identity of the analyte (e.g. nature and speciation), concentration range covered, kind of the matrix of material considered for validation, the corresponding protocol (describing equipment, reagents, analytical procedure, including calibration, as well as quality procedures and safety procedures) and the intended application with its critical requirements [33, 34]. The method applicability must be consistent with the “golden rules” for method validation namely:

- The analytical procedure has to be validated covering the full range of analyte concentrations specified in the method scope; the analytical procedure has to be validated as a whole, including sample treatments prior to analysis.
- The analytical procedure has to be validated for each kind of matrix where it will be applied.
- Fitness for purpose is the extent to which the method performance matches the agreed criteria or requirements. The accuracy can be assessed in a global way by the concept of acceptability  $\lambda$ . The analytical result  $z$  may differ from the “unknown true value”  $T$  to an extent less than the acceptability limit:

$$|Z-T| < \lambda \quad (2)$$

Where  $\lambda$  is the acceptability limit depends on the goals of the analytical procedure.

- A procedure can be validated if it is very likely that the requirement given above is fulfilled i.e.:  $P(|Z-\mu| < \lambda) \geq \beta$ , ( $\beta$  being the probability that a future determination falls inside the acceptability limits).

- It is possible to compute the so-called “ $\beta$ -expectation tolerance interval” ( $\beta$ ETI) i.e. the interval of future results that meets the above equation by using the accuracy profiles which is devoted to accuracy study and measurement uncertainty.
- The use of acceptability limits together with accuracy profiles is an excellent way of checking the fitness for purpose of the validated method [6, 7, and 8].

## 2.2. Intermediate precision and trueness studies

Both intermediate precision and trueness studies can be performed using the prediction of actual concentrations from the validation standards (VS) selected for the analytical assay. Certified or internal reference materials represent the best way to obtain VS, but spiked samples can be considered as a suitable alternative. VS's must be stable, homogeneous and as similar as possible to the future samples to be analysed, and they represent, in the validation phase, the future samples that the analytical procedure will have to quantify. Each (VS) must be prepared and treated independently as a future sample. This independence is essential for a good estimation of the between-series variance. Indeed, the analytical procedure is not developed to quantify routinely with the same operator and on the same equipment a single sample unknown on one day but a very large number of samples through time, thus often implying several operators and several equipments [8]. A suitable way to perform the intermediate precision study is to consider a single sample matrix and a range of analyte concentrations. It is advisable to be at least three concentrations levels  $m$  (low, medium, and high) covering the dynamic working range, with a number of  $n$  replicates at each concentration. The ICH [35] document recommends three replicates and the Food and Drug Administration (FDA) document on bio-analytical validation considers five replications [36], therefore 3-5 replicates is acceptable.

Calculations on intermediate precision must be carried out on results instead of responses. Considering the different conditions  $p$  (here, the days) chosen as the main source of variation, an analysis of variation (ANOVA) is then performed for each concentration. In actual fact, for each concentration level,  $m_e$ , we consider the results of the analysis, according to the inverse prediction,  $Y = \frac{Y - A}{b_{SAM}}$  (8)

Where Y is the measured response, A is the true value, and  $b_{SAM}$  being previously established in method scope and applicability,  $Z_{ij}$  with two indices: I (from 1 to p) corresponding to the different days and j (from 1 to n) accounting for the repetitions.

From ANOVA, we can easily obtain estimations of within condition variance ( $S^2_w$ ) and between-condition variance ( $S^2_B$ ) is given by:

$$S^2_w = S^2_r = \frac{\sum_{i=1}^p \sum_{j=1}^n (Z_{ij} - \bar{Z}_i)^2}{p(n-1)} \quad (9)$$

$$\text{With } \bar{Z}_i = \frac{\sum_{j=1}^n Z_{ij}}{n} \quad (10)$$

The between-condition variance is estimated according to:

$$S^2_B = \frac{\sum_{i=1}^p \bar{Z}_i^2}{p} - \bar{Z}^2 \quad (11)$$

$$\text{With } \bar{Z} = \frac{\sum_{i=1}^p \sum_{j=1}^n Z_{ij}}{Pn} \quad (12)$$

The intra-laboratory reproducibility or intermediate precision can be taken as:

$$S^2_R = S^2_r + S^2_B \quad (13)$$

From these data, the corresponding relative standard deviations,  $RSD_r$ , and  $RSD_R$  are calculated. These values can be compared with the expected values issued from the Horwitz's equation [36]. Horwitz devised an expression to predict the expected value of the relative standard deviation for inter-laboratory reproducibility ( $PRSD_R$ ) according to:  $PRSD_R = 2^{(1-0.5 \log C)}$  (14)

Where, C is the analyte concentration in decimal fraction units. The Horwitz value is now widely used as a benchmark for the performance of analytical methods via a measure called the "Horrat", which is defined as the ratio of the actual relative standard

deviation, RSD(either for repeatability or reproducibility), calculated from the analytical data to the Horwitz value:

$$Horrat = \frac{RSD}{PRSD_R} \quad (15)$$

Apart from Horwitz's parameters, expected values of RSD according to the Association of Official Analytical Chemists (AOAC) peer verified methods program are considered. These two approaches, as a function of the analyte concentration, are present in Table 1.

Some practical requirements concerned with inter-laboratory studies are:

$$RSD_r = 0.5-0.6 \text{ times } PRSD_R \quad (16)$$

$$RSD_R = 0.5-2 \text{ times } PRSD_R \quad (17)$$

**Table 1: Accepted RSD percentages obtained from the Horwitz function and from the AOAC peer verified methods (PMV) programme with respect to the analyte level of a sample in question.**

Analyte %	Analyte fraction	Unit	Horwitz %RSD	AOAC PMV % RSD
100	1	100%	2	1.3
10	10-1	10%	2.8	1.8
1	10-2	1%	4	2.7
0.1	10-3	0.1%	5.7	3.7
0.01	10-4	100ppm	8	5.3
0.001	10-5	10ppm	11.3	7.3
0.0001	10-6	1ppm	16	11
0.00001	10-7	100ppb	22.6	15
0.000001	10-8	10ppb	32	21
0.0000001	10-9	1ppb	45.3	30

For intra-laboratory validations, a quick rule is to divide the interval by 2, leading to:

$$RSD_r = 0.2-0.3 \text{ times } PRSD_R \quad (18)$$

$$RSD_R = 0.2-1 \text{ times } PRSD_R \quad (19)$$

In the bias calculation, the  $\bar{\bar{Z}}$  value is taken as the final result for z corresponding to the VS of estimated “true” concentration T, so the average bias for the VS is obtained from the elemental bias  $\delta_{ij} = Z_{ij} - T$  according to:

$$\delta = \frac{1}{pn} \sum_{i=1}^p \sum_{j=1}^n \delta_{ij} = \frac{1}{pn} \sum_{i=1}^p \sum_{j=1}^n Z_{ij} - T = \bar{\bar{Z}} - T \quad (20)$$

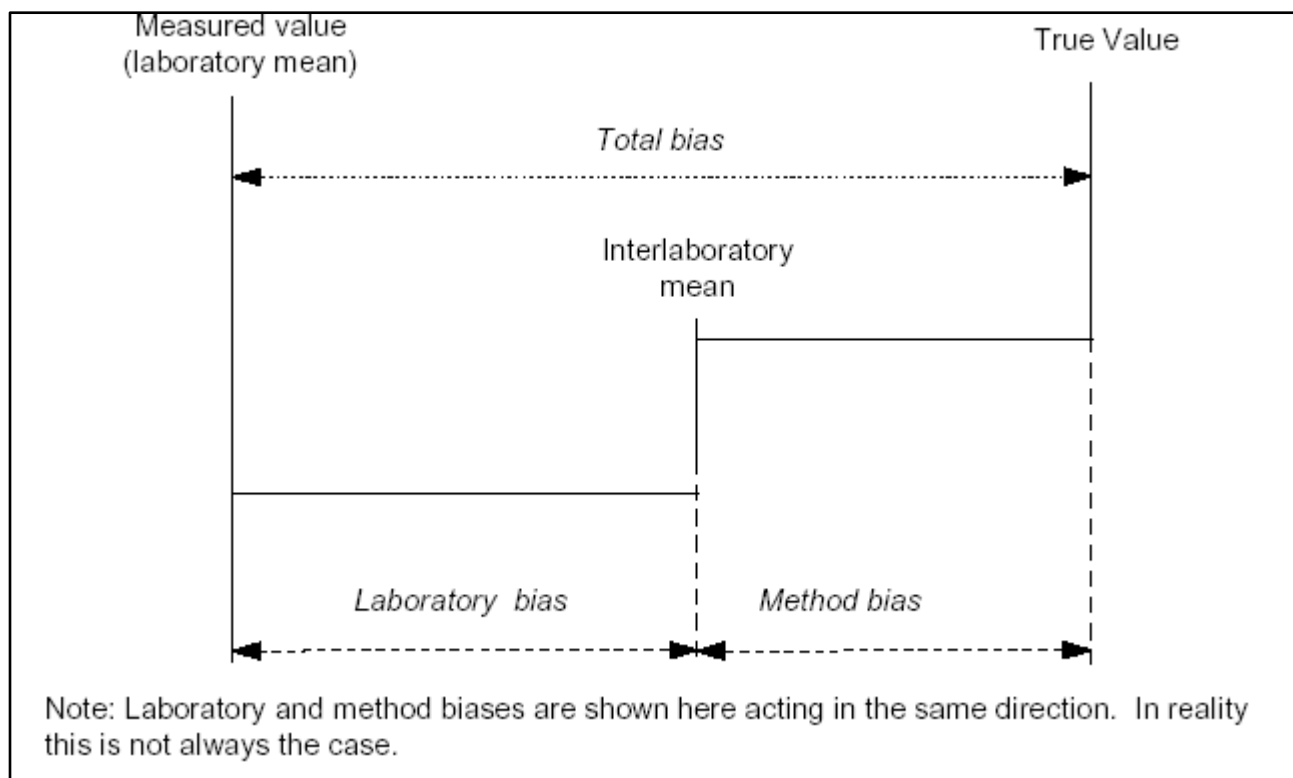
The bias uncertainty can be estimated from the same ANOVA design, according to ISO/DTG 21748 guide as:

$$\mu^2(\sigma) = \frac{S^2_R \left( 1 - Y + \frac{Y}{n} \right)}{p} \quad (21)$$

$$\text{With } Y = \frac{S^2_r}{S^2_R} \quad (22)$$

The practical assessment of trueness depends on comparison of mean results from a method with known values, that is, trueness is assessed against a reference value (i.e. true value or conventional true value) [2]. Trueness is expressed in terms of bias or percentage of error. Bias is the difference between the mean value determined for the analyte of interest and the accepted true value or known level actually present [2, 5]. According to the Analytical Methods Committee (AMC) technical brief [37] trueness is the closeness of agreement between the average value obtained from a large series of test results and an accepted reference value. Method trueness is also an indicator of utility and applicability of that method with real samples [5]. Different sources of systematic errors contribute to the overall bias ‘persistent’ bias affecting all data of the analytical system, over longer periods of time and being relatively small but continuously present. Different components contribute to the persistent bias, such as laboratory bias, method bias and the matrix-variation effect. Next to persistent bias, the larger ‘run effect’ is the bias of the analytical system during a particular run [5]. The overall bias determined by a particular laboratory during validation should then be compared with any reported bias for the regulatory method.





**Figure 3: Types of biases that are usually obtained in a laboratory which can be described as the laboratory and the method bias in a recovery experiment by virtue of using the CRM either fortified with the sample matrix or unfortified standard.**

More of these bias components are encountered when analyzing RM. Generally, RM's are divided into certified RM's (CRM's), either pure substances/solutions or matrix CRM's and non-certified laboratory RM's also called quality control samples as shown in figure 3. CRM's can address all aspects of bias (method, laboratory and run bias); they are defined with a statement of uncertainty and traceable to international standards. CRM's are therefore considered as a useful tool to achieve traceability in analytical measurements, to calibrate methods to monitor laboratory performance, to validate methods and to allow comparison of methods. Over and above that, the use of CRM's does not necessarily guarantees trueness of the results. The best way to assess bias practically is indeed by using replicate analysis of samples with known concentration, such as RM's. The ideal RM is a matrix CRM, as this is very similar to the samples of interest (the latter is called matrix matching). However, a correct result obtained with a matrix CRM does not guarantee that the results of unknown samples with other matrix compositions will be correct. If no such (certified) RM's are available, a blank sample matrix of interest can be 'spiked' or 'surrogate'. The recovery is then calculated as the percentage of the measured spike of the matrix sample relative to the measured spike of the blank control or the amount of spike added to the sample. The smaller the percentage recovery the larger the bias, that is affecting the method and thus, the lower the trueness [5].

### **2.2.1 Precision**

Precision is the closeness of agreement between independent test results obtained under stipulated conditions [37]. It has to be noted that precision depends only on the distribution of random errors and does not relate to the true value or the specific value. Independent test results means results obtained in a manner not influenced by any previous results on the same or similar test object. The measure of precision usually is expressed in terms of imprecision and computed as a standard deviation of the test result. Less precision is reflected by a larger standard deviation. The two most common precision measures are 'repeatability' and 'reproducibility'. They represent the two extreme measures of precision which can be obtained. Repeatability (the smallest expected precision) will give an idea of the sort of variability to be expected when a method is performed by a single analyst on one piece of equipment over a short time scale, i.e. the sort of variability to be expected between results when a sample is

analysed in duplicate. If a sample is to be analysed by a number of laboratories for comparative purposes then a more meaningful precision measure is to use the reproducibility (this is the largest measure of precision normally encountered, although it does formally exclude variation with respect to time) [2]. According to AMC technical brief [37], repeatability conditions where independent test results are obtained with the same method on identical test items in the same laboratory by the same operator using the same equipment within short intervals of time where as reproducibility conditions are conditions where test results are obtained with the same method on identical test items in different laboratories with different operators using different equipment [37].

## **2.3 Harmonization of a method**

Harmonization is generally used for criteria unification between organizations (e.g. standards, guides, etc.). For instance, nowadays it is well documented that ‘fit-for-purpose’ internal method validation (i.e. establishing method’s requirements-features relationships), Quality Assurance (including Internal Quality Control, IQC) and sample uncertainty estimations (normally based on adapting the ‘Bottom-up’ approach to the chemical analysis) are mandatory for accredited routine testing laboratories [2, 12, 38]. However, more harmonization on (detailed) technical protocols and statistics is necessary. Recently, an effort to harmonize the internal validation-control-uncertainty process (i.e. the so called u-approach [23]) has been made, since in the past, these internal quality aspects have been treated separately, producing inconsistencies.

Another aspect that deserves more harmonization is the different perspectives of research laboratories performing method development tasks (prior to publishing almost ‘ready-to-use’ analytical methods, but, focusing mainly on establishing method’s features) and routine testing laboratories (usually under ISO 17025 accreditation processes) where practical, e.g. economical aspects use to be mandatory. In most cases, the lack of harmonization of analytical methods (published in prestigious journals) are not easily implemented by testing laboratories.

### **2.3.1. Internal Quality protocol**

Most standard operation procedures (SOPs) written by testing laboratory managers tend to be ambiguous (i.e. they specify what must be done but not how it

have to be done, in detail; as most standards and guides). However, most of them pass the auditing processes from national accreditation entities. In view of harmonization, it should be suitable to account with more detailed step-by-step protocols covering the different internal (preferably also external) quality tasks to be performed. Moreover, if harmonized statistics are used, the same information may cover all the decision-making aspects, assuring consistent results. The following sections shows an attempt to harmonize the internal quality tasks related to quantitative assay methods. This protocol could offer guidance to routine and research laboratories as a way to encourage convergence on their viewpoints. Obviously, technical SOPs derived from it, would need to be adjusted for each particular method/assay/laboratory and circumstances.

### **2.3.2 Protocol for harmonizing the quality tasks (quantitative methods).**

#### **(i) Method's scope (Previous requisite for ISO 17025 accreditation)**

Fix analytes and matrices (Scope-samples) to be analysed. Decide on the concentration range (minimum and maximum Scope-concentrations, Sc). Sc should be simultaneously consistent with requisites (e.g. legal levels) and method's features (e.g. limit of quantification, LOQ, linear range), but also with practical considerations. Normally this aspects are accounted by testing (but overlooked by research) laboratories (e.g. presumable repeated problems to fit IQC limits or excessive cost associated to the chemicals/materials due to the selection of (unnecessary) extremely low or high Sc, respectively) [12].

#### **(ii) Method's requisites**

Decide on external (legislation, client) and/or internal (prefixed by the laboratory) requisites for the candidate method. Consider the accuracy (in intermediate precision conditions; intermediate accuracy), a mandatory aspect (e.g. fix the limits for the error,  $E_{LIM}$ , and the intermediate relative standard deviation,  $RSDi_{LIM}$ , to be satisfied). Intermediate accuracy must account for bias (trueness) and precision (batch/run repeatability and between-batch/run) values [2, 12, 38].  $RSDi_{LIM}$  can be derived from repeatability ( $RSDr_{LIM}$ ) and between-run ( $RSDrun_{LIM}$ ) relative standard deviation (assimilating the run to a batch) limits as:

$$RSDi_{LIM} = (RSDr_{LIM}^2 + RSDrun_{LIM}^2)^{0.5}$$

### **(iii) Feature selection**

Select those method's features that need to be validated in order to fit the requisites ('fit-for-purpose' concept). Include always intermediate accuracy, by planning (experimental design) a number of runs (Ns, sessions combining e.g. days/analysts/instruments; at least equivalent to that planned during the future routine analysis stage) and replicates per run, Nr, (i.e. intermediate accuracy  $\mathbf{X}_{Nr \times Ns}$  validation matrix) [2, 12, 39].

### **(iv) Reference values**

From an academic and literature research point of view, CRMs are considered the highest achievable real reference in the metrological hierarchy [15] and their use is widely suggested in ISO 17025 [38]. However, ISO 17025 auditors (currently) use to consider that validation and IQC tasks must involve the Scope-samples (i.e. available real samples appearing in the accreditation scope). A compromise (which also allows covering the entire Sc-range) could be using real Scope-samples fortified with CRM at different levels [12, 40]. Select (preferably; if possible) a certified reference material or solution (CRM) to fortify (i) an available Scope-sample blank, an available 'negative Scope-sample' (concentration under the LOQ or minimum-Sc with accepted negligible signal effect) or an available 'low level positive Scope-sample' (low-concentration sample, i.e. the 'sample is considered as a blank' requiring a recovery experiment [41]). This must provide the accepted reference value and the uncertainty e.g. of the (adjusted-to-concentration) CRM-validation solutions ( $\mu_{CRM} \pm u_{CRM}$ ), where  $u_{CRM}$  can be estimated by the (classical) bottom-up approach [40, 42].

### **(v) Accuracy pre-validation (suggested)**

Recommendable if (i) no previous information on intermediate accuracy is available, (ii) It is supposed that intermediate accuracy statistics are sufficiently within the limits (e.g.  $E \ll E_{LIM}$  and  $RSDi \ll RSDi_{LIM}$ ), (iii) the behaviour of the method with real (e.g. different matrices) samples is unknown [13, 43]. Performs a relatively low  $\mathbf{X}_{2 \times 12}$  experimental design could be a reasonably starting point if atypical data are not expected (otherwise,  $Nr \geq 3$  is preferable).  $\mathbf{X}_{Nr \times Ns}$  experiment using e.g. CRM-validation solutions to obtain provisional estimates of method's intermediate accuracy (e.g. E, RSDi). Validation statistics can be obtained from  $\mu$  (the 'run mean'; grand

mean of the  $\mathbf{X}$  validation matrix) and ANOVA of the  $\mathbf{X}$  validation matrix estimates (residual mean square, MSr, and the between-run mean square MSrun):

$$E = 100 (\mu - \mu_{\text{CRM}}) / \mu_{\text{CRM}} \quad (3)$$

$$\text{RSDr} = 100 (\text{MSr})^{0.5} / \mu \quad (4)$$

$$\text{RSDrun} = 100 ((\text{MSrun} - \text{MSr}) / \text{Nr})^{0.5} / \mu$$

$\text{RSDi} = 100 ((\text{MSrun} - \text{MSr}) / \text{Nr} + \text{MSr})^{0.5} / \mu = (\text{RSDr}^2 + \text{RSDrun}^2)^{0.5}$  could be avoided if  $E \ll E_{\text{LIM}}$  and  $\text{RSDi} \ll \text{RSDi}_{\text{LIM}}$  (see Fig. 3).

Perform a relatively short  $\mathbf{X}_{\text{Nr} \times \text{Ns}}$  experiment using Scope samples to obtain a realistic view of method behaviour in routine conditions (e.g. RSDi) or to investigate matrix effect. If it is necessary (E is not estimated).

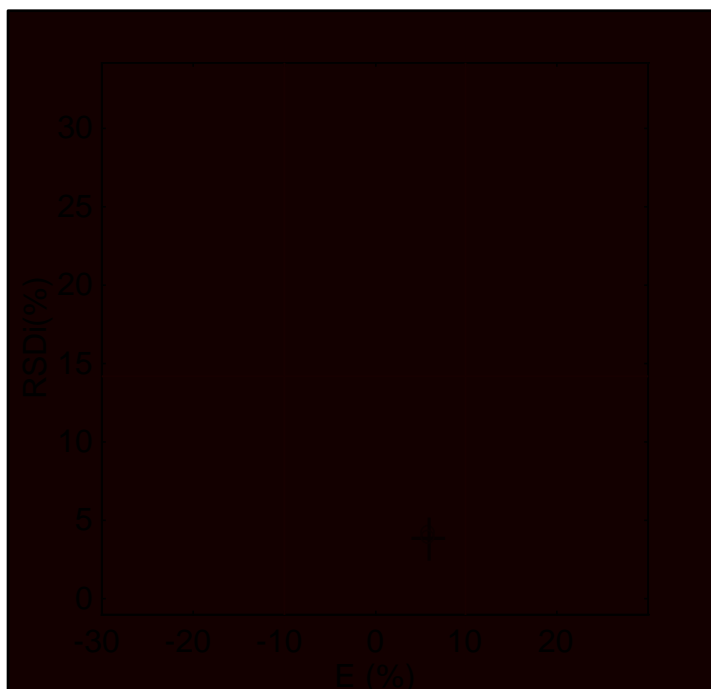
#### (vi) Accuracy statistics estimation

Perform a relatively large i.e. a safe experiment could be  $\mathbf{X}_{4 \times 25}$  [23]. However, it could be adjusted to previous information on intermediate accuracy  $\mathbf{X}_{\text{Nr} \times \text{Ns}}$  experiment using e.g. CRM-validation solutions. After appropriate outlier detection/elimination not a harmonized task. Results will depend on the criterion used. Inconsistencies between classical outliers test (e.g. Grubbs test) and control plots used on the validation data (i.e.  $\mathbf{X}$  validation matrix used as control data) can be found [42]. The z-score approach is a recognised tool in collaborative studies [13]).  $z_{ij} = |(x_{ij} - \mu_0) / \sigma_0|$ , where  $z_{ij}$  are the absolute z-scores for each estimated concentration ( $x_{ij}$ ) in the  $\mathbf{X}$  validation matrix,  $\mu_0 = \mu_{\text{CRM}}$  and  $\sigma_0 = \text{RSDr}_{\text{LIM}} \cdot \mu_{\text{CRM}} / 100$  could be consistent values; so,  $z_{ij} = 100 |((x_{ij} / \mu_{\text{MRC}}) - 1) / \text{RSDr}_{\text{LIM}}|$ . An  $x_{ij}$  with  $z > 2$  in a run (j) could be considered outlier (more study is required). This strategy (convenient for testing laboratories) still deserves more study/discussion/harmonization (e.g. z-score approach), obtain intermediate accuracy estimates (e.g. E and RSDi). Perform the process at e.g. 3 concentration levels (close to the extreme Scope and an intermediate level). ISO 17025 auditors (currently) use to demand accuracy validation covering the entire Sc range. This allows establishing the precision-concentration relationships and can serve to detect the type of systematic (constant or proportional) error in cases of large Scope ranges [23, 42].

**(vii) Accuracy validation assessment**

Perform a massive (say  $10^4$ ) simulated validations (i.e. Monte Carlo simulation generating new  $\mathbf{X}_{N_{rx}N_s}$  data) each simulated  $x_{i,j}$  can be derived from the model (see figure 4) [43]:

$x_{ij} = \mu_{CRM} (1 + E/100) + f_j + e_{ij}$ ; where  $x_{ij}$  is the computed result related to the  $i$ th replicate of the  $j$ th run;  $f_j$  is the random run effect for the  $j$ th run ( $f_j \sim N(0, \sigma_{run}^2)$ ; where  $\sigma_{run}^2$  corresponds to  $RSD_{run}$ ) and  $e_{ij}$  is the random repeatability effect for the  $i$ th replicate and  $j$ th run ( $e_{ij} \sim N(0, \sigma_r^2)$ ; where  $\sigma_r^2$  corresponds to  $RSD_r$ ) consistent with the experimental  $E$ ,  $RSD_r$  and  $RSD_{run}$  estimates . Calculate uncertainty intervals for the intermediate accuracy estimates. The expanded uncertainty interval ( $\pm U$ ) can be estimated from two percentiles of the simulated results [44]. For the  $RSD_i \pm U(RSD_i)$  uncertainty interval the 2.5%- and 97.5%-percentiles of the  $RSD_i$  results were used. For the  $E \pm U(E)$  uncertainty interval the 5%- and 95%-percentiles of the  $E$  results has been recommended in [45] and verify that they are within the limits (e.g.  $|E \pm U(E)| < |E_{LIM}|$  and  $[RSD_i + U(RSD_i)] < RSD_{iLIM}$ ) or perform the intermediate accuracy validation plot (See Figure 4) in order to declare the accuracy validation as valid [13].



**Figure 4: Intermediate accuracy validation plot for  $X_{6 \times 12}$ , (o) estimated and ( $\phi$ ) simulated**

**E and RSDi values. Theoretical estimations of  $E_{LIM}$  and  $RSDi_{LIM}$  are included. Since  $|E \pm U(E)| < |E_{LIM}|$  and  $[RSD_i + U(RSD_i)] < RSD_{iLIM}$  the method accuracy can be declared as validated.**

#### **(viii) Harmonized statistics for sample uncertainty and IQC**

Classical approaches to estimate the process standard deviation ( $S$ ; used to fix the mean control chart limits [46]) but used to be inconsistent with chemical analysis processes [47]. The Analytical Methods Committee proposal [48], based on the average moving range ( $S = 0.8865 \text{ mean}(\text{MR})$ ), works reasonably well using the ‘run mean’ as centre line ( $CL = \mu$ ), if  $RSD_{run} > RSD_r$  [20]. The u-approach [13] establishes that  $S_u$  (to distinguish it from  $S$ ) is equivalent to the combined uncertainty ( $u$ ; definition of absolute uncertainty related to the use of CRM [17]). So, assuming that each CRM control result corresponds to the mean of  $N_r$  replicates in a single run:  $S_u = u = (s_{run}^2 + s_r^2 / N_r + s_\mu^2 / N_s + u_{MRC}^2)^{0.5}$ , where the variances can be estimated from ANOVA of the  $\mathbf{X}$  validation matrix:  $s_\mu^2 = MS_{run} / N_r$ ,  $s_r^2 = MS_r$ ,  $s_{run}^2 = (MS_{run} - MS_r) / N_r$ ,  $s_i^2 = s_{run}^2 + s_r^2$ . Then the u-approach allows harmonizing the validation-control-uncertainty process.  $S_u$ , that could be used combined with  $CL = \mu$ , seems also to work well with  $CL = \mu_{CRM}$  for methods with moderate bias and without limitation on the  $RSD_{run}$  and  $RSD_r$  values. Note that  $u$  could depend on concentration level. Calculate (harmonized



with intermediate accuracy validation) statistics for Scope-sample result uncertainty (i.e. expanded uncertainty, U). During the routine analysis period, after performing Nr' replicates of Scope-samples, the expanded uncertainty of the sample result (Res; usually the mean) can be calculated as  $U = k u$  (k = coverage factor; for k = 2, a common option, U is roughly equivalent to half the length of a 95%-confidence interval [3, 40]).

For comparing the uncertainty at different concentration levels, the relative uncertainty in percentage can be calculated:  $U_r = 100 U / \mu$ . Note that u could depend on concentration level. In some instances, a practical approach a simplification could be using the 'worst case' (i.e. the highest U) for all the Scope (Sc) range. For reporting purposes the proposed format can be used:  $Res \pm U(Res)$ ; k = 2, and IQC stages (e.g. method verification). Method Verification is currently an ambiguous and not harmonized task. Verification solutions should be performed at least two levels (e.g. close to the extreme Sc values; to agree with ISO 17025 auditors' common requirements) which could be alternated between runs and performed before starting each run (e.g. Nr' replicates), as a rule to start the sample analysis period. They could be made exactly as the CRM-validation solutions. A logical approach should be relating the Verification limits to uncertainty, e.g.  $\mu_{CRM} \pm U$ . A practical approach (a simplification) could be using the 'worst case' (i.e. the highest U) instead of using the adjusted-to-concentration U values. For the low concentration level k = 3 could be considered reasonable and harmonized mean control chart. Along a routine analysis session (run), control solutions should be used to monitor method stability (and bias, using the u-approach), according to a pre-fixed IQC plan (e.g. Nr' replicates of the control solution each 10 samples or 2 h). 'Stable' control solutions could be performed at a single concentration level (e.g. intermediate Sc level complementing the method verification stage; which should agree with ISO 17025 auditors' preferences). They could be made exactly as the CRM-validation solutions. According to the harmonized u-approach, Action limits (AL) and Warning limits (WL) could be fixed for a harmonized mean control chart:

$$AL = CL \pm 3 S_u = \mu_{CRM} \pm U (k=3) \quad (5)$$

$$WL = CL \pm 2 S_u = \mu_{CRM} \pm U (k=2) \quad (6)$$

These limits could be combined with classical rules to declare the out-of-control situation (e.g. Rule 1 corresponds to a point outside either AL. Rule 2 to two consecutive points between the WL and AL on either side of the CL [36, 43]). NOTE: other control plots, e.g. Range or Moving Range if  $Nr' = 1$ , CUSUM, MA, EWMA

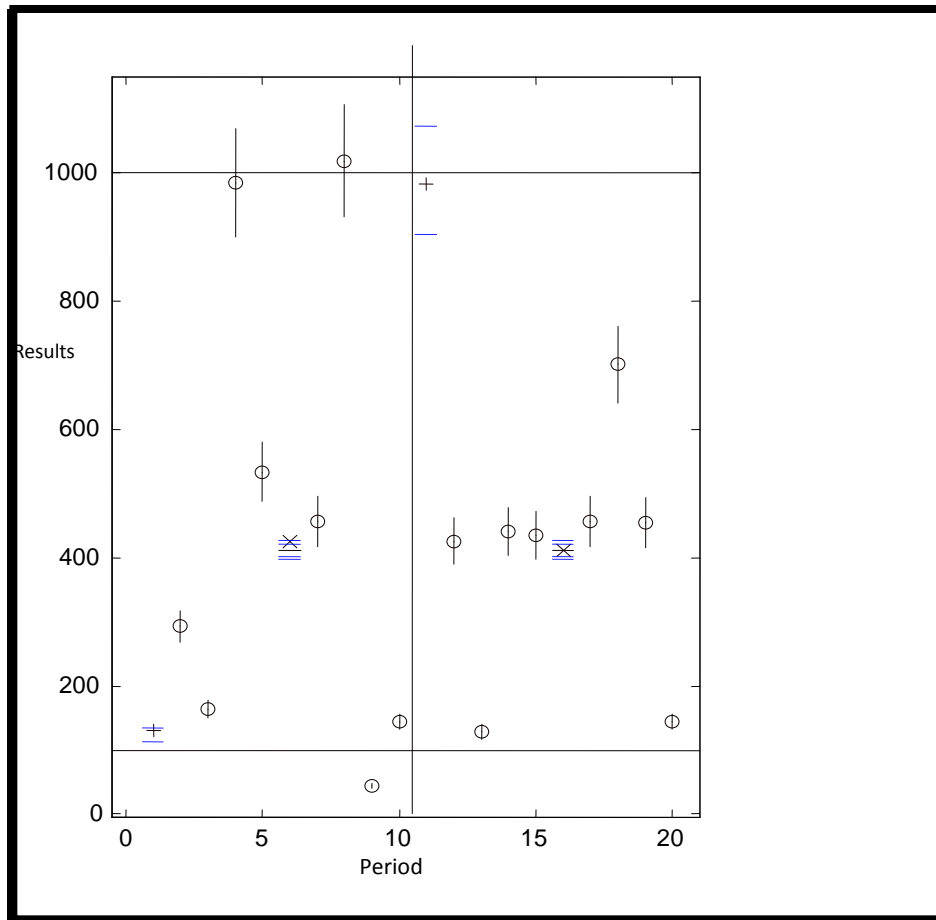
[35, 42] could complement the mean control charts. A practical approach (a simplification) could be using the ‘worst case’ (i.e. the highest U) instead of using adjusted-to-concentration U values), corresponding to the routine analysis period. The U, statistic depends on the number of replicates ( $Nr'$ ) used.  $Nr'$  could be (but it is not necessary) the same during sample analysis and IQC stages. Alternatively, it could be modified according to a ‘Progressive routine work’ scheme.  $Nr'$  concerns the method rapidity (and cost) and should be adjusted to the method’s features. ‘Progressive routine work’ (an internal rather than a described/normalised practice in some testing laboratories) involves fixing a low  $Nr'$  value (i.e. in some instances  $Nr' = 1$  could be reasonable) for routine work (i.e. sample analysis and IQC stages) and only when necessary (e.g. punctual high disperse results, samples close or above critical levels, etc.),  $Nr'$  is progressively increased (e.g. up to a maximum value of  $Nr$ ). Note that U depends on  $Nr'$ , although for simplicity, the corresponding value to the minimum  $Nr'$  (the worst case) could be used. This strategy which is convenient for testing laboratories) still deserves more study, discussion, and harmonization. **NOTE:** that when  $Nr'$  is punctually increased, appropriate outliers detection/elimination tools could be used (e.g. z-score approach; for samples, where  $\mu_{CRM}$  is not available, the mean or median could be used) [23].

#### **(ix) Validation of other features**

If necessary, perform the validation of the rest of fit-for-purpose method’s features (apart from accuracy). If all requisites are satisfied, declare the method as valid (in the corresponding validation report from the Quality System for accredited laboratories). Optionally (or if it is required by ISO 17025 auditors), non fit-for-purpose features could be established [12, 38].

**(x) Short-term routine work**

This step is important for testing laboratories (e.g. it can provide initial information that can be used to develop unambiguous consistent SOPs, when the results conforms with the requirements, or to advise of inconsistencies that would require further method development or relaxing on requirements [2]), but also for research teams (method development tasks) intending to offer ‘ready-to-use’ methods. In order to pre-examine the reliability of the overall criteria described in this protocol, a short-term routine analysis period (verification solutions, Scope-sample analysis and control solutions) can be initiated (Figure 5).



**Figure 5: Short-term routine work where(x): Method Verification: Period 1 (run-1, low-Sc) and period 11 (run-2; high-Sc); including verification limits. Sample analysis: Periods 2-5 and 7-10 (run-1); Periods 12-15 and 17-20 (run-2); including uncertainty estimations. Method Control (intermediate-Sc): Period 6 (run-1) and Period 16 (run-2); including AL and WL. Horizontal lines represent the extreme Sc. Vertical line represent the run separation. Comments: Verification is ‘conform’ in the 2 run periods. Control is ‘conform’ (when more data are available, the harmonized mean control plot could be used). Sample in period 8 must be reported as > 1000 and in period 9 as < 100 (arbitrary units).**

In figure 5 experimental results in different stages/periods (an example of short-term routine work; (x)): Method Verification: Period 1 (run-1, low-Sc) and period 11 (run-2; high-Sc); including verification limits. Sample analysis: Periods 2-5 and 7-10 (run-1); Periods 12-15 and 17-20 (run-2); including uncertainty estimations. Method Control (intermediate-Sc): Period 6 (run-1) and Period 16 (run-2); including AL and WL. Horizontal lines represent the extreme Sc. Vertical line represent the run separation. Comments: Verification is ‘conform’ in the 2 run periods. Control is

‘conform’ (when more data are available, the harmonized mean control plot could be used). Sample in period 8 must be reported as  $> 1000$  and in period 9 as  $< 100$  (arbitrary units). Alternatively, it could be extracted from previous information obtained during the application of the protocol (e.g. The **X** validation matrix data could be used, as IQC data, to check the established limits or, if Scope-sample were analysed as was stated in section (v),  $U$  can be computed and compared with legal or critical levels, if available).

**(xi) How harmonization can be improved (theoretical simulation)**

The above feature includes some decision-making aspects, aimed to improve harmonization, and some innovative (or known but unusual) criteria. Some of them must be proved (in future) in different situations (methods, analysis and laboratories) in order to be adjusted when necessary. Prediction using simulation can be used to predict the behaviour of the method to be validated, supposing reference and method data exist, method parameters are as follows: i.e. [ $\mu_0 = 10$ ,  $E_0 = 5$ ,  $RSDr_0 = 20$ ,  $RSDrun_0 = 10$ ,  $E_{LIM} = 15$ ,  $RSDi_{LIM} = 30$ ,  $Nr = 6$ ,  $Ns = 12$ ,  $Nmc = 10$ ] according to the limits set by AOAC and US EPA used in case study 2 ( see figure 4). Where  $\mu_0$  refers to an internal requisite i.e.  $10 \text{ ugL}^{-1}$  for case study 2 is an internal requisite,  $E_0$  refers to the estimated bias of a method,  $E_{LIM}$  corresponds to acceptability limit for the relative error in trueness assessment,  $RSDi_{LIM}$  corresponds to acceptability limit for RSDi in precision assessment,  $Nr$  corresponds to the number of independent replicates per run in the validation stage,  $Ns$  refers to the number of runs during the validation stage,  $Nmc$  corresponds to the number of theoretical simulation.

One aspect that deserves more attention is the impact on laboratory staff (usually with limited statistical skill). It is convenient to integrate/automate all aspects considered, which could be easily done via software (e.g. MATLAB®) facilitating report archive (e.g. for auditing tasks). For instance, normally distributed validation results, consistent with  $E$ ,  $RSDr$  and  $RSDrun$  estimates can be easily obtained by means of the RANDOM function (Statistic toolbox in MATLAB [27]). Once the algorithm is programmed, e.g. providing graphical outputs (as Figures 4 and 5), the process become easy to interpret without special statistical qualification.

## 2.4 Sensitivity, detection limit and quantification limit

### 2.4.1 Sensitivity

Sensitivity is the change in the analytical response divided by the corresponding change in analytical concentration i.e. at a given value of analytical concentration  $Z_o$  :

$$Sensitivity = \left( \frac{dy}{dZ} \right) Z_o \quad (7)$$

Sensitivity is not always mentioned as a validation parameter in official guidelines. It is useful in validation because it is usually arbitrary, depending on the instrument settings. In addition to sensitivity, there are two parameters reciprocally derived from sensitivity, much more used for performance characteristics the limit of detection (LOD) and the limit of quantitation or determination (LOQ).

### 2.4.2 Limit of detection

There is no analytical term or parameter for which there is a greater variety of terminology and formulations than for LOD and LOQ [5]. Measurements made at low analyte or property levels, e.g. in trace analysis, it is important to know the lowest concentration of the analyte or property value that can be confidently detected by the method [2]. The limit of detection, or detection limits, is the terminology most widely used as accepted by the EURACHEM. ISO uses ‘minimum detectable net concentration’, while IUPAC prefers ‘minimum detectable (true) value’. The ICH guide prefers to define LOD as the lowest amount of analyte in a sample which can be detected but not necessary quantitated as an exact value. This is consisted with what other official organizations refer.

Generally, the LOD is expressed as a concentration  $C_L$  or a quantity  $q_L$ , defined from the smallest signal  $x_L$  which can be detected with a reasonable certainty for a given analytical procedure. The lowest signal  $x_L$  is the signal that lies  $k$  times  $SD_{blank}$  above the mean blank value, whereby  $k$  is a numerical factor chosen according to the level of confidence required [2, 4, 7, 47]. The larger the value of  $k$ , the larger is the level of confidence. Eurachem and IUPAC recommend a value of 3 for  $k$ , meaning that the chances that a signal more than  $3s$  above sample blank value is originating from the blank is less than 1%. The LOD is thus the concentration or amount corresponding to the measurement level (response, signal)  $3S_{bl}$  (standard deviation of the blank) in balnk

units above the value for zero analyte. At the concentration or amount three times the  $S_{bl}$ , the relative standard deviation or coefficient of variation on the measured signal is 33% (measure of uncertainty). According to USP/ICH, the LOD corresponds to that signal where the ‘signal-to-noise ratio’ is 2:1 or 3:1.

**Table 2: Limit of detection and limit of quantitation quick reference [2].**

What to analyse	What to calculate
10 independent sample blanks measured once each or	Sample standard deviation of a sample blank values, or fortified sample blank values
10 independent sample blanks fortified at lowest acceptable concentration measured once each.	Express LOD as the analyte concentration corresponding to a mean sample blank value $+3s$ or for voltammetry $0+3s$ . LOQ expressed as sample blank value $+10s$

It is not true as often thought- that detection or quantification is impossible below the detection limit; but, at these lower levels, the uncertainty of the detection or quantification of a measurement is higher than the actual value itself. In this context, LOD can be defined as the point at which a measured value is larger than the uncertainty associated with it. The LOD is a concentration point where only the qualitative identification is possible but not accurate and precise quantification.

#### **2.4.3 Limit of quantification**

The ‘limit of quantitation’ is strictly the lowest concentration of analyte that can be determined with an acceptable level of repeatability precision and trueness. For LOQ,  $k$  is taken to be 5.6 or even 10 [2, 4, 7, 26]. A value of 10 for  $k$  means that the relative standard deviation (%RSD) at the LOQ is 10%. The LOQ thus corresponds to that concentration or amount of analyte, quantifiable with a variation coefficient not higher than 10%. The LOQ is always higher than LOD and is often taken as a fixed multiple (typically 2) of the LOD. The determination limit is referred to as the signal 10 times above the noise or background signal, corresponding to a ‘signal-to noise ratio) of 10:1 [2, 4, 7, 26]. Neither LOD nor LOQ represent levels at which quantitation is impossible. It is simple that the size of the associated uncertainties approach comparability with the actual result in the region of the LOD.

## **2.5 Matrix effects on sample response**

Sample components different from analyte (s) are denoted by the term matrix. Matrices can affect the analyte response by two different mechanisms. Some of the matrix components present at the time of measuring can somehow modify the analyte response. Or some sample treatment, indispensable because of the matrix presence, can affect the response; any extraction, precipitation, adsorption or the like operations to which the sample is subjected prior to the measurement can result in a matrix effect [48]. The use of external calibration enormously simplifies the protocol because calibration standards are prepared as simple solutions of the analyte. However, the effects of possible matrix effects coming from the sample material must be checked. A very useful tool for testing constant and proportional bias due to matrix effects is the standard addition method (SAM) [7, 48]. Recovery is sometimes considered a separate validation parameter, but, in any case, it should always be established as a part of method validation. Apart from the statistical significance, there are published acceptable percentages as a function of the analyte concentration as shown in Table 3 [4, 49].



**Table 3: Acceptable recovery percentages depending on the analyte concentration level [7]**

Analyte	Analyte fraction	unit	Recovery (%)
100	1	100%	98-102
10	$10^{-1}$	10%	98-102
1	$10^{-2}$	1%	97-103
0.1	$10^{-3}$	0.1%	95-105
0.01	$10^{-4}$	100ppm	90-107
0.001	$10^{-5}$	10ppm	80-110
0.0001	$10^{-6}$	1ppm	80-110
0.00001	$10^{-7}$	100ppb	80-110
0.000001	$10^{-8}$	10ppb	60-115
0.0000001	$10^{-9}$	1ppb	40-120

In principle, recoveries could be estimated by the analysis of matrix reference materials. The recovery is the ratio of the concentration of analyte found to that stated level of concentration to be present. If suitable materials are present it is possible to carry out recovery studies on reference materials. Provided that they have been produced by characterisation of natural materials rather than by characterisation of synthetic materials into which the analyte has been spiked, the recovery study should accurately represent (as shown in Table 4) the extraction of real test portions [2, 49].

**Table 4: Calculation of recoveries using the fortified as well as unfortified samples [2]**

What to analyse	Replicates	What to calculate	Comments
Matrix blanks or samples fortified or unfortified with the analyte of interest at a range of concentrations	6	<p>Determine the recovery of analyte at various concentrations:</p> $\text{Recovery\%} = (C_1 - C_2)/C_3 \times 100$ <p>Where <math>C_1</math> = concentration determined in fortified sample</p> <p><math>C_2</math> = concentration determined in unfortified sample</p> <p><math>C_3</math> = concentration of fortification for the determination of the recovery of analyte relative to the certified value.</p>	Fortified samples should be compared with the same sample unfortified to assess the net recovery of the fortification.
Certified reference materials (CRM), depending on how the CRM was produced and characterised, it may possible to get >100% recovery	6	Determine recovery of analyte relative to the certified value	Recoveries from fortified samples or matrix blank will usually be better than real samples in which the analyte is more closely bound

The recovery obtained in the routine method is estimated by using both methods to analyse a large set of typical test materials, a set that covers the required range of matrices and analyte concentration. This gives the recovery (and its uncertainty) for the routine method for any conceivable situation. In practice there may be no such definitive method available for reference, so reference materials or surrogate studies have to be used for the estimation of recovery. However, reference materials are few, and lack of resources restricts the range of test materials that can be used to estimate recovery by using surrogate spike. In addition, the use of surrogates

itself adds an uncertainty to a recovery estimate because it may not be possible to determine whether some proportion of the native analyte is covalently or otherwise strongly bound to the matrix and hence not recoverable. In order to overcome this, the recovery is estimated during the process of method validation. Recoveries are determined over a wide range of pertinent matrices and analyte concentration as resources allow. These values are then held to apply during subsequent use of the analytical method. To justify that assumption, all routine runs of the method must contain a reference material (or spiked samples) to act as an internal quality control. This helps to ensure that the analytical system does not change in any significant way that would invalidate the original estimates of recovery [49].

## **2.6 The study of accuracy**

Accuracy is the closeness of agreement between a test result and the accepted reference value. This definition is consistent with the ISO 3534:311 definition which states that the term accuracy, when applied to a set of test results, involves a combination of random components and a common systematic error or bias component [2, 37]. Accuracy is essentially the absence of error. A result of higher accuracy has a smaller error. Accuracy should not be used in contrast to precision a result is unlikely to be accurate if the results in general are not precise. It should be noted that, strictly, accuracy applies to results and not more general entities such as analytical methods laboratories or individuals, and should only be used that way in formal writing [37]. Method validation scrutinises the accuracy of results by considering both systematic and random errors. Accuracy is therefore studied as an entity with two components trueness and precision-but considered as a global entity, the uncertainty, from which the  $\beta_{eti}$  ( $\beta$ -expectation tolerance interval) will be estimated as well as the accuracy profiles once the acceptability limits have been established. Here, the more focus of attention is on estimating the uncertainty from the accuracy profile of the analytical assay from validation data [2, 8].

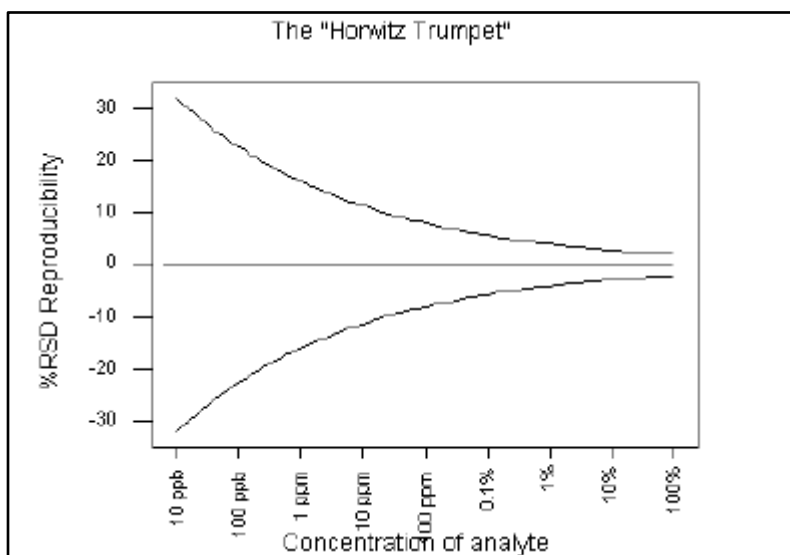
Precision and bias studies, which form part of the measurement uncertainty estimate, are the most important validation criteria.

Precision measures are divided into:

- Repeatability precision measures S or SD ( $S_r$  or  $SD_r$ ) and RSD ( $RSD_r$ );

- Intra-laboratory reproducibility precision or ‘intermediate precision’, measures SD and RSD; and, Inter-laboratory reproducibility precision S or SD ( $S_R$  or  $SD_R$ ) and RSD ( $RSD_R$ ).

Besides standard deviation and coefficient of variation, repeatability or reproducibility values or limits ( $r$ ,  $R$ ) are additional parameters of high value in the assessment of precision. These criteria mean that the absolute variation between two independent results is often the results-obtained within the same laboratory and between different laboratories-may exceed the value of  $r$  and  $R$  respectively in a maximum of 5% of the cases. Another measure of precision is the confidence interval, in which all measurements fall with a certain probability or confidence level  $1-\alpha$  (where,  $\alpha$  is often 0.05, giving a probability here of 95%). Calculated repeatability, intermediate precision and reproducibility values can be compared with those of existing methods (see figure 6). If there are no methods with which to compare the precision parameters, theoretical relative reproducibility and repeatability standard deviations can be calculated from the Horwitz equation [5, 49].



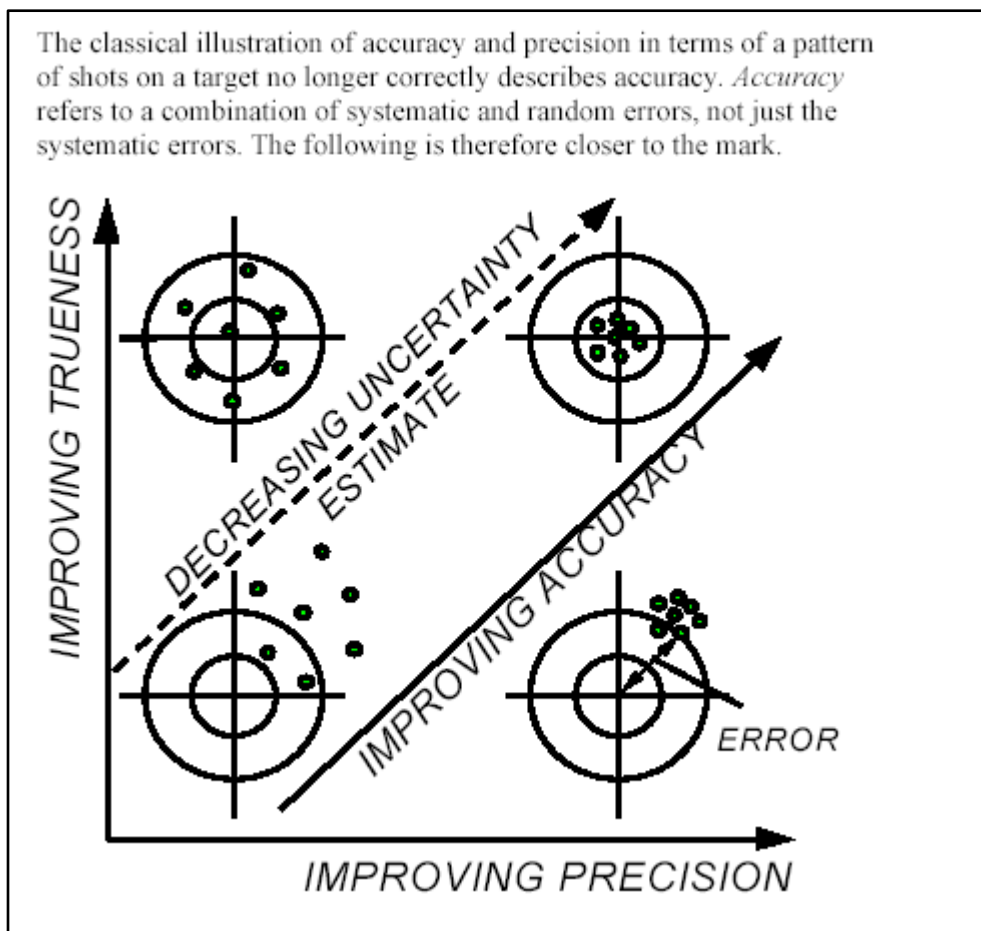
**Figure 6: The Horwitz function for comparing the %RSD v.s. analyte concentration**

## 2.7 Measurement uncertainty

Expressing the measurement uncertainty is now a well known problem for most analytical chemists who want to be accredited according to the ISO 17025 standard [38]. All results of a measurement have an associated uncertainty. (The loose term ‘margin of error’ conveys a rough idea of what analytical chemists mean by exactly defined term ‘uncertainty’). Every time an analytical measurement on a particular material is repeated we get a different result, even when it is repeated by the same person, in the same laboratory, using the same equipment, on the same day. This is not the outcome of carelessness: it is simply a reflection of uncontrolled variation in the measurement, which is usually a complex multistage procedure. In chemical analysis the uncertainty relative to the result could be as low as 0.1% or, for very difficult analysis, as high as 20% [50].

Uncertainty is an estimate attached to a test result which characterises the range of values within which the true value is asserted to lie (ISO 3534:3.25). The uncertainty of a measurement parameter, associated with the result of a measurement that characterises the dispersion of the values that could reasonable be attributed to the measurand (GUM: 2, 2.3). An estimate of uncertainty must incorporate both what is known about random effects and what is known about uncertainties associated with systematic effects on the measurement process. Because uncertainty estimates incorporates uncertainties from all possible effects, an estimate of uncertainty is probably the most appropriate means of expressing the accuracy of the results [37].

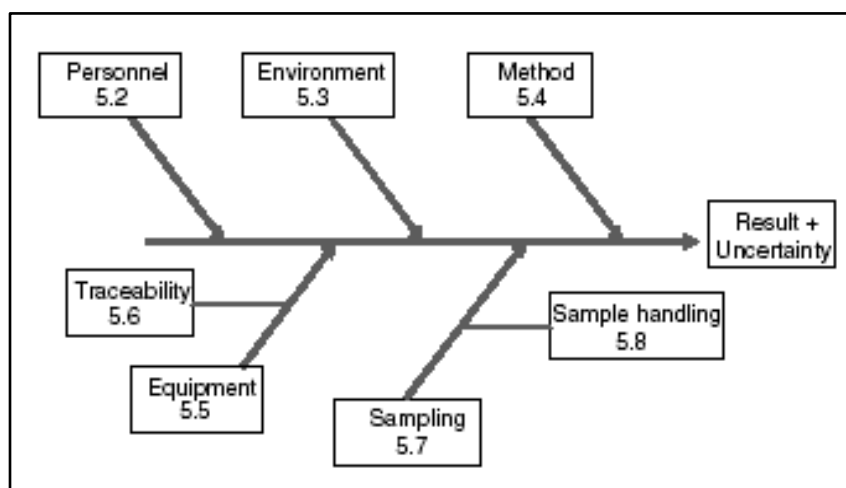
In routine analysis, the uncertainty associated with an individual result will usually be estimated from prior studies, including validation studies, and will not involve an individual estimate for each result (see figure 7).



**Figure 7:** Illustration of accuracy, precision and measurement uncertainty which states that in order to improve trueness the uncertainty estimate should be decreased thereby improving accuracy and precision

Different guides lay down general rules for the expression and evaluation of measurement uncertainty across a wide range of chemical measurements. According to Eurachem Guide [50, 51] the word “uncertainty”, when used outside the world of science, indicates doubt. Thus uncertainty of measurement could be understood to mean that the analyst is unsure about the validity and exactness of his results. In the Eurachem guide, the definition of “uncertainty” is a parameter associated with the result of a measurement that characterises the dispersion of the values that could reasonably be attributed to the measurand [4, 51, and 52]. The strategy proposed by the ISO guide for the expression of the uncertainty of a measurement Guidelines for the expression of the uncertainty of a measurement (GUM), is a widely adopted standard approach is reproduced below and demonstrates its scope. “The knowledge of the uncertainty of measurement results is essential to the interpretation of the results”.

Without quantitative assessment of uncertainty, it is impossible to decide whether observed differences between results reflects more than experimental variability, whether tests items comply with specifications, or whether laws based on limits have been broken. Without information on uncertainty, there is a real risk of either over-or under-interpretation of results. Incorrect decisions taken on such a basis may result in unnecessary expenditure in industry, incorrect prosecution in law, or adverse health or social consequences [52]. The GUM guide (see figure 8) has presented the most applicable procedure to identify uncertainty sources by cause-and-effect diagram (sometimes known as Ishakawa or fishbone diagram) and then combine them). This is likely to be the convenient procedure for an in-house determination of uncertainty.



**Figure 8: The relationship between the cause-and-effect diagram from the ISO 17025 standards [3]**

The traditional cause and effect diagram with its five “bones” can be related to some important chapters of the ISO 17025 standard. Therefore, it is clear that most of the examples presented in the Eurachem guide deal mainly with the influence of the method: Test and calibration methods and method validation or the equipment and consequently underestimate uncertainty. There is still a lot of debate among analysts about whether sources of uncertainty that are not taken into account may have a strong influence. The goal at stake is not significant, because it is clear that it is tempting for some laboratories to use uncertainty as a commercial argument: The smaller the

uncertainty, the better the measurement (or the laboratory). Therefore it is necessary to give an unambiguous response to this question in order to avoid any flaw in the use of uncertainty. The adoption of this new concept can be a great opportunity for analytical laboratories because it gives a real meaning to chemical measurements. Considering the problems raised by the expression of uncertainty and the need to have realistic values, new proposals have been made. They consist of experimental data obtained from precision studies. A recent draft of the ISO/DTS 21748 Guide suggests using repeatability, reproducibility and trueness estimates to estimate measurement uncertainty. Considering the data collected by analysts when developing an accuracy profile, it seems sensible to also use the data to estimate uncertainty. Depending on the way the experimental design is envisaged-several conditions within a laboratory, or several laboratories-the accuracy profile corresponds to either an intra-laboratory validation or inter-laboratory validation (reproducibility).

The guide only refers to inter-laboratory precision data. Intra-laboratory validation can also precisely reflect the actual capability of the method when applied by a given laboratory. When choosing between the intra-laboratory and inter-laboratory approaches it is important to note that the goal of the validation process must be to reflect the way an analytical procedure will be used in the future. If an analytical procedure is only intended to be used within a laboratory of interest such as days, runs, operators or batches. The estimated total variance or intermediate precision  $\sigma_w^2$  will be the sum of the between-condition variance and the within-condition variance or repeatability  $\sigma_w^2$ . On the other hand, if an analytical procedure is intended for use by many laboratories, then the important conditions to vary in experimental design are the between-laboratory conditions. In such a design, the between-condition (runs, days, batches) variances are by default integrated into, or more precisely not differentiated from, the within laboratory variance (repeatability). With such a design, the total variance, called the “reproducibility”, will be the sum of the between-laboratory variance and the within-laboratory variance (repeatability). To summarise, the only difference between the inter-laboratory and the intra-laboratory studies is the way the experiments are designed, and the definition that is given to the “condition” when estimating the “between-condition” variance components or uncertainties. According



to the recommendations of the ISO/DTS 21748 guide [53], a basic model for the uncertainty in a measurand Y:

$$\mu(Y)^2 = S^2_R + \mu(\hat{\sigma})^2 + \sum c_i \mu(x_i)^2 \quad (23)$$

Where  $S_R$  is the reproducibility standard deviation,  $\mu(\hat{\sigma})$  is the uncertainty associated with the bias  $\sigma$  of the method, and  $\sum c_i^2 \mu(x_i)^2$  is the sum of all effects due to other deviations. This equation can be simplified if only a statistical estimation of approach is used:  $\mu(Y)^2 = S^2_R + \mu(\hat{\sigma})^2$  (24)

According to the same guide ISO/DTS 21748 [53], the bias uncertainty can be estimated as:

$$\mu(\hat{\sigma}) = \frac{\sqrt{S^2_R(1-Y + \frac{Y}{n})}}{p} \quad (25)$$

Where n is the number of replicates (within-condition), p the number of varied conditions (the number of laboratories when reproducibility is the objective) and

$Y = \frac{S^2_r}{S^2_R}$ , where  $S^2_r$  is an estimate of the repeatability (within-condition) variance, and

$S^2_R = S^2_r + S^2_B$  is the estimate of the reproducibility conditions (the sum of the repeatability and the between-condition variance  $S^2_B$  component B. From the equation

$[\hat{\delta}_m \pm Q_1 K_s \hat{\delta}_m]$ , it is possible to write that the variance used to estimate the

$\beta$  -expectation tolerance interval which is equal to:

$\sigma_{Tot}^2 = K_s^2 \sigma_m^2$ , we can develop this equation by using  $\sigma_m^2 = \sigma_w^2 + \sigma_B^2$  and

$$K_s = \sqrt{1 + \frac{1}{pnB^2}}, \text{ with } B = \sqrt{\frac{A+1}{nA+1}} \text{ and } A = \frac{\sigma_B^2}{\sigma_w^2} \text{ yielding;} \quad (26)$$

$$\sigma_{Tot}^2 = \left[ \sqrt{1 + \frac{1}{pnB^2}} \sigma_m \right]^2 = \sigma_m^2 + \left( \frac{1}{pnB^2} \right) \sigma_m^2 \quad (27)$$

$$\hat{\sigma}_{Tol}^2 = \frac{\hat{\sigma}_m^2 + \left(1 + \frac{nS_B^2}{S_w^2}\right) S_B^2 + S_w^2}{np \left(1 + \frac{S_B^2}{S_w^2}\right)} \quad (28)$$

$$\hat{\sigma}_{Tol}^2 = \frac{\hat{\sigma}_m^2 + nS_B^2 + S_w^2}{np} \quad (29)$$

From the classic theory of ANOVA models for random effects, we know that

$\frac{nS_B^2 + S_w^2}{nP}$  is an estimator of the uncertainty (variance) of the overall mean (or bias)

$\hat{\sigma}_m$  when a nested design (with p conditions for experiments and n replicates within each condition) is envisaged as it is here. The above equation can be simplified as follows:

$$\sigma_{Tol}^2 = \sigma_m^2 + \sigma_{\sigma_m}^2 \quad (30)$$

Where  $\sigma_{\sigma_m}^2$  is the estimated uncertainty (variance) of the estimated bias  $\sigma_m$ . So the above equation shows that the variance used for computing the  $\beta$ -expectation tolerance interval is equal to the sum of the total variance of the method plus the variance of the bias. As long as the “between-condition” are the same (laboratories or days), it is clear that equation 29 & 30 account for the same sources of uncertainties: the estimated total variance plus the estimated variance of the bias, meaning that it is then possible to use the standard deviation of the  $\beta$ -expectation tolerance interval as an estimate of the standard uncertainty in the measurements [4, 53, 54].

All of our contributory error distributions have been converted, be they normal, rectangular or triangular, into equivalent normal distributions and combined them by one means or another. The combined standard uncertainty is equivalent to one standard deviation of a normal distribution. As it is known that 1 standard deviation of a normal distribution covers 68.3% of the values in the distribution. By convention, the true value of a measurement result is taken to lie within the uncertainty limits with a probability of 95%. This is approximately equivalent to 2 standard deviations (actually 1.96). In order to bring our reported measurement uncertainty into line with accepted

practice, it is necessary to multiply it by 2. The factor 2 is known as a coverage factor **k**. We can then add to our report that the confidence level of the quoted uncertainty is approximately 95%. If we wanted to be especially cautious, we could use a coverage factor of 3 to get a confidence level of 99.7%. The standard uncertainty is denoted by the symbol  $\mu$ . Expanded uncertainties are usually denoted by a symbol  $U$ . The concept of uncertainty has been introduced as a suitable measure of quality, so how then do we put it to practical use? Chemists are users of as well as a producer of chemical results. Putting on the 'user' hats it can be seen how an uncertainty value helps to interpret an associated chemical measurement. Supposing that Chemists are responsible for accepting or rejecting batches of a certain material used in a manufacturing process. The decision will be based upon a chemical analysis of the material and one of the criteria for acceptance is that the concentration of compound **X** in the material shall not exceed a specified level. Given that we have a number of reports (i.e. five cases) in front of us from the laboratory. The measured concentration of compound **X** in each of the five cases together with the expanded uncertainty  $U$  ( $k = 2$ ) and the reference value that must not be exceeded is given. Looking at five given cases, let assume that in case (a). We see that the measured value is less than the reference value. The upper extreme of the expanded uncertainty is also less than the reference value. We can therefore safely conclude that the concentration of compound **X** is less than the reference value in case (a) so this particular batch of material can be accepted. In case (b), the measured value exceeds the reference value, as does the lowest extreme of the expanded uncertainty. There is therefore no doubt that the concentration of **X** exceeds the reference value and the batch of material must therefore be rejected [3]. In case (c), although the measured value is less than the reference value, if the expanded uncertainty is taken into account it is possible that the actual concentration of **X** could exceed the reference value. In case (d), the measurement result equals the reference value. Although the measured value does not exceed the reference value, the expanded uncertainty means that the true value of the concentration of **X** could exceed the reference value. Finally, in case (e) the measurement value exceeds the reference value but if the expanded uncertainty is taken into account, it is possible that the true value of the concentration of **X** could be below the reference value. To decide whether to reject or accept these batches, there is a need to know how the reference value was set (e.g.

was it set taking into account the possible uncertainty?) and the ‘criticality’ of the measurement. It is important to discuss such ‘boderline’ cases with the end-user of the data, and agree how they are to be handled, before the measurements are made [3].

A second use for uncertainty values lies in their potential for helping us to improve our experimental procedures. In calculating the uncertainty for a measurement, we will have assembled a list of standard uncertainties for the variables of the measurement model. In order to improve the quality of the measurement, the Chemist must first look at the components of the measurement system contributing to the largest uncertainty. If this is the dominant contribution to the combined uncertainty, then any attempt to improve other aspects of the measurement process will be a waste of time. By attempting to reduce the size of the dominant uncertainty first, the greatest return from the effects of the efforts will then be produced [24].

## CHAPTER 3

### THEORETICAL PRINCIPLES OF VOLTAMMETRY

#### 3.1 Differential pulse voltammetry/polarography

Differential Pulse voltammetry (DPV) is the most universal and frequently used voltammetric measurement mode. It is equally well suited for irreversible and reversible systems and offers a high sensitivity. In digital instruments the excitation signal consists of a staircase-shaped increasing direct potential, to which small square wave pulses with a constant potential (pulse amplitude) are applied in periodic succession. The superimposition is synchronized using the drop time and takes place when the electrode surface no longer changes [55]. Polarography is defined as the technique used when the current-potential curve is recorded by using a liquid working electrode whose surface can be renewed periodically or continuously (e.g. by drops). This includes the classical dropping mercury drop electrode (DME) and the subsequently developed static mercury drop electrode (SMDE). Voltammetry on the other hand includes all methods in which the current-potential measurements are made at stationary and fixed working electrodes (irrespective of their material composition). These include the hanging drop mercury electrode (HMDE), the thin mercury film electrode (TMFE), glassy carbon electrode (GCE) and carbon paste electrodes (CPE).

In the Pulse mode the current is measured twice at each mercury drop, before each pulse and at the end of the pulse time (see figure 9). The difference between the measurements is plotted against the direct potential and produces peak-shaped polarograms, as the change in current is the largest for the potential alterations in the region of the half-wave potential (see figure 10). The formation of this difference also leads to a further reduction of the capacitive current contribution and therefore to an increase in sensitivity, even when compared with determinations by normal pulse polarography.

According to the Ilkovich equation 
$$i_p = \frac{n^2 F^2}{4RT} A \cdot c_a \cdot \Delta E_A \sqrt{\frac{D}{\pi t_p}}$$
, for reversible electrode processes the peak height  $i_p$  in the DP polarograms is proportional to the analyte concentration  $c_a$  and is determined by the amplitude  $\Delta E_A$  of the square wave

pulses as well as by the pulse time  $t_p$  among other factors. The detection limit for determinations by differential pulse polarography is similar to that for square wave polarography at about  $10^{-7}$ - $10^{-8}$  mol/L; however, the decrease in sensitivity resulting from irreversibility is lower. However according to Msangati et al this is not true in reality the latter is more sensitive than the former even though the former is the most frequently used in routine analysis as DPV [55, 56].

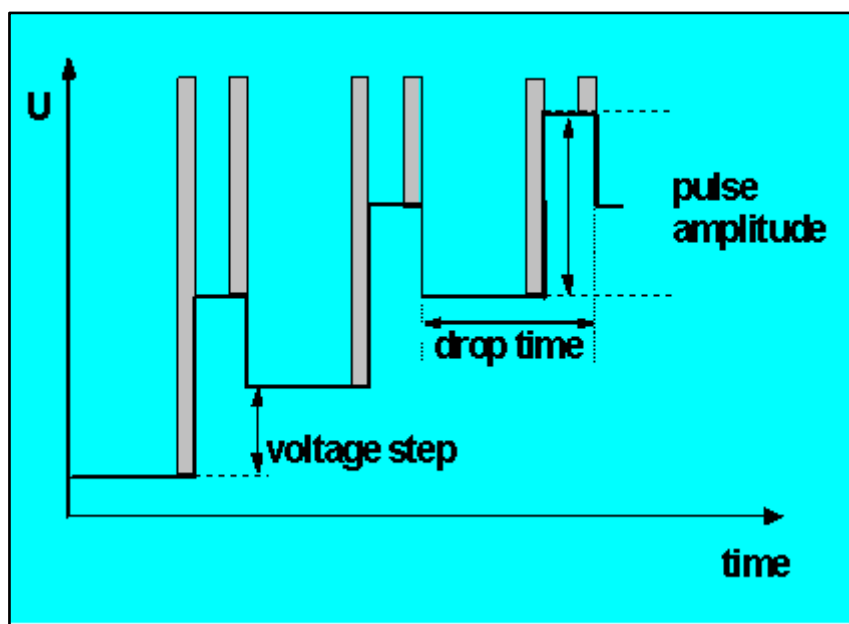


Figure 9: A typical DPV measurement process whereby the current is measured twice at each mercury drop, before each pulse and at the end of the pulse time. The difference between the measurements is plotted against the direct potential and produces peak-shaped polarograms, as the change in current is the largest for the potential alterations in the region of the half-wave potential.

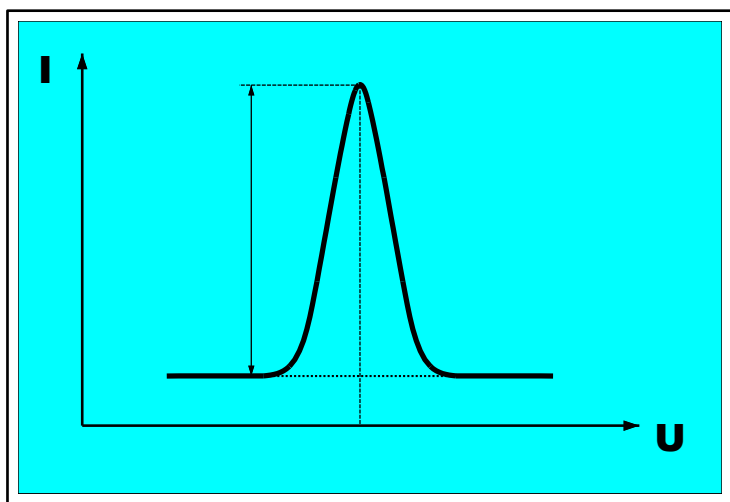


Figure 10: Typical current-potential curve expected for a DP measurement

### 3.1.1 Voltammetry

Electrochemical techniques are also suitable for the determination of trace metals in sea water and voltammetry is the most common method [57, 58]. In sea water, trace metals often occur at low concentrations ( $<10^{-8}\text{M}$ ) and in case these concentration are below the detection limit of the technique used, a pre-concentration step is required prior to analysis. Moreover, the pre-concentration step isolates the metal from the matrix and thereby enhances the selectivity of the analysis [58, 59]. Voltammetry is based on the measurement of a current response as a function of the potential applied to an electrochemical cell. In stripping voltammetry a pre-concentration step and pulse measurement techniques that generate an extremely high signal-to-background ratio are combined with a stripping step, thereby enhancing sensitivity and selectivity. During the pre-concentration step, the trace metal of interest is collected onto or in a working electrode and during the stripping step the collected metal is oxidised or reduced back into solution.

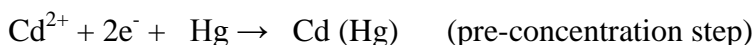
The stripping voltammetric techniques suitable for determination of ultra-low levels of trace metals in sea water include anodic stripping voltammetry (ASV) and adsorptive cathodic stripping voltammetry (AdCSV). Important advances have been made during the past 25 years in the application of stripping voltammetry to marine trace metal measurements. The strength of stripping voltammetry is in its extremely low detection limits ( $10^{-10}$ - $10^{-12}\text{M}$ ), its multi-element and speciation capabilities and its suitability for on-line, ship-board and in-situ applications. Analytical developments

have resulted in our ability to determine a wide range of trace metals in sea water (over 20 metals), and the instrumentation has been computerised, automated and become portable. Two electrode systems, the mercury film electrode and hanging mercury drop electrode are the electrodes that have traditionally been used to determine the extent of interaction between metal ions and particles, since in natural water conditions the measurements that can be made without previous filtration, or centrifugation that can cause laboratory artefacts [59, 60]. However, besides the measurements in the water column, determination at the sediment water interface and in the sediment cores are of great interest, since sediments play an important role releasing or trapping vital or toxic compounds depending on the environmental conditions. For these determinations, microelectrodes (characteristic dimension  $<500\mu\text{m}$ ) may be very useful, since due to their size they can be used in extremely small volumes. The coupling of micro-electrodes to stripping methods may present several advantages for speciation studies of metal ions in natural waters [60]. Despite the excellent performance of mercury electrodes, however, future regulations and occupational health considerations may severely restrict (or even ban) the use of mercury as an electrode material because of its high toxicity. New alternative electrode materials are, therefore, highly desired to develop “environmentally friendly” stripping sensors suitable for on-site monitoring of heavy metals. It was reported in the work done by S. Legeai et al [59] that Bismuth can be used as an alternative to mercury because it was observed that it can be compared favourably with that of mercury electrodes, with its attractive properties including high sensitivity, well defined stripping signals, large cathodic potential range, and insensitive to dissolved oxygen in contrast with mercury electrodes [59]. In this work the above mentioned modified micro electrode was not used because we wanted to develop methods that are ready to be validated and being incorporated into an SOP (case study). For future purposes Bismuth and Stannum are possible alternatives to mercury electrodes, one of their advantage is that they are insensitive to oxygen i.e. no reoxidation takes place in the cell, therefore nitrogen becomes unnecessary which minimises the cost of buying a gas and at the same time being environmentally friendly. Therefore, the developed methods using mercury film will be compared with the methods developed using bismuth film and Stannum film electrodes.



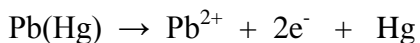
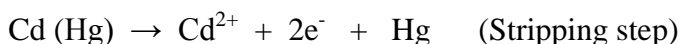
### 3.2 Anodic Stripping Voltammetry

Anodic stripping voltammetry (ASV) involves three main steps: electro deposition (pre-concentration), equilibration and stripping. Its remarkable sensitivity is attributed to the deposition (pre-concentration) step in which metal ions in the sample solution are reduced at negative potential and concentrated as amalgams into a mercury electrode:



After a predetermined time, the stirring of the solution is turned off. The solution is allowed to become quiescent and the concentration of the metal in the amalgam to reach uniformity. The rest period extends for about 30sec, during which the applied potential remains unchanged, thus ensuring that no reoxidation of the metal by traces of oxygen takes place. The electro-deposition current decreases during the rest period [57, 58].

The amalgamated metals are measured in the third (stripping) step by applying a positive potential scan and measuring the peak currents produced as the system reaches the oxidation potentials of the metals. During this scan, the amalgamated metals are re-oxidized, stripped out of the electrode and current flows:



Stripping analysis permits extremely sensitive determination of trace metals in the environment. As a result of its inherent sensitivity, it is one of the few techniques that are sufficiently sensitive to determine heavy metals directly in natural waters [61]. The deposition step makes stripping analysis much more sensitive than direct polarographic techniques, because the concentration of the metal in the mercury is 100-1000 times greater than that of metal ion in the sample solution [58, 61].

### 3.3 Standard Addition

The method of standard addition is used in instrument analysis to determine concentration of substance (analyte) in an unknown sample by comparison to a set of samples of known concentration, similar to using an external calibration curve. Standard addition can be applied to most analytical techniques and is used instead of a calibration (external) curve to solve the matrix effect problem. The matrix effect problem occurs when the unknown sample contains many impurities. If impurities present in the unknown sample interact with the analyte to change the instrumental response, then a calibration curve based on pure analyte samples will give incorrect determination. One way to solve this problem is to use standard addition method. The standard solution (solution of known concentration of analyte) is added to the known solution, so any impurities in the unknown are accounted for in the calibration [57]. For the calibration of the linear standard addition curve, the parameters (**a** and **b**) of the linear regression curve  $y = a + bx$  are calculated. The parameters (**a** & **b**) has the following meaning:

**a** = Y regression/offset – intercept of standard addition curve.

**b** = Slope – slope of the standard addition curve.

Then the concentration of the analyte can be calculated by:- Concentration =  $a/b$

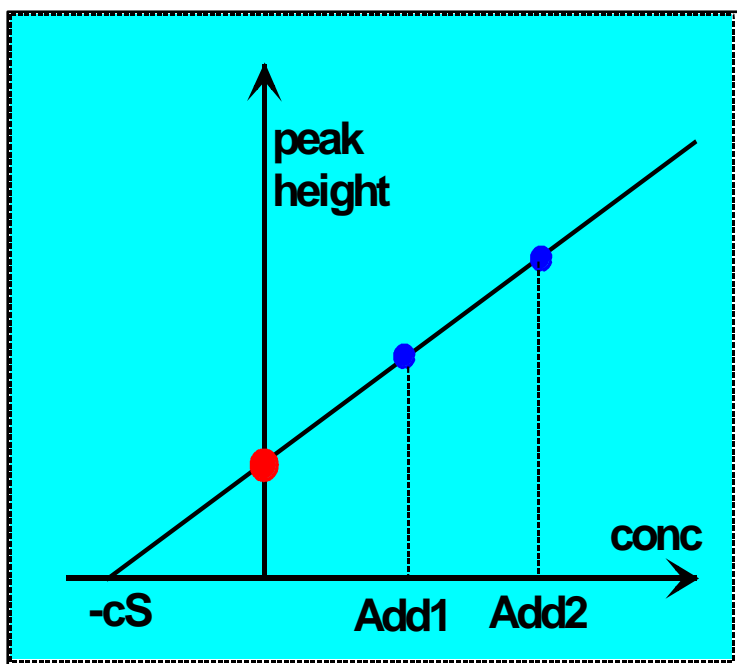


Figure 11: Typical standard addition curve

### 3.4 Rotating disc electrode (RDE)

The limited anodic potential of mercury electrodes has precluded their utility for monitoring oxidizable compounds. Accordingly, solid electrodes with extended anodic potential windows have attracted considerable analytical interest. Of the many different solid materials that can be used as working electrodes, the most often used is carbon, platinum, and gold. Silver, nickel and copper can also be used for specific applications. An important factor in using solid electrodes is the dependence of the response on the surface state of the electrode. Accordingly, the use of such electrodes requires precise electrode pre-treatment and polishing to obtain reproducible results. The nature of these pre-treatment steps depends on the materials involved. Mechanical polishing (to a smooth finish) and potential cycling are commonly used for metal electrodes, while various chemical, electrochemical, or thermal surface procedures are added for activating carbon-based electrodes. Unlike mercury electrodes, solid electrodes present a heterogeneous surface with respect to the electrochemical activity. Such surface heterogeneity leads to deviations from the behaviour expected for homogenous surfaces. Solid electrodes can be stationary or rotating, usually in a planar disk configuration. Such electrodes consist of a short cylindrical rod of the electrode material embedded in a tightly fitting tube of an insulating material (Teflon, Kel-F etc.).

It is essential to avoid crevices between the sleeve and the electrode materials, and thus to prevent solution creeping (and an increase in background response). Electrical contact is made at the rear face. The RDE provides an efficient and reproducible mass transport and hence the analytical measurement can be made with high sensitivity and precision. Such well-defined behavior greatly simplifies the interpretation of the measurement. The convective nature of the electrode results also in very short response times. The detection limits can be lowered via periodic changes in the rotation speed and isolation of small mass transport-dependent currents from simultaneously flowing surface-controlled background currents [58, 61].

### 3.4.1 Glassy-Carbon Electrodes

Glassy (or “vitreous”) carbon (GC) has been very popular because of its excellent mechanical and electrical properties, wide potential window, chemical inertness (solvent resistance), and relatively reproducible performance (see figure 12). The material is prepared by means of a carefully controlled heating program of a premodeled polymeric (phenol-formaldehyde) resin body in an inert atmosphere. The carbonization process is carried out very slowly over the 300-1200° C temperature range to insure the elimination of oxygen, nitrogen, and hydrogen. The structure of glassy carbon involves thin, tangled ribbons of cross-linked graphite-like sheets. Because of its high density and small pore size, no impregnated procedure is required. However, a surface pre-treatment is usually employed to create active and reproducible glassy-carbon electrodes and to enhance their analytical performance. Such pre-treatment is usually achieved by polishing (to a shiny “mirror-like” appearance) with successively smaller alumina particles (down to 0.05µm) on a polishing cloth. The electrode should then be rinsed with deionised water before use. Additional activation step, such as electrochemical, chemical, heat, or laser treatments, have also been used to enhance performance. The improved electron-transfer reactivity has been attributed to the removal of surface contaminants, exposure of fresh carbon edges, and an increase in the density of surface oxygen groups (which act as interfacial surface mediators). Several reviews provide more information on the physical and electrochemical properties of glassy carbon electrodes [57, 62, and 63].

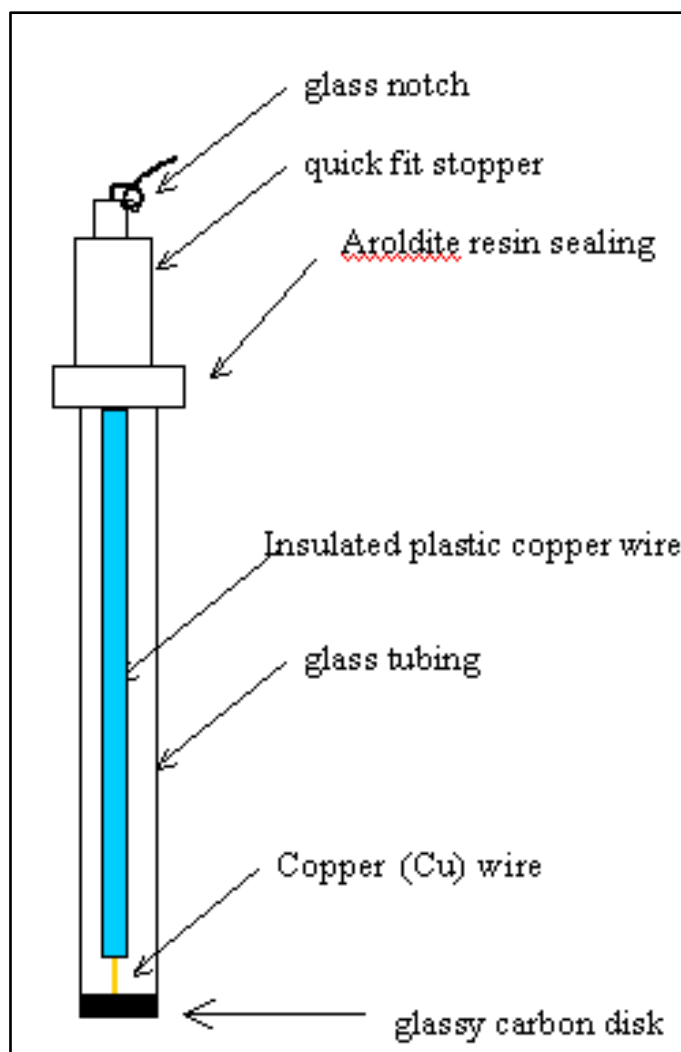


Figure 12: Schematic diagram of a glassy carbon electrode

### 3.4.2 Metal Electrodes

There is a wide choice of noble metals that is available, platinum and gold have proven to be the most widely used metallic electrodes. Such electrodes offer very favorable electron-transfer kinetics and a large anodic potential range. By contrast, the low hydrogen overvoltage at these electrodes limits the cathodic potential window (from -0.2 to -0.5 regions, depending upon the pH). The high background currents associated with the formation of surface-oxide or adsorbed hydrogen layers are more problematic. Such films can also strongly alter the kinetics of the electrode reaction, leading to irreproducible data [62]. Compared to platinum electrodes, gold electrodes are more inert, and hence are less prone to formation of stable oxide films or surface

contamination. Gold electrodes are also widely used as substrates for self-assembled organosulfur monolayers or for stripping measurements of trace metals. Other metals, such as copper, nickel, or silver have been used as electrode materials in connection specific applications, such as the detection of amino acids or carbohydrates in alkaline medium (copper and nickel) and of cyanide or sulphur compounds (silver). Unlike platinum or gold electrodes, the copper electrode offers a stable response for carbohydrates at constant potential [62]

### **3.4.3 Mercury Film Electrode**

The mercury film electrode, used for stripping analysis or flow amperometry, consists of a very thin (10-100 $\mu$ m) layer of mercury covering a conducting support. Because of the adherent oxide films on metal surfaces, and the interaction of metals with mercury, glassy carbon is most often used as a substrate for the MFE. The mercury film formed on a glassy carbon support is actually composed of many droplets. As a result of not being a pure mercury surface, such film electrodes exhibit lower hydrogen over voltage and higher background currents. Another useful substrate for the MFE is iridium (because of its very low solubility in mercury and the excellent adherence of the resulting film). Mercury film electrodes are commonly pre-plated by cathodic deposition from a mercuric nitrate solution. An in-situ-plated MFE is often employed during stripping analysis. This electrode is prepared by simultaneous deposition of the mercury and the measured metals. Most commonly, a disk-shaped carbon electrode is used to support the mercury film. Mercury film ultra microelectrodes, based on coverage of carbon fiber or carbon micro disk surfaces, have also received increasing attention in recent years [58].

### **3.4.4 Chemically modified Electrodes**

Chemically modified electrodes (CMEs) represent a modern approach to electrode systems. These rely on the placement of a reagent onto a surface, to impart the behaviour of that reagent to the modified surface. Such deliberate alteration of electrode surfaces can thus meet the needs of many electroanalytical problems, and may form the basis for new analytical application and different sensing devices [62]. An example of such an application is the work done by Msangati et al [55, 56]. There are various ways in which CMEs can benefit analytical applications. These include

acceleration of electron-transfer reactions, preferential accumulation, or selective membrane permeation. Such steps can impart higher selectivity, sensitivity, or stability to electrochemical devices. Many important applications, such as electrochromic devices, controlled release of drugs, electro synthesis, and corrosion protection, should also benefit from the rational design of electrode surfaces [62]. One of the most common approaches for incorporating a modifier onto the surface has been the coverage with an appropriate polymer film. Polymer modified electrodes are often prepared by casting a solution containing the dissolved polymer onto the surface and allowing the solvent to evaporate, or via electro-polymerization in the presence of the dissolved monomer. The latter offers precise control of the film thickness and is particularly attractive in connection with miniaturized sensor surfaces [62].

## CHAPTER 4

### GENERAL EXPERIMENTAL WORK AND DESIGN.

#### 4.1 Laboratory Experimental Work (Instrumental)

##### 4.1.1 Apparatus (Case study 1)

Polarographic measurements were conducted by using a Metrohm model 797 VA Computrace (Herisau, Switzerland) with a 797 VA electrode stand with a multimode electrode in a static mercury drop electrode (SMDE) mode. All experiments were carried out in a three compartment, three electrode cell. The reference electrode was Ag/AgCl, saturated with 3M KCl and the auxiliary electrode was a platinum wire. Solutions were stirred during the purging by a rotating polytetrafluoroethylene (PTFE) rod. All sample solutions were stored mainly in low density polyethylene (LDPE) and polystyrene flasks. The available software provided with the equipment allows automatic successive signal trace (e.g. standard addition mode) treatment (i.e. 'peak' base line assignment, 'peak' height calculation) and concentration estimation (an example can be found in [32]).

##### 4.1.2 Apparatus (Case study 2)

Voltametric measurements were conducted by using a Metrohm model 797 VA Computrace (Herisau, Switzerland) with a 797 VA electrode stand with a rotating disc electrode (glassy carbon) were used for the voltammetric measurements. All experiments were carried out in a three compartment, three electrode cell. The reference electrode was Ag/AgCl, saturated AgCl, 3M KCl and the auxiliary electrode was a platinum wire. Obtained scans were evaluated with the Metrohm 797 VA potentiostat. All sample solutions were stored in the fridge at 4°C mainly in low density polyethylene (LDPE) and polystyrene flasks before analysis. The available software provided with the equipment allows automatic successive signal trace (e.g. standard addition mode) treatment (i.e. 'peak' base line assignment, 'peak' height calculation) and concentration estimation an example can be found in [14].



#### **4.1.2.1 Reagents and reference solutions (Case study 1)**

All solutions were prepared by using Milli-Q quality water (Millipore, Randburg, South Africa). Analytical grade phenol and o-nitro phenol were purchased from Aldrich Sigma (Aston Manor, South Africa). 98% H<sub>2</sub>SO<sub>4</sub> was purchased from Fluka, (Buchs, Switzerland). Working solutions of 200 mg.L<sup>-1</sup> o-nitro phenol and 0.1 M phenol were prepared by using deionised water (Milli-Q quality water; Millipore, Randburg, South Africa).

#### **4.1.2.2 Reagents and reference solutions (Case study 2)**

All solutions were prepared by using Milli-Q quality water (Millipore, Randburg, South Africa). The reagents used were: 1000 mg.L<sup>-1</sup> ( $U = 2$ ,  $k=2$ ; see appendix 2) Pb and Cd Certified reference material (CRM) from Fluka (Buchs, Switzerland), NIST-BAM-CRM traceable, Analytical grade 32% Hydrochloric acid from Fluka, (Buchs, Switzerland), 98% H<sub>2</sub>SO<sub>4</sub> from Fluka, (Buchs, Switzerland), Analytical grade mercury (II) nitrate monohydrate from Merck (Darmstadt, Germany).

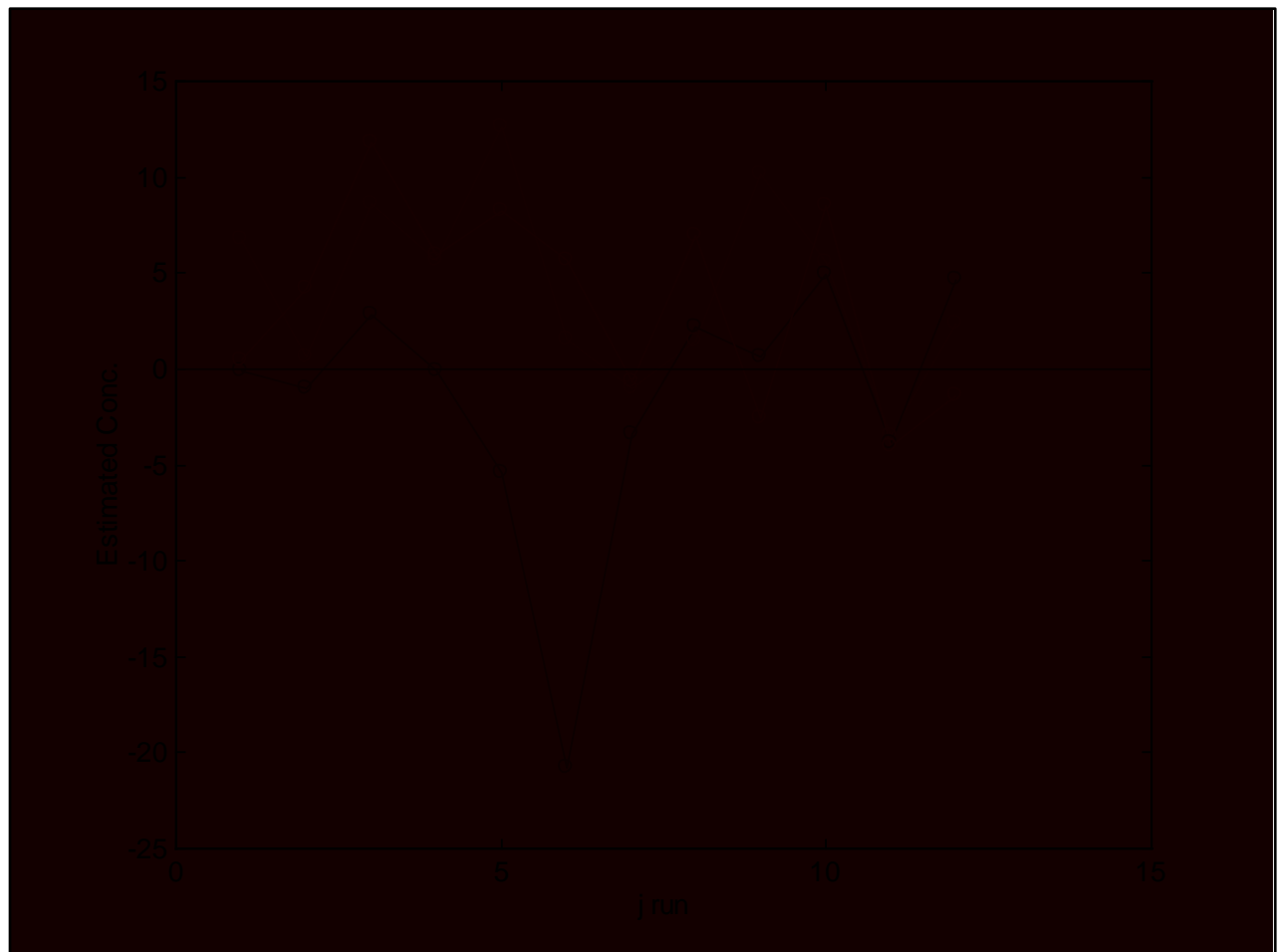
### **4.2 Experimental Design**

#### **4.2.1 Pre-validation of the method and standardization of the signal**

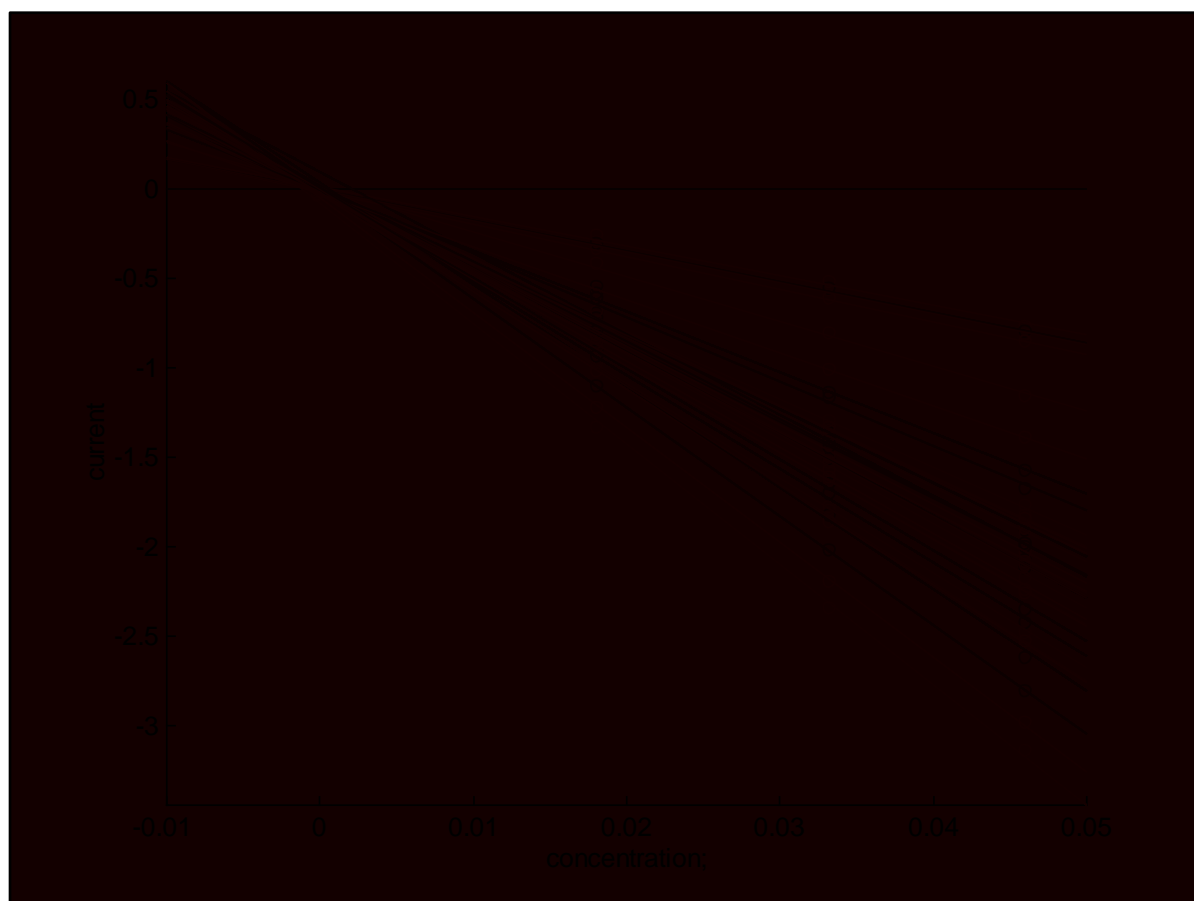
Sample matrices including Tap water, Mineral water and an industrial treated waste-water outer fall (Umhlathuzana) were used for the pre-validation study. The standard addition procedure [32] was followed, A modification of the protocol published by Metrohm [32] was followed. During the analysis 1 mL of the sample, 1 mL of 0.1 M phenol and 4 mL of 98% sulphuric acid was were transferred into a voltammetric cell. After cooling, 4 mL of water was carefully added to the voltammetric cell, and de-aerated by purging for 3 min with high purity inert nitrogen gas. Following the signal obtained, a single standard addition of 0.25 mL of a 200 mg.L<sup>-1</sup> o-nitro phenol solution was added to the voltammetric cell and a second signal was recorded. In all cases, independent replicate assays were performed. The  $X_{3 \times 12}$  Experimental design was used for the pre-validation study, a run was considered as a session as 3 replicates per day for 12 days. The data analysis revealed four main observations: (i) the occasional presence of atypical concentration estimates (an anticipated situation) (see figure 13). Signal inspection showed some 'peak' shape shifts in the voltage domain, between sample and successive standard additions and

between-replicates (see figure 14), (ii) the (provisional) repeatability effect was larger than the between-run imprecision ( $RSD_r > RSD_{run}$ ). (iii) Similar signal behaviour was observed between different types of samples, so matrix effect could be neglected. (iv) Sample mean concentrations (in the 6 – 10 mg.L<sup>-1</sup> range) were within the 5 – 50 mg.L<sup>-1</sup> EC directive expected concentration interval.

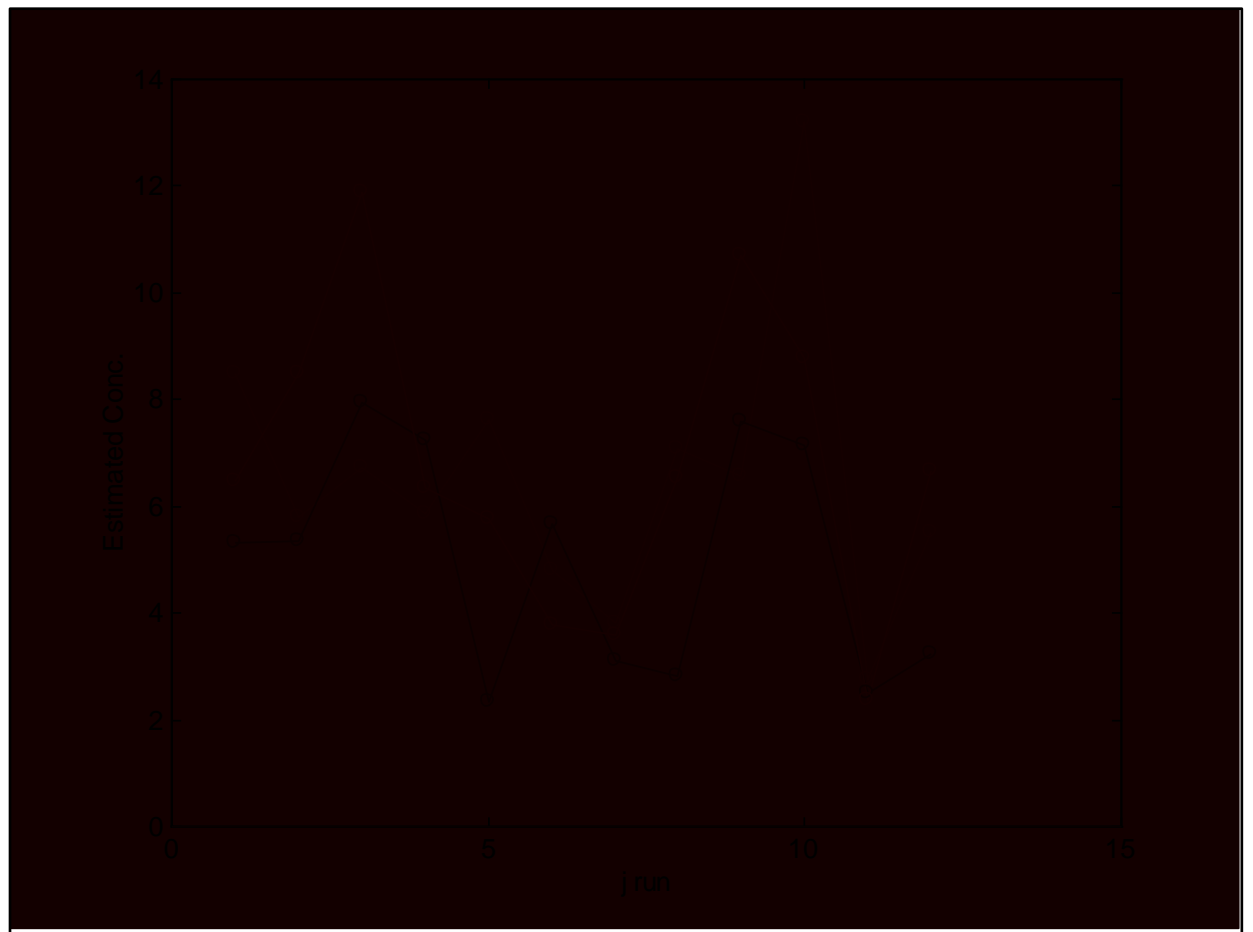
Some decisions were adopted in view of adjusting the validation protocol. The possibility of examining each signal and/or its mathematical treatment was considered not practical; mainly because the difficulties encountered prevented the introduction of the protocol into an SOP to be audited and the increase of analysis times for routine work. Therefore, the automatic calculation (equipment software) was preferred for the subsequent validation protocol. This means that since the standard addition technique was employed the instrument software automatically calculates the concentration of the sample for every standard addition. Due to signal variability, the need of three o-nitro phenol standard additions was evaluated.



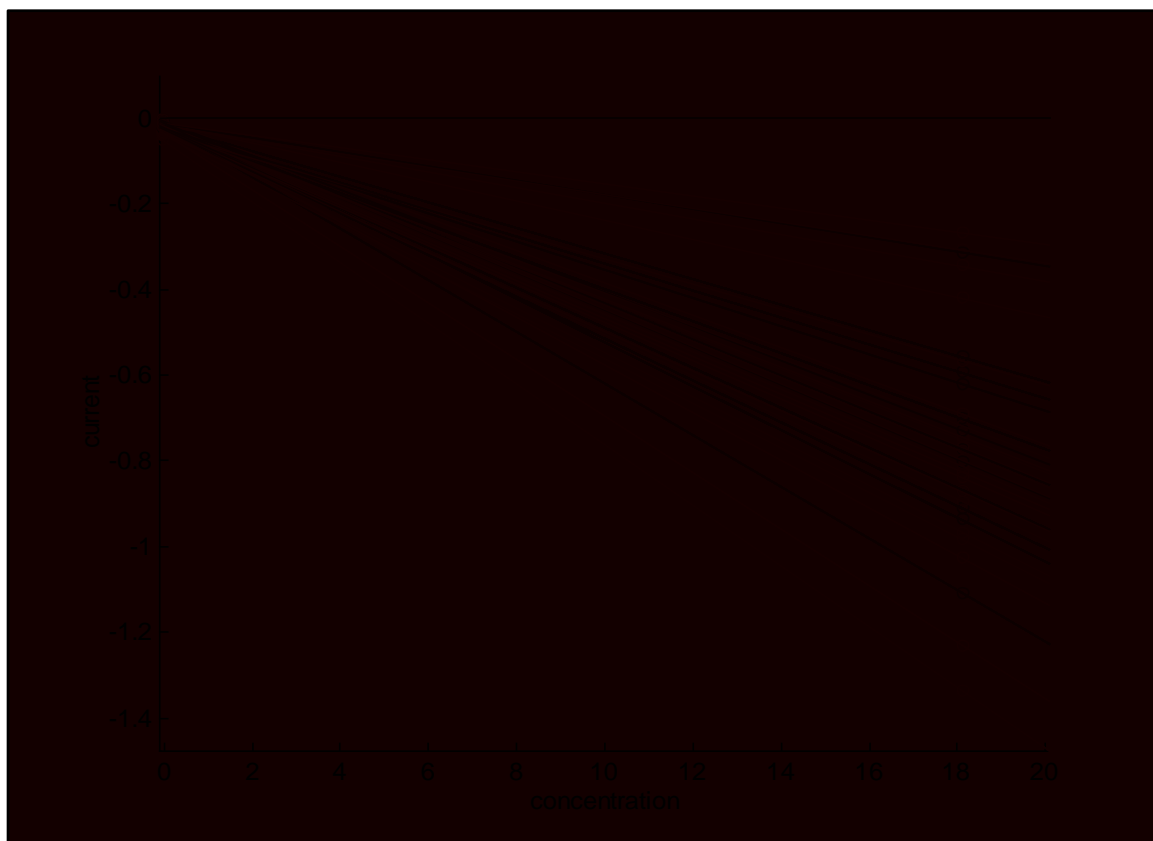
**Figure 13: The 3-point graph obtained for the estimated concentration in 12 runs which shows the signal variability and concentrations.**



**Figure 14:** The regression graph obtained for 3-point calibration curve using the Mathlab's regression tool which shows the signal and concentration variability.



**Figure 15:** Estimated concentration obtained in 12 runs using a 1-point approach which shows no concentration and signal variability.



**Figure 16: Regression graph obtained for a 1-point calibration plot which shows no variation in the signal obtained.**

The one-point calibration results agrees with those obtained by using the 3-point calibration curve (in the three samples data matrices) by using only the first standard addition signal, the obtained results were even slightly better than using the 3-standard addition signals (see figure 15 and 16). When using the Metrohm software and removing the first and second standard addition, the results are similar to the above results obtained by using the Matlab's regression tool to validate the pre-validation model. The advantage is that the instrument's internal software calculates the concentration of each standard added in the cell taking into account the dilution factors, this might not be confused with an external calibration curve where the concentration is measured by extrapolation of the analyte signal v.s. calibration standard concentration. The standard addition in this sense is different from the external calibration curve to an extent that the sample is measured and following the sample measured to the same solution a certain concentration of the standard is added as an addition and the analyte concentration is obtained from that addition instead of measuring a series of standards.

Therefore, the ‘1-standard addition’ approach was selected and used for the entire work. It provides half analysis time per sample (a practical advantage fitting the internal requisites. The 1-point calibration approach is suitable for routine laboratory environment because it is not time consuming, minimal use of the reagents, minimal waste etc.

#### 4.2.2 Experimental design: Monte Carlo Simulation

Uncertainty calculations are based on Monte Carlo calculation. A bivariate plot is presented for assessing whether the uncertainty interval for the bias  $[E \pm U(E)]$  and intermediate precision  $(RSD_i \pm U)$ ,  $(RSD_i)$  are included within prefixed limits. This plot facilitates visual interpretation even for unqualified laboratory staff.

It was reported in by Sagrado et al [13], that the  $X_{2 \times 12}$  experimental design for few days, used as a starting experimental design provides acceptable results for most analytical methods. Simulation could also help to find more appropriate designs i.e. increasing the data matrix with new validation results or using preliminary validation data as initial estimates (pre-validation stage) for simulating more complex designs. In order to statistically compare the proposed  $U$ , approach [12], with other previously proposed approaches [44]. A computer application is programmed to perform Monte Carlo simulations (here, independent validations). This guarantees that the assumptions on the data quality are satisfied. The first step involves generating the validation  $X_{n_r \times n_s}$  data set. The model proposed by Kuttatharmakul [44] was followed. It assumes normally distributed validation results, consistent with the acceptable true values ( $U_o$ ,  $E_o$ ,  $RSD_{ro}$ , and  $RSD_{runo}$ , the zero in the subscripts is used to differentiate the true values from the estimated ones) and uses the RAND function in MATLAB [64]:

$$x_{ij} = \mu_o(1+E_o/100) + f_j + e_{ij} \quad (31)$$

Where  $x_{ij}$  is the result related to the  $i$ th replicate of the  $j$ th run (i.e. a batch characterized by a given day, analyst, and equipment),  $E_o$  represents the true bias in percentage,  $f_j$  is the random run effect for the  $j$ th run ( $f_j \approx N(0, \sigma_{run}^2)$ ), where  $\sigma_{run}^2$

corresponds to  $RSD_{run_o}$ , and  $e_{ij}$  is the random error under repeatability conditions for the  $i$ th replicate and  $j$ th run ( $e_{ij} \approx N(0, \sigma_r^2)$ ), where  $\sigma_r^2$  corresponds to  $RSD_{r_o}$ ). From  $X_{n_r \times n_s}$  the accuracy statistics (simulating the results from one validation): error in percentage,  $E = 100(\mu - \mu_o)/\mu_o$  and the relative standard deviation in intermediate precision conditions as a percentage,

$$RSD_i = (RSD_r^2 + RSD_{run}^2)^{0.5}. \quad (32)$$

Finally, a number of simulations (for example  $10^4$  different validations) can be defined by the user to obtain uncertainty estimates related to  $E$  and  $RSD_i$ . The expanded uncertainty interval ( $\pm U$ ;  $CL \approx 95\%$ ) can be estimated from two percentiles of the simulated results of a given statistic [12]. Real water samples could be used for validation purposes if they have been previously been assayed by an alternative (i.e. reference) method or they come from a collaborative study. In these cases, matrix influence could affect uncertainty estimations, and if possible, such contribution should be estimated [23].



## CHAPTER 5

### CASE STUDY 1

#### 5.1 Purpose statement

The main aim is to propose a step-by-step protocol to evaluate and harmonize internal quality aspects of a method. Such criterion involves a statement of method's requisites (external and internal), method's 'fit-for-purpose' features selection, pre-validation to adjust the validation protocol, accuracy validation and assessment, harmonization of the validation-verification-control-uncertainty process, validation of other required method's features, validity statement in terms of 'fit-for-purpose' decision and short-term routine work.

#### 5.2 Mixture Reference solution

A certified 1000 mg.L<sup>-1</sup> Sodium nitrate (NaNO<sub>3</sub>) solution (1001 ± 5 mg.L<sup>-1</sup> Merck, Darmstadt, Germany; traceable to SRM from NIST) was used. From it a CRM solution was prepared by pipetting 5 (± 0.015) mL and diluting into a 100 (± 0.08) mL volumetric flask. Strictly speaking, this provides a 50.05 mg.L<sup>-1</sup> solution (from this particular CRM batch); however, in order to have a long-term value to be used with future CRM batches, the accepted reference value was set to  $\mu_{\text{CRM}} = 50 \text{ mg.L}^{-1}$ . The combined standard uncertainty ( $u_{\text{CRM}}$ ) was estimated by the bottom-up approach [13]. That means from the standard uncertainty of the original certified solution  $(5/(3))^{0.5}$ ; assuming rectangular distribution, since the limits do not specify a level of confidence). In addition the terms corresponding to the volumetric material used for pipetting and dilution, according to the information from their calibration certificates  $(0.015/(6))^{0.5}$  and  $0.08/(6)^{0.5}$ , respectively, assuming triangular distribution as ISO recommends), were calculated as follows:  $u_{\text{CRM}} = 50 \times [(5/(3))^{0.5}/1000]^2 + (0.015/(6))^{0.5}/5)^2 + (0.08/(6))^{0.5}/100)^2]^{0.5} = 0.1576 \sim 0.16 \text{ mg.L}^{-1}$ .

### 5.2.1 Samples

Sample matrices including Tap water, Mineral water and an industrial treated waste-water outer fall (Umhlathuzana) were used for the pre-validation study. Two sample matrices were used to illustrate the routine work protocol included Umhlathuzana industrial treated waste-water outer fall, and New Umbilo treated sewerage waste-water outer fall.

### 5.2.2 Sample analysis Procedure

A modification of the protocol published by Metrohm [32] was followed. During the analysis 1 mL of the sample, 1 mL of 0.1 M phenol and 4 mL of 98% sulphuric acid were transferred into a voltammetric cell. After cooling, 4 mL of water was carefully added to the voltammetric cell, and de-aerated by purging for 3 min with high purity inert nitrogen gas. Following the signal obtained, a single standard addition of 0.25 mL of a 200 mg.L<sup>-1</sup> o-nitro phenol solution was added to the voltammetric cell and a second signal was recorded. In all cases, independent replicate assays were performed.

### 5.2.3 Accuracy validation

The procedure used was similar to the sample analysis described in section (5.2.2), except that the sample was substituted by the CRM solution ( $\mu_{\text{CRM}} = 50 \text{ mg.L}^{-1}$ ). This procedure was repeated over several runs. For the present case study (see section 5.2.5), a run was considered a validation session in the morning (performed by one analyst) or in the afternoon (performed by a second analyst), so a maximum of 2 runs per day can be achieved. Single equipment was used and turned off between runs. This procedure is also valid for the accuracy pre-validation experiment.

### 5.2.4 Verification

Similar procedures were used as described in sample analysis of section (5.2.3), except that the sample was substituted by two 'verification solutions' (the CRM solution of 50 mg.L<sup>-1</sup>, used as high-level verification solution and a 5 mg.L<sup>-1</sup> from 1/10 diluted CRM solution, used as low-level verification solution) used in alternate working routine sessions. This procedure was performed at the start of each session.

### **5.2.5 Internal Quality Control (IQC)**

The procedure used was the same as that in sample analysis described in section (5.2.2), except that the sample was substituted by a ‘control solution’ (the CRM solution of 50mg.L<sup>-1</sup>). It was performed approximately every 2 hours along the sample analysis periods, in each working routine session.

### **5.2.6 Limit of detection (LOD) estimation**

The procedure used was similar to the sample analysis described in section (5.2.2), except that the sample was substituted by ten independent nitrate solutions of 1 mg.L<sup>-1</sup> (1/50 diluted CRM solution; ‘lowest acceptable concentration’ used as blank [2]).

## **5.3 Results and discussion**

A step-by-step protocol was designed for harmonizing quality control tasks in testing laboratories. Obviously, some decision-making aspects could depend on the method’s purpose and features. As a case study, the protocol was adjusted for a differential-pulse polarographic method to perform an indirect determination of nitrate using o-nitro phenol (the ‘candidate method’) in water samples. Since the method was adapted from an application bulletin (from the equipment supplier) [32] it should be considered as an internal Standard Operation Procedure (SOP), which requires validation according to the standard ISO 17025 [38]. The protocol described in the sections (5.2.2 to 5.2.6), could be adapted to other methods. The aim is to harmonize the work of research teams and routine laboratories, by taking into account that the different aims, strategies and practical viewpoints exist. According to the following points the protocol could be seen as: A starting point, and, necessarily, definitive (harmonized) protocols must be established by international normalisation/accreditation entities. This work is a contribution to show possible solutions to this problem (lack of harmonization). Evidently, the protocol should be tested in the future with more methods in order to check its general applicability. However, parts of the protocol have already been used satisfactorily in some previous studies related to water analysis methods exhibiting different features and requisites [13, 23, 43]; as well as in this paper (section vii) with a case study, quite different from the previous ones.

### 5.3.1 External and internal method requisites and method's features

Currently, the accepted philosophy in method validation involves adjustment of this stage to the 'fit-for-purpose' concept, resulting in a comparison between requirements and estimated method's features. Therefore, the laboratory must decide on external (legislation, client) and/or internal (prefixed by the laboratory) requisites for the candidate method. This allows deciding which method's features need to be validated (e.g. the 'fit-for-purpose' features). Accuracy (in intermediate precision conditions), should be considered as a mandatory aspect, so the limits for bias and intermediate relative standard deviation (e.g.  $E_{\text{LIM}}$  and  $RSDi_{\text{LIM}}$ ; see appendix 2) should be fixed in this step. All previous information available on the method's features could be used to anticipate the possibilities of the candidate method.

### 5.3.2 Accuracy pre-validation

Realistic internal method's accuracy (bias and precision) validation suggests a performance of the validation experiments under intermediate precision conditions, e.g. designing an  $\mathbf{X}_{Nr \times Ns}$  ( $Nr$  replicates  $\times$   $Ns$  runs; appendix 2) data matrix, whose analysis allows accuracy estimation [44]. This implies a definition of which factors (time periods, analyst, and equipments) to be changed between-runs (ideally, the same expected between working routine sessions or between-batches in a testing laboratory (appendix 2). Certified reference materials (CRMs) are regarded as among the best available references for validating the accuracy of a method [44], resulting in the accepted true values of analyte contents. A 'pre-validation study' [45] could give provisional estimates of method's accuracy (precision and, if e.g. a CRM is used, bias).  $\mathbf{X}_{2 \times 12}$  could be a reasonable starting point [13], if atypical data are not expected; otherwise,  $Nr > 2$  could be preferable. In some cases, this strategy could constitute the definitive validation stage (e.g. if 'accuracy validation assessment' with these data is to conform to requisites). Alternatively, real samples could be used (instead of e.g. CRM) to obtain a realistic view of the method behaviour in routine conditions (e.g.  $RSDi$ ) or to investigate matrix effect. However,  $E$  could not be estimated at this point.

### 5.3.3 Accuracy validation using a CRM solution

The same principles described for the accuracy pre-validation stages ( $\mathbf{X}_{Nr \times Ns}$  validation matrix, run definition, use of a CRM) are valid. A previous theoretical study suggests that an  $\mathbf{X}_{4 \times 25}$  matrix should provide safe results [23]. However,  $Nr$  and  $Ns$  could be adjusted (i.e. augmented or reduced) if previous information on the method's accuracy is available (e.g. pre-validation study; section ii). It has been demonstrated that the experimental effort can consistently concentrate in the validation stage, i.e. relatively high  $Nr$  and  $Ns$  (but reduced along future routine work, i.e. low  $Nr'$ ) [23]. Accordingly, a relatively large  $\mathbf{X}_{Nr \times Ns}$  experiment, using a CRM solution adjusted, at least, to the critical concentration level, could be performed. A check for outliers is always a recommendable task, particularly when normal distribution statistics are used [12]. Unfortunately, outlier detection is not a harmonized task and different results can be obtained depending on the criterion used. Also, inconsistencies between conventional outliers test (say the Grubbs test) and control plots on the validation data have been described [43]. Finally, some criteria could be difficult to automate. Looking for acceptance (e.g. for auditing purposes) we suggest a widely accepted criterion (i.e. used in proficiency testing, an obligatory task for laboratories under ISO 17025 accreditation) and therefore, well-known by laboratory's staff; the z-score calculation (see appendix 1) on each  $\mathbf{x}_j$  vector of  $\mathbf{X}$ :

$$z_{i,j} = (x_{i,j} - \mu_0) / \sigma_0 \quad (33)$$

Where  $z_{i,j}$  is the z-score for each estimated concentration ( $x_{i,j}$ ) in the  $\mathbf{X}$  validation matrix data.  $\mu_0 = \mu_{\text{CRM}}$  and  $\sigma_0 = RSDr_{\text{LIM}} \cdot \mu_{\text{CRM}} / 100$  could be reasonable assigned values for the validation stage. Normally,  $|z| > 2$  values are considered questionable [13]; so it can be used as an accepted criterion to automatic outlier elimination. However, in our opinion, this matter still deserves more study including harmonization by normalization or accreditation entities.

Most testing laboratories prefer to use accepted criteria instead of statistical criteria for decision-making purposes. For instance, it is common to calculate the method's relative error and to compare it with a prefixed limit than to assess if there are statistical differences between the experimental and the accepted true value (by means of a hypothesis test). In the same way, the z-score approach is not really an

outlier test (e.g. it does not account with the number of degrees of freedom and uses an assigned value for the accepted standard deviation); in contrast, is a generalised acceptance limit for routine laboratories. Moreover, it is simple and easy to automate (e.g. in MATLAB®) to be incorporated into an SOP. Validation statistics can be obtained from the (remaining  $|z| \leq 2$ ) **X** data.  $E$  and  $RSDi$  can be estimated from the run mean ( $\mu$ ) and the precision statistics after ANOVA analysis [44] using the residual mean squares ( $MS_r$ ) and the between-run mean squares ( $MS_{run}$ ); see appendix 1:

$$E = 100 (\mu - \mu_{CRM}) / \mu_{CRM} \quad (34)$$

$$\begin{aligned} RSDi &= 100 s_i / \mu = 100 (s_{run}^2 + s_r^2)^{0.5} / \mu \\ &= 100 ((MS_{run} - MS_r)/Nr + MS_r)^{0.5} / \mu \end{aligned} \quad (35)$$

### 5.3.4 Accuracy validation assessment by means of Monte Carlo simulation

The objectives of method validation should be to obtain not only, estimates of bias and precision, but also to evaluate their confidence limits [42]. A single validation experiment only provides information on bias and intermediate precision estimates, but not on the  $E$  and  $RSDi$  variability. Direct comparison between  $E$  and  $RSDi$  and the prefixed limits ( $E_{LIM}$  and  $RSDi_{LIM}$ ) in deciding on method's fit-for-purpose could be risky [13]. More validation experiments are impractical; however, alternatively, they can be simulated to obtain estimates on bias and precision confidence. In practice, Monte Carlo simulations offer an attractive yet simple alternative to this practice [13, 65]. Simulation could be performed following the model proposed by Kuttatharmmakul [44]. This model is able to generate  $\mathbf{X}_{N_r \times N_s}$  data sets assuming normally distributed validation results, consistent with  $E$ ,  $RSDr$  and  $RSDrun$  estimates by means of the RANDN function in MATLAB [64].

$$x_{ij} = \mu_{CRM} (1 + E / 100) + f_j + e_{ij} \quad (36)$$

where  $x_{ij}$  is the result related to the  $i$ th replicate of the  $j$ th run;  $f_j$  is the random run effect for the  $j$ th run ( $f_j \sim N(0, s_{run}^2)$ ; where  $s_{run}^2$  corresponds to  $RSDrun$ ; see appendix 2) and  $e_{ij}$  is the random error under repeatability conditions for the  $i$ th replicate and  $j$ th run ( $e_{ij} \sim N(0, s_r^2)$ ; where  $s_r^2$  corresponds to  $RSDr$ ; see appendix 2).

The model could be applied to generate say  $10^4$   $\mathbf{X}$  new data matrices. The expanded uncertainty intervals ( $\pm U$ ;  $CL \sim 95\%$ ) can be estimated from two percentiles of the simulated  $E$  and  $RSDi$  estimates [45]. For the  $RSDi \pm U(RSDi)$  uncertainty interval the 2.5%- and 97.5%-percentiles of the  $RSDi$  results can be used. For the  $E \pm U(E)$  uncertainty interval the 5%- and 95%-percentiles of the  $E$  results were used,  $CL \sim 90\%$ , has been recommended [45]. Once uncertainty intervals are estimated it must verify that they are within the limits (e.g.  $|E \pm U(E)| < |E_{LIM}|$  and  $[RSDi + U(RSDi)] < RSDi_{LIM}$ ) in order to declare the validity statement for accuracy [13]. In view of facilitating the decision task to the laboratory staff, the  $RSDi \pm U(RSDi)$  vs.  $E \pm U(E)$  'Validation assessment' plot can be used. In this case, it simply must be verified, graphically, that uncertainty intervals (e.g. error bars) are inside the area defined by  $E_{LIM}$  and  $RSDi_{LIM}$  (lines) into the plot.

### 5.3.5 QA and uncertainty statistics related to accuracy validation estimates

IQC and sample uncertainty statistics (appendix 2) can be harmonized with respect to the accuracy ones according to the u-approach [23]:

$$S_u = u = (s_{run}^2 + s_r^2 / Nr' + s_\mu^2 / Ns + u_{CRM}^2)^{0.5} \quad (37)$$

Note that  $Nr'$  (instead of  $Nr$ ) is used since, the number of replicates must be adjusted to those used in the different routine stages.  $Nr'$  could not (necessarily) be the same during sample analysis, method verification and method control (e.g. control charts). From a practical point of view, it could be modified according to a 'progressive' scheme (an internal procedure in some testing laboratories rather than a normalized practice), consisting of fixing a low  $Nr'$  value and increasing it only when necessary (e.g. punctual high disperse results, samples close or above critical levels, etc.) up to a maximum of e.g.  $Nr$ .

From  $u$ , the expanded uncertainty can be used for *mean* sample concentration estimates (for simplicity, multiplying  $u$  by  $k = 2$  [3, 12]; see appendix 2):

$$U = k u \sim 2 u \quad (38)$$

$S_u$  could be combined with  $\mu_{CRM}$  as centre line to fix the control limits ( $AL$  and  $WL$ ; see appendix 1) of the mean control chart [23].

$$AL = \mu_{CRM} \pm 3 S_u \quad (39)$$

$$WL = \mu_{CRM} \pm 2 S_u \quad (40)$$

Method verification is currently an ambiguous and not a harmonized task. Also, there is little guidance on which acceptance limits should be used. It has been recommended to relate these limits to the process standard deviation [12]. A logical approach should be relating the verification limits to uncertainty:

$$\text{Verification Limit} = \mu_{CRM} \pm U \quad (41)$$



### 5.3.6 Initial routine work

Routine work is mainly related to the sample analysis and IQC tasks, such as control charts. However, ISO 17025 auditors used to demand method verification at the start of a working routine session, before processing the samples. A practical strategy could consist of preparing verification solutions at two extreme concentration levels on the scope range (complementing the accuracy validation scheme, particularly if it is performed at a single concentration), that could be alternated between working routine sessions. Unfortunately, method verification (e.g. IQC) has traditionally been avoided in research papers involving method development.

In our opinion, initial results in the routine work stage, aimed at the pre-examination of the reliability of the overall criteria described in the protocol, should be performed as part of the internal quality harmonized protocol. A plot of experimental results in different stages/periods along few (at least 2) working routine sessions (e.g. method verification, followed by sample analysis and method quality control periods), according to the pre-fixed plan, is proposed. Limits for verification (Eq. 41) and control ( $AL$  and  $WL$ ; Eqs. 40 and 41), as well as the uncertainty intervals for sample results could be included. Obviously, when more IQC data are available (e.g. when it is applied by the testing laboratory in the future), conventional control charts and application of rules to out-of-control situation (preferable harmonized) are recommendable. In that case, typical rules to monitor the method could be applied (e.g. one point above  $AL$  and two consecutive points between  $WL$  and  $AL$ ; where the  $u$ -approach has demonstrated to be consistent [23]). Therefore, the proposed plot should not be confused with a substitution of the control charts.

### 5.3.7 Application of the protocol to a case study

Table 5 illustrates some of the task related to the first step (section i) as well as some decision-making aspects related to the candidate differential-pulse polarographic method. In order to anticipate the possibilities of the method, previous information on it was used. Additional favourable characteristics of the selected methodology [32] are: (i) Voltammetry is a recognised technique with many ISO, EPA, ASTM, AOAC and DIN standards, (ii) versatility (e.g. besides nitrate, quantification of heavy metals, organics and other anions using the same instrument is possible), (iii) small footprint

(i.e. less than a metre), so do not require change to existing laboratory infrastructure or the installation of extra-devices.

**Table 5: Requisite-feature relationships. In footnotes a priori advantages and possible cautions related to the candidate differential-pulse polarographic method, from available/inferred information.**

Requisites	Concentration	Feature	Limits and comments
External (from [39])		Fit-for-purpose	
Bias and precision (batch and between-batch) values of 10%	Parametric level (50 mg.L <sup>-1</sup> )	Accuracy (trueness and intermediate precision)	$E_{\text{LIM}}=10\%$ <sup>a</sup> $RSDi_{\text{LIM}} = (RSDr_{\text{LIM}}^2 + RSDrun_{\text{LIM}}^2)^{0.5} \sim 14\%$ (see appendix 1)
Limit of detection 10% of the parametric level	5 mg.L <sup>-1</sup>	Limit of detection (LOD)	$LOD_{\text{LIM}} = 5 \text{ mg.L}^{-1}$ <sup>b</sup>
Internal		Non fit-for-purpose	
High selectivity		Confirmation of identity (selectivity)	Wide applicability (different types of water matrices and accreditation scope)  No problems related to the ‘confirmation of identity’ feature [34, 45], if it is required by ISO 17025 auditors
Limit of quantification lower than the real sample concentration levels		Limit of quantification (LOQ)	Generally, <i>LOQ</i> instead of <i>LOD</i> used to be required by ISO 17025 auditors
No special staff’s statistical qualification		-	Simple and/or recognizable statistical approaches  Automatization of the validation and routine (QA, uncertainty) calculations <sup>c</sup>  Avoid non-systematic mathematical signal treatment <sup>d</sup>

<sup>a</sup> Calibration on every sample using standard addition could eliminate many of the matrix effects seen e.g. in spectroscopic techniques, favouring accuracy (advantage in terms of trueness). However, the treatment of several signals to estimate each concentration could provide high *RSDr* or punctual atypical estimations (caution)<sup>b</sup> Expected *LOD* in the mg.L<sup>-1</sup> region (advantage in terms of fitting  $LD_{\text{LIM}}$ )<sup>c</sup> MATLAB® algorithms (for all automatic statistical calculations), could provided tabulated results (e.g. Table 7.) and graphical outputs (e.g. Figures 11 and 12) easy to interpret without special statistical qualification. A priori, signal examination/treatment is undesirable since it becomes time-consuming and would require special staff formation (that anyway, could increases the risk of mistakes, prejudgement, etc.). <sup>d</sup> Unpredictable staff errors occasioned by non-systematic signal treatment were avoided. For simplicity and rapidity, automatic (a practical decision; usually preferred by testing laboratories) calculation provided by the software equipment was chosen together with an automatic criterion to detect/eliminate punctual outliers was selected.

Accuracy pre-validation (section ii) was performed with real samples (instead of CRM solution) in order to obtain previous information on the real method

applicability, samples' concentrations levels, and possible matrix effects. In contrast, provisional trueness information cannot be obtained. An  $\mathbf{X}_{3 \times 12}$  experimental design was used (i.e. assuming possible atypical results, see Table 5,  $Nr = 3$  was preferred), for each of the three samples (tap water, mineral water and an industrial treated waste-water outer fall). The run (factors involved) was defined in section (vi). Initially, the recommended standard addition procedure [32] was followed, so after the sample measurement three successive additions of o-nitro phenol were used. The data analysis (shown in appendix 3 in the disc provided), corresponding to the three samples, revealed four main provisional observations: (i) the occasional presence of atypical concentration estimates (an anticipated situation). 'Manual' signals inspection showed some 'peak' shape shifts in the voltage domain, between sample and successive standard additions and between-replicates, (ii) the repeatability effect was more important than the between-run one ( $RSD_r > RSD_{run}$ ). (iii) Similar behaviour between different types of samples, so matrix effect could be neglected (an anticipated result; Table 6). (iv) Sample *mean* concentrations (in the 6 – 10 mg.L<sup>-1</sup> range) were in the lower part of the 5 – 50 mg.L<sup>-1</sup> EC directive concentration range (see Table 6).

**Table 6: Results for tap water related to the  $X_{3 \times 12}$  experimental designs for a pre-validation study**

Day	Batch	Replicate	Replicate	Replicate	Mean (mg/L)	Std deviation	% RSD
		1	2	3			
1	1	8.276	5.114	6.145	6.511	0.099	1.521
2	2	5.114	5.15	8.321	6.195	0.116	1.872
3	3	6.786	7.558	11.72	8.688	0.100	1.151
4	4	5.794	6.888	6.225	6.302	0.089	1.412
5	5	7.78	2.174	5.794	5.249	0.088	1.677
6	6	4.651	4.615	3.852	4.373	0.213	4.871
7	7	3.559	2.791	3.532	3.294	0.078	2.368
8	8	6.791	2.77	6.498	5.353	0.211	3.942
9	9	6.668	7.247	9.88	7.932	0.238	3.001
10	10	12.547	6.795	8.706	9.349	0.14	1.497
11	11	2.536	2.488	2.243	2.422	0.093	3.840
12	12	5.443	3.356	6.175	4.991	0.093	1.863

The next step of the protocol entailed some decisions regarding the adjustment of the accuracy validation experiment. The possibility of examining each signal and/or its mathematical treatment was avoided; mainly because the difficulties of introducing it into an SOP to be audited and the increase of analysis times for routine work will be compromised. Therefore, the automatic calculation (equipment software) was preferred for the subsequent validation protocol. Due to the intrinsic signal variability, the need of three o-nitro phenol standard additions was evaluated. It was observed (in the three samples data matrices) that using only the first standard addition signal the obtained precision statistics were similar (even slightly better) than using the 3-standard addition signals. So the ‘1-standard addition’ approach was selected. It provides half the analysis time per sample (a practical advantage). In such conditions, the increase of  $N_r$  and the systematic application of the z-score approach (as a part of the accuracy validation SOP) were decided. Figure 17 shows the polarograms obtained for the accuracy validation studies using a CRM solution (for all polarograms measured see appendix 3-5)

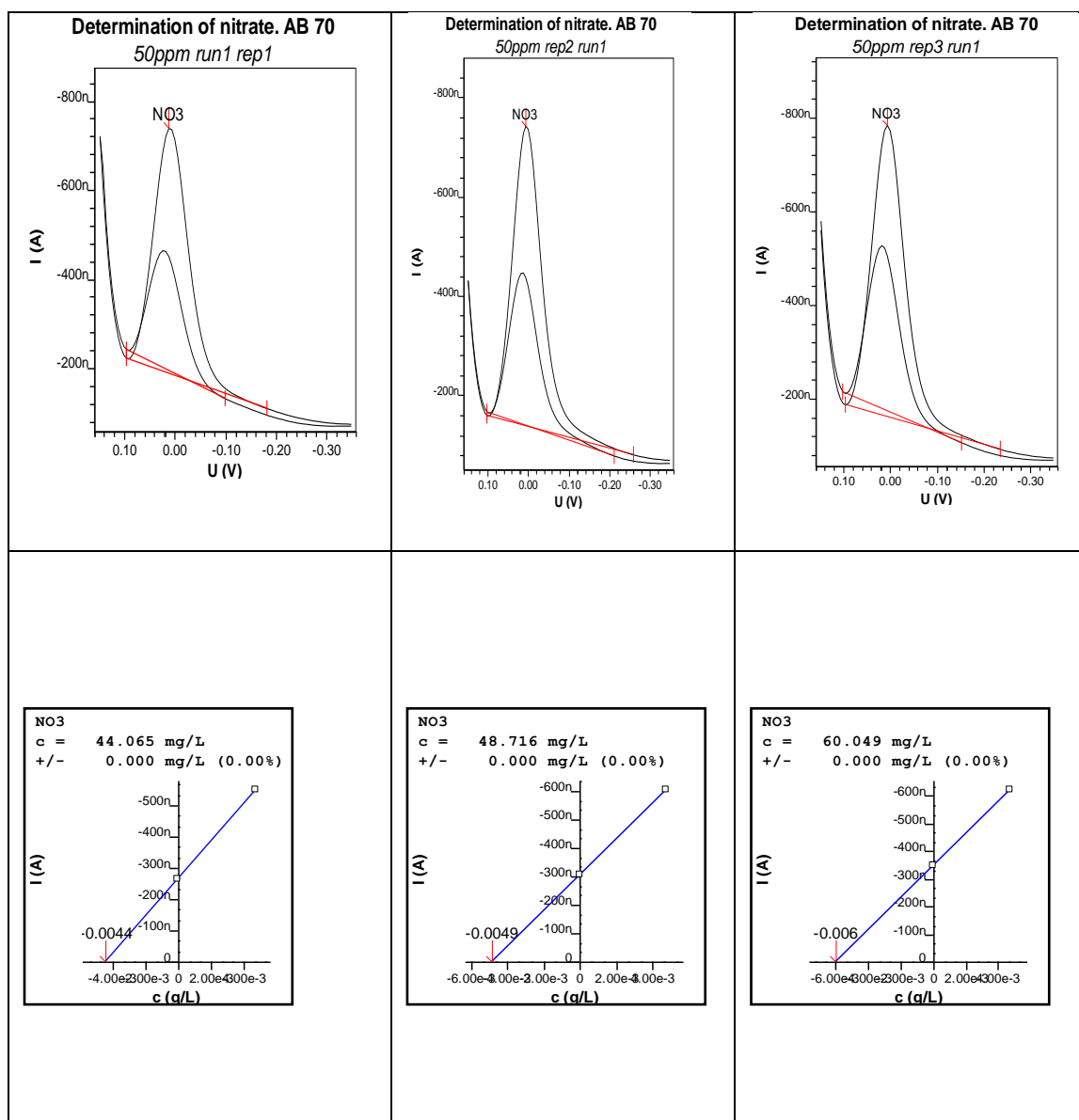


Figure 17: Summary of polarograms obtained in three replicates for Case Study 1

Accuracy validation matrix (section 5.3.3), was performed following the procedure indicated in section (5.3.5).

**Table 7:  $\mathbf{X}_{6 \times 12}$  data matrix (experimental nitrate concentration in  $\text{mg.L}^{-1}$ ) corresponding to the analysis of a reference nitrate solution ( $50 \text{ mg.L}^{-1}$ )**

		Run											
		1	2	3	4	5	6	7	8	9	10	11	12
Replicate	1	44.07	57.18	46.90	55.41	34.84	57.83	43.81	42.56	55.02	48.75	42.33	52.75
	2	48.72	56.20	54.24	55.04	43.17	45.47	49.12	50.62	53.24	61.75	38.56	62.05
	3	60.05	53.47	57.78	57.25	41.75	48.39	34.11	55.35	40.57	42.50	44.42	58.22
	4	55.91	42.31	49.45	39.13	45.37	45.07	43.48	53.77	51.04	55.83	47.65	39.85
	5	59.26	36.17	43.73	54.64	43.43	57.16	42.79	36.18	49.69	51.98	56.04	58.96
	6	44.76	55.42	40.52	52.45	46.00	59.32	41.62	38.46	42.66	43.20	46.73	49.92

The CRM solution adjusted to the parametric level to fit the EC requisite (Table 7). Consistently with the provisional pre-validation observations,  $Nr = 6$  replicates per run was selected, assuring more reliable *means* and facilitating the outlier detection/elimination tasks; in contrast  $Ns = 12$  runs was considered acceptable for the expected  $RSDr > RSDrun$  situation. Table 8 shows the  $\mathbf{X}_{6 \times 12}$  data matrix obtained.

**Table 8: Harmonized statistics from the  $X_{6 \times 12}$  data matrix (Table 11), after eliminating those values with  $|z| > 2$**

Stage	Statistic	Value
<b>Validation</b>		
	$\mu (\pm s_\mu)$	49.71 ( $\pm 3.28$ ) mg.L <sup>-1</sup>
	$E$	- 0.58 %
	$RSDr$	10.67 %
	$RSDrun$	4.95 %
	$RSDi$	11.76 %
<b>IQC</b>		
	$S_u = u$ (u-approach) <sup>a</sup>	3.41 mg.L <sup>-1</sup>
<b>Uncertainty<sup>b</sup></b>		
	$U (Nr'=3)$	8 mg.L <sup>-1</sup>
	$U (Nr'=4)$	8 mg.L <sup>-1</sup>
	$U (Nr'=5)$	7 mg.L <sup>-1</sup>
	$U (Nr'=6)$	7 mg.L <sup>-1</sup>

<sup>a</sup> Estimated for  $Nr' = 6$ .

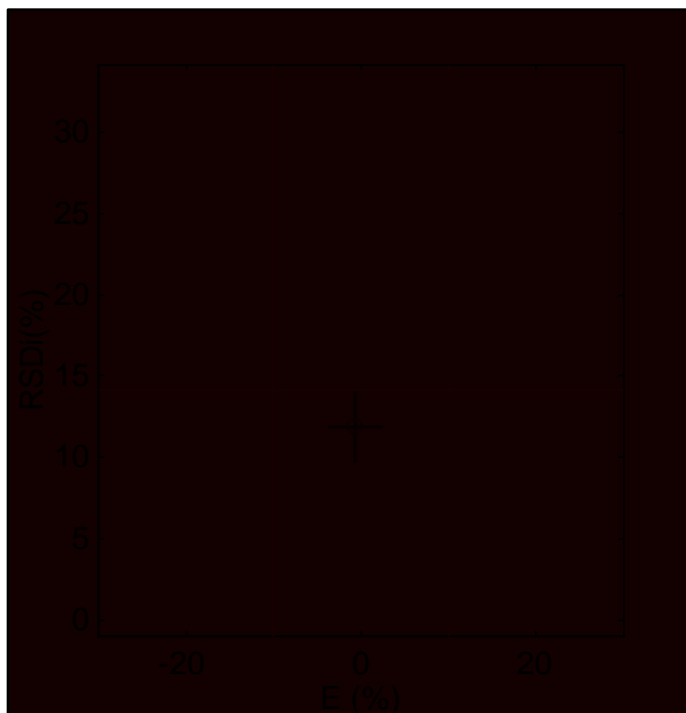
<sup>b</sup> Estimated (using  $k = 2$  and rounded). Useful for samples containing  $\sim 50$  mg.L<sup>-1</sup>.

As expected, some atypical estimated concentrations were obtained; however, the run mean ( $\mu$ ; see appendix 2) was 48.8 mg.L<sup>-1</sup> (using all the data) which agrees well with  $\mu_{CRM}$  (50 mg.L<sup>-1</sup>). This suggests good method's trueness (an anticipated result; Table 12). The systematic (automatic) application of the z-score approach (Eq. 33) provides the elimination of some  $x_{ij}$  data. Table 8 shows the main statistics (appendix 2) related to accuracy validation associated to the remaining data.

$E = - 0.58 \%$  (Eq. 34) represent the overall method's bias, obtained in intermediate precision conditions. Stating that  $E = - 0.58\%$  discards the uncertainty of the bias and then it becomes inconsistent to compare it directly with the  $E_{LIM}$  value. Therefore, trueness assessment requires the inclusion of an uncertainty interval for  $E$ . This is also

valid for  $RSDi$  (Eq. 35), and then a global accuracy assessment strategy becomes evident. Accuracy validation assessment (section iv) was performed by means of Monte Carlo simulation, based on the model of Eq. 36, from  $E$ ,  $RSDr$  and  $RSDrun$  estimates (Table 11; appendix 2).

Figure 18 shows the  $RSDi \pm U(RSDi)$  vs.  $E \pm U(E)$  ‘Validation assessment’ plot. As can be seen, both uncertainty intervals were within the EC-limits (i.e. error bars are inside the area defined by  $E_{LIM}$ - and  $RSDi_{LIM}$ -lines), so the validity statement for accuracy could be declared.



**Figure 18:** Accuracy assessment plot showing  $RSDi \pm U(RSDi)$  and  $E \pm U(E)$  Uncertainty intervals (wide lines) corresponding to  $10^4$  simulated validations related to the estimated  $RSDi$  and  $E$  (o) values from experimental validation data. Limiting values (thick lines) consistent to EC regulation are included.

To complete the method validation stage under the fit-for-purpose viewpoint (in this case to fit EC requisites of Table 9), validation of  $LOD$  must be performed. Since the blank does not give any response, ten independent blank solutions spiked with  $1 \text{ mg.L}^{-1}$  (as described in section 5.3.7). The z-score was calculated for the signal vector using its mean and standard deviation,  $s$ , as  $\mu_0$  and  $\sigma_0$ , respectively, in Eq. 33. A  $|z| > 2$



case was observed, so the corresponding signal was eliminated. The experimental statistics obtained at 1 mg.L<sup>-1</sup> level ( $mean = 1.60 \text{ mg.L}^{-1}$ ,  $s = 0.49 \text{ mg.L}^{-1}$ ;  $Nr = 9$ ) provide an idea of the expected accuracy for the method at low concentrations ( $E \sim 60\%$ ;  $RSD \sim 30\%$ ).

Obviously those values are notably larger than those at the required (50 mg.L<sup>-1</sup>) parametric level (a common outcome for most methods). Although this value corresponds to repeatability conditions, they partially complement the lack of information on bias-concentration and precision-concentration relationships during the accuracy validation step performed only at the parametric level to fit the EC requisite.

According to the experimental  $s$  value found at 1 mg.L<sup>-1</sup>, the ‘blank + 3  $s$ ’ criterion provides  $LOD = 1.4 \text{ mg.L}^{-1}$  (lower than  $LOD_{LIM} = 5 \text{ mg.L}^{-1}$ ) so this feature can be considered as validated. Strictly speaking, this level does not allow for the analyte’s quantification, and the limit of quantification ( $LOQ$ ) is normally required by ISO 17025 auditors instead of  $LOD$  (Table 6). A similar calculation, but using the ‘blank + 10  $s$ ’ criterion provides  $LOQ = 4.5 \text{ mg.L}^{-1}$ , which was lower than the sample range observed during the pre-validation step (6-10 mg.L<sup>-1</sup>) and could become satisfactory, as an initial estimation (Table 7). In the future, if  $LOQ$  must be estimated (e.g. by testing laboratories for accreditation purposes), alternative estimation based on fixing the maximum  $RSD$  value for the method [2], could be more recommendable.

Once method validation stage was completed satisfactorily, QA and uncertainty statistics harmonized with those from accuracy validation (section v) were calculated. Since Eq. 37 depends on  $Nr'$ , this value was pre-fixed as 3 independent replicates (as a minimum) for all the routine stages (sections 5.3.4, 5.3.6 and 5.3.7). This reduces the experimental effort (in the long-term routine analysis stage) with respect to that performed in the ‘punctual’ method validation stage, as suggested in [23]. Table 8 shows the expanded uncertainty ( $U$ ; Eqs.5 and 6) values, corresponding to different  $Nr'$  values (from 3 to 6; considering the possibility of punctually increases  $Nr'$ ; progressive work strategy), consistent with the parametric level. Table 8 also shows  $Su$  (Eq. 37), corresponding to  $Nr' = 6$ , since it provides narrower control limits to better control the possibility of atypical results during the method control stage. Therefore, action ( $AL \sim 40$  and  $\sim 60$ ; Eq. 39) and warning ( $WL \sim 43$  and  $\sim 57$ ; eq. 40) limits were obtained (i.e.

to be used in control charts in the future). The final step of the protocol corresponds to the initial routine work (vi). It was performed on an  $Nr' = 3$  basis (although the progressive work strategy was activated) along two working routine sessions.

**Table 9: Initial routine work step**

Working session	routine	Code	Replicate	Concentration
1		High-level verification solution	1	55.12 <sup>a</sup>
			2	56.43 <sup>a</sup>
			3	60.15 <sup>a</sup>
		Sample	1	10.22
			2	11.88
			3	13.73
		Control solution	1	49.67
			2	47.69
			3	54.03
		Sample	1	11.13 <sup>a</sup>
			2	8.88 <sup>a</sup>
			3	10.17 <sup>a</sup>
2		Low-level verification solution	1	6.13
			2	4.94
			3	5.53
		Sample	1	9.81 <sup>a</sup>
			2	11.50 <sup>a</sup>
			3	11.49 <sup>a</sup>
		Control solution	1	58.78
			2	35.08
			3	62.16
		Sample	1	9.86 <sup>a</sup>
			2	12.87 <sup>b</sup>
			3	9.70 <sup>b</sup>

<sup>a</sup> The data from these four industrial samples (in the  $\sim 10 \text{ mg.L}^{-1}$  level) were arranged into a  $X_{3 \times 4}$  data matrix. The protocol (section ii to v) was applied in order to have an initial estimation of the expanded uncertainty at this level (the sample level  $U_s$ ). Using  $Nr' = 3$  in eq.37,  $U_s$  was  $1.85 \sim 2 \text{ mg.L}^{-1}$ . Since sample concentration is close to the 'low-level verification solution' ( $5 \text{ mg.L}^{-1}$ ),  $\mu_{CRM} \pm U_s$  (eq. 41) was used as verification limits at this concentration level.

Table 9 shows the initial results obtained. The four industrial samples (from two different origins) show a concentration in the  $\sim 10 \text{ mg.L}^{-1}$  level (five times lower than the parametric level), this results agree with those obtained in the pre-validation study. It should be noticed that the accuracy (bias and precision statistics), and therefore uncertainty estimations usually depend on the concentration level.  $U = 8 \text{ mg.L}^{-1}$ , corresponding to  $Nr' = 3$  (Table 8) is consistent for the  $50 \text{ mg.L}^{-1}$  level ( $Ur = 16 \%$ ), but not necessarily with the current sample level. Table 8 (footnote) indicates an initial estimation of the sample expanded uncertainty ( $Us$ ) at this level. Also, for method verification purposes,  $U = 8 \text{ mg.L}^{-1}$  is consistent for the  $50 \text{ mg.L}^{-1}$  level (i.e. the high-level verification solution); but inconsistent for the low-level verification solution ( $5 \text{ mg.L}^{-1}$ ). For this level,  $Us$  was also used (Table 12, footnote).

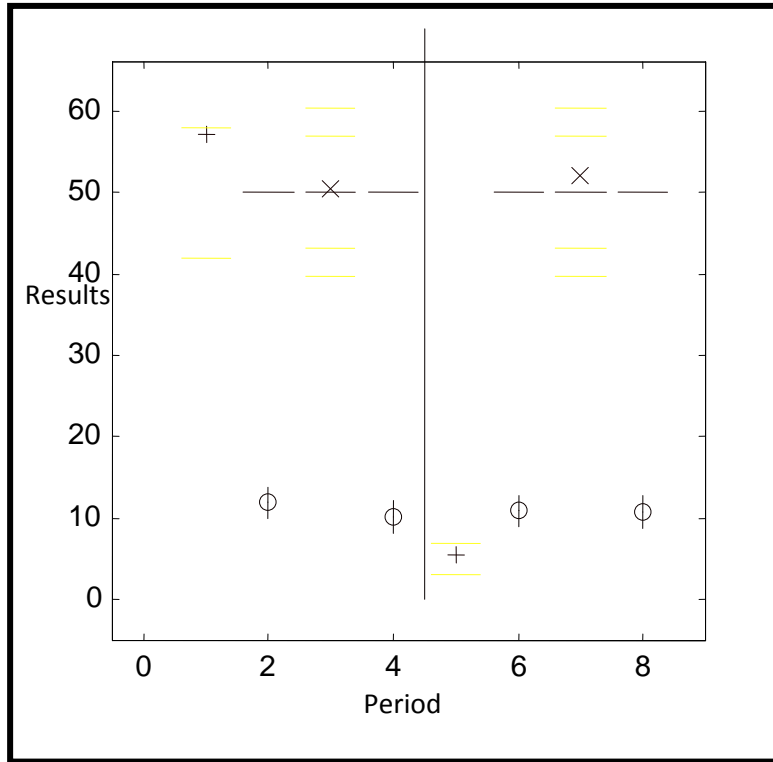


Figure 19: illustrate the routine work stage results on Table 13. Period 1 ('high-level verification solution') indicates that the mean of three replicates are within the accepted limits, so the method is ready to be used for sample analysis in this session. Periods 2 and 4 correspond to triplicate sample analysis, along this session. In both cases the  $mean \pm U_s$  intervals are within the EC parametric value. Similar results were found in the case of working routine session 2, this time using the 'low-level verification solution' (Period 5) and showing the results of two new industrial samples (Periods 6 and 8). Acceptance of these results depends also on IQC (Periods 3 and 7, in working routine sessions 1 and 2, respectively). In this case the method is initially in control; so *WL* and *AL* seem to be adequate for use in conventional control charts.

To verify the adequacy of derived limits, Figure 19 illustrate the routine work stage results on Table 13. Period 1 ('high-level verification solution') indicates that the mean of three replicates are within the accepted limits, so the method is ready to be used for sample analysis in this session. Periods 2 and 4 correspond to triplicate sample analysis, along this session. In both cases the  $mean \pm U_s$  intervals are within the EC parametric value. Similar results were found in the case of working routine session 2, this time using the 'low-level verification solution' (Period 5) and showing the results of two new industrial samples (Periods 6 and 8). Acceptance of these results depends also on IQC (Periods 3 and 7, in working routine sessions 1 and 2, respectively). In this case the method is initially in control; so *WL* and *AL* seem to be adequate for use in conventional control charts.

## 5.4 Conclusions

As accreditation standards (e.g. ISO 17025), most SOPs written for testing laboratory managers tend to be ambiguous (i.e. they specify what must be done but not how it has to be done, in detail). This work demonstrated that a detailed step-by-step protocol can be defined, thus covering the different quality tasks to be performed (e.g. method validation, QA, sample analysis). Moreover, if harmonized statistics are used, the same information may cover all the decision-making aspects, assuring consistent results. Outlier detection/elimination is a recommendable task; however, identification of the assignable cause is important too. In this case study, the main reason for outliers could be attributed to intrinsic method variability under repeatability conditions (due to signals), which explains the infrequent  $RSD_r > RSD_{run}$  situation observed experimentally when data on Table 8 is submitted to ANOVA. The individual ('manual') signal treatment become a non-practical solution (probably unacceptable for routine testing purposes; e.g. to be incorporated into an SOP). Then, the use of a relatively large number of replicates and the application of a systematic outlier detection/elimination approach, as a part of the method, become a more practical solution (at least for testing laboratories). Validation in this condition provides an estimation of the method variability obviously lower than that without outlier elimination, but adjusted to this particular case study.

The analysis of the uncertainty components in Eq. 37 reveal that the term  $u_{CRM}^2$  ( $\sim 0.026$ ; section ii) is negligible (e.g.  $s_r^2 \sim 28$ ; Table 9 and appendix 1). Strictly speaking, the contribution of the spiked o-nitro phenol solution (i.e. a new  $u_{spike}^2$  term) should be considered; however, habitually  $u_{spike}^2$  is neglected [15] and, as  $u_{CRM}^2$ , it can therefore be avoided. In the case of Figure 18 triplicate assays provided 'conform' observations in all stages; however, the occurrence of atypical results in the further routine work would require the increase of  $Nr'$  ('progressive routine work' strategy). In such a case, since no reference concentration is available for samples, the *mean* or *median* of the  $Nr'$  estimated concentration could be used as  $\mu_0$  in Eq. 33, if outlier detection/elimination is needed. Even if no outlier is detected, the use of larger  $Nr'$  (e.g. up to 6) values should provide a more reliable result, according to the good trueness exhibited by the method. Figure 12 just represent the initial results in the routine work stage to show to the testing laboratory, that the developed method and all

limits included into the protocol are consistent with the different quality tasks to be performed (i.e. the method is almost ready-to-use). Figure 12 is therefore an extra-work that should be done by research laboratories prior to publishing a new method (normally not performed).

Obviously, technical protocols should be adjusted for each particular method, according to their requisite-feature relationships. This work can offer guidance not only to routine laboratories, but also to research groups intending to propose almost ‘ready-to-use’ methods, to develop unambiguous consistent protocols. For instance, in view of the current method evaluated in this work, we can conclude that it can be considered ‘fit-for-purpose’, with respect to the EC requisites, and exhibit adequate features for QA tasks. Moreover, this study has revealed their critical aspects and guidance on how to adjust the protocol in order to solve eventual problems.

## CHAPTER 6

### CASE STUDY 2

#### 6.1 Purpose Statement:

The purpose of the case study; is to use voltammetry as a complimentary technique for the determination of trace metals in harbour water to be used in routine and research laboratories. The main goal of this work is on method validation, and harmonization of the internal quality tasks under routine conditions. Currently the ‘fit-for-purpose’ internal method validation (i.e. establishing method’s requirements-features relationships), Internal Quality Control (IQC) and sample uncertainty estimations are mandatory for accredited routine testing laboratories [2, 38, and 39]. However, more harmonization on (detailed) technical protocols and statistics is necessary. Recently, an effort to harmonize the internal validation-control-uncertainty process (i.e. the so called u-approach [13]) has been made, since in the past, these internal quality aspects have been treated separately, producing inconsistencies. An aspect that deserves more attention is the different perspectives of research laboratories performing method development tasks (mainly focused on establishing method’s features) and routine testing laboratories (usually under ISO 17025 accreditation processes and where practical/economical aspects use to be mandatory) with respect to the internal quality tasks. The lack of harmonization provides analytical methods (published in prestigious journals) non-practical or difficult to be adapted by testing laboratories.

#### 6.2 Preparation of the electrodes

Before the analysis commences the surface of the working electrode was checked thoroughly. Especially the interface between the electrode and the holder should have no defects. After polishing the electrode surface with Aluminium oxide, it was rinsed with high purity water and was dried with a paper towel. The solution was purged with nitrogen prior to analysis, therefore, for that reason the polishing step with Aluminium oxide was repeated only once a day.

### **6.2.1 Cell pre-treatment procedure**

To remove all traces of metals, the voltammetric cell was rinsed first with small amounts of water and 0.5-1ml concentrated Sulphuric Acid, then with high purity water. This procedure was repeated twice. Prior to every analysis, the voltammetric vessel was filled with 0.1M Hydrochloric Acid and the electrodes, mounted in the stand, immersed in the solution for approximately 40 to 60 seconds in order to clean them. Then Hydrochloric acid solution was replaced with the sample to be analyzed and the determination was carried out.

### **6.2.2 Sampling**

Samples were collected at Durban Harbour in 14 different locations at a depth of approximately 30cm. The samples were sampled using a boat by CSIR personnel. The harbour water samples for metal ion determination were collected in high density polyethylene sample bottles, which had been soaked in 1.0M  $\text{HNO}_3$  (Merck, Darmstadt, Germany), for two weeks and then rinsed and filled with deionised water and left again for two weeks. After sampling the sea water was filtered using cellulose filters ( $0.45\mu\text{m}$ ), acidified with (1ml conc. HCl or conc.  $\text{HNO}_3$  pH 2.0) and subsequently stored in a freezer.

### **6.2.3 Sample analysis procedure**

A modification of the protocol published by Metrohm (Application Bulletin) [14] was followed. 10ml of the sample was added to the voltammetric cell followed by 0.1ml of 1g/L  $\text{Hg}(\text{NO}_3)_2$  and 0.1ml of 30% HCl (supporting electrolyte). Lead and Cadmium can be concentrated on the glassy carbon electrode with mercury plated in-situ. Mercury is deposited on the surface of the electrode simultaneously with elements to be determined (Hg film in situ). During the subsequent stripping process they are again removed from the electrode and thus determined. This was followed by purging with inert nitrogen (high purity) gas for 5 min. Following the signal obtained, a single standard addition of 0.1mL of a mixture of Cd and Pb ( $1\mu\text{g.L}^{-1}$  mixture prepared by dilution of the CRM) was introduced into the cell and a second signal was recorded. Currents measured in the Differential Pulse (DP) voltammetric mode are proportional to the metal ion concentration in solution.



#### 6.2.4 Mixture reference solution

A 10  $\mu\text{g.L}^{-1}$  Cd and Pb mixture CRM solution was prepared by taking a suitable aliquot from the certified 1000  $\text{mg.L}^{-1}$  solutions and dilution with ultra-pure water as follows: pipetting 1 ( $\pm 0.008$ ) mL of 1000  $\text{mg.L}^{-1}$  Pb and Cd and diluting into a 500 ( $\pm 0.2$ ) mL and pipetting 1 ( $\pm 0.008$ ) mL of this mixture solution and diluting into a 200 ( $\pm 0.1$ ) mL. This provides a  $\mu_{\text{CRM}} = 10 \mu\text{g.L}^{-1}$  for both metals. The combined standard uncertainty ( $u_{\text{CRM}}$ ) was estimated by the bottom-up approach [16]. From the standard uncertainty of the original certified solution ( $u = U/k = 1 \text{ mg.L}^{-1}$ ) and the terms corresponding to the volumetric material used for pipetting and dilution, according to the information from their calibration certificates ( $u = 0.008/(6)^{0.5}$ ,  $u = 0.2/(6)^{0.5}$  and  $u = 0.1/(6)^{0.5}$ ; assuming triangular distribution as ISO recommends),  $u_{\text{CRM}}$  was calculated as:  $10 [ (1/1000)^2 + 2 (0.008/(6)^{0.5}/1)^2 + (0.2/(6)^{0.5}/500)^2 + (0.1/(6)^{0.5}/200)^2 ]^{0.5} = 0.04729 \sim 0.05 \mu\text{g.L}^{-1}$ .

#### 6.2.5 Accuracy validation and IQC studies

The procedure used was similar to the sample analysis described in section (6.2.4), except that 10ml of the sample was substituted by 10ml of the mixture CRM solution ( $\mu_{\text{CRM}} \pm u_{\text{CRM}} = 10 \pm 0.05 \mu\text{g.L}^{-1}$  Pb and Cd). The step-by-step protocol used is illustrated in Figure 20. It combines conventional tasks usually done by testing laboratories with recent approaches (the corresponding box appears highlighted) to improve and harmonize the internal quality process [12, 13]. To illustrate its practical application, validation and sample data obtained from a differential-pulse anodic stripping voltammetric method which is able to perform simultaneous direct determination of Cadmium and Lead (the ‘candidate method’) in marine water samples were used. Since the method was adapted from an application bulletin [14], it should be considered as an internal Standard Operating Procedure (SOP) which requires validation according to the standard ISO 17025 [38].

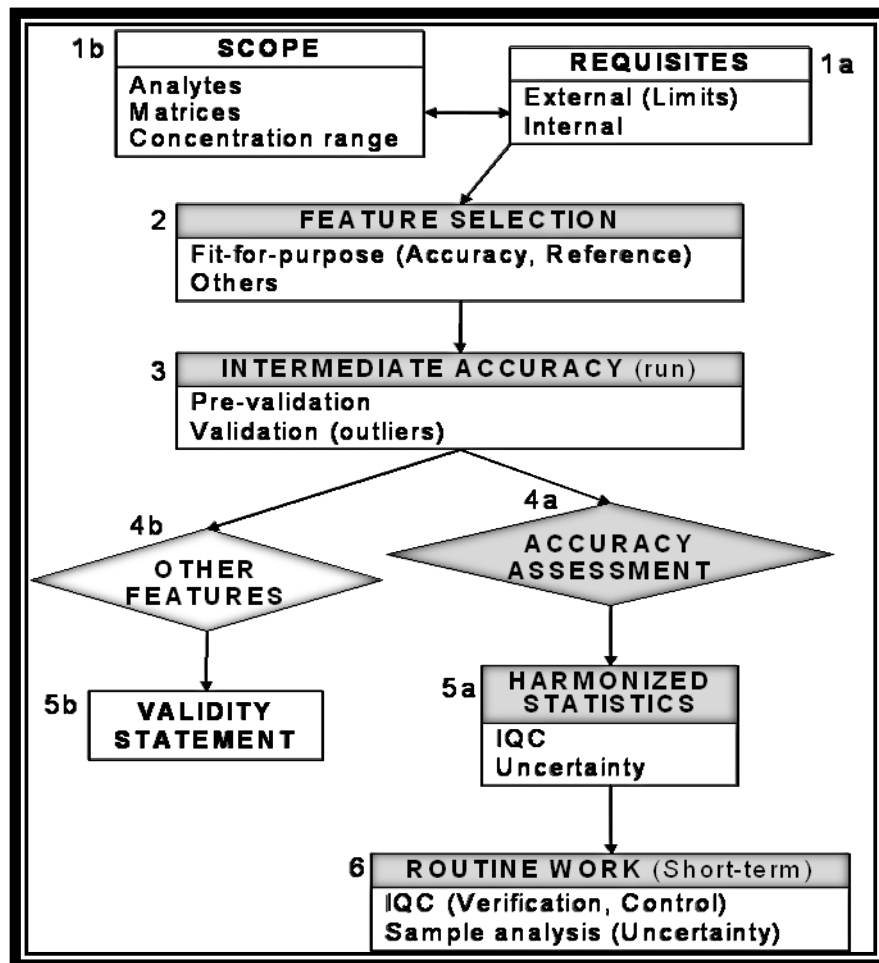


Figure 20: Harmonized protocol for internal quality aspects. In box 1a, external requisites are preferable to fix acceptability limits (e.g.  $E_{\text{LIM}}$  and  $RSDi_{\text{LIM}}$ ; see appendix 2). In box 2, accuracy (trueness and precision), related to a selected reference value (e.g.  $\mu_{\text{CRM}} \pm u_{\text{CRM}}$ ; see appendix 2) must be always considered a fit-for-purpose feature. In box 3, Intermediate accuracy estimates (i.e. in intermediate precision conditions, e.g.  $E$  and  $RSDi$ ; see appendix 2) are preferable. In box 4, massive (say  $10^4$ ) simulated validations (i.e. Monte Carlo simulation generating new  $X_{N_{\text{TXN}}}$  data) consistent with the experimental  $E$ ,  $RSDr$  and  $RSDr_{\text{run}}$  estimates allows the calculation of uncertainty intervals and verify that they are within the limits (e.g.  $|E \pm U(E)| < |E_{\text{LIM}}|$  and  $[RSDi + U(RSDi)] < RSDi_{\text{LIM}}$ ; see appendix 2). In box 5a, IQC (e.g.  $S_u$ ,  $AL$ ,  $WL$ ) and uncertainty (e.g.  $U$ ) statistics (see appendix 2) can be harmonized with accuracy validation statistics (u-approach). In box 6, routine work (at least in a short-term; to pre-examine the reliability of the criteria used) combining sample analysis (and uncertainty estimations,  $U$ ) and IQC tasks (e.g. method verification and control charts) is recommended in view to provide a virtually ‘ready-to-use’ method.

In box 1a figure 20, external requisites are preferable to fix acceptability limits (e.g.  $E_{\text{LIM}}$  and  $RSDi_{\text{LIM}}$ ; see appendix 2). In box 2, accuracy (trueness and precision), related to a selected reference value (e.g.  $\mu_{\text{CRM}} \pm u_{\text{CRM}}$ ; see appendix 2) must be always considered a fit-for-purpose feature. In box 3, Intermediate accuracy estimates (i.e. in intermediate precision conditions, e.g.  $E$  and  $RSDi$ ; see appendix 2) are preferable. The run concept (e.g. factors days/analysts/instruments that have to be varied between-validation sessions; at least equivalent to that planned during the working-sessions in

the routine analysis stage) must be defined. Relatively short (pre-validation) or large (validation)  $X_{NrxNs}$  experiment designs (see appendix 2) are possible. In the second case, outliers' detection/elimination schemes, preferably using widely accepted criteria (e.g. z-score approach; see appendix 2) are recommendable. In box 4, massive (say  $10^4$ ) simulated validations (i.e. Monte Carlo simulation generating new  $X_{NrxNs}$  data) consistent with the experimental  $E$ ,  $RSD_r$  and  $RSD_{run}$  estimates allows the calculation of uncertainty intervals and verify that they are within the limits (e.g.  $|E \pm U(E)| < |E_{LIM}|$  and  $[RSD_i + U(RSD_i)] < RSD_{iLIM}$ ; see appendix 2). In box 5a, IQC (e.g.  $S_u$ ,  $AL$ ,  $WL$ ) and uncertainty (e.g.  $U$ ) statistics (see appendix 2) can be harmonized with accuracy validation statistics (u-approach). In box 6, routine work (at least in a short-term; to pre-examine the reliability of the criteria used) combining sample analysis (and uncertainty estimations,  $U$ ) and IQC tasks (e.g. method verification and control charts) is recommended in view to provide a virtually 'ready-to-use' method. Besides the analytes (Pb and Cd) and sample matrix (marine water), the concentration range must be selected to complete the method's scope (box 1a in Figure 20). The target concentration was fixed at  $10 \mu\text{g.L}^{-1}$ , for both metals, close to the US EPA [11] guideline values ( $8.1$  and  $8.8 \mu\text{g.L}^{-1}$  for Pb and Cd, respectively); corresponding to the Criteria Continuous Concentration (CCC, Table 5), which is an estimate of the highest concentration in salt water to which an aquatic community can be exposed indefinitely without resulting in an acceptable effect [11]. The lowest concentration level was initially established at  $0.05 \mu\text{g.L}^{-1}$  corresponding to the limit of quantification ( $LOQ$ ) that is indicated in the application bulletin [14] (no criterion was provided). Therefore, a provisional concentration range could be set at  $0.05 - 10 \mu\text{g.L}^{-1}$ . The accepted philosophy in method validation involves the adjustment of this task to the 'fit-for-purpose' concept [2, 38], i.e. to fix method's requisites (box 1b in Figure 19). No external (i.e. legislation) requisites for accuracy were found for marine water samples. Alternatively, the Horwitz equation to estimate the expected reproducibility relative standard deviation ( $RSD_R$ ) is widely used for analytical methods [64]. For  $10 \mu\text{g.L}^{-1}$  level, an  $RSD_R = 32\%$  is expected, and from it,  $RSD_{iLIM}$  up to this value can be used as a requirement [65]. Similarly, AOAC Peer Verified Methods program [66] (involving 2-3 laboratories, and then close to the single-laboratory case), provides expected  $RSD_R$ , but also recovery, values (e.g. For  $10 \mu\text{g.L}^{-1}$ , recovery: 60-115% and  $RSD$ : 21%).

Assuming  $RSDr_{LIM} = RSDrun_{LIM} = 21\%$ ,  $RSDi_{LIM} = (RSDr_{LIM}^2 + RSDrun_{LIM}^2)^{0.5} \sim 29.7\%$  (See appendix 2), which is close to the value provided by Horwitz equation.  $RSDr_{LIM} = 29.7$  and  $E_{LIM} = \pm 15\%$  could be selected as external requisites (limits) for accuracy. Accuracy (trueness and precision), is fixed as a mandatory ‘fit-for-purpose’ feature (box 2 in Figure 20), in this case, with respect to the AOAC requirements. In a testing laboratory, other features could be taken into account, if additional requisites, external (i.e. from clients) and or internal (e.g. if they appear in a technical validation SOP), are stipulated. In the present case, since no additional requisites are available, trueness and precision, which could be validated altogether, were selected in this step. Realistic internal method validation implies to perform accuracy validation experiments (e.g. using a CRM), under intermediate precision conditions, e.g. designing an  $\mathbf{X}_{Nr \times Ns}$  ( $Nr$  replicates  $\times$   $Ns$  runs; see appendix 2) data matrix, whose analysis (ANOVA) permits method ‘intermediate’ accuracy estimation [44].

At this point, to define the run conditions becomes necessary (box 3 in Figure 20). For the present case study, a run was considered as a validation session along one day. Since no previous information on the method is available, a ‘pre-validation study’ [23] could give provisional estimates of method’s accuracy (box 3 in Figure 20). An  $\mathbf{X}_{3 \times 7}$  experimental design was used. Accuracy pre-validation was performed with the mixture CRM solution of  $10\mu\text{g.L}^{-1}$  (Pb and Cd). The analysis of these data revealed the occasional presence of atypical concentration estimates. ‘Manual’ signals inspection showed some ‘peak’ shape shifts in the voltage domain, between sample and successive standard additions and between-replicates. In both cases the infrequent situation,  $RSDr > RSDrun$ , was observed, indicating that repeatability is the predominant source of variability in the data. This suggests some strategies for the subsequent accuracy validation step. The intermediate accuracy validation (box 3 in Figure 20) experimental design was:  $Nr = 6$  replicates per run, which assures more reliable *means* and facilitates the outlier detection/elimination tasks, and  $Ns = 12$  runs (acceptable for the expected  $RSDr > RSDrun$  situation). An automatic outlier detection criterion was included (The z-score approach; see appendix 2). It is a generalised acceptability limit for routine laboratories [39]. Moreover, it is simple and easy to automate (e.g. in MATLAB®) to be incorporated into an SOP. This allows eliminating eventual estimated concentrations (into the  $\mathbf{X}_{6 \times 12}$  validation matrix), that could be

inconsistent due to the intrinsic signal variability under repeatability conditions. Table 10 and 11 shows the main accuracy validation statistics (see appendix 2) for Lead and Cadmium (after outliers' elimination). The results confirm the infrequent  $RSDr > RSDrun$  situation for this method.

**Table 10:  $X_{6 \times 12}$  data matrix (experimental Lead concentration in  $\mu\text{g.L}^{-1}$ ) corresponding to the analysis of a mixture CRM solution containing  $10 \mu\text{g.L}^{-1}$  Pb.**

Cadmium		Run											
		1	2	3	4	5	6	7	8	9	10	11	12
Replicate	1	10.146	10.561	7.99	9.895	7.553	7.495	3.231	9.886	16.466	7.558	5.533	6.239
	2	9.374	8.435	8.03	10.199	10.402	12.413	3.947	6.511	9.765	7.659	6.03	5.577
	3	9.617	8.388	8.015	12.256	7.364	11.454	3.949	8.437	10.419	7.847	5.98	7.245
	4	10.146	10.561	7.99	9.895	7.553	7.495	3.231	9.886	16.466	7.558	5.533	6.239
	5	9.374	8.435	8.03	10.199	10.402	12.413	3.947	6.511	9.765	7.659	6.03	5.577
	6	9.617	8.388	8.015	12.256	7.364	11.454	3.949	8.437	10.419	7.847	5.98	7.245

**Table 11:  $X_{6 \times 12}$  data matrix (experimental Cadmium concentration in  $\mu\text{g.L}^{-1}$ ) corresponding to the analysis of a mixture CRM solution containing  $10 \mu\text{g.L}^{-1}$  Cd.**

Cadmium		Run											
		1	2	3	4	5	6	7	8	9	10	11	12
Replicate	1	4.75	7.973	7.07	7.791	7.701	5.783	4.778	6.297	16.678	8.14	7.024	8.573
	2	9.439	8.994	9.02	9.891	10.091	9.237	6.164	5.296	5.094	11.26	8.33	7.339
	3	9.847	9.102	4.892	10.888	8.235	8.647	5.016	8.585	8.349	8.592	8.071	8.973
	4	4.75	7.973	7.07	7.791	7.701	5.783	4.778	6.297	16.678	8.14	7.024	8.573
	5	9.439	8.994	9.02	9.891	10.091	9.237	6.164	5.296	5.094	11.26	8.33	7.339
	6	9.847	9.102	4.892	10.888	8.235	8.647	5.016	8.585	8.349	8.592	8.071	8.973

Monte Carlo simulation has been proposed as a consistent way to assessing accuracy [23] (box 4a in Figure 20). Figure 21 shows the voltammograms obtained for the accuracy validation study for Cd and Pb.

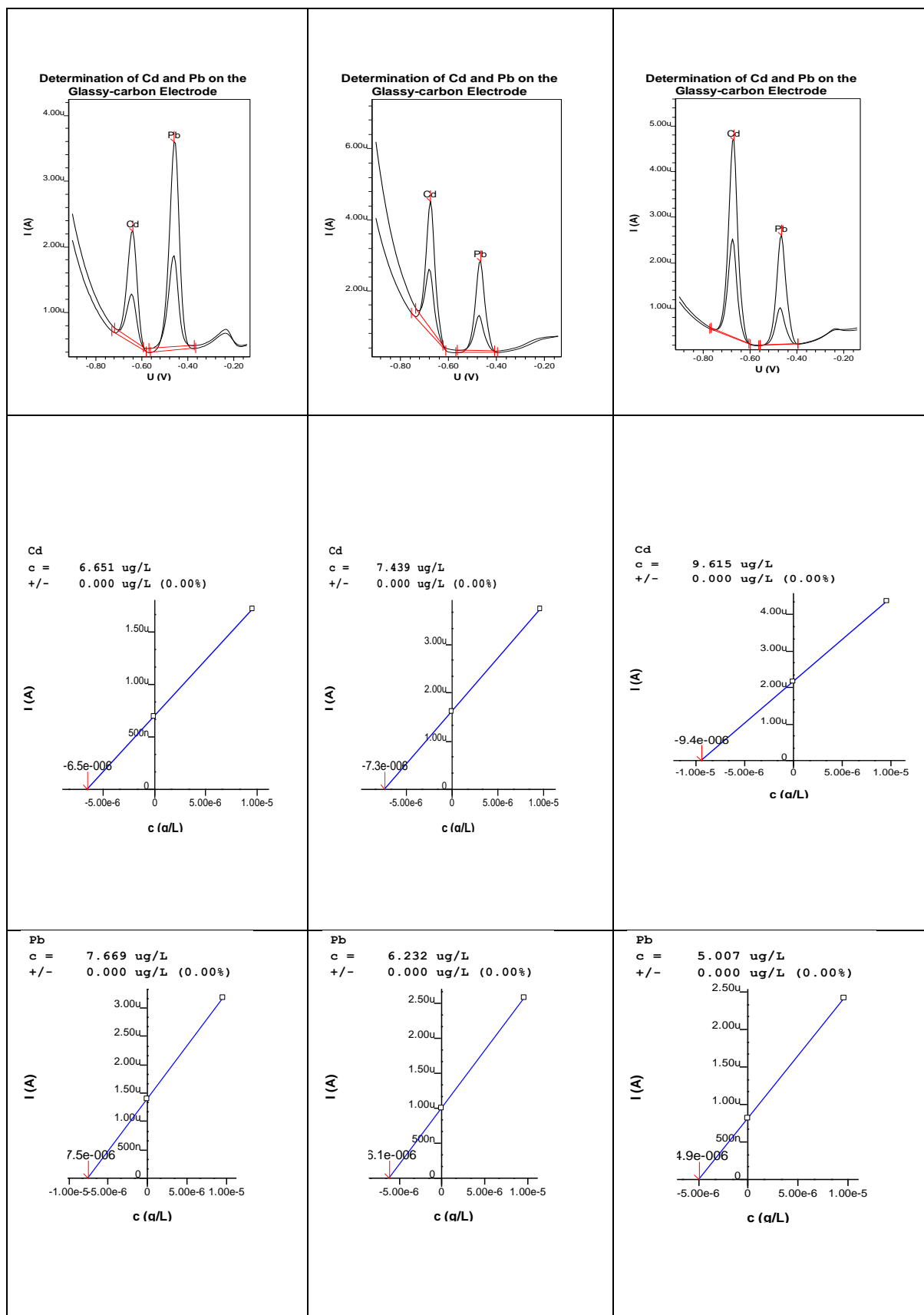
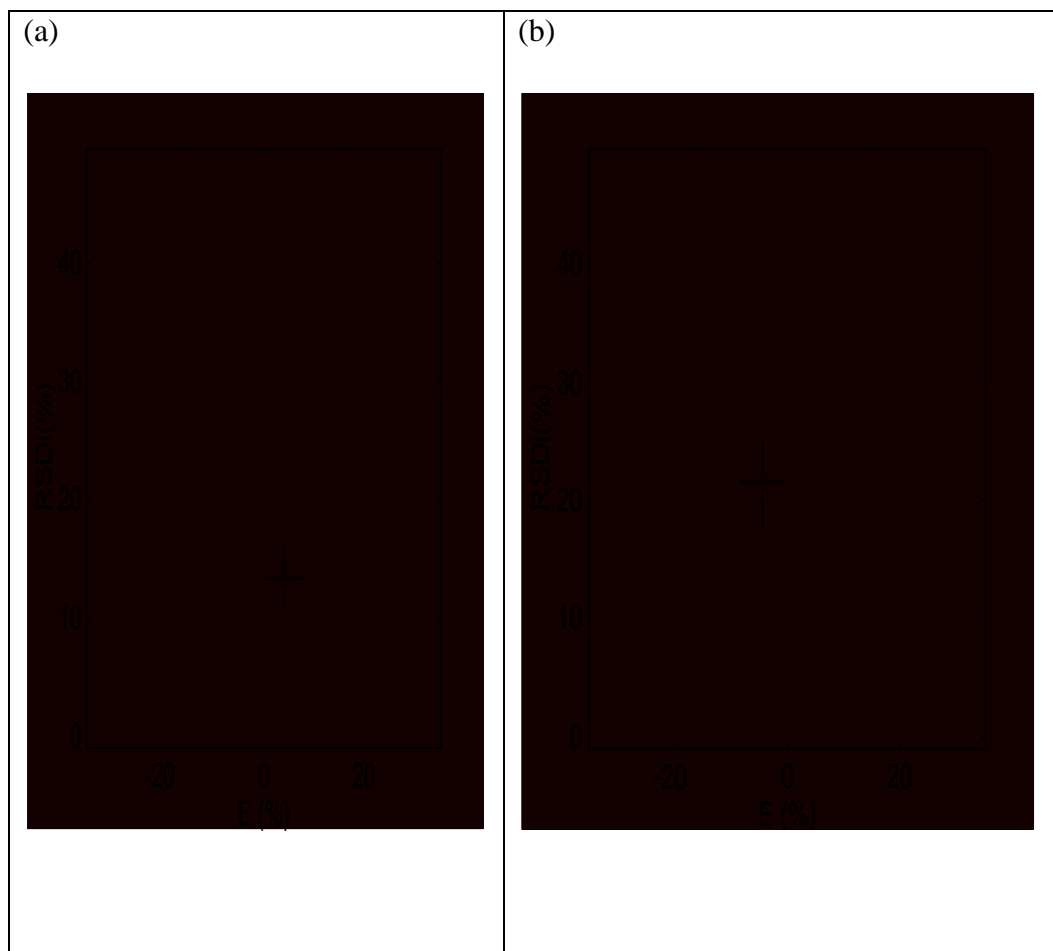


Figure 21 Summary of the voltammograms for case study 2

Figure 22a (Pb data) and 22b (Cd data) show the  $RSDi$  vs.  $E$  ‘Validation assessment’ plot corresponding to the intermediate accuracy validation experiment. For both analytes  $RSDi \pm U(RSDi)$  and  $E \pm U(E)$  uncertainty intervals are within the limits (i.e. inside the area defined by  $\pm E_{LIM}$  and  $RSDi_{LIM}$ ). In this condition, intermediate accuracy could be declared as validated.



**Figure 22:** Accuracy assessment plot for the validated data plot showing  $RSDi \pm U(RSDi)$  and  $E \pm U(E)$  uncertainty intervals ( $\oplus$  ; wide lines) corresponding to  $10^4$  simulated validation data consistent with  $E$  and  $RSDi$  estimates (o). Limiting values,  $E_{LIM}$  and  $RSDi_{LIM}$  (thick lines) are included. (a) Pb validation data after outlier elimination (z-score approach); fit-for-purpose result ( $RSDi \pm U(RSDi)$  and  $E \pm U(E)$  uncertainty intervals are within  $E_{LIM}$  and  $RSDi_{LIM}$  limits). (b) Cd validation data after outlier elimination (z-score approach); fit-for-purpose result ( $RSDi \pm U(RSDi)$  and  $E \pm U(E)$  uncertainty intervals are within  $E_{LIM}$  and  $RSDi_{LIM}$  limits).

Accuracy assessment plot showing  $RSDi \pm U(RSDi)$  and  $E \pm U(E)$  uncertainty intervals ( $\oplus$  ; wide lines) corresponding to  $10^4$  simulated validation data consistent with  $E$  and  $RSDi$  estimates (o). Limiting values,  $E_{LIM}$  and  $RSDi_{LIM}$  (thick lines) are included. (a) Pb validation data after outlier elimination (z-score approach); fit-for-purpose result ( $RSDi \pm U(RSDi)$  and  $E \pm U(E)$  uncertainty intervals are within  $E_{LIM}$  and  $RSDi_{LIM}$  limits). (b) Cd validation data after outlier elimination (z-score approach); fit-for-purpose result ( $RSDi \pm U(RSDi)$  and  $E \pm U(E)$  uncertainty intervals are within  $E_{LIM}$  and  $RSDi_{LIM}$  limits). It should be mentioned that the same conclusion was achieved avoiding the outlier elimination step; however, it is considered a safe practice for the present method. Since no other fit-for-purpose features were prefixed, no more validation experiment were necessary (box 4b in Figure 20). So, the current method can be declared validated (box 5b in Figure 20). The precedent ‘fit-for-purpose’ validity statement, allows continuing the protocol, by estimating uncertainty and IQC statistics harmonized with intermediate accuracy estimates (box 5a in Figure 20). At this point, the number of replicates to be used during the routine work stage ( $Nr'$ ) must be prefixed.  $Nr' = 3$ , as initial value, was considered reasonable. Table 12 shows the IQC and uncertainty harmonized statistics (see appendix 2) for the target concentration level.  $S_u$  could be used to establish control limits for the mean control chart using the mixture CRM solution [13]. As can be seen,  $U$  for samples close to the target level could be set at  $\pm 2$  (for  $Nr'$  of 3 to 6 replicates).



**Table 12: Harmonized statistics from the  $X_{6 \times 12}$  data matrix for a validation study, after eliminating those values with  $|z| > 2$  (see appendix 2)**

Stage	Statistic	Pb	Cd
<b>Validation</b>			
	$\mu (\pm s_\mu)$	10.4 ( $\pm 0.9$ ) $\mu\text{g.L}^{-1}$	9.6 ( $\pm 0.7$ ) $\mu\text{g.L}^{-1}$
	$E$	4.1 %	-4.1 %
	$RSDr$	11.93 %	20.31 %
	$RSDrun$	7.16 %	0 %
	$RSDi$	13.91 %	20.31 %
<b>IQC</b>			
	$S_u = u$ (u-approach) <sup>a</sup>	0.94 $\mu\text{g.L}^{-1}$	0.82 $\mu\text{g.L}^{-1}$
<b>Uncertainty <sup>b</sup></b>			
	$U$ ( $Nr'=3$ )	2 $\mu\text{g.L}^{-1}$	2 $\mu\text{g.L}^{-1}$
	$U$ ( $Nr'=4$ )	2 $\mu\text{g.L}^{-1}$	2 $\mu\text{g.L}^{-1}$
	$U$ ( $Nr'=5$ )	2 $\mu\text{g.L}^{-1}$	2 $\mu\text{g.L}^{-1}$
	$U$ ( $Nr'=6$ )	2 $\mu\text{g.L}^{-1}$	2 $\mu\text{g.L}^{-1}$

<sup>a</sup> Estimated for  $Nr' = 6$ .

<sup>b</sup> Estimated (using  $k = 2$  and rounded). Useful for samples containing  $\sim 10 \mu\text{g.L}^{-1}$ .

The final step of the protocol corresponds to the initial routine work (box 6 in Figure 20). Samples were sampled in 14 different locations and analyzed along 7 days (2 samples per day at different hours). At the start of each working session in a day, a verification solution (the mixture CRM solution) was analyzed. In the middle of each session an IQC solution (the mixture CRM solution) was assayed. The means (from  $Nr' = 3$  replicates) of all these results are shown in Table 13.

**Table 13: Results for 14 harbour water samples and independent mixture CRM solutions (10 µg.L<sup>-1</sup> Pb and Cd level) used as IQC (verification and control) solutions along 7 working sessions. Results for verification and control were rounded according to the estimated expanded uncertainty ( $U = 2 \mu\text{g.L}^{-1}$ ; see Table 16). Verification range limit (7 and 13;  $\mu_{\text{CRM}} \pm U$ ;  $k=3$ )**

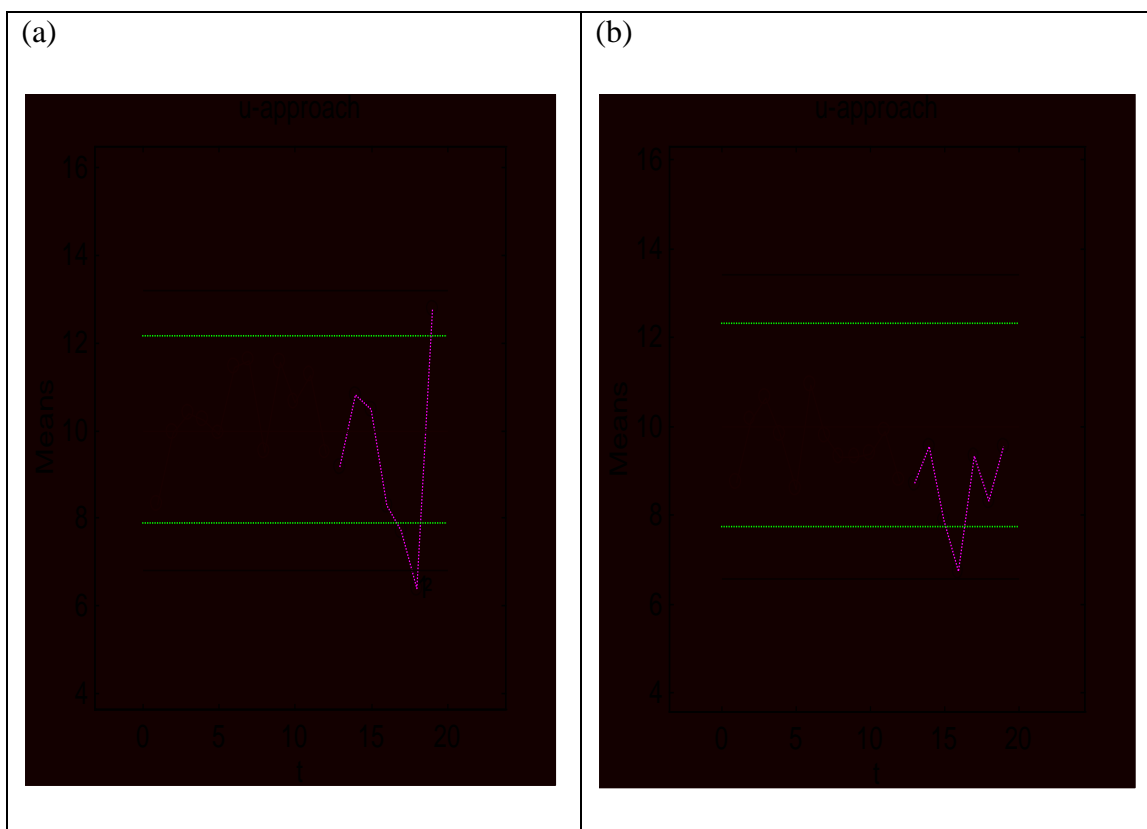
Session	Code	Pb, µg.L <sup>-1</sup>	Cd, µg.L <sup>-1</sup>	Comment
1	Verification 1	10	8	Conform
	Sample 1	0.51	<0.05	
	Control 1	9	9	
	Sample 2	3.40	<0.05	
2	Verification 2	8	7	Conform
	Sample 3	1.26	0.45	
	Control 2	11	10	
	Sample 4	0.27	<0.05	
3	Verification 3	8	9	Conform
	Sample 5	1.02	<0.05	
	Control 3	10	8	
	Sample 6	2.14	0.3465	
4	Verification 4	4	5	Out-of-Verification limit
	Sample 7	0.85	<0.05	
	Control 4	8	7	
	Sample 8	1.82	0.36	
5	Verification 5	12	10	Conform
	Sample 9	0.47	<0.05	
	Control 5	8	9	
	Sample 10	0.77	0.26	
6	Verification 6	6	8	Out-of-Verification limit (Pb)
	Sample 11	0.26	<0.05	
	Control 6	6	8	(See Figure 15)
	Sample 12	0.59	0.06	
7	Verification 7	7	9	Conform
	Sample 13	0.25	<0.05	
	Control 7	13	10	
	Sample 14	6.00	<0.05	Sample close to the CCC limit

It should be noticed that all the experiments were done as they were planned (i.e. no action were taken as a function of the IQC results), in view to realise the method's behaviour at this stage. As can be seen, samples show relatively low level of both analytes (Pb, from 0.25 to 6  $\mu\text{g.L}^{-1}$  and, particularly, Cd from  $<0.05$  to 0.45  $\mu\text{g.L}^{-1}$ ); in all cases under the critical level (CCC 8.1-8.8  $\mu\text{g.L}^{-1}$ , respectively Table 5). Therefore, a priori, with no risks in the area studied must be reported. On the other hand, from a quality point of view, the reliability of these results should be analysed. Method verification conform to the limits except for days 4 (both analytes were below the verification limit) and 6 (Pb was below the verification limit). Strictly speaking, this does not allow sample analysis until corrective actions (e.g. pre-fixed into the Quality System SOPs; say repeating the method verification with a new verification solution, recalibrate the equipment, etc.) are applied. In this particular case, other considerations could be taken into account:

- The AOAC recovery expected at this concentration level at 60-115 %. Thus, a testing laboratory interested in this method could consider a lower 'low-verification' limit.
- A practical approach could be the systematic application of a 'progressive routine work' strategy, which can be incorporated into an SOP. In this example, it could be applied as follows: once the verification limit is violated (after performing the prefixed  $Nr' = 3$  replicates), start increasing  $Nr'$ , e.g. up to 6 if necessary, and apply systematically the z-score approach as part of the SOP (see appendix 2) to eliminate punctual outliers affecting the *mean*.

Figure 22 shows the typical mean control chart where the control results (7 last *means* in the plot) were incorporated together with the 12 validation *means* used for calculating the control limits according to the u-approach (Table 13). Only in the Pb-chart an out-of-control case was found. As before, this would imply to stop the routine work and to apply SOP actions (e.g. in the worst case, the sample analysis could be repeated once the method becomes in-control). However, the same comments

mentioned before could be considered here. At this stage ‘progressive work strategy’ can be applied and the z-score approach as a first step in the SOP.



**Figure 23:** Mean control chart for the progressive routine work for (a) Pb and (b) Cd. The 12 first points (joined by a line) correspond to the validation data *means* used for obtaining *AL* and *WL* limits (see appendix 2). The subsequent 7 points (joined by a line) correspond to IQC data *means* along the initial working sessions. 1 and 2 refers to violation of typical rules [17]: rule-1(one *mean* above *AL*) and rule-2 (two consecutive *means* between *WL* and *AL*), respectively.

Mean Control Charts (see figure 23) for (a) Pb and (b) Cd. The 12 first points (joined by a line) correspond to the validation data *means* used for obtaining *AL* and *WL* limits (see appendix 2). The subsequent 7 points (joined by a line) correspond to IQC data *means* along the initial working sessions. 1 and 2 refers to violation of typical rules [17]: rule-1(one *mean* above *AL*) and rule-2 (two consecutive *means* between *WL* and *AL*), respectively. Sample 14 in Table 13 deserves more attention. The estimated value for lead ( $6.00 \mu\text{g.L}^{-1}$ ) is relatively close to the critical and target levels, and then, an uncertainty close to that reported in Table 16 could be approximated (i.e., Pb concentration =  $6 \pm 2$ ;  $k=2$ ). This means that this sample uncertainty interval is close to the CCC limit. In such a case, the ‘progressive work strategy’ is also recommended, at least to take a more reliable decision related to this sample.

### 6.3 Conclusion

Fit-for-purpose ‘partial’ validation could offer to testing laboratories almost ready-to-use (or ready-to-validate) methods to be more easily implemented, with a prior knowledge on the main method’s features (mainly accuracy), IQC behaviour, but also presumable critical aspects and possible solutions to be included as part of the SOP. Following a systematic protocol, harmonization between research laboratories performing method development task and testing laboratories could be improved. Finally, the use of harmonized validation-IQC-uncertainty statistics provides consistency on the decision making tasks. Currently, ISO 17025 auditors accept validation (and IQC results) obtained in the same laboratory to be accredited. In view of method validation, the current work can be considered just a ‘partial method validation’ (i.e. a fit-for-purpose validation of those features subjected to available requisites; this concept still requires more harmonization from normalisation/accreditation entities). This means that if a complete validation is required (say by a particular ISO 17025 auditor); extra work must be performed by a testing laboratory adapting (and then validating) the present method. By contrast, at present the analytical method in question can be considered as fit for its intended purpose, according to the AOAC criteria [68], and consistent with the US EPA critical levels for the analytes/matrix [11]. Normally, validating the entire concentration range (e.g. covering the method scope) used to be required. This would imply repeating the proposed protocol at least at a low level (say, close to the experimentally determined *LOQ*). In our opinion, validating the method at a very low level, far from the critical level of the analyte, although informative, has no practical interest (e.g. in terms of assessing the risk associated to an analyte); and should be re-considered in some instances (another view of the non-harmonized partial validation concept).

ISO 17025 auditors use to demand method verification at least at two extreme concentration levels on the scope concentration range, that could be alternated between working routine sessions (non-harmonized task). As before, there must be less interest in verifying the method in the non-risky level. Similarly, with the current experimental design, no information on the uncertainty of samples below the target level exists. However, the real practical interest on assessing the uncertainty level, affects mainly

the samples close or above the critical level (e.g. sample 14 in Table 13). Again, this point deserves more attention and revision (harmonization).

## CHAPTER 7

### SUMMARY OF THE TWO CASE STUDIES

#### **7.1 Summary of the two case studies demonstrating Monte Carlo simulation's applicability in Quality control of Analytical Laboratory analysis**

##### **7.1.1 Summary of Case study 1**

In this case study, a holistic approach to validate analytical methods was assessed by virtue of Monte Carlo simulation. The main aim was to propose a step-by-step protocol to evaluate and harmonize internal quality aspects of a method. Such criterion involves a statement of method's requisites (external and internal), method's 'fit-for-purpose' features selection, pre-validation to adjust the validation protocol, accuracy validation and assessment, harmonization of the validation-verification-control-uncertainty process, validation of other required method's features, validity statement in terms of 'fit-for-purpose' decision making and short-term routine work. The proposed approach can be considered as an alternative to other well established schemes such as hypothetical testing, "bottom up" and "top down" approaches. The protocol, as described here, could be transferred to other methods. Its aim is to harmonize the work to be done by research teams and routine laboratories (assuming that different aims, strategies and practical viewpoints exist). Consequently, the recommended protocol should be seen as a starting point. It is necessary to propose definitive (harmonized) protocols that must be established by international normalisation/accreditation entities. Over and above that, this is just a contribution to show possible solutions to this global problem. Evidently, the protocol should be tested in the future with more methods in order to check its general applicability. However, parts of the protocol have already been used satisfactorily in some previous studies related to water analysis methods exhibiting different features and requisites [12, 13, 47 and our recent publication [69].

Outlier detection/elimination is a recommendable task. However, identification of the assignable cause is important too. In this case study, the main reason for outliers could be attributed to intrinsic method's variability under repeatability conditions (due to signals), which explains the  $RSD_r > RSD_{run}$  situation observed experimentally

when data on Table 7 is submitted to ANOVA. The signal treatment, become a non-practical solution (for routine testing purposes). Then, the use of a relatively large number of replicates (high  $Nr$  value) and the application of a systematic outlier detection/elimination approach, as part of the method, becomes a more practical solution (e.g. for testing laboratories). Validation in this condition provides an estimation of the method's variability obviously lower than that without outlier elimination, but adjusted to this particular protocol. The results in the progressive routine work step just represent the initial results in the routine work stage with the aim to pre-examine the reliability of the overall criteria described in this protocol (e.g. to show to the testing laboratory, that the developed method and all limits included into the protocol are consistent with the different quality tasks to be performed; i.e. the method is almost ready-to-use). Obviously, when more IQC data are available (e.g. when it is applied by the testing laboratory), control charts and application of rules to out-of-control situation (preferable harmonized) are recommendable. The progressive routine work step is therefore an extra-work to be done by research laboratories planning to publish a new method (normally not performed), and should not be confused with a substitution of the control charts, that should be used in the future by the testing laboratories. The experimental data for samples obtained from two different origins show concentration in the  $10 \text{ mg L}^{-1}$  level with an expanded uncertainty ( $U_s$ ) of  $1.85 \sim 2 \text{ mg.L}^{-1}$ . The expanded uncertainty ( $U$ ) for method verification purposes is  $8 \text{ mg L}^{-1}$  which is consistent with the parametric value of  $50 \text{ mg L}^{-1}$  (i.e. the high-level verification solution). The latter is not consistent with the low-level verification solution ( $5 \text{ mg L}^{-1}$ ). For this level,  $U_s = 1.85 \sim 2 \text{ mg.L}^{-1}$  was used. Most testing laboratories prefer to use acceptance criteria instead of statistical criteria for decision-making purposes. For instance, it is common to calculate the method's relative error and to compare it with a prefixed limit than to assess if there are statistical differences between the experimental and the accepted true value (by means of a hypothesis test). In the same way, the z-score approach is not really an outlier test (e.g. it does not account for the number of degrees of freedom and uses an assigned value for the accepted standard deviation); in contrast, it is a generalised acceptance limit accepted for routine laboratories.  $E = -0.58 \%$  (Table 8) represent the overall method's bias, obtained in intermediate precision conditions. Stating that  $E = -0.58\%$  discards the



uncertainty of the bias and then it becomes inconsistent to compare it directly with the  $E_{\text{LIM}}$  value. Therefore, trueness assessment requires the inclusion of an uncertainty interval for  $E$ . This is also valid for  $RSDi$ , and then a global accuracy assessment strategy becomes evident. Figure 18 shows a practical solution of this question. This could be substituted by numerical criteria, e.g.  $|E \pm U(E)| < |E_{\text{LIM}}|$  and  $[RSDi + U(RSDi)] < RSD_{\text{LIM}}$ . However, an advantage of Figure 18 is that it becomes more intuitive for testing laboratory staff (which sometimes could have limited statistical skill) which normally performs the validation studies.

The experimental result at 1 mg.L<sup>-1</sup> level (mean = 1.60 mg.L<sup>-1</sup>,  $s = 0.49$  mg.L<sup>-1</sup>;  $Nr = 9$ ) provides an idea of the expected accuracy for the method at low concentrations ( $E \sim 60\%$ ;  $RSD \sim 30\%$ ). obviously, those values are notably larger than those at the required (50 mg.L<sup>-1</sup>) parametric level (a common outcome for most methods). Although this value corresponds to repeatability conditions, they partially complement the lack of information on bias-concentration and precision-concentration relationships during the accuracy validation step performed only at the parametric level to fit the EC requisite. According to the experimental  $s$  value found at 1 mg.L<sup>-1</sup>, the ‘blank + 3  $s$ ’ criterion provides  $LOD = 1.4$  mg.L<sup>-1</sup> (lower than  $LOD_{\text{LIM}} = 5$  mg.L<sup>-1</sup>) so this feature can be declared as validated. Strictly speaking, this level does not allow analyte’s quantification, and limit of quantification ( $LOQ$ ) is normally required by ISO 17025 auditors instead of  $LOD$  (more harmonization is needed at this point). A similar calculation, but using the ‘blank + 10  $s$ ’ criterion provides  $LOQ = 4.5$  mg.L<sup>-1</sup>, which was lower than the sample range calculated previously (6-10 mg.L<sup>-1</sup>; pre-validation), and could become satisfactory, as an initial estimation. If  $LOQ$  must be estimated (e.g. by testing laboratories for accreditation purposes), alternative criteria based on fixing the maximum RSD value Eurachem Guide [2] could be more recommendable.

### 7.1.2 Summary for Case study 2

In this case study, a holistic approach to validate analytical methods was assessed by virtue of Monte Carlo simulation. Such an approach involves a statement of the methods’s scope (analytes, matrices, concentration level) and requisites (internal or external); selection of the method’s (fit-for-purpose) features; pre-validation and

validation of the intermediate accuracy (under intermediate precision conditions) and its assessment (by means of Monte Carlo simulation); validation of other method's features (if applicable); and a validity statement in terms of a "fit-for-purpose" decision, harmonized validation-control-uncertainty statistics (the "u-approach") and short-term routine work (with the aim of proposing virtually "ready-to-use" methods). In the first step the scope was clearly outlined which involves the analytes (Cd and Pb), sample matrix (marine water), the target concentration was fixed at  $10 \text{ ug L}^{-1}$ , for both metals close to the US EPA guidelines ( $8.1$  and  $8.8 \text{ ug L}^{-1}$ ) for Cd and Pb respectively. To fit the method's fit-for purpose concept the requisites for marine water were found to be  $E_{\text{LIM}} = \pm 15\%$  and  $RSD_{\text{rLIM}} = 29.7 \%$ . A bivariate plot was prepared and presented to assess whether the uncertainty  $E \pm U(E)$  and  $RSDi + U(RSDi)$  intervals are totally within the  $\pm E_{\text{LIM}}$  and  $RSDi_{\text{LIM}}$ . This approach does not entail extra work for the laboratory. Since the validation data was found to be between the intermediate accuracy and precision limits the method was considered as valid. After elimination of the outliers by a z-score approach the run mean  $u_{\text{CRM}}$  was  $10.4$  and  $9.4 \text{ ug L}^{-1}$  and an uncertainty ( $S_u$ ) of  $0.9$  and  $0.7 \text{ ug L}^{-1}$  for Pb and Cd respectively. It should be noticed that all the experiments were done as they were planned (i.e. no action were taken as a function of the IQC results), in view to realise the method's behaviour at this stage. As can be seen, samples show relatively low level of both analytes (Pb, from  $0.25$  to  $6 \text{ ug.L}^{-1}$  and, particularly, Cd from  $<0.05$  to  $0.45 \text{ ug.L}^{-1}$ ); in all cases under the critical level (CCC;  $8.1$ - $8.8 \text{ ug.L}^{-1}$ ). A mean control chart was plotted using the "u-approach" it can be seen that only the Pb chart shows the results that are out of the range. The worst case should be applied as it was explained in earlier sections.

## 7.2 CONCLUSIONS

In this study we have tried to harmonize the work to be done by research teams and routine laboratories (assuming that different aims, strategies and practical viewpoints exist). Consequently, the recommended protocol should be seen as a starting point, and, necessarily, definitive (harmonized) protocols must be established by international normalisation/accreditation entities. So, this is just a contribution to show possible solutions to this global problem. Ideally, the method should be validated at two (extreme values) or three (including an intermediate value) concentration levels covering the whole concentration range prefixed (e.g. accreditation scope). This is not a problem when bias and precision are determined under repeatability conditions (which could provide unrealistic accuracy estimations in the  $RSD_{run} > RSD_r$  case; the habitual one). However, it becomes inconvenient when the work is done under intermediate precision conditions (in practice probably unacceptable for routine laboratory, due to the experimental effort involved). To offer protocols that could be accepted, cost-benefit solutions are necessary.

Therefore, an in-between strategy could consist of validating the target/critical concentration level (e.g. to fit external method requirements; the fit-for-purpose concept) and cover the other levels along QA tasks (e.g. method verification during the routine stage; a task required by ISO 17025 auditors). Obviously, this strategy implies that bias-concentration and precision-concentration relationships are not estimated during the validation stage (a limitation that could be assumed from a practical point of view; e.g. we accept that the bias and imprecision of the method can be larger for concentration level far from the target/critical level). Method validation, Quality Assurance and sample uncertainty estimation are essential internal tasks for accredited laboratories. In the past the statistical aspects involving these tasks have been treated separately, however, recently (in this thesis), an approach to harmonize the internal validation-control-uncertainty process has been reported. In this thesis, an unambiguous step-by-step protocol to evaluate and harmonize internal quality aspects of a method is defined. Such protocol involves a statement of method's scope (analytes, matrices, concentration level) and requisites (external and internal), method's 'fit-for-purpose' features selection, pre-validation (to adjust the validation protocol), accuracy validation (in intermediate precision conditions) and assessment

(via Monte Carlo simulation), harmonization of the validation-verification-control-uncertainty process (u-approach), validation of other required method's features, validity statement in terms of 'fit-for-purpose' decision, Internal Quality Control tasks (including Method Verification and harmonized mean control charts), harmonized sample uncertainty estimations and short-term routine work (intending to propose 'ready-to-use' methods). Decision-making aspects (in view of harmonization), innovative criteria and impact on laboratory staff are outlined. Some steps of the protocol have been already used satisfactorily in some studies related to water analysis methods exhibiting different features and requisites. For instance: Almost all steps are followed in case study 1. The use of z-score and the 'Progressive routine work' become essential in this method. The risks of using classical (non-harmonized) mean control plots become evident in this example. Also, it is demonstrated that the experimental effort could rely more in the (punctual) validation stage (e.g. large  $N_s$ ) and reduced in the (continuous) routine stages (i.e. low  $N_r$ ). As a conclusion, the case studies have shown that an extra-effort could be done for harmonizing the internal quality aspects in laboratories.

However, lack of harmonization can only be corrected when standardisation and/or accreditation organizations assume responsibility for deriving unified/unambiguous criteria for testing laboratories. Moreover, referees/editors, responsible for publishing new or modified analytical methods, must demand researchers to include evidences (preferable using the same quality schemes that testing laboratories) of real applicability in routine conditions (e.g. In order to pre-examine the reliability of the overall criteria described in this protocol, a short-term routine analysis period (verification solutions, Scope-sample analysis and control solutions) can be initiated. Alternatively, it could be extracted from previous information obtained during the application of the protocol (e.g. The  $\mathbf{X}$  validation matrix data could be used, as IQC data, to check the established limits or, if Scope-sample were analysed,  $U$  can be computed and compared with legal or critical levels, if available. One of other advantages of using Monte Carlo simulation is that you can perform a simulation study in order to predict the behaviour of the validation estimates.

### **7.3 Recommendations**

Some miss-directions and areas to be explored are outlined below:

#### **7.2.1. Miss-directions related to method validation**

The resistance to accept partial validation instead of complete, with all features for some methods under concrete circumstances. This is due to the resistance to accept validation- control -uncertainty estimations at one concentration level, and the level of interest for some methods/analyses (e.g. a clearly defined ‘parametric’ or critical concentration level for a given analyte). In general, validating only at the low-Scope level (the worst case; a simplification) should cover the other Scope levels (reducing the experimental effort). Resistance to accept CRMs e.g. prepared in natural matrices without the need of fortifying ‘local’ Scope-samples, to validate some methods/analyses. The general use or exigency of the ‘Bottom-up’ approach to estimate sample result uncertainty even in the case of complex methods is not really a recommended task. Adapting the ‘Top-down’ philosophy to the internal validation information could become easier and more realistic.

#### **7.2.2 Areas to be explored**

The use of outliers’ detection criteria (validation and routine analysis periods) and consistency with IQC limits. In general, appropriate validation designs, should account for the intrinsic variability factors that can affect the candidate method’s results, including the controlled ones, through intermediate precision conditions (day, analyst, equipment), but also uncontrolled ones, e.g. roughness, re-calibrations, etc.

Introduction of multivariate statistics into the qualimetric process (classically dominated by univariate statistics), trying to find better decision criteria, for example case study 2, whereby bianalytes statistics were used for harmonization of internal quality tasks in analytical laboratories. Uncertainty estimation based on Monte Carlo simulation (in order to displace tedious conventional estimations, based on error propagation). The methodology developed here can be applied to other disciplines of analytical chemistry such as HPLC, GC, GFAAS, ICP, NIR and Capillary electrophoresis etc.

## REFERENCES

1. I. Taverniers, M. De. Loose. E. Van. Bockstaele. Trac Trends in Analytical Chemistry 23 (2004) (7): 480-490.
2. EURACHEM/CITAC Guide (1998). The Fitness for Purpose of Analytical Methods: A Laboratory Guide to Method Validation and Related Topics.  
<http://www.eurachem.ul.pt/index.htm>. [Accessed June 2007]
3. E. Prichard, V. Barwick, (2007). Quality assurance in analytical chemistry, John Willey & Sons Pty Ltd, Teddington, UK, Pg. 156-176.
4. M. Feinberg, B. Boulanger., W. Dewe, P. Hubert. Anal Bioanal Chem 380 (2004): 502-514.
5. I. Taverniers, M. De Loose, Erik. Van. Bockstaele. Trac Trends in Analytical Chemistry 23 (2004) (8): 535-552.
6. P.H. Hubert, J.J. Nguyen-Huu, B. Boulanger, E. Chapuzet, P. Chiap, N. Cohen, P.A. Compagnon, W. Dewe, M Feinberg, M. Lallier, M. Laurentie, N. Mercier, G. Muzard, C. Nivet, and L. Valat. J. of pharmaceutical and Biomedical Analysis 36 (2004) (3): 579-586.
7. B. Boulanger, P. Chiap, W. Dewe, J. Chrommen, P.H. Hubert, J. of pharmaceutical and biomedical analysis, 32 (2003) 4-5: 753-765.
8. A. Gustavo Gonzalez, M. Angeles. Herrador. Trac Trends in Analytical Chemistry 26 (2007) (3): 227-238.
9. J.M. Hassan, H.E.M. Sayour, S.L. Al-Mehrezi. Analytica Chimica Acta 581 (2007) 1: 13-18.
10. Thesis: S. Maharaj. (2003). the accumulation of heavy metals by aquatic Plants, Durban Institute of Technology, 15-17.

11. US EPA, (2005) water quality criteria documents for the protection of aquatic life in ambient water, GLI Guidelines (EPA-822-R-01-001, 65FR31682).  
<http://www.epa.gov/waterscience/criteria/wqcriteria.html> [Accessed: March 2008]
12. L. Escuder-Gilabert, E. Bonet-Domingo, M.J. Medina-Hernández, S. Sagrado, *Anal Bioanal Chem* 387 (2007): 619-625.
13. E. Bonet-Domingo, L. Escuder-Gilabert, M.J. Medina-Hernández, S. Sagrado. *Anal Chem* 78 (2006): 8113-8120.
14. Metrohm Application Bulletin 241/1e Cadmium and Lead at Ultra Trace Graphite Electrode by anodic stripping voltammetry.  
<http://www.metrohm.com//infocenter/download.php4?doc=applicationsbulletins/pdf/241e.pdf> [Accessed: March 2008].
15. R.L.C Thiago, J.L.C. Paixao, M. Bertoti. *Talanta* 71 (2007) 1: 186-191
16. S. Prasad, A. A. Chetty. *Food Chemistry* 106 (2008): 772-780.
17. O.V. Kaminskaya, E.A. Zakharova, G.B. Slepchenko. *Springer link* 59 (2004) 11: 1090-1096.
18. A. Manova, M. Strelec, F. Cacho, J. Lehotay, E. Beinrohr. *Analytica Chimica Acta* 588 (2007): 16-19.
19. M. Grabarczyk, M. Korolczuk. *Anal Bioanal Chem* 376 (2003): 1115-1118.
20. B. Sliukic, C. E. Banks, A. Crossley, R. G. Compton. *Analytica Chimica Acta* 587 (2007): 240-246.
21. Z. Gao, G. Wang, Z. Zhao. *Analytica Chimica Acta* 230 (1990): 105-112.
22. Official Journal of the European Communities, (1998) COUNCIL DIRECTIVE 98/83/EC. On the quality of water intended for human consumption.  
[http://eurlex.europa.eu/LexUriServ/site/en/oj/1998/l\\_330/](http://eurlex.europa.eu/LexUriServ/site/en/oj/1998/l_330/)

- [.l\\_33019981205en00320054.pdf](#) [Accessed: 6 November 2007].
23. B. Boulanger, P. Chiap, W. Dewé, J.Crommen, Ph. Hubert. J. Pharm Biomed Anal 32 (2003): 753-765.
  24. J.C. Huarte-Mendicoa, I. Astiasaran, J. Bello. Food chemistry 58 (1997) (1-2): 39-42.
  25. P. Nagaraja, M.S.H. Kumar. Analytical sciences 18 (2002): 355-360.
  26. R. Perez-Olmos, R. Herrero, J.L.F.C. Lima, M.C.B.S.M. Montenegro, Food chemistry, 59 (1997) 2: 305-311.
  27. S.E. McMullen, J.A. Cassanova, L.K. Gross, F.J. Schenck. Journal of AOAC International 88 (2005) 6: 1793-1796.
  28. M. M.M. Alawi, M.A. Fresenius'Z. Anal.Chem 377 (1984): 372-375.
  29. P. Reece, H. Hird, Food additives and contaminants 17 (2000) 3: 219-222.
  30. N. Oztekin, M.S. Nutku, F.B. Erim, Food chemistry, 76 (2002): 103-106.
  31. S. Da Silva, D. Shan and S. Cosnier, Sensors and Actuators B: Chemical Volume 103 (2004) (1-2): 397-402.
  32. Metrohm Application Bulletin 70/2e Polarographic determination of nitrate.  
[http://www.metrohm.com/infocenter/download.php4?doc=applications/bulletins/pdf\\_e/070\\_e.pdf](http://www.metrohm.com/infocenter/download.php4?doc=applications/bulletins/pdf_e/070_e.pdf) [Accessed: 6 November 2007].
  33. Food and drug administration, "International conference on Harmonization" Fed registry 62 (1997): 27463-27467.
  34. M. Thompson, S. L. R. Ellison, R. Wood. Pure and Applied Chemistry 74 (2002) 5: 835-855.
  35. <http://www.fda.gov/cder/guidance/1320fnl.pdf>. [accessed: August 2007]
  36. <http://www.fda.gov/cder/guidance/4252fnl.htm>. [Accessed: August 2007]



37. AMC Technical briefs, (2003). Terminology the key to understanding analytical. "Part1: Accuracy, precision, and uncertainty, analytical methods committee." Royal society of chemistry, Volume 13.
38. ISO/IEC 17025, (2005). *General Requirements for the Competence of Testing and Calibration Laboratories*. ISO, Geneva.
39. S. Sagrado, E. Bonet-Domingo, M.J. Medina, Y. Martín-Biosca (2005). *Manual Práctico de Calidad en los Laboratorios. Enfoque ISO 17025 2<sup>nd</sup> ed*; AENOR, Madrid.
40. E. Hund, D.L. Massart, J. Smeyers-Verbeke. *Trends in analytical chemistry* 20 (2001) 8: 394-406.
41. 2002/675/EC Council directive (2002). Implementing Council Directive 96/23/EC concerning the performance of analytical methods and the interpretation of results.
42. M.A. Herrador, A.G. Asuero, A.G. González *Chemom Intell Lab Syst* 79 (2005): 115-122.
43. A. Maroto, J. Riu, R. Boqué, F.X. Rius. *Anal Chim Acta* 391 (1999):173-185
44. S. Kuttatharmmakul, D.L. Massart, J. Smeyers-Verbeke. *J Chemom Intell Lab Syst*. 52 (2000): 61-73.
45. E. Bonet-Domingo, S. Grau-González, Y. Martín-Biosca, M.J. Medina-Hernández, S. Sagrado. *Anal Bioanal Chem* 387 (2007): 2537-2545.
46. D.L. Massart, B.M.G. Vandeginste, L.M.C. Buydens, S. De Jong, P.J. Lewi, J. Smeyers-Verbeke. (1997) *Handbook of chemometrics and Qualimetrics Part A*. Amsterdam: Elsevier.
47. M. Valcàrcel, A. Ríos, *Accred. Qual. Assur* 8 (2003): 78-81.
48. RSC Analytical Methods Committee. (2003) AMC Technical Briefs No.12, March.

- <http://www.rsc.org/Membership/Networking/InterestGroups/Analytical/AMC/technicalBriefs.asp> [Accessed: 11 January 2008].
49. M. Thompson, (2004). The amazing Horwitz function. Royal society of chemistry No.17.
50. Eurachem Guide. (2007). the use of uncertainty information in compliance assessment, 1<sup>st</sup> edition. <http://www.eurachem.ul.pt/index.htm>. [Accessed: October 2007]
51. Eurachem/CITAC Guide. (2000) Quantifying uncertainty in analytical measurement. 2nd edition. <http://www.eurachem.ul.pt/index.htm>. [Accessed: October 2007]
52. (ECC), European Council Regulation. (1993) Report on the relationship between analytical results, measurement uncertainty, Recovery factors and the provisions of EU food and feed legislation, with particular reference to community legislation concerning contaminants in food.
53. ISO/DTS 21748. (2003). Guide to the use of repeatability, reproducibility and trueness estimates in measurement uncertainty estimates. ISO, Geneva.
54. S. N. Deming, S. L. Morgan. Experimental design: A chemometric approach. Amsterdam: Elsevier. 346-356.
55. T.A.M. Msangati, J.C. Ngila, South African Journal of Chemistry 56 (2003): 5-9.
56. T.A.M. Msangati, J.C. Ngila, Talanta 58 (2002): 605-610.
57. G. Henze, (2007). Introduction to Polarography and Voltammetry, Metrohm Monograph: (3-43), (49-52).
58. E.P. Achterberg, C. Braungardt, Analytica Chimica Acta 400 (1999): (1-3) 381-397.
59. S. Legeai, K. Soropogui, M. Cretinon, O. Vittori, A. Hereen De Oliveira, F.

- Barbier, M.F. Grenier-Loustalot, *Anal Bioanal Chem*, 383 (2005): 839-847.
60. M.M. Correia dos Santos, M.F. Vilhena, M.L. Simões Gonçalves, *Analytica Chimica acta* 441 (2001) 2: 191-200.
61. L.C. Mello, A. Claudino, I. Rizatti, R.L. Bortoluzzi, D.R. Zanette, *Journal of Brazilian Chemical Society*, 16 (2005) 3A: 308-315.
62. Wang J. *Analytical electrochemistry*: VCH publishing, 2<sup>nd</sup> edition, 68-76
63. E. P. Achterberg, C. Braungardt, D. Whitworth. *J. Ocean Sci. Technol*, 1 (2000): 227-247.
- 64 The Math Works Natick MA USA  
  
[http://www.mathworks.com/products/product\\_listing/index.html](http://www.mathworks.com/products/product_listing/index.html)  
  
[Accessed: 6 November 2007].
- 65 E .Mullins, (2003) *Statistics for the Quality Control Chemistry Laboratory*; Royal Society of Chemistry, Cambridge.
- 66 Codex Alimentarius Commission. (2001) Codex Committee on Methods of Analysis and sampling CX/MAS 01/04, Criteria for evaluating acceptable methods of analysis for Codex purposes, Agenda Item 4a, 23<sup>rd</sup> session, Budapest, Hungary. [ftp://ftp.fao.org/codex/ccmas26/ma26\\_05e.pdf](ftp://ftp.fao.org/codex/ccmas26/ma26_05e.pdf) [Accessed: March 2008]
- 67 E. Prichard. (1995). *Quality in the Analytical Laboratory*, ACOL series, Wiley, Chichester, West Sussex, UK.
- 68 AOAC International, Method Validation Programs (OMA/PMV Department), Including appendix D (2000): Guidelines for Collaborative study procedures to Validatate characteristics of a method of analysis, <http://www.aoac.org/vmeth/devmethno.htm> [Accessed: March 2008]
- 69 K. Bisetty, N. J. Gumede, L. Escuder-Gilabert and S. Sagrado. *Anal Bioanal Chem* (2008) 392: 277-286.

## APPENDICES

### Appendix 1: Definitions

**Accuracy:** ‘The closeness of agreement between a test result and the accepted reference value’

- ‘A quantity referring to the differences between the mean of a set of results or an individual result and the value which is accepted as true or correct value for the quantity measured.’

### Accuracy (of a Measuring Instrument):

‘The ability of a measuring instrument to give responses close to a true value’.

### Bias:

‘The difference between the expectation of the test results and an accepted reference value.’

### Calibration curve:

‘The graphical representation of measuring signal a function of quantity of analyte’

### Chemometrics:

‘Is the application of mathematical or statistical methods to chemical data. The International Chemometrics Society (ICS) offers the following definition: Chemometric research spans a wide range of different methods which can be applied in chemistry’

### Error of measurement:

‘The result of a measurement minus the true value of the measurand’

**Qualimetrics:**

‘is an area of great importance to laboratories opting for compliance with a quality system (GLP: ISO 9000). It is the synergy between chemometrics and quality assurance in demonstrating and maintaining quality measurements’

**Random Error:**

‘Result of a measurement minus that would result from an infinite number of measurements of the same measurand carried out under repeatability conditions.

**Systematic error:**

‘Mean that would result from an infinite number of measurements of the same measurand.’

**Fitness for purpose:**

‘Degree to which data produced by measurement process enables a user to make technically and administratively correct decision for a stated purpose’.

**Limit of detection:**

‘The lowest content that can be measured with reasonable statistical certainty’

**Limit of quantitation:**

‘The content equal to or greater than the lowest concentration point on the calibration curve’

**Measurand:**

‘Particular quantity subject to measurement’

**Measurement:**

‘Set of operations having the object of determining a value of a quantity’

**Precision:**

‘The closeness of agreement between independent test results obtained under stipulated conditions’

**Intermediate precision:**

‘Intermediate precision expresses within laboratories variation: different days, different analysts, different equipment, etc.’

**Proficiency Testing:**

‘A periodic assessment of the performance of individual laboratories and groups of laboratories that is achieved by the distribution by an independent testing body of typical materials or unsupervised analysis by the participants’

**Reference Material (RM):**

‘Material or substance one or more of whose property values are sufficiently homogeneous and well established to be used for the calibration of an apparatus, assessment of a measurement method, or for assigning values to materials’

**Certified Reference Materials:**

‘Reference material, accompanied by a certificate, one or more of whose property values are certified by a procedure, which establishes its traceability to an accurate realization of the unit in which the property values are expressed, and for which each certified value is accompanied by an uncertainty at a stated level of confidence’.

**Repeatability:**

‘Precision under repeatability conditions, i.e. conditions where independent test results are obtained with the same method on identical test items in the same laboratory by the same operator using the same equipment within short intervals of time’.

**Repeatability (of results of measurements):**

‘Closeness of the agreement between the results of successive measurement of the same measurand carried out in the same condition of measurement.’

**Repeatability standard deviation:**

‘The standard deviation of test results obtained under repeatability conditions.’

**Reproducibility:**

‘Precision under reproducibility conditions, i.e. conditions where test results are obtained with the same method on identical test items in different laboratories with different operators using different equipment.’

**Reproducibility standard deviation:**

‘The standard deviation of test results obtained under reproducibility conditions.’

**Robustness:**

‘The robustness of an analytical procedure is a measure of its capacity to remain unaffected by small, but deliberate variations in method parameters and provides an indication of its reliability during normal usage.’

**Sensitivity:**

‘The change in the response of a measuring instrument divided by the corresponding change the stimulus.’

**Specificity:**

‘The ability of a method to measure only what it is intended to measure.’

**Traceability:**

‘Property of the result of a measurement or the value of a standard whereby it can be related with a stated uncertainty, to stated references, usually national or international standards (i.e. through an unbroken chain comparison).’

**Trueness:**

‘The closeness of agreement between the average values obtained from a large set of test results and an accepted reference value.’

**Uncertainty (of a measurement) i.e. measurement uncertainty:**

‘Parameter associated with the result of a measurement, that characterizes the dispersion of the values that could reasonable be attributed to the measurand.’

**Expanded uncertainty:**

‘U- A quantity defining an interval about a result of a measurement that may be expected to encompass a large fraction of the distribution of values that could reasonable be attributed to the measuarand.’

**Validation:**

Confirmation by examination and provision of objective evidence that the particular requirements for a specified intended use are fulfilled.

**Method validation:**

1. The process of establishing the performance characteristics and limitations of a method and the identification of the influences which may change these characteristics and to what extent. Which analytes can it determine in which matrices in the presence of which interferences? Within these conditions what level of precision and accuracy can be achieved?
2. The process of verifying that a method is fit for its purpose, i.e. for use in solving a particular analytical problem.

**True Value:**

Is the value consistent with the definition of a given particular quantity.

**Verification:**

Confirmation by examination and provide objective evidence that specified requirements have been fulfilled.



## Appendix 2 Symbols for the variables used in the text.

Symbol	Meaning
$AL$	Action limits for the mean control chart
$CL$	Confidence level
$E$	Estimated relative error (bias) for $\mu$
$[E \pm U(E)]$	Uncertainty interval for $E$ (e.g. estimated by means of Monte Carlo simulation)
$E_{LIM}$	Acceptation limit for the relative error in trueness assessment
$e_{i,j}$	Random error under repeatability conditions for the $i$ th replicate and $j$ th run
$f_j$	Random run effect for the $j$ th run
$i$	Index of replicate (from 1 to $Nr$ )
$j$	Index of run (from 1 to $Ns$ )
$k$	Coverage factor ( $k = 2$ is a common option)
$LD$	Limit of detection
$LD_{LIM}$	Acceptation limit for $LD$
$LQ$	Limit of quantification
$\mu \pm s_\mu$	Experimental mean (run mean) and standard deviation corresponding to the vector of $j$ means. $s_\mu$ can be estimated the between-run mean square ( $MS_{run}$ ) from ANOVA of $\mathbf{X}$ : $s_\mu^2 = MS_{run}/Nr$ [41]
	Certified value of a CRM
$\mu_{CRM}$	
$\mu_{CRM} \pm u_{CRM}$	Certified value and its combined standard uncertainty related to a CRM or its solution, where $u_{CRM}$ can be estimated by the (classical) bottom-up approach

[36]

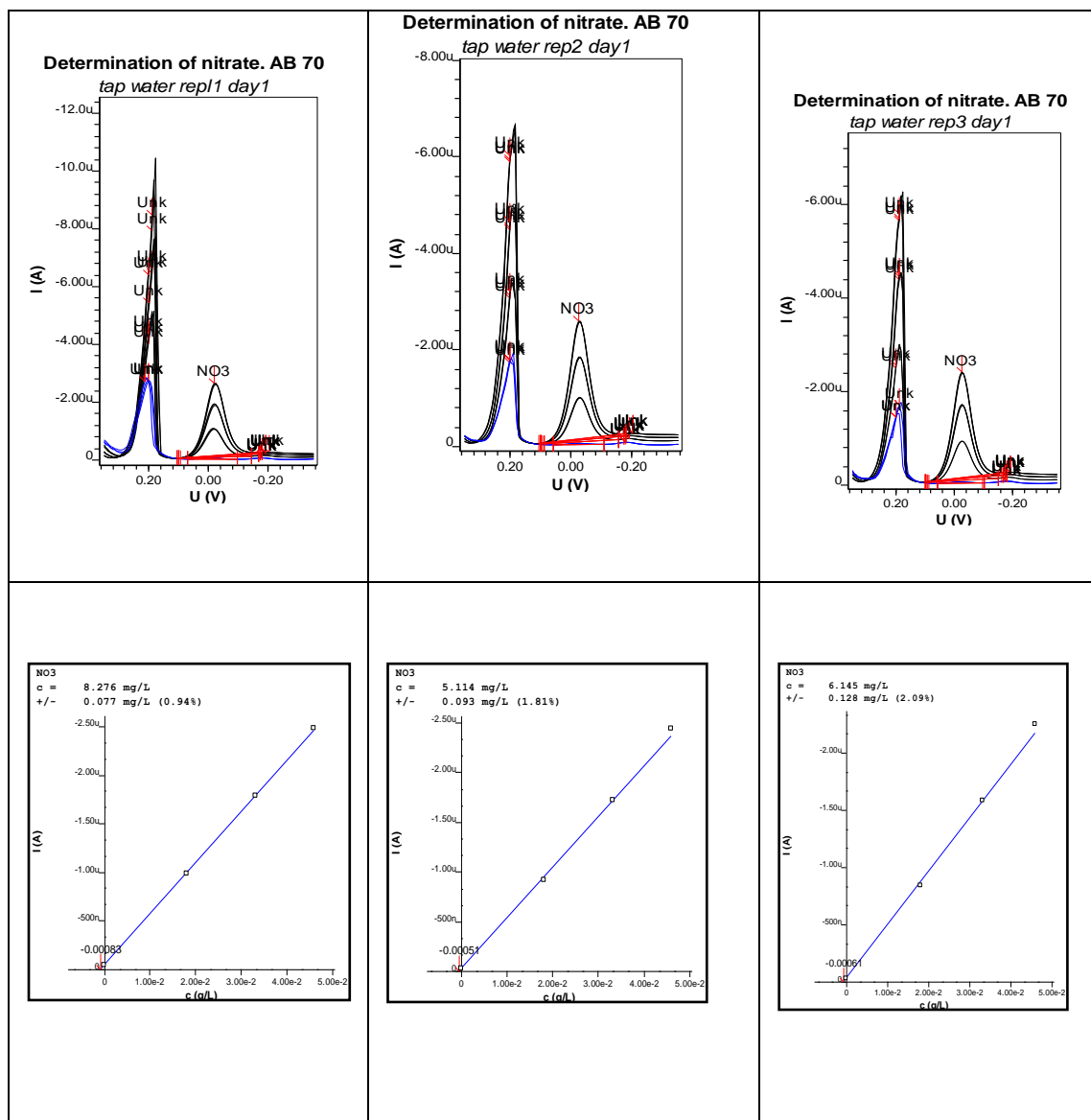
$Nr$	Number of independent replicates per run in the validation stage
$Nr'$	Number of independent replicates per working session in the routine work stage (It could be different for verification, control and sample analysis periods)
$Ns$	Number of runs during the validation stage. The same changes (time periods, analyst, and equipments) expected between working routine sessions or between-batches should be accounted between-runs. During a run $Nr$ independent replicates are performed.
$RSDr$	Relative standard deviation in repeatability conditions
$RSDrun$	Between-run relative standard deviation
$RSDi$	Relative standard deviation in intermediate precision conditions. $RSDi$ can be calculated from repeatability ( $RSDr$ ) and between-run ( $RSDrun$ ) relative standard deviations: $RSDi = (RSDr^2 + RSDrun^2)^{0.5}$ , which can be derived from the definition of $s_i^2$
$[RSDi \pm U(RSDi)]$	Uncertainty interval for $RSDi$ (e.g. estimated by means of Monte Carlo simulation)
$RSDi_{LIM}$	Acceptation limit for $RSDi$ in precision assessment. $RSDi_{LIM}$ can be calculated from repeatability ( $RSDr_{LIM}$ ) and between-run ( $RSDrun_{LIM}$ ) relative standard deviation limits: $RSDi_{LIM} = (RSDr_{LIM}^2 + RSDrun_{LIM}^2)^{0.5}$ , which can be derived from the definition of $s_i^2$
$s_i^2$	Variance in intermediate precision conditions. $s_i^2 = s_{run}^2 + s_r^2$ [23]
$s_r^2$	Variance in repeatability conditions. It can be estimated from the residual mean square ( $MS_r$ ) from ANOVA of $\mathbf{X}$ : $s_r^2 = MS_r$ [23]
$s_{run}^2$	Between-run variance. It can be estimated from the residual and between-run mean squares ( $MS_r$ and $MS_{run}$ ) from ANOVA of $\mathbf{X}$ : $s_{run}^2 = (MS_{run} - MS_r)/Nr$

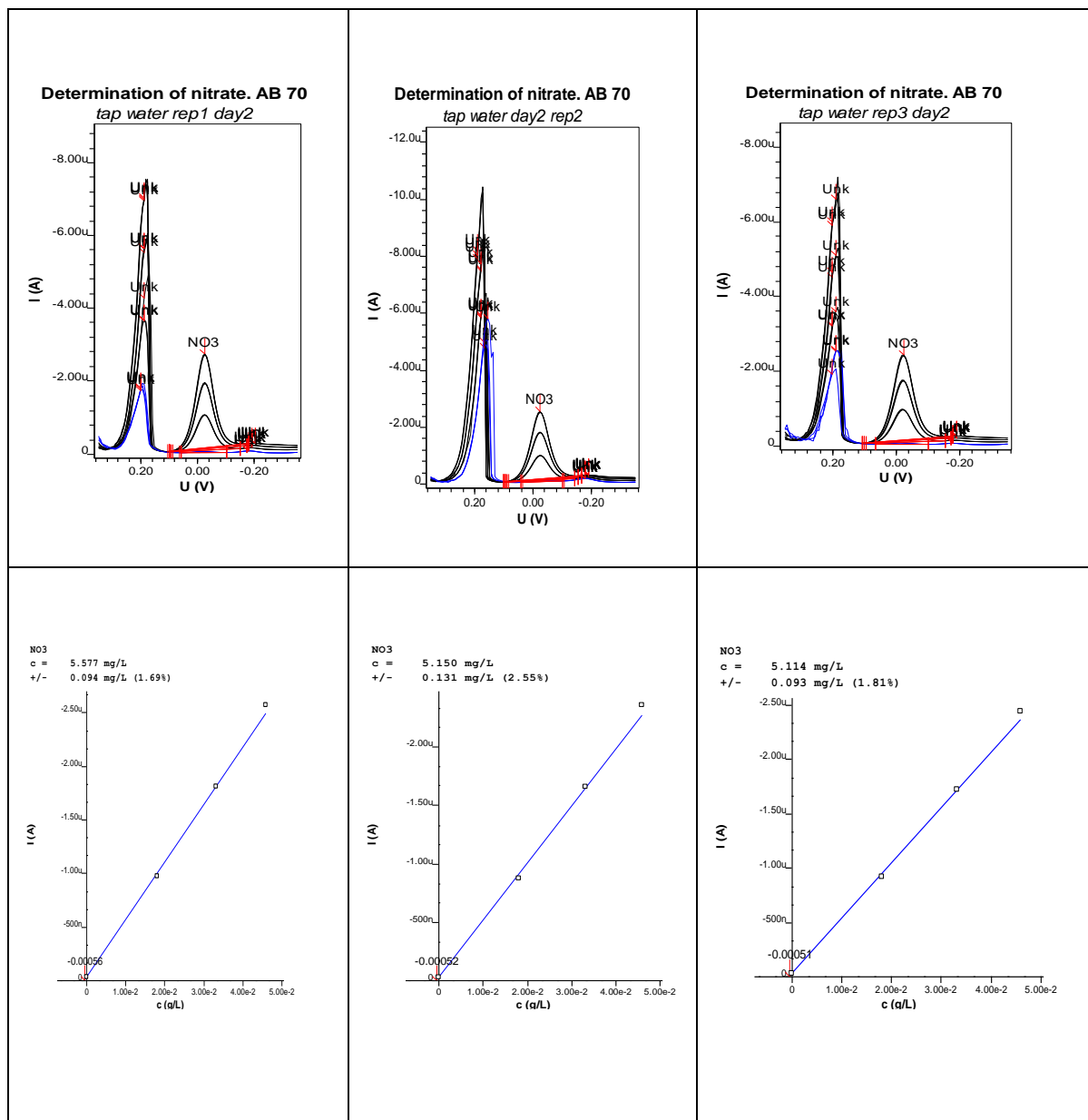
[23]

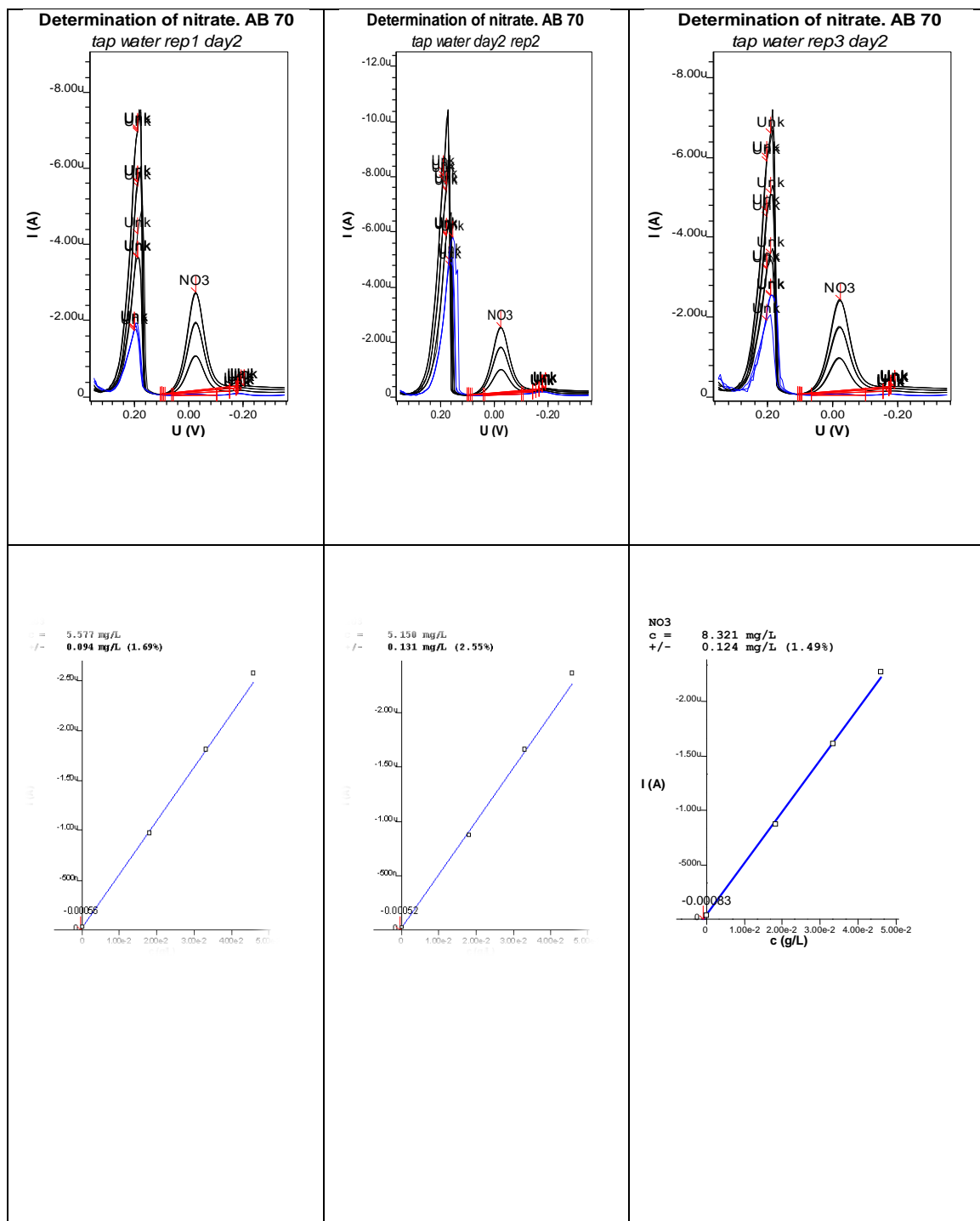
$S_u$	Process standard deviation (IQC) estimated according the u-approach [23]
$\mathbf{X}$	Matrix of size $N_r \times N_s$ (rows by columns) representing the experimental design to perform accuracy validation in intermediate precision conditions
$\mathbf{x}_j$	Vector of replicate values in a given conditions (run $j$ )
$x_{ij}$	Data related to the $i$ th replicate of the $j$ th run (or batch)
$U$	Estimated combined standard uncertainty for the CRM concentration level
$u_{\text{CRM}}$	Combined standard uncertainty related to a CRM or its solution
$u$	Estimated combined standard uncertainty. It can be estimated according the u-approach [12, 13]: $S_u = u = (s_{\text{run}}^2 + s_r^2 / N_r + s_\mu^2 / N_s + u_{\text{MRC}}^2)^{0.5}$
$U$	Estimated expanded uncertainty for the CRM concentration level
$U_s$	Estimated expanded uncertainty for samples concentration level
$WL$	Warning limits for the mean control chart
$z$ (z-score)	$z = (x - \mu_0) / \sigma_0$ ; where $\mu_0$ and $\sigma_0$ are the assigned analyte value and the accepted standard deviation for the method, respectively [34].

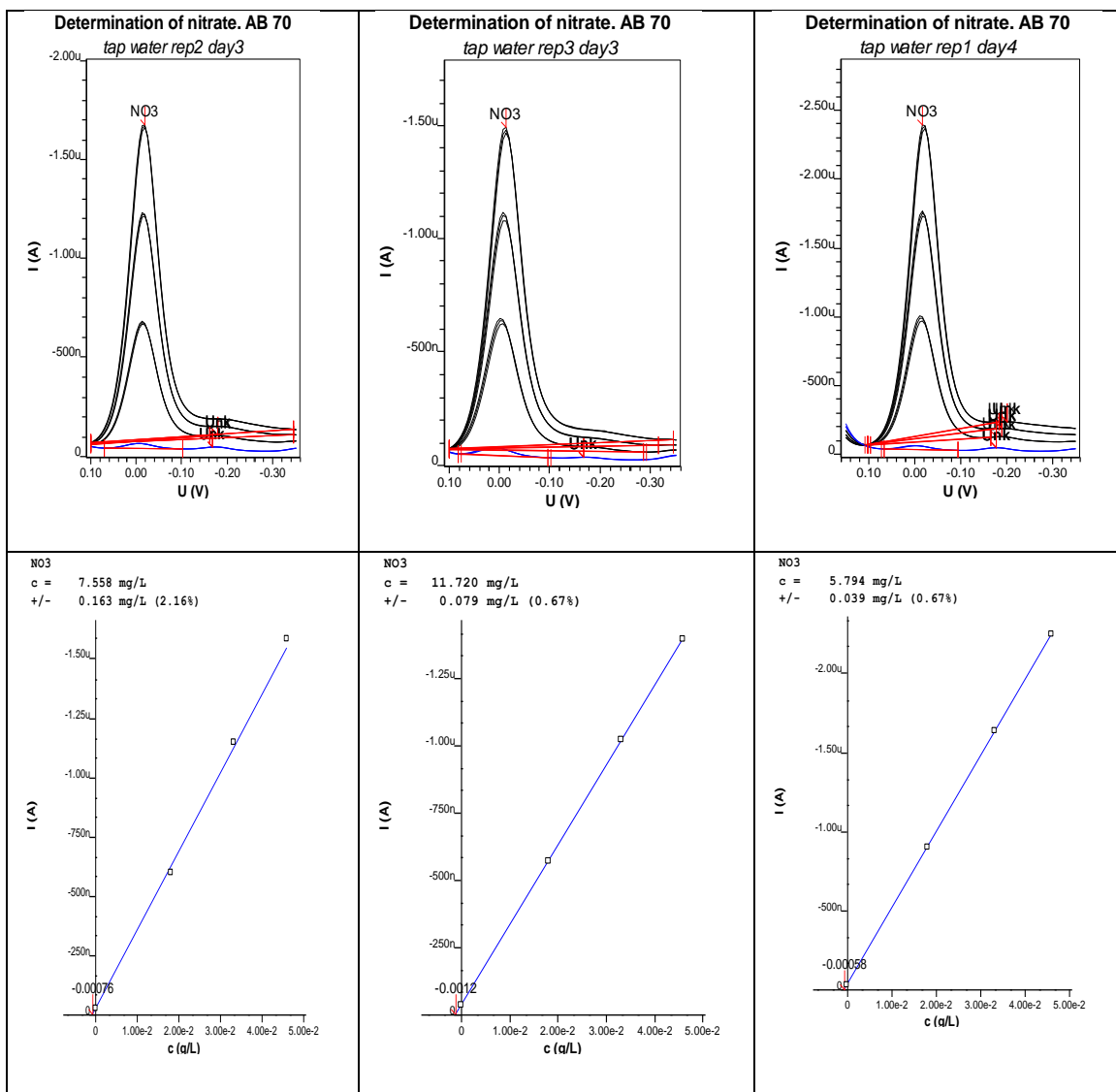
---

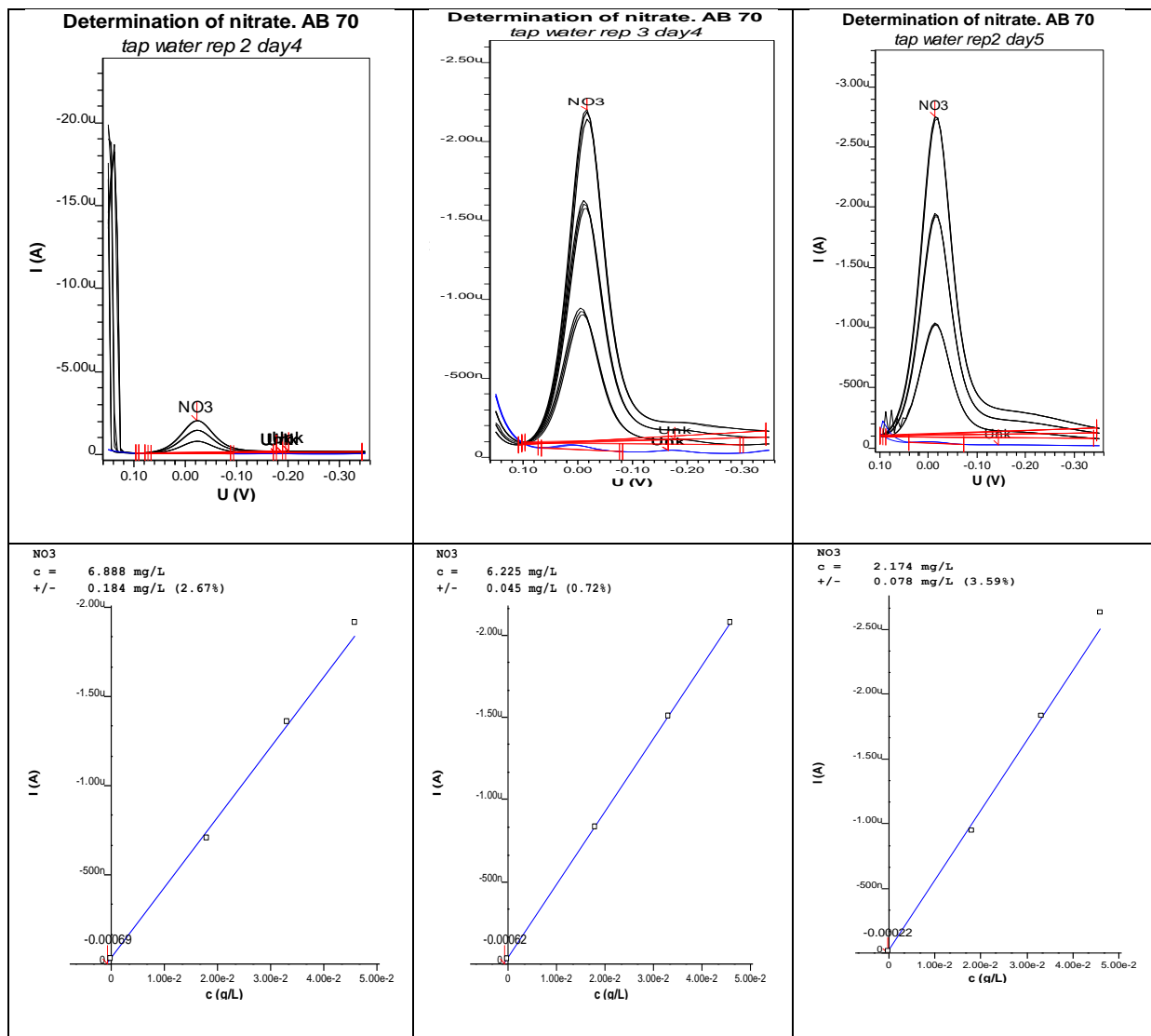
### Appendix 3 – Results for pre-validation (case study 1)



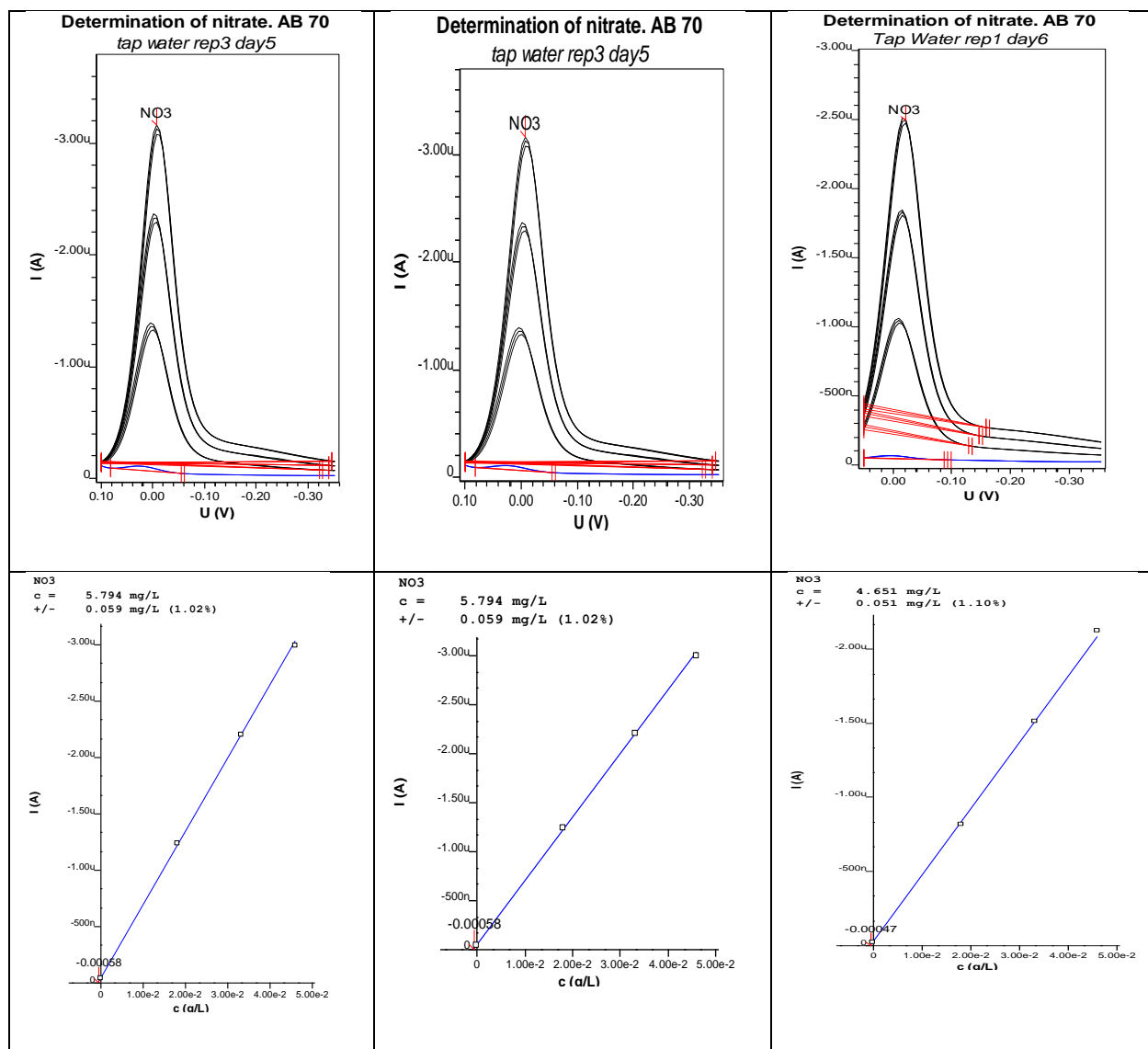


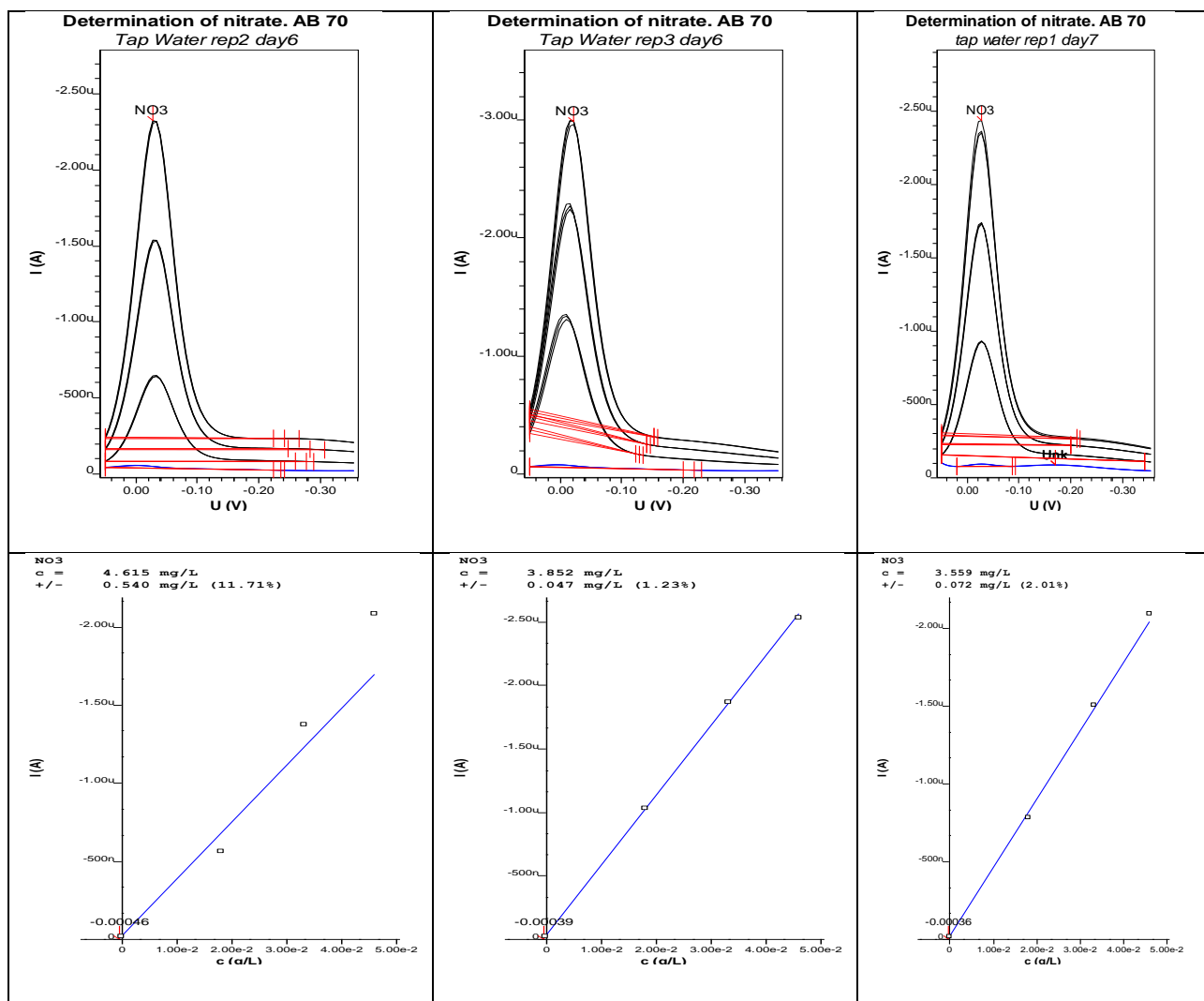


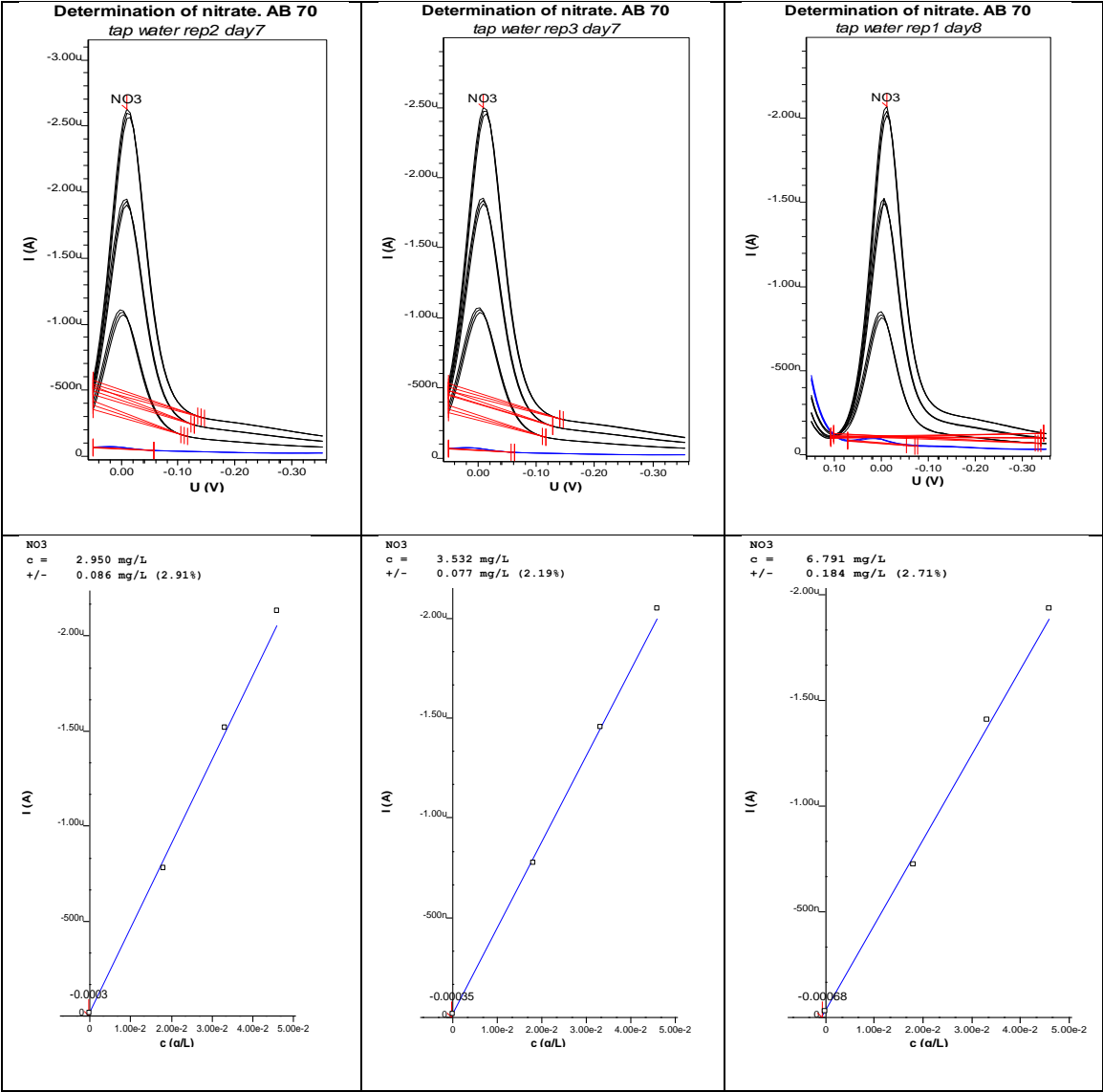


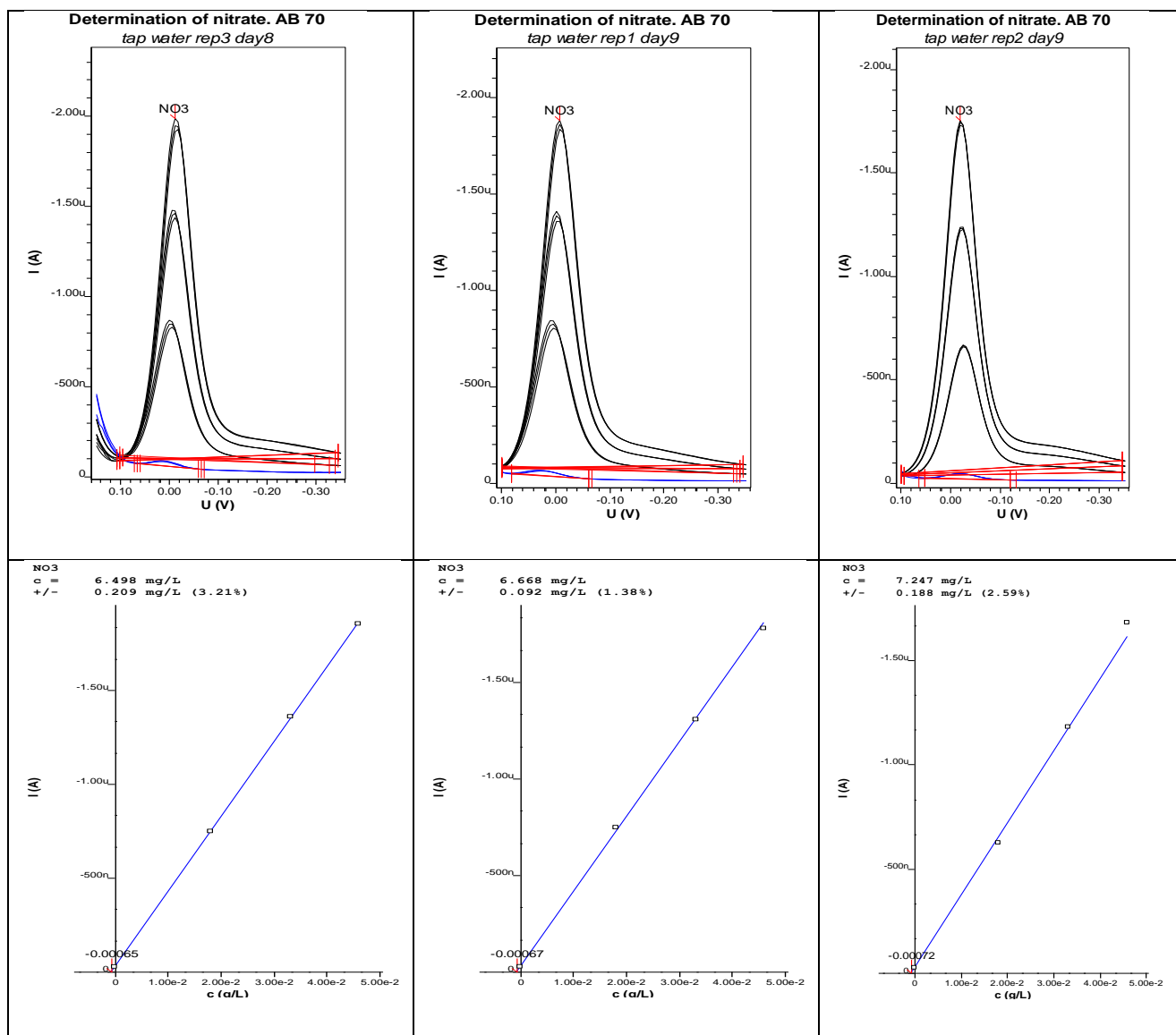


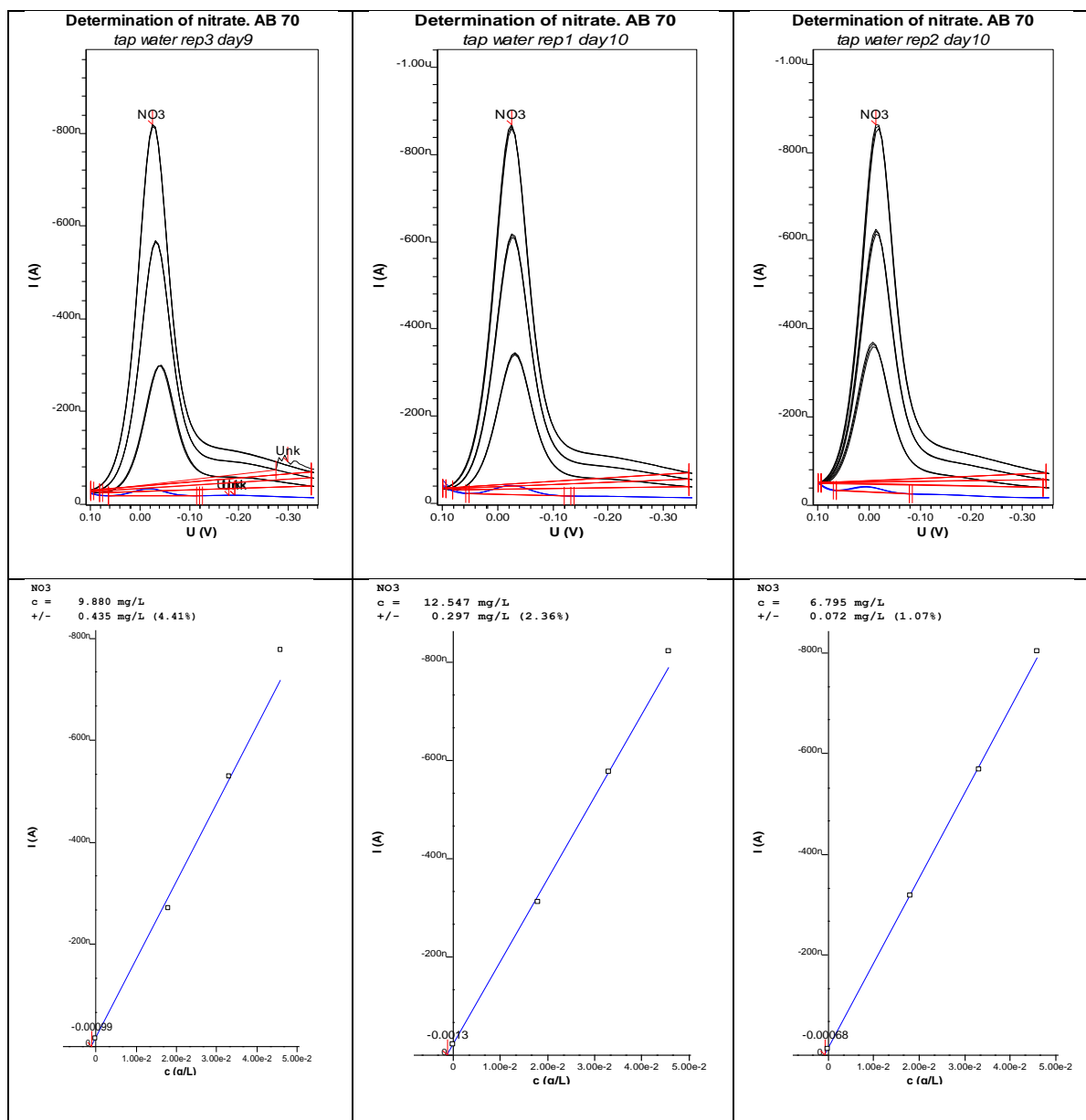


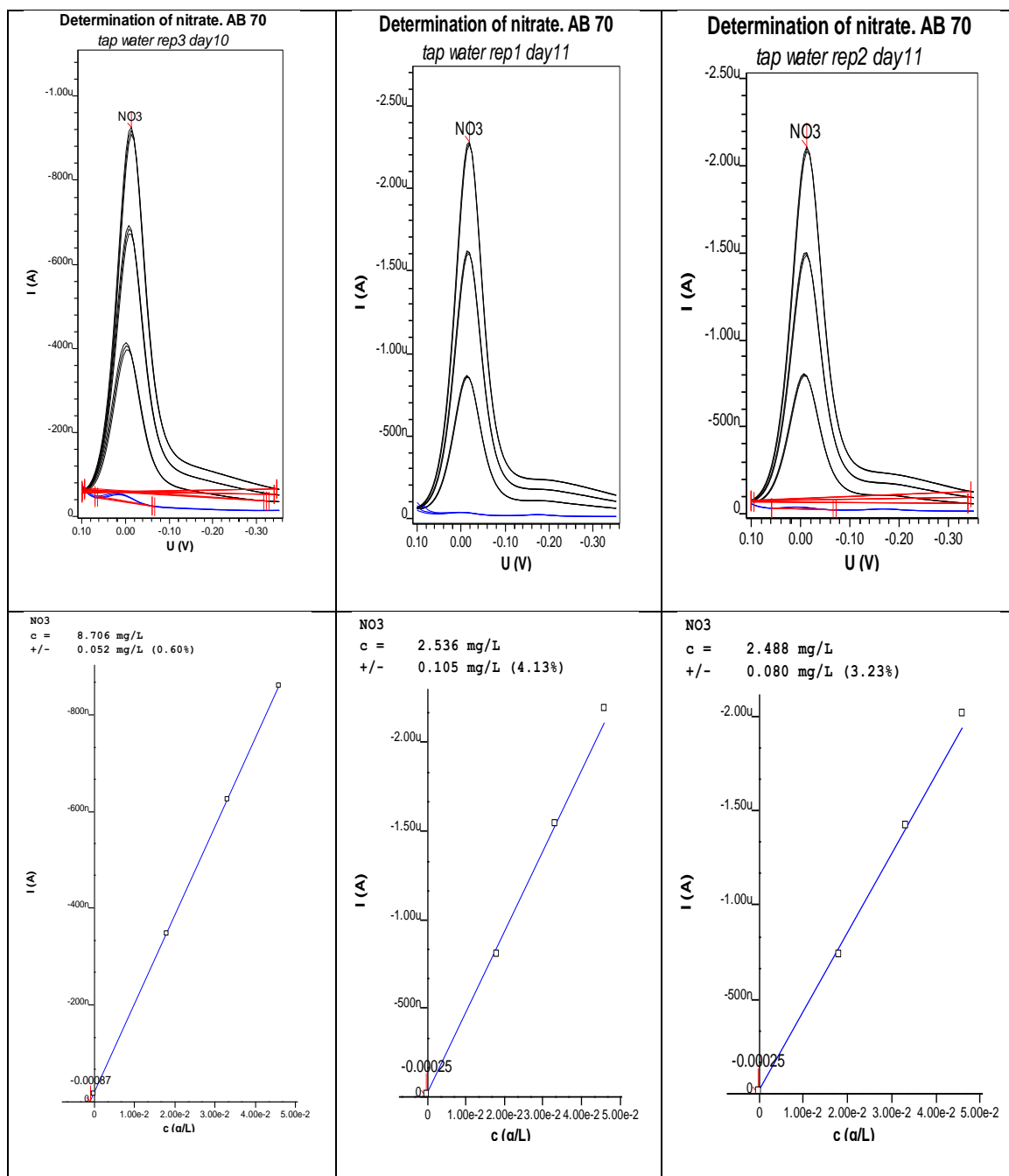


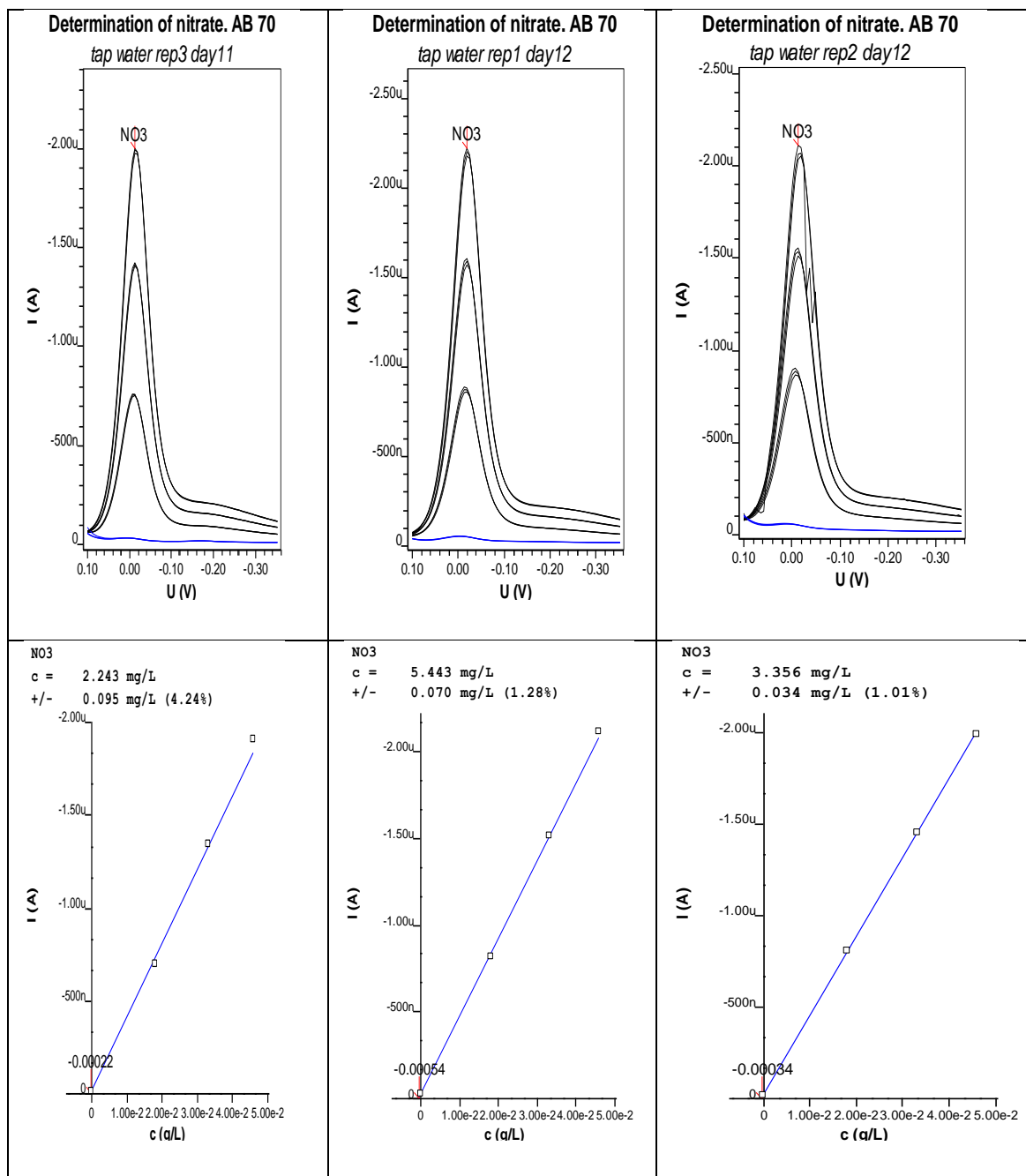


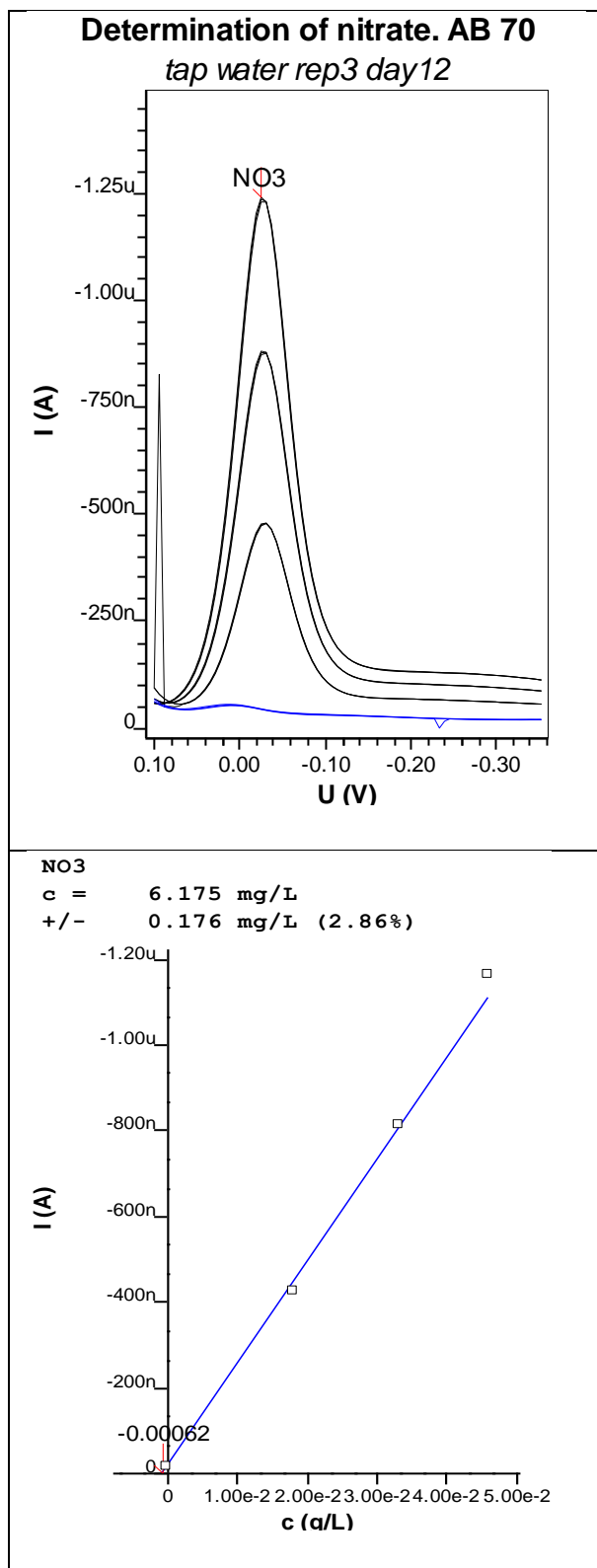






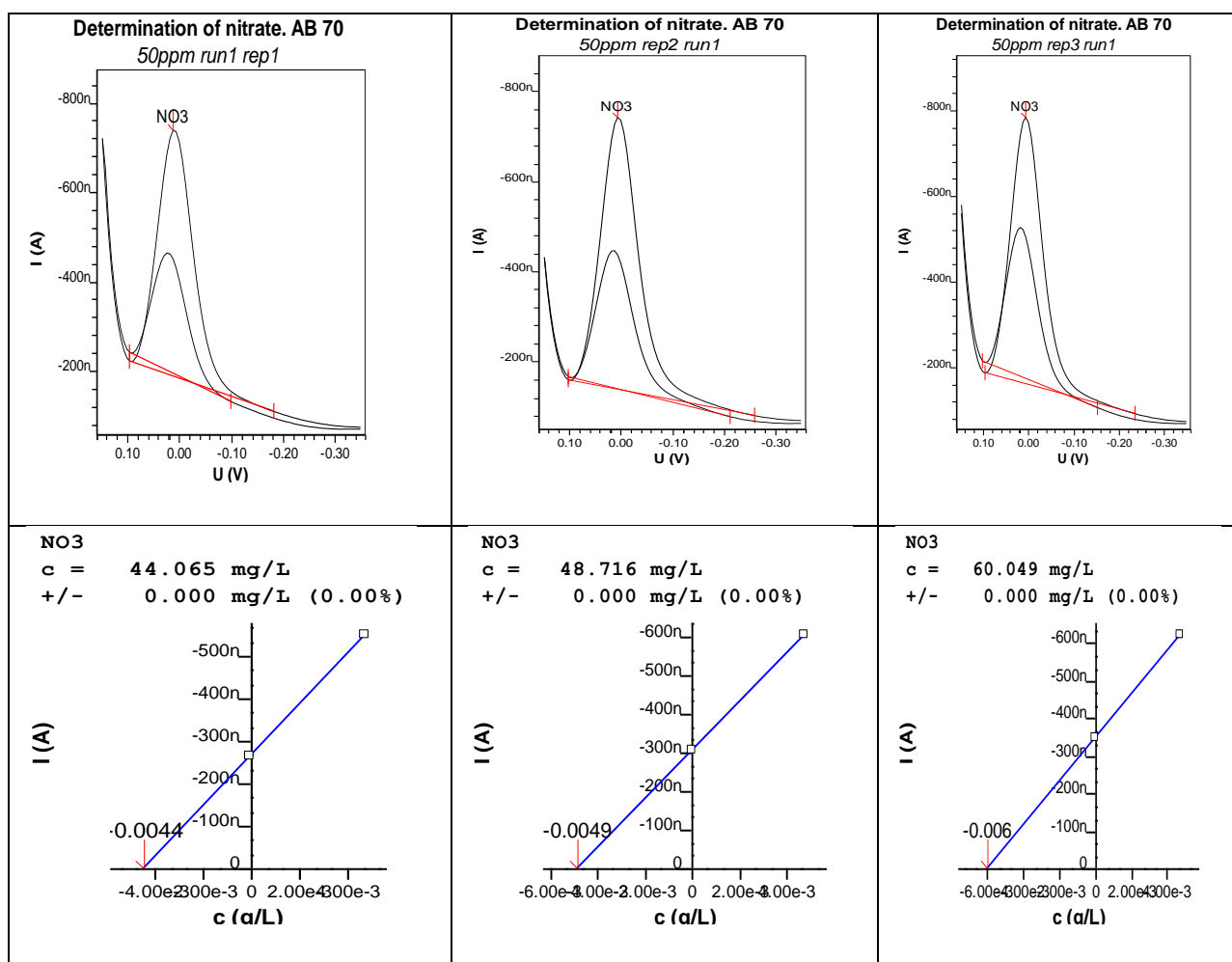


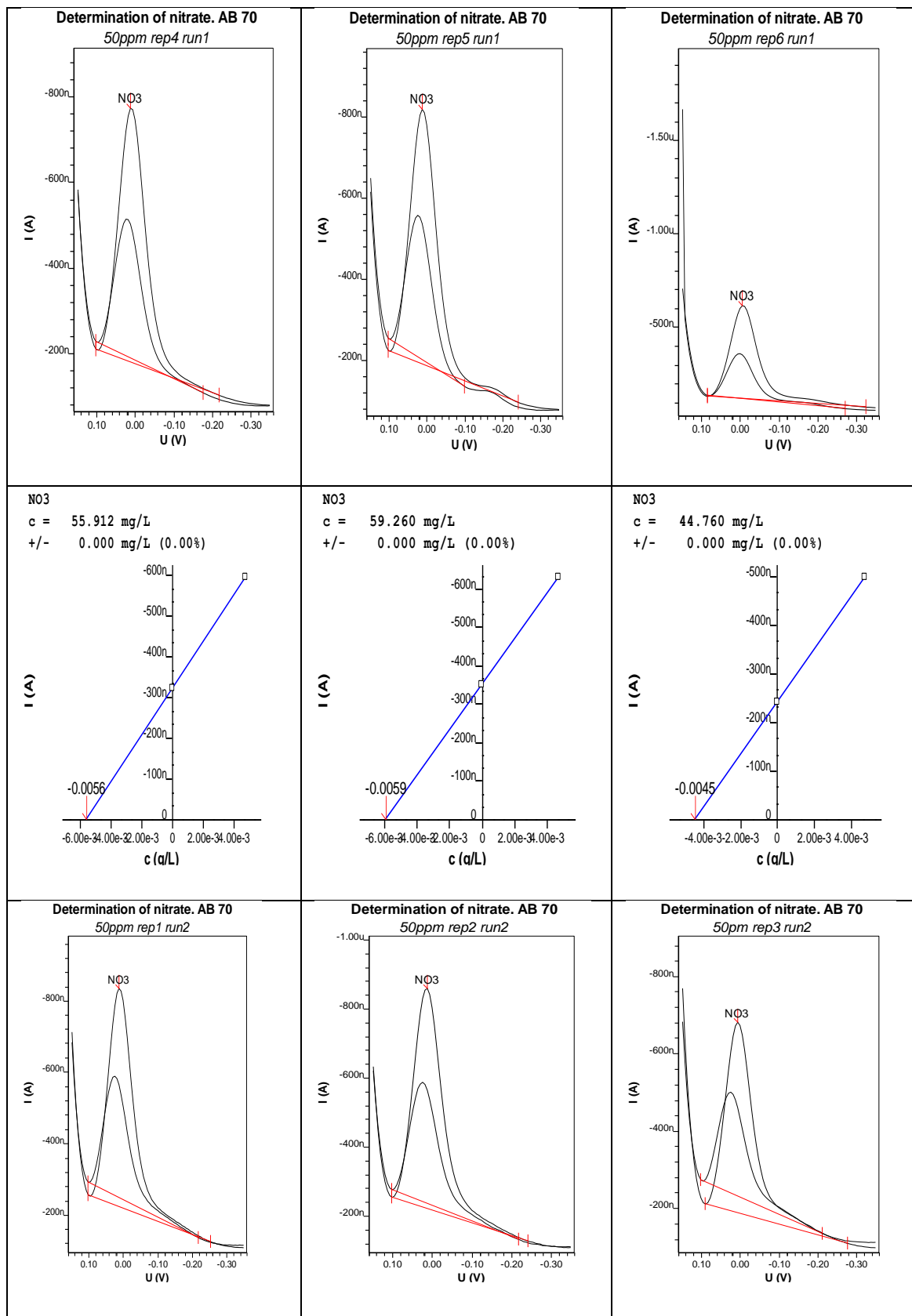


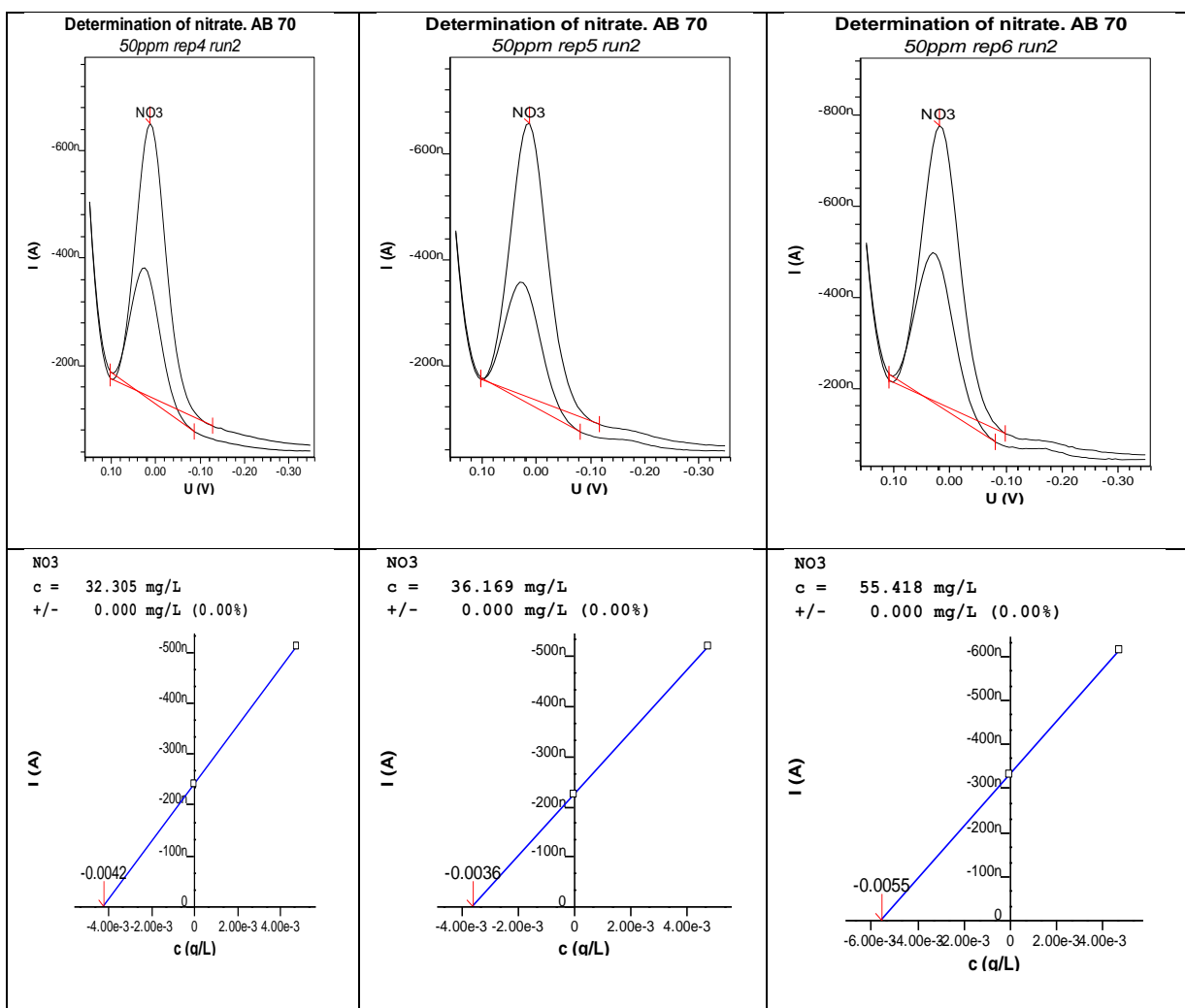
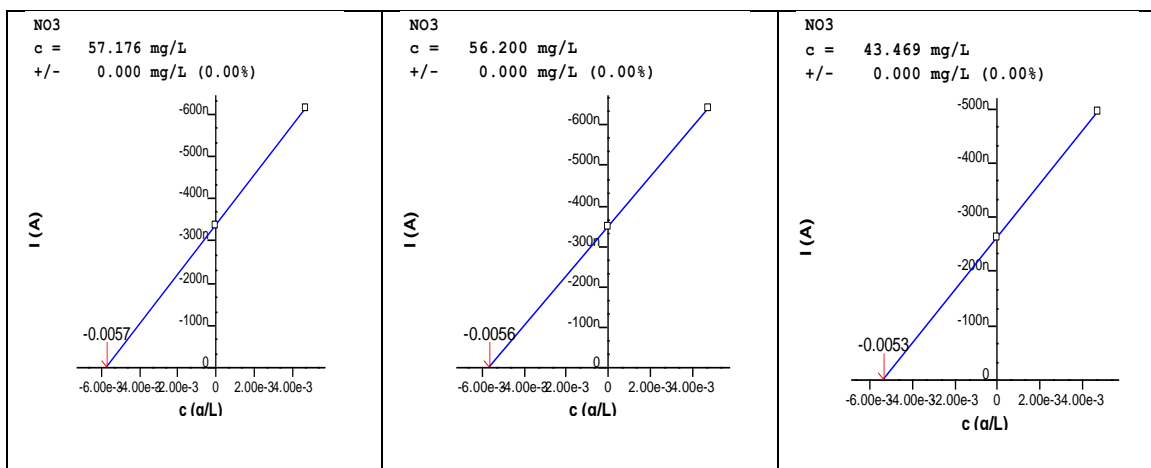


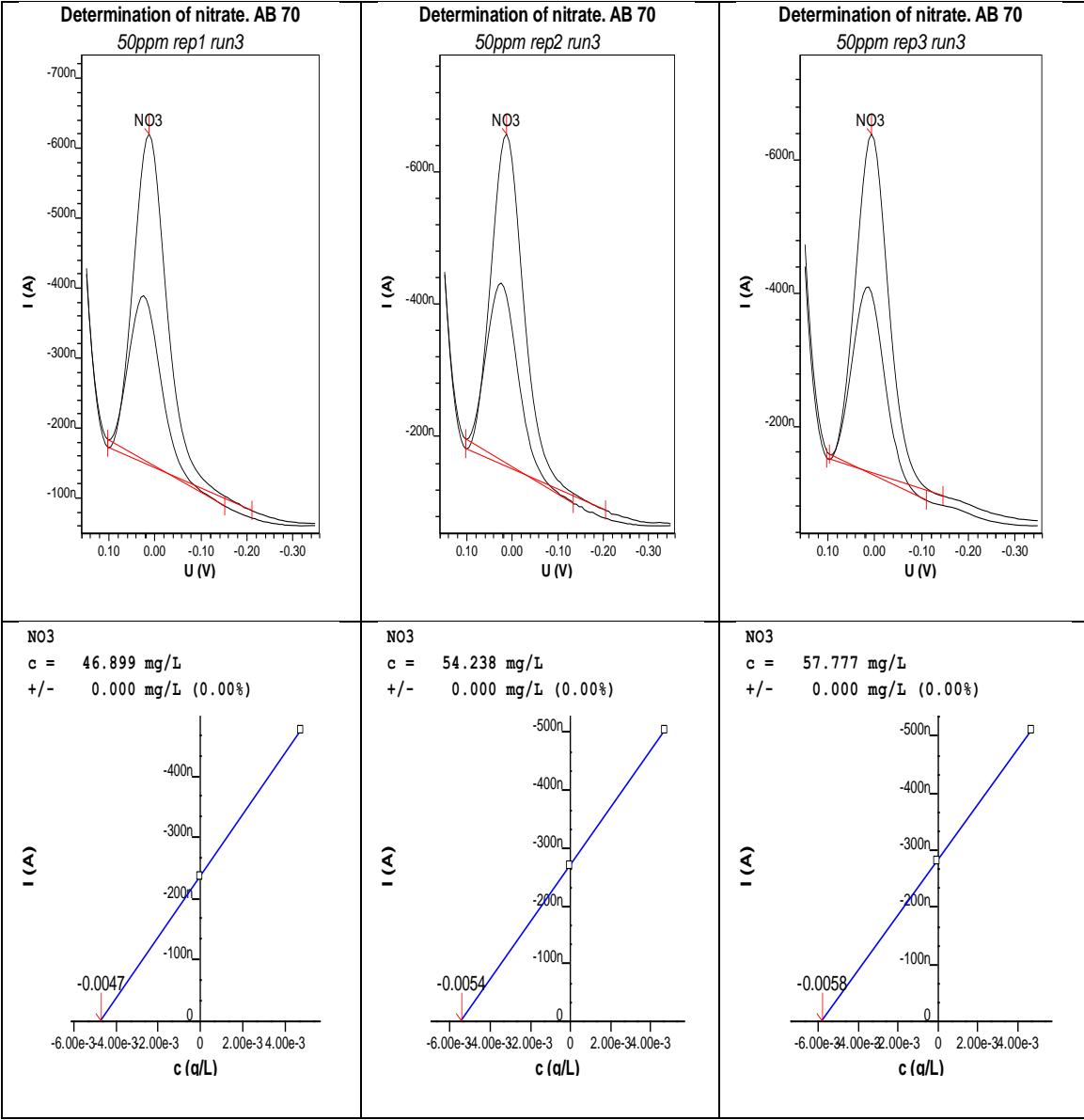


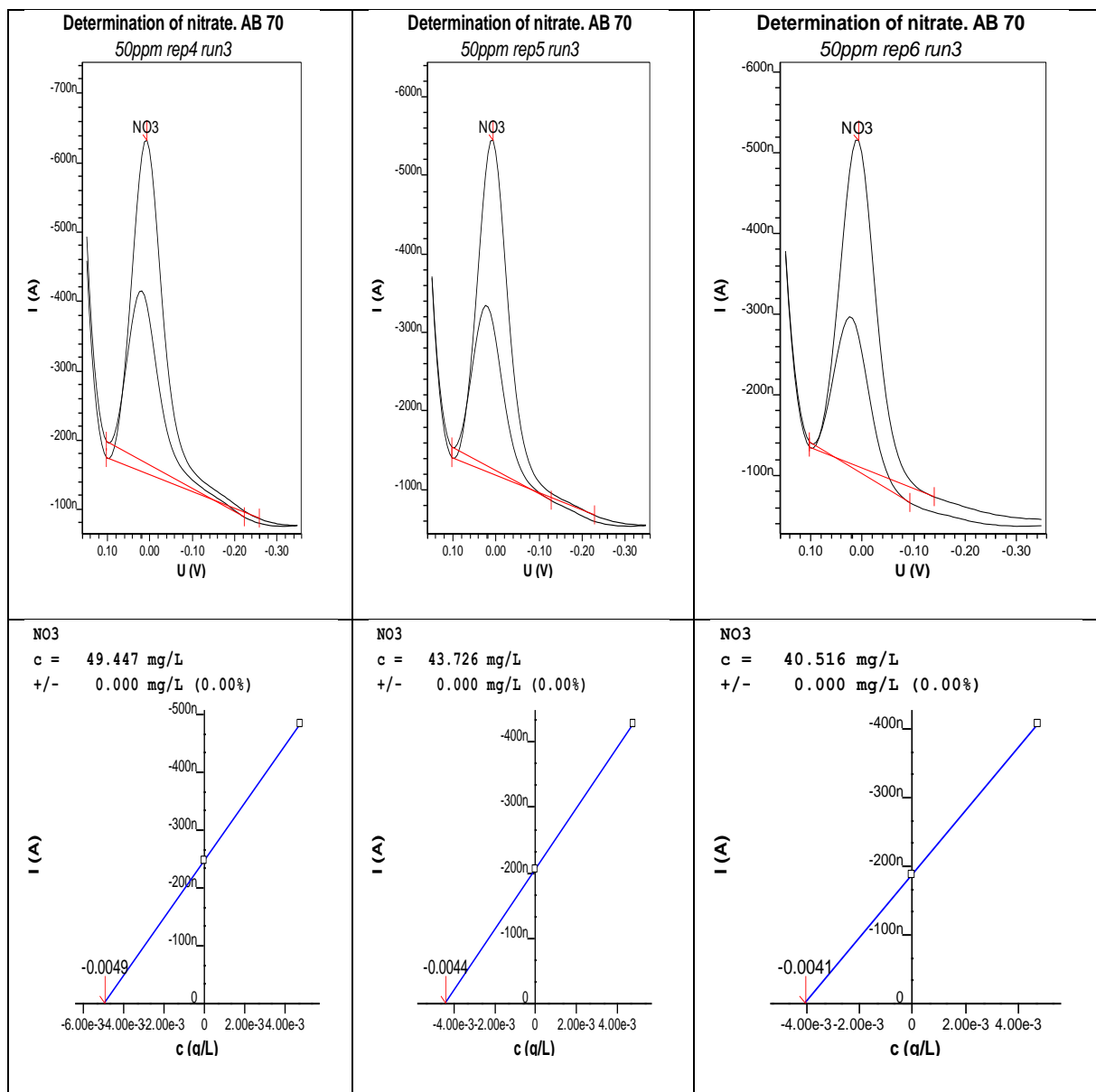
## Appendix 4 – Results for a validation protocol of 50ppm nitrate EU parametric value

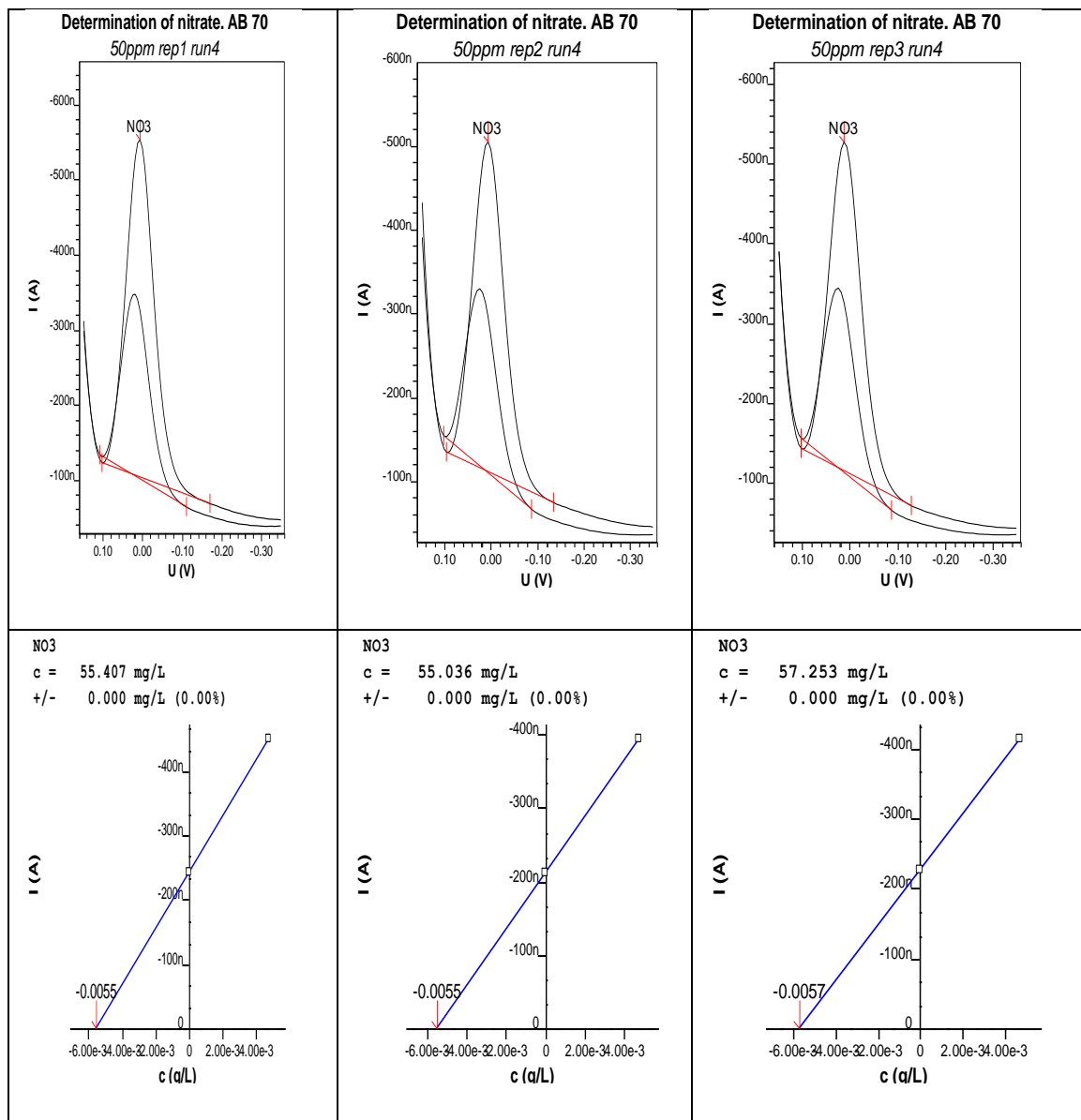


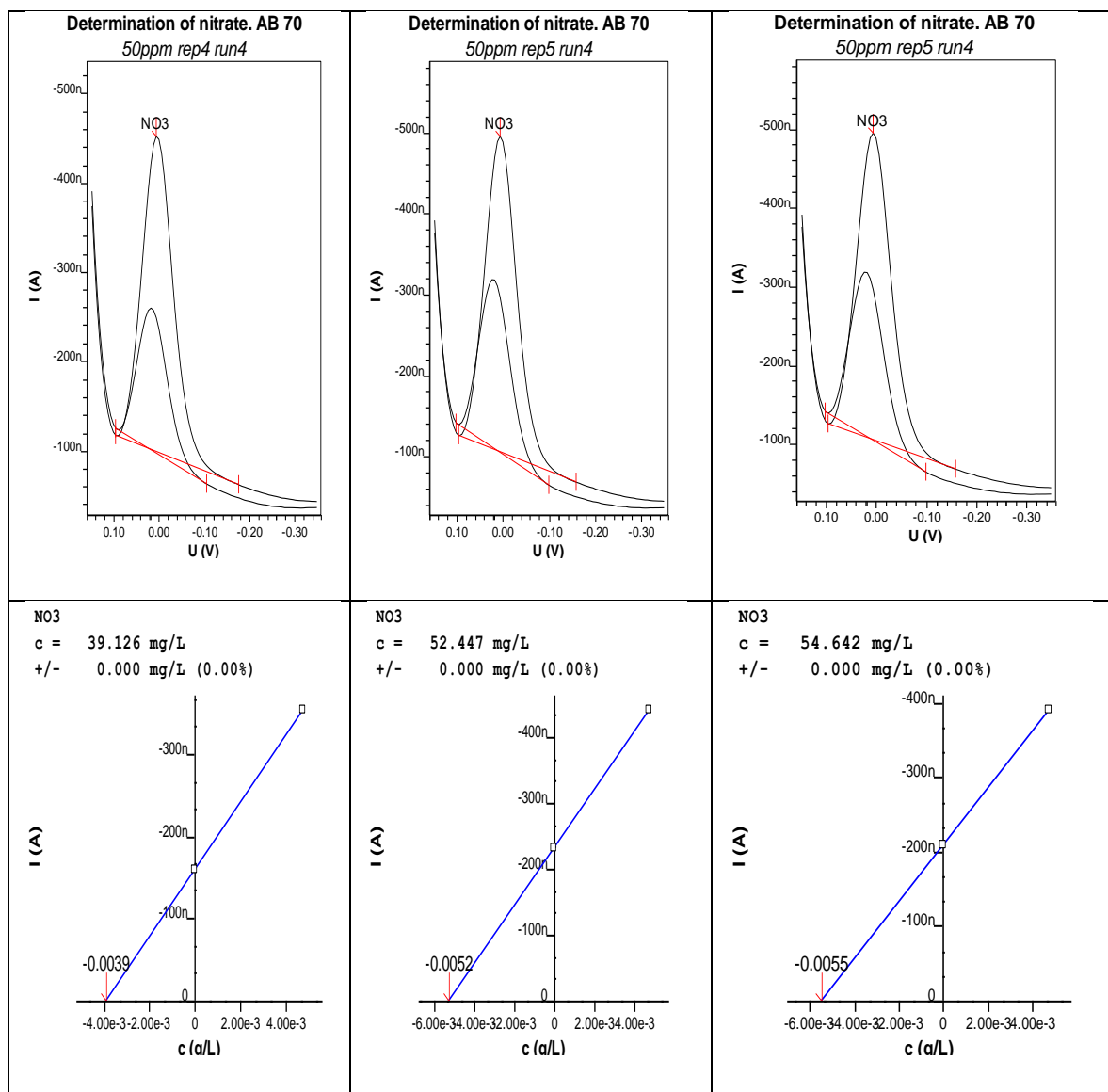


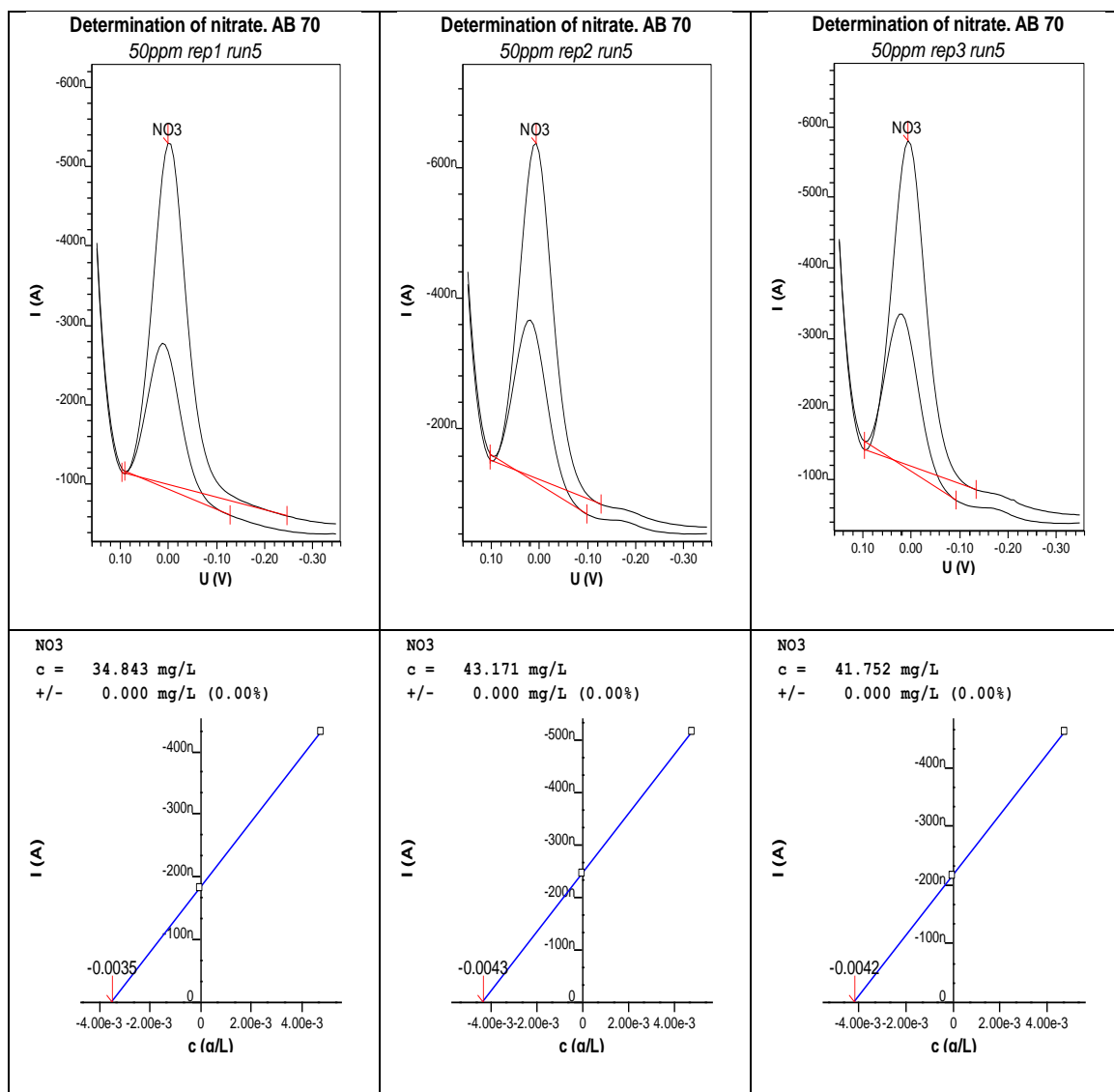




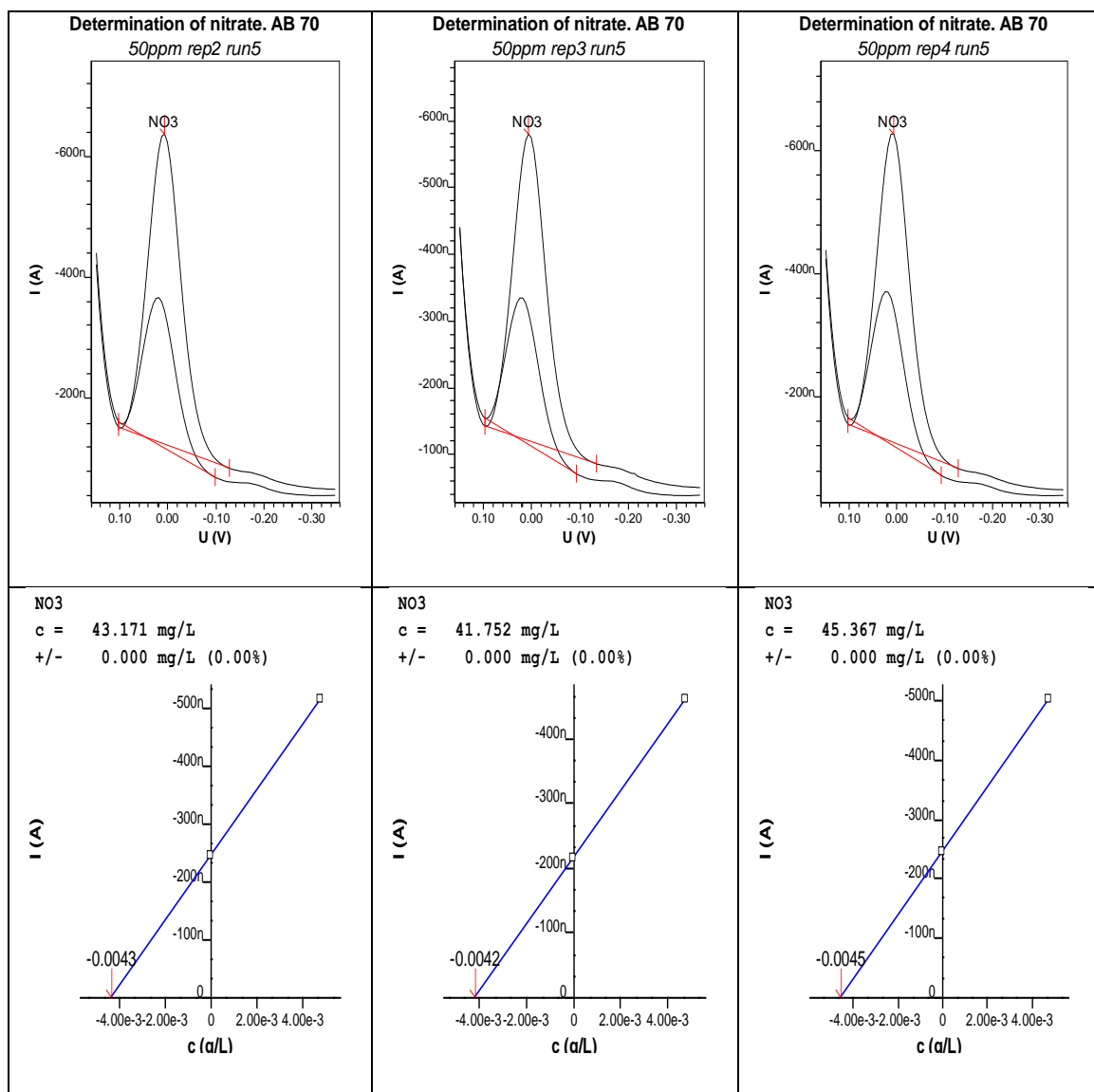


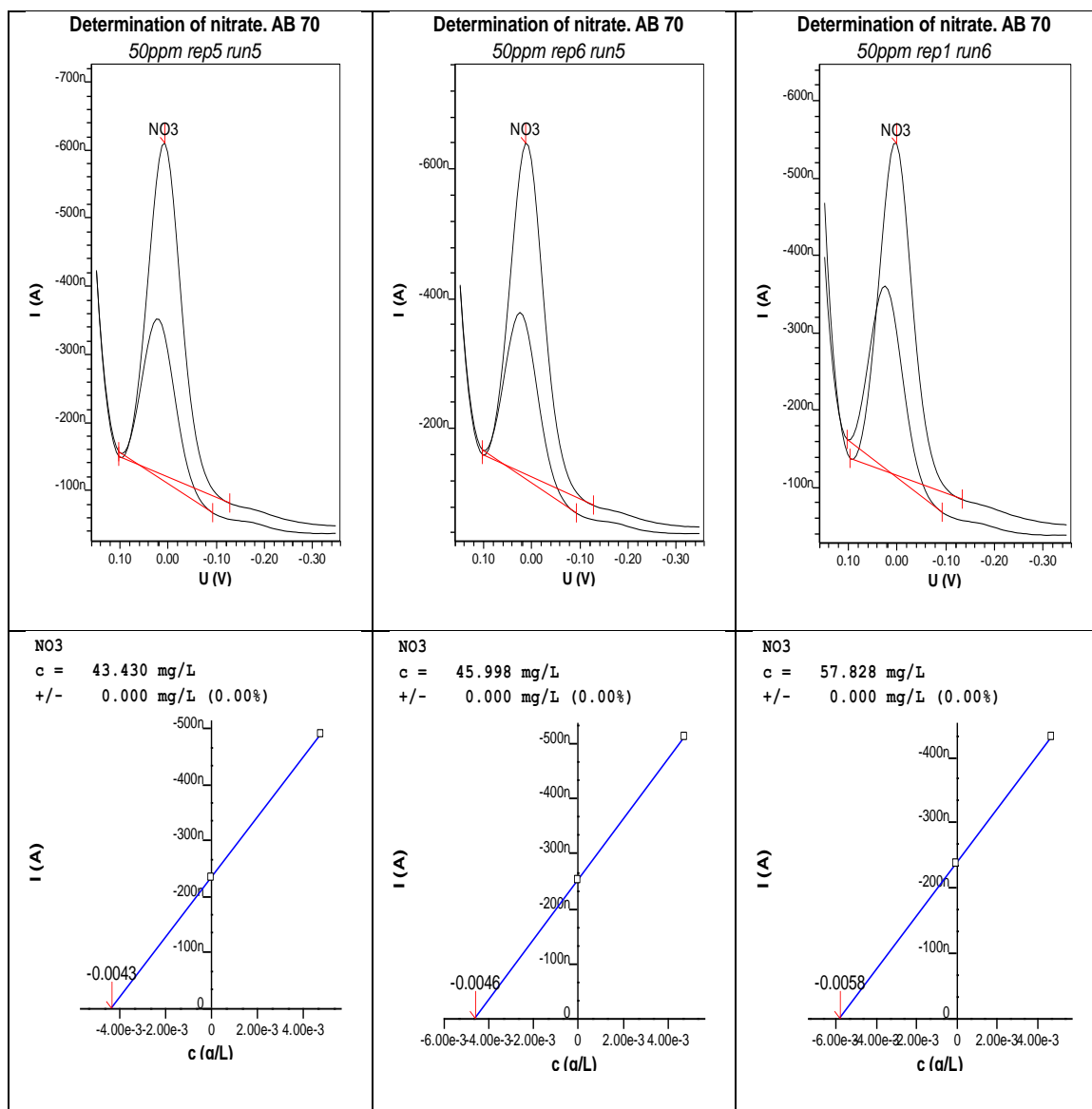


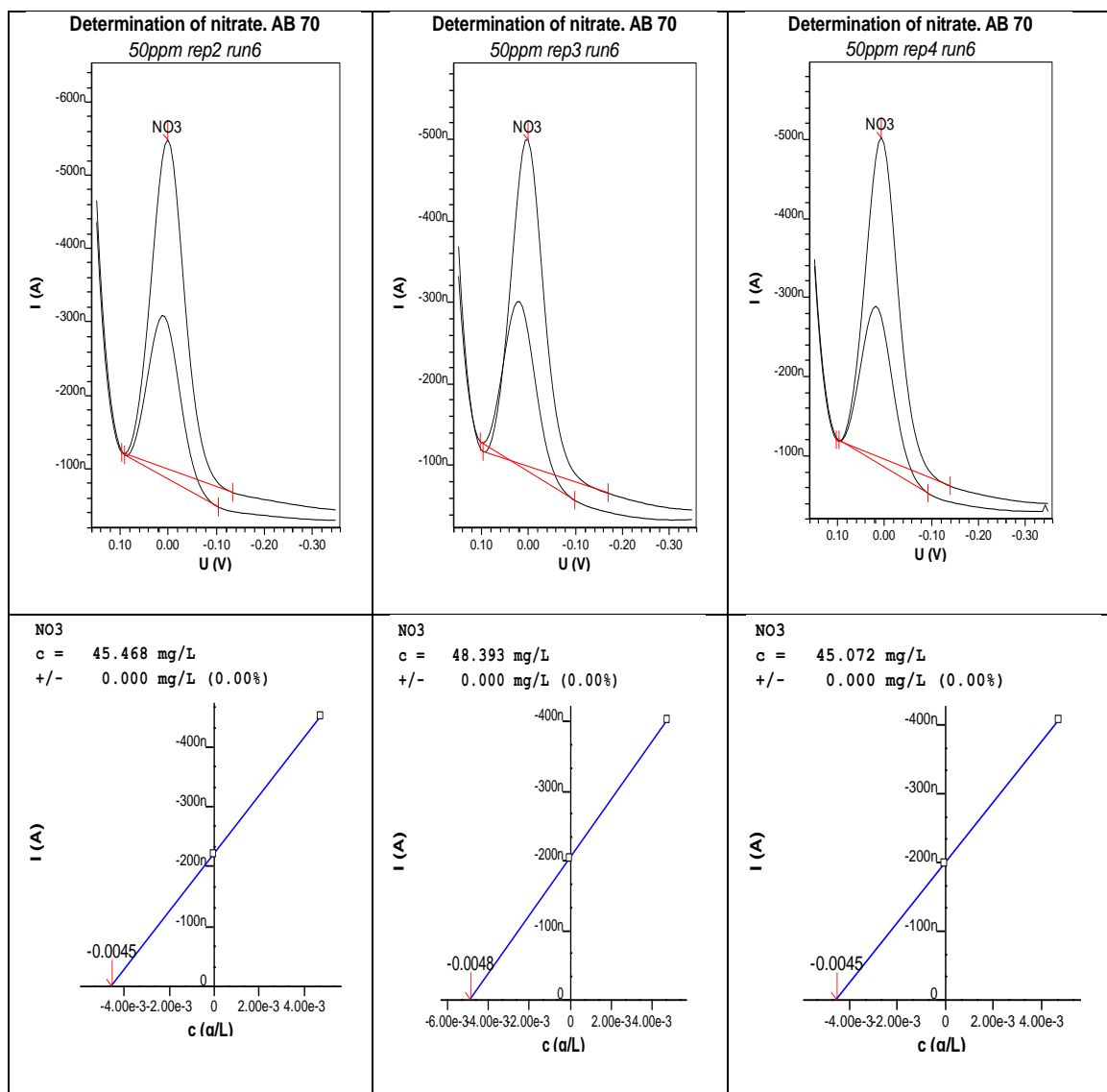


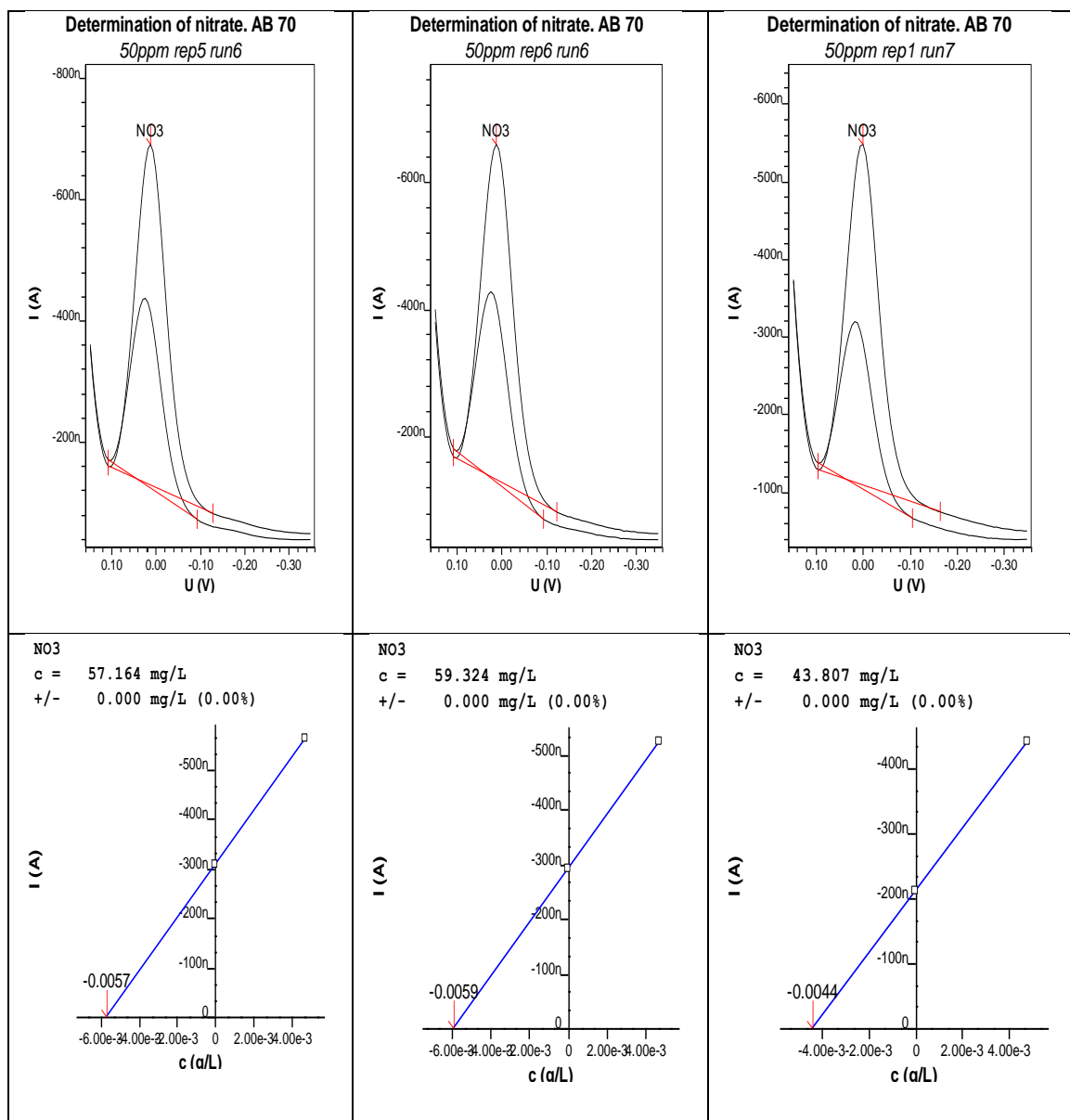


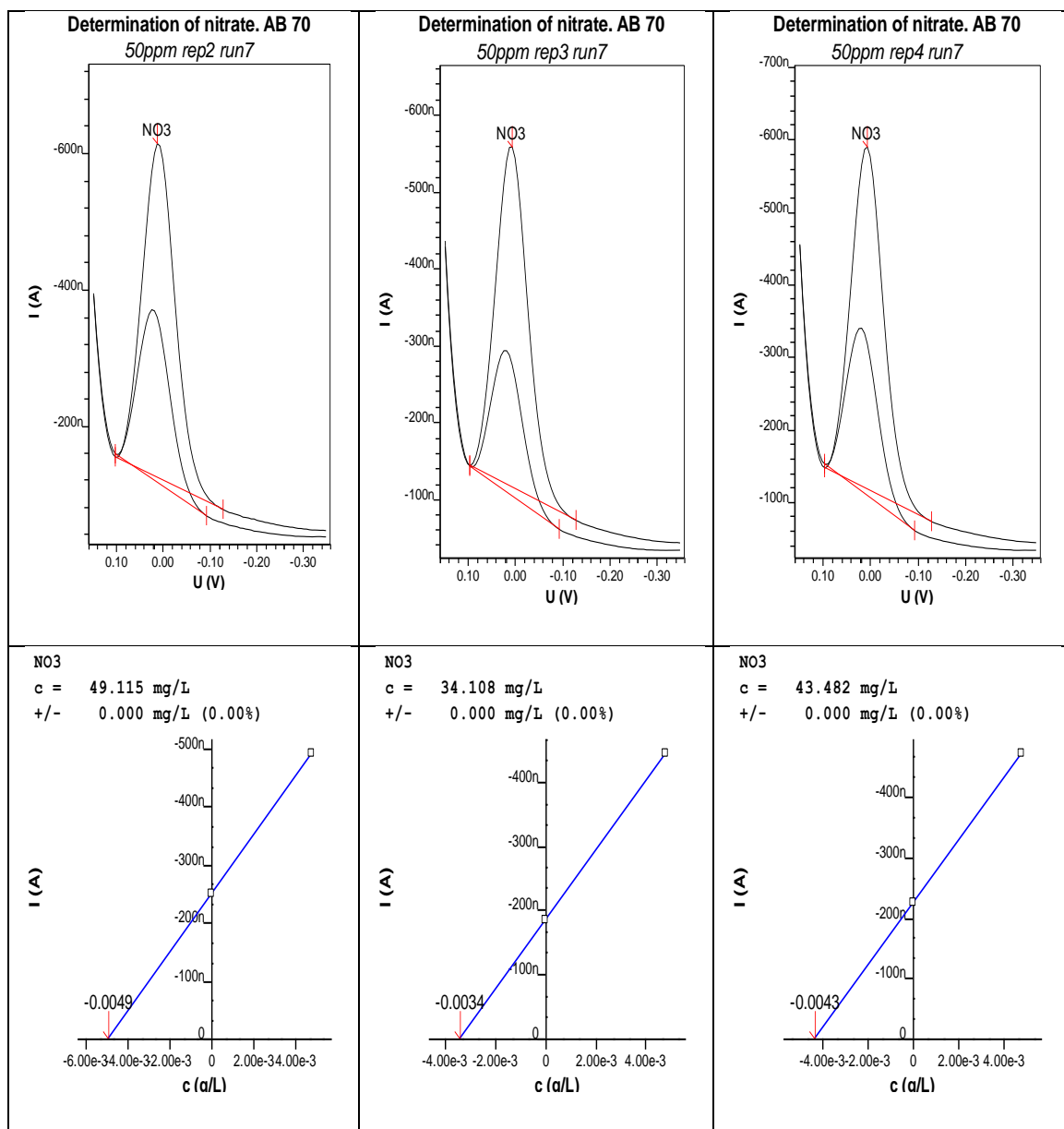


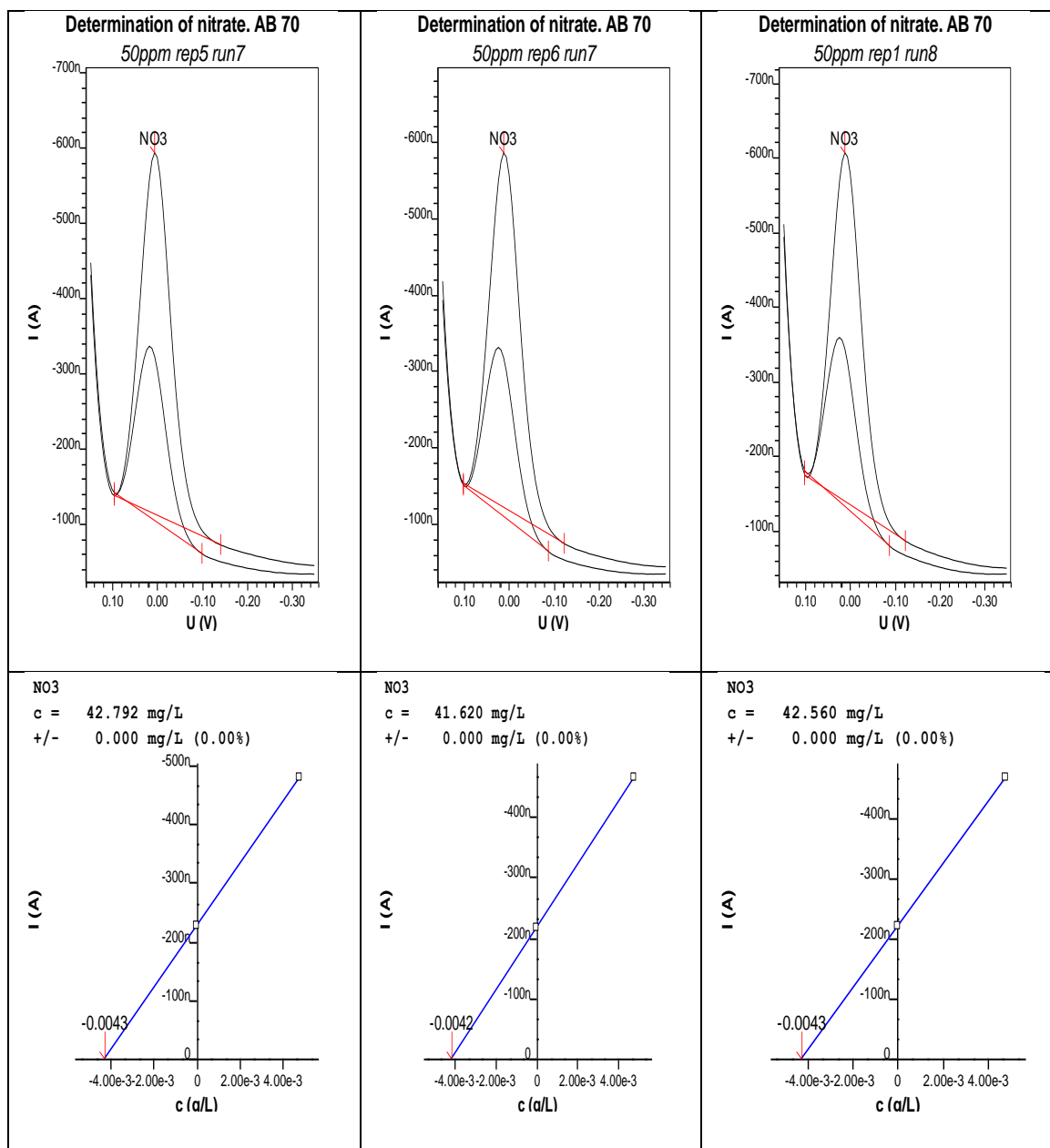


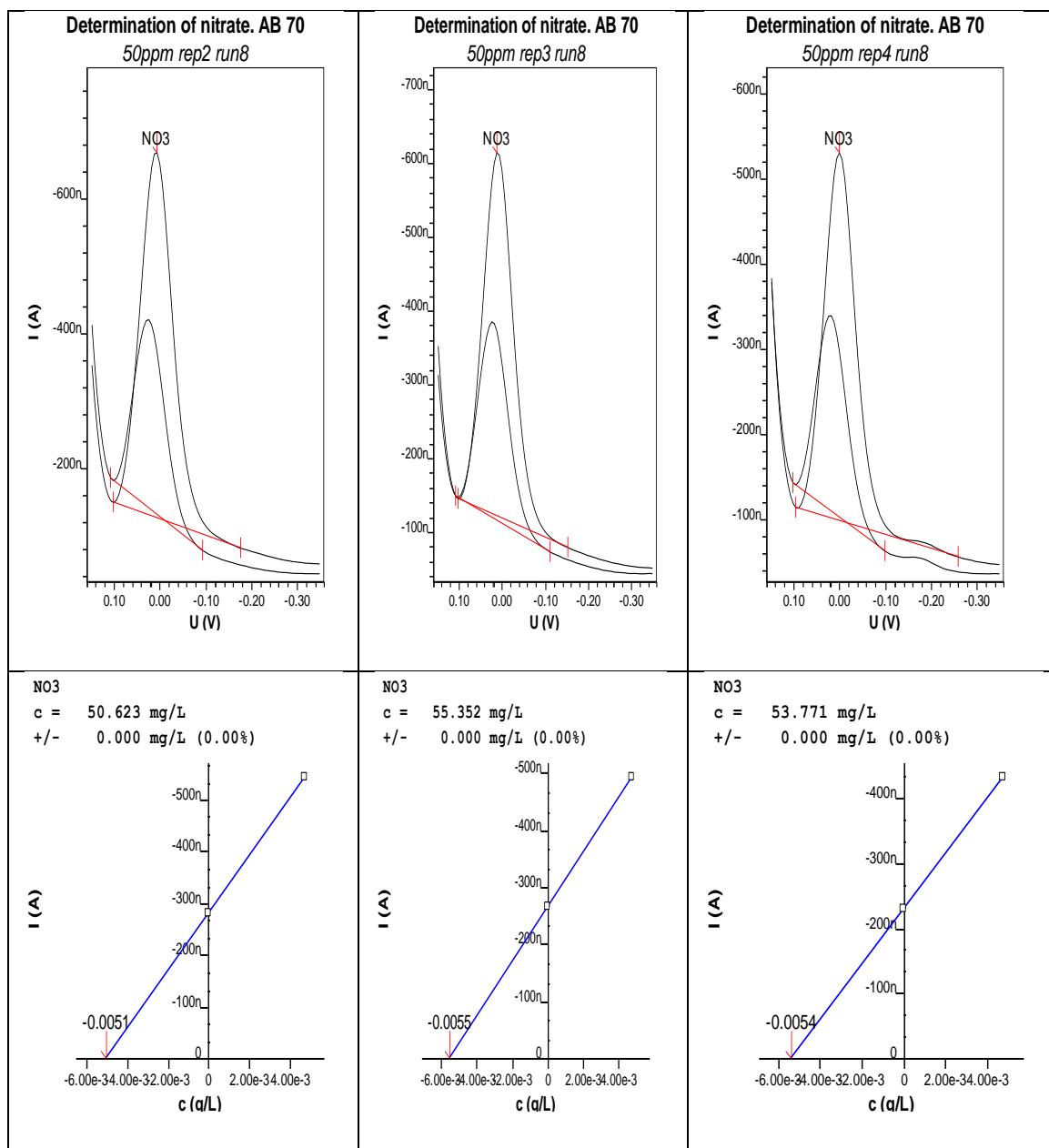


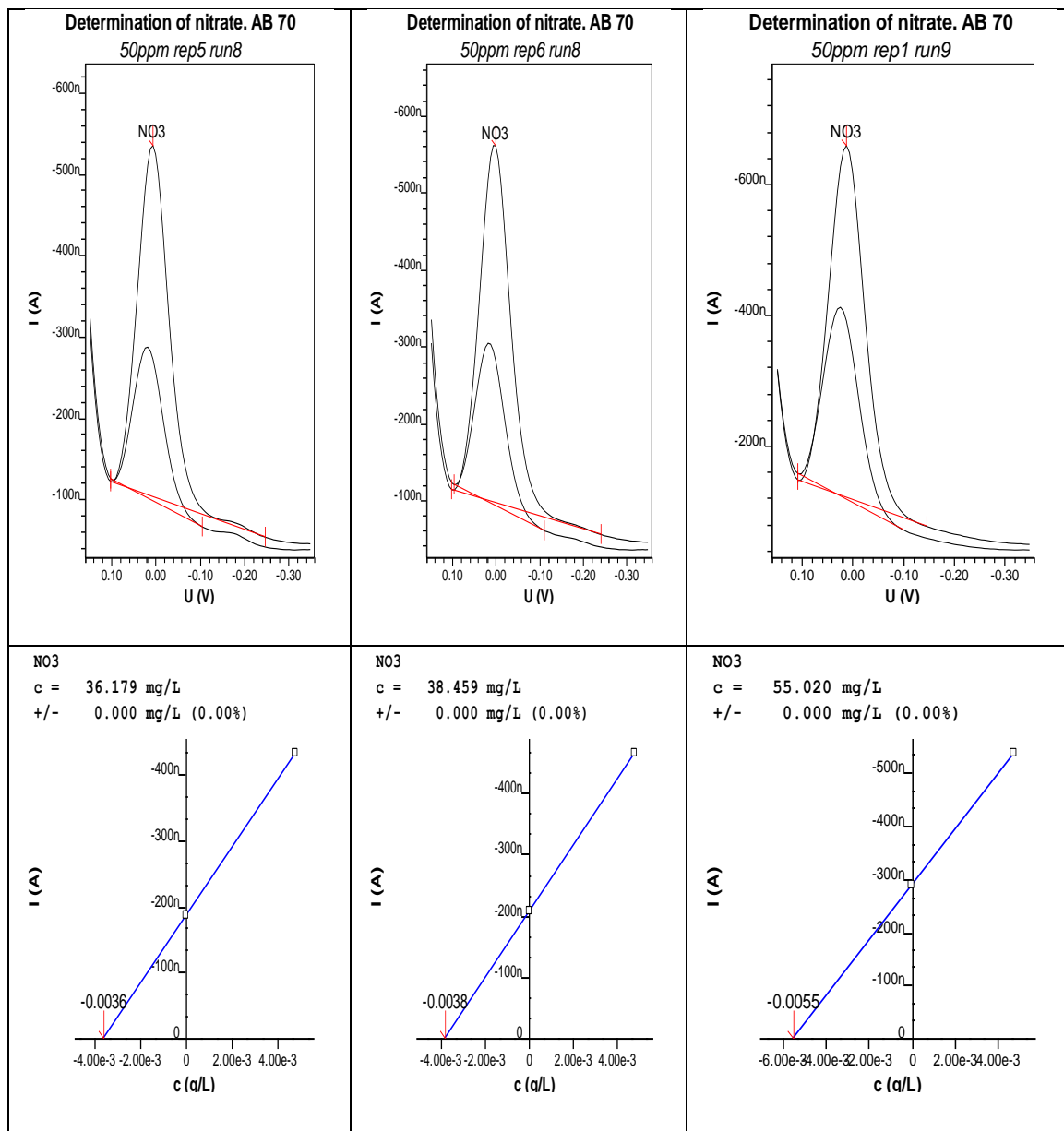




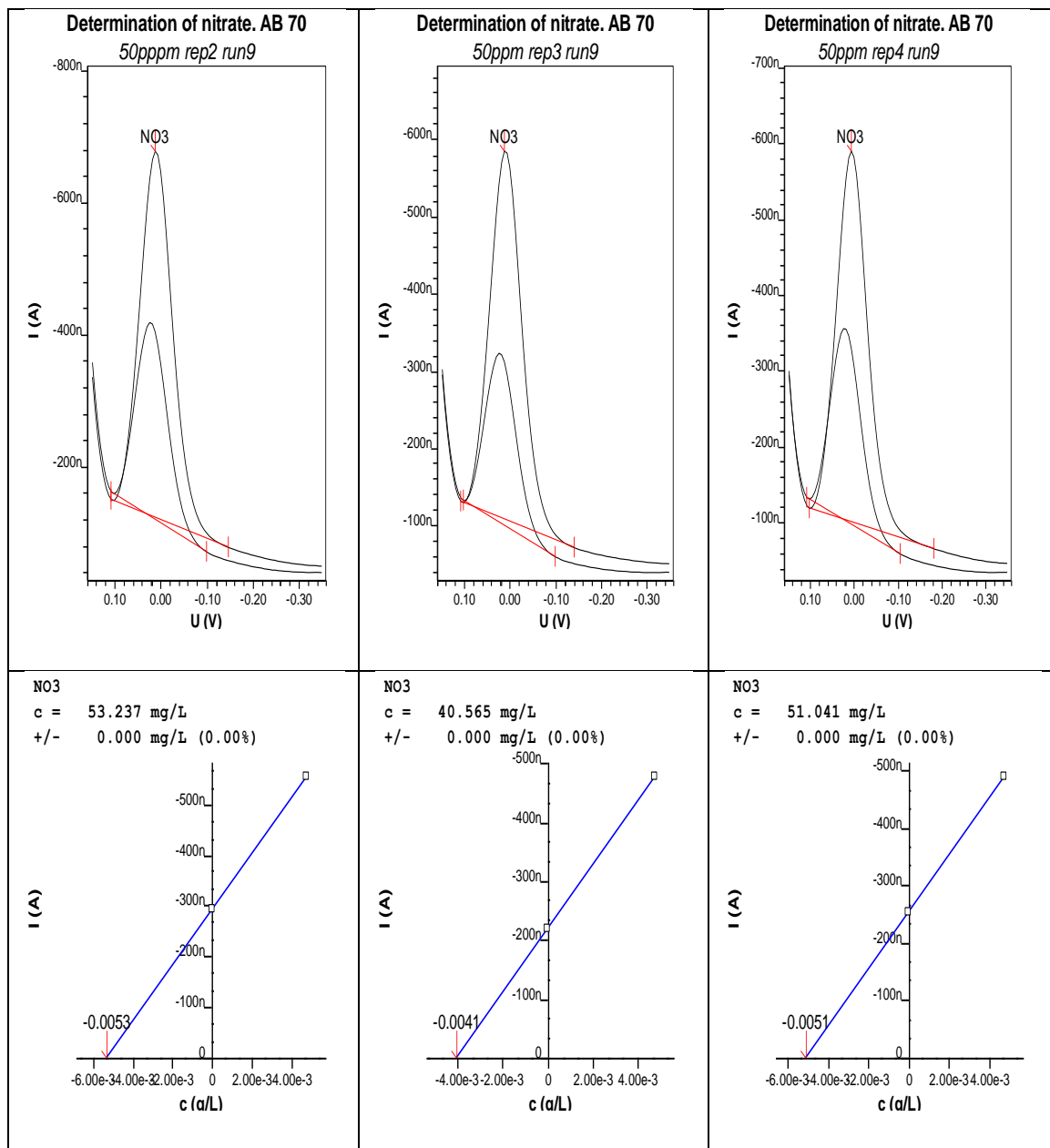


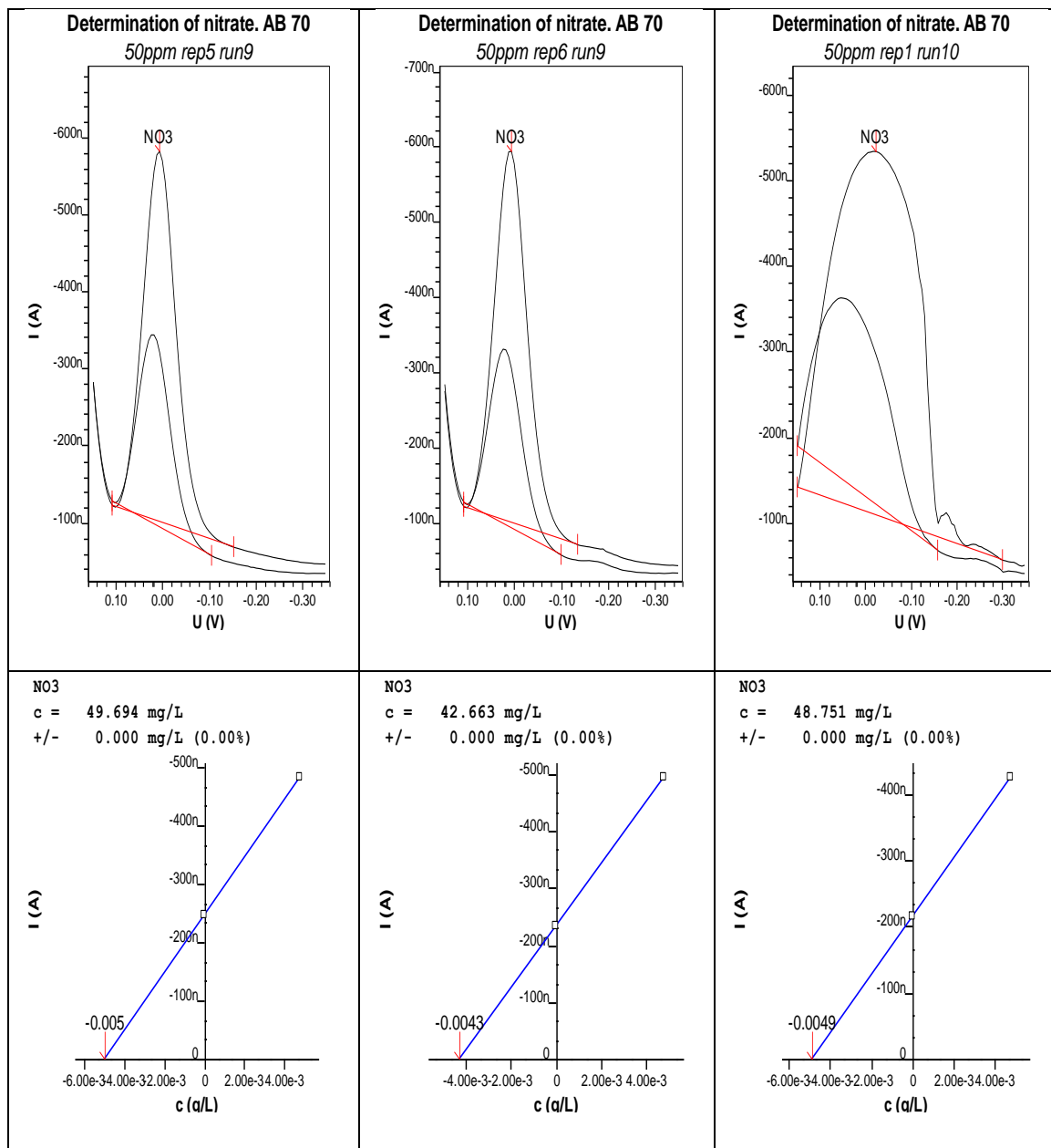


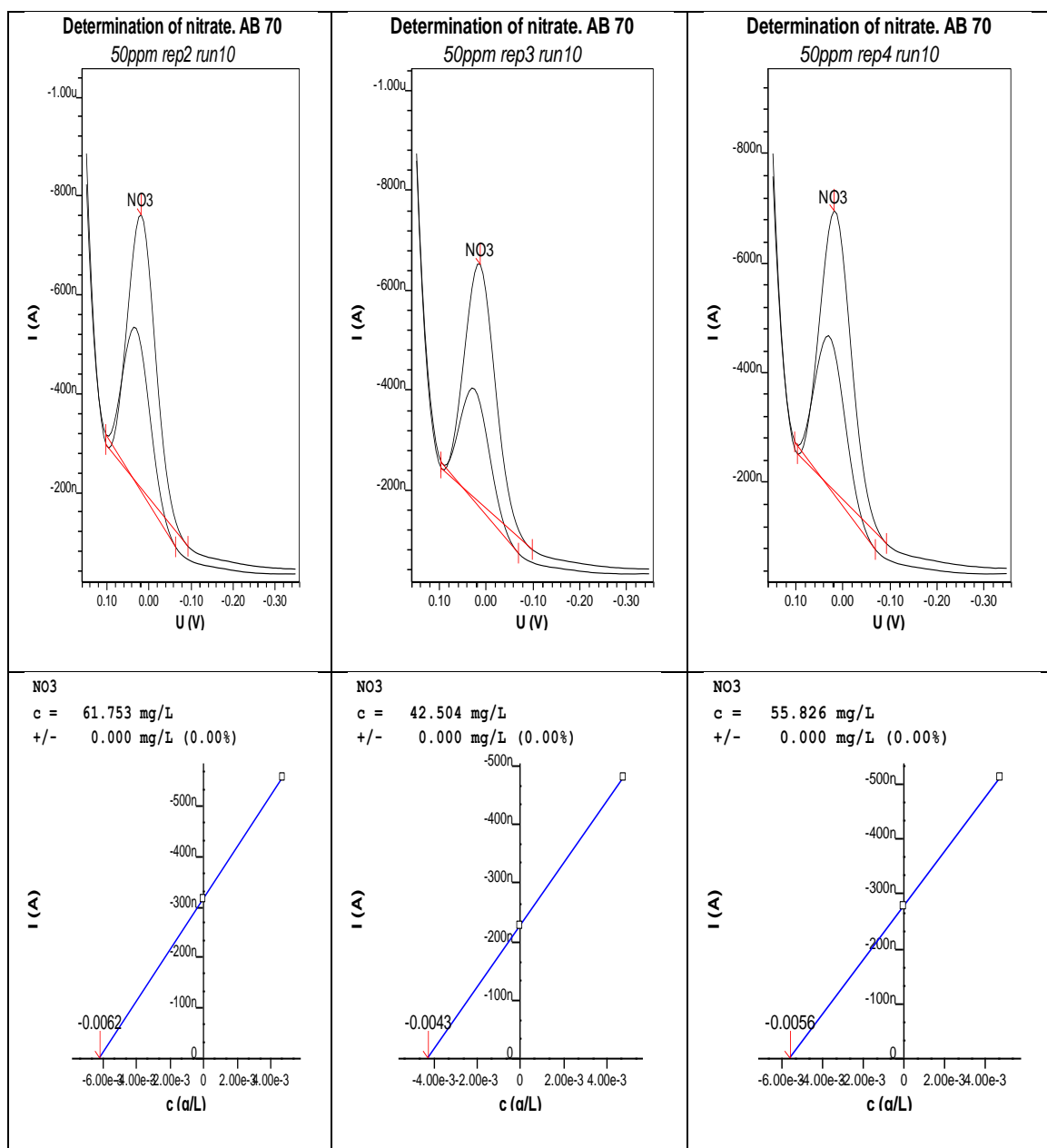


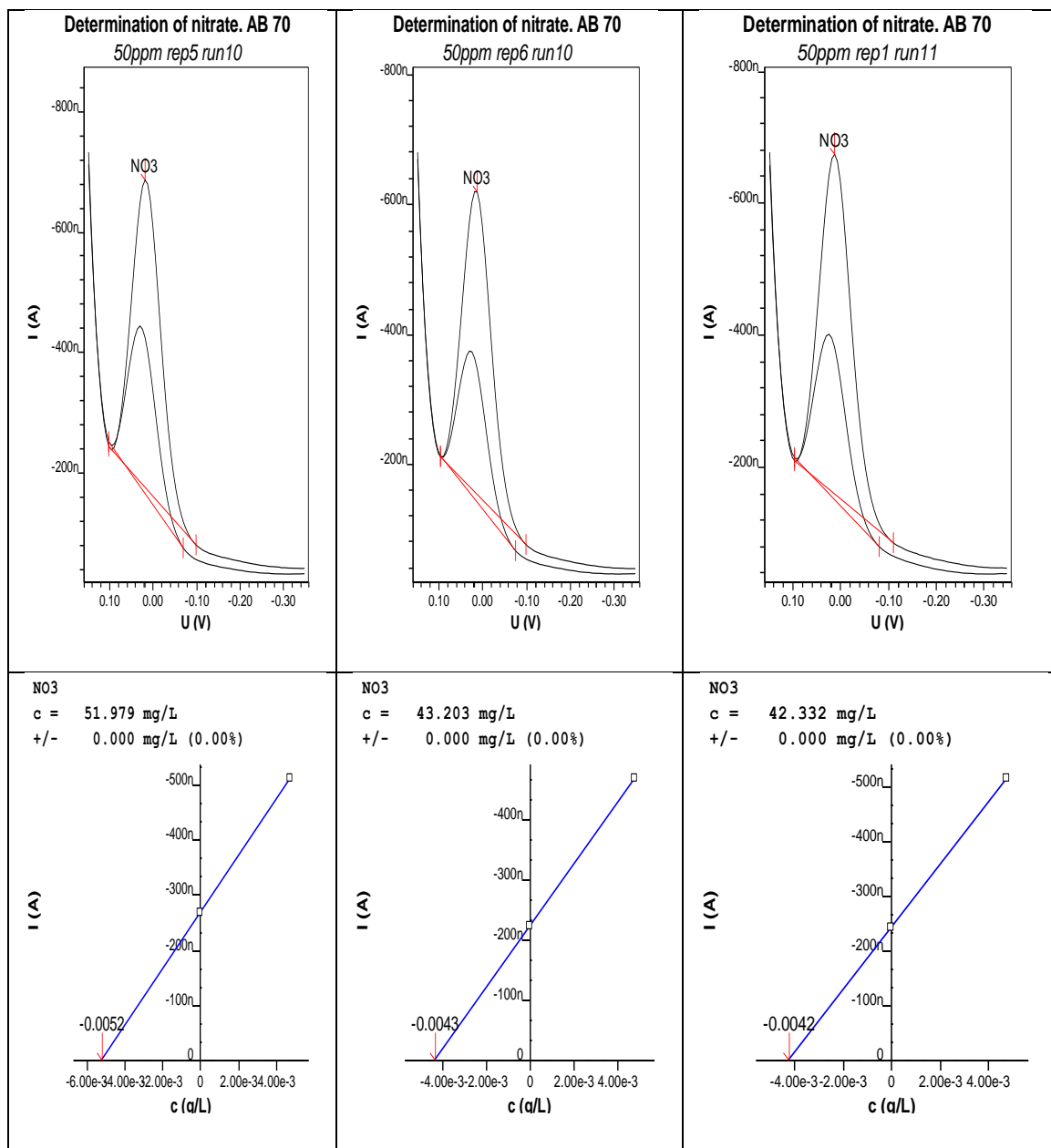


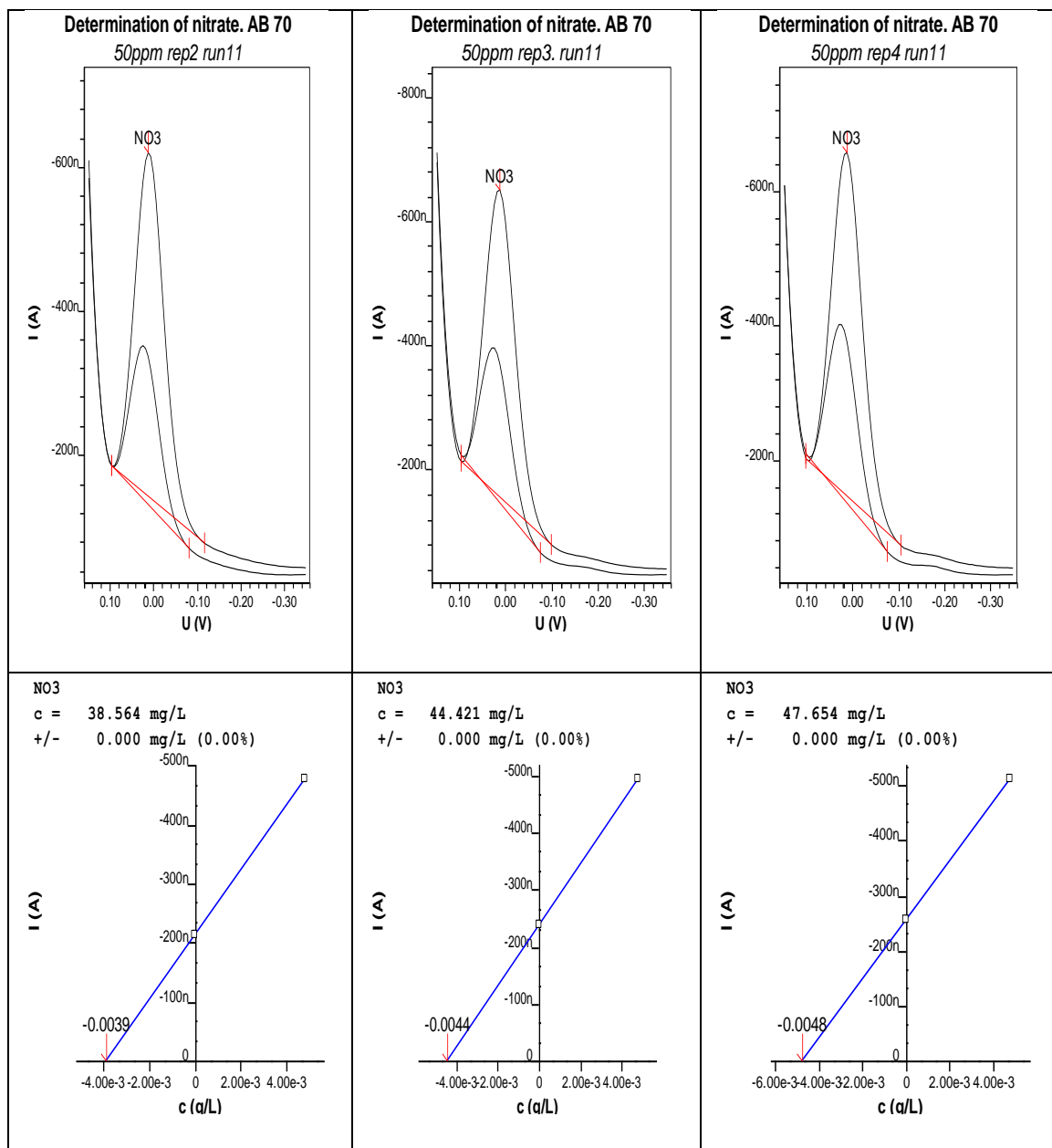


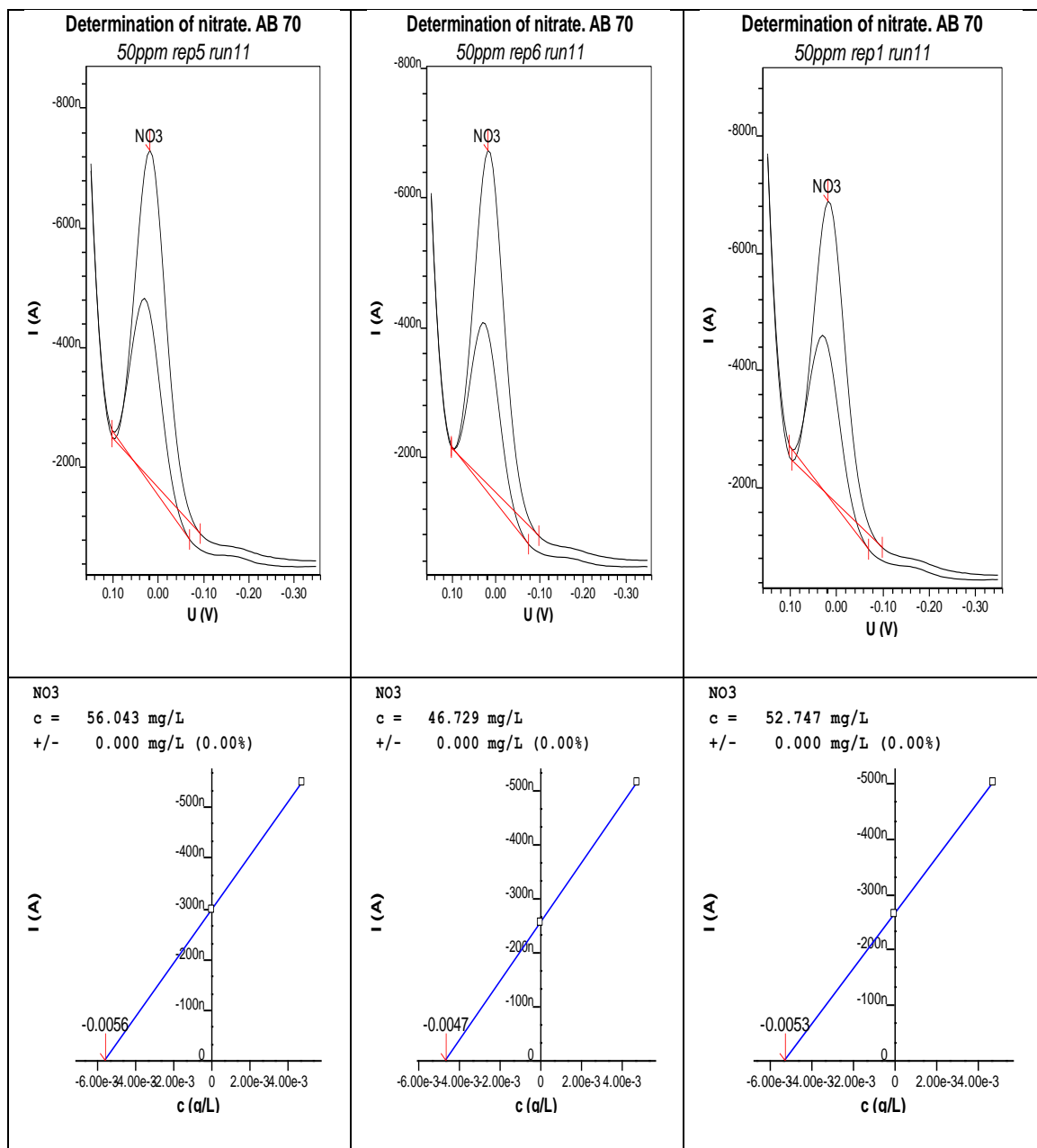


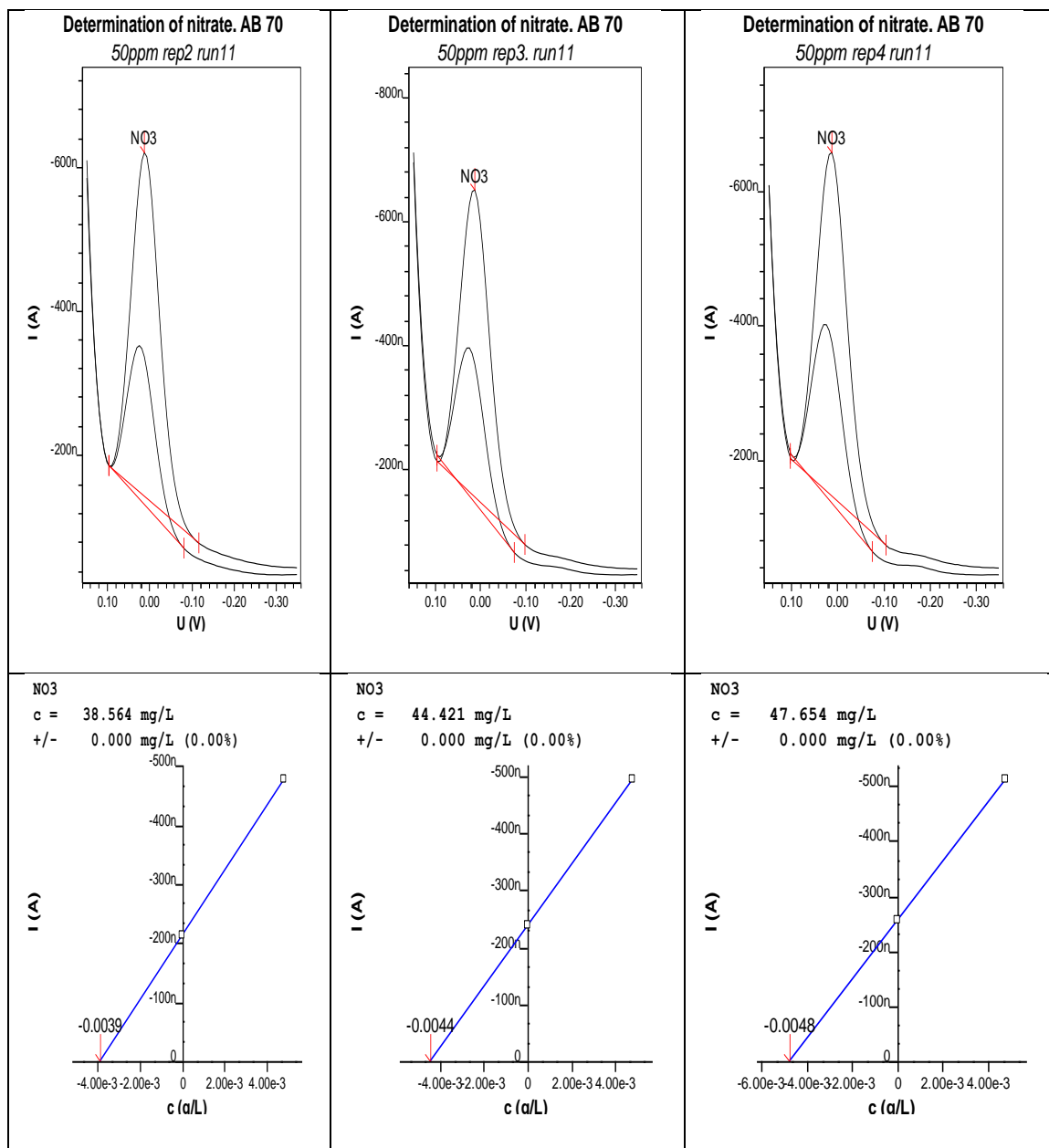


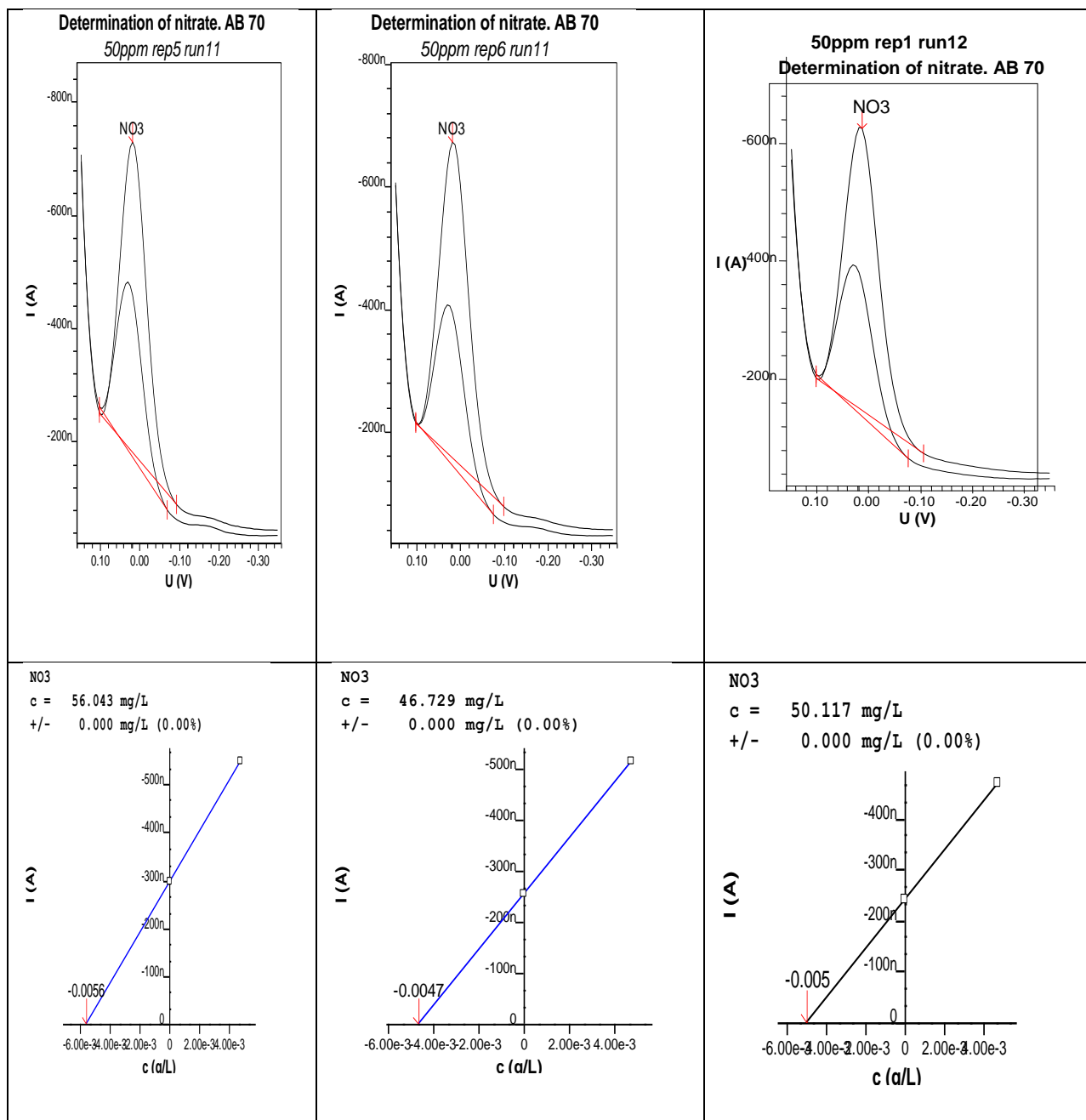






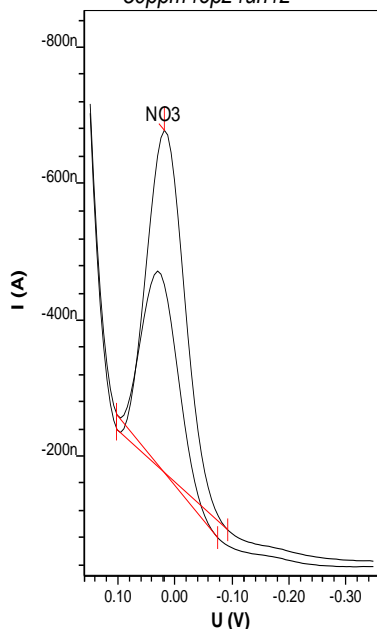




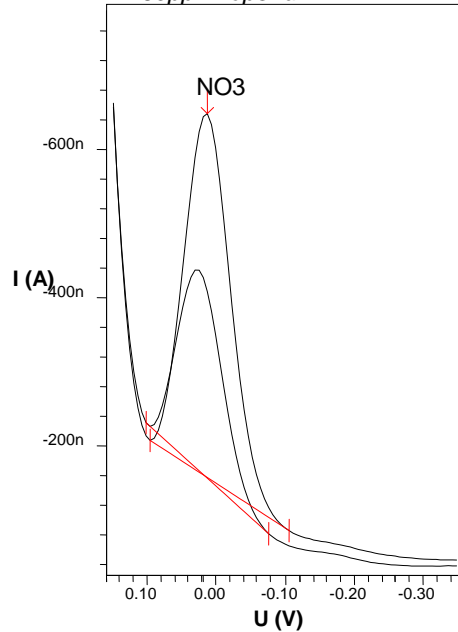




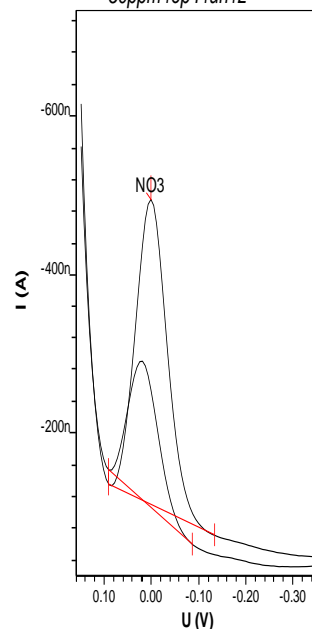
Determination of nitrate. AB 70  
50ppm rep2 run12



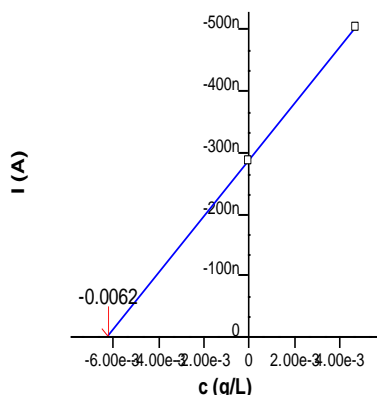
Determination of nitrate. AB 70  
50ppm rep3 run12



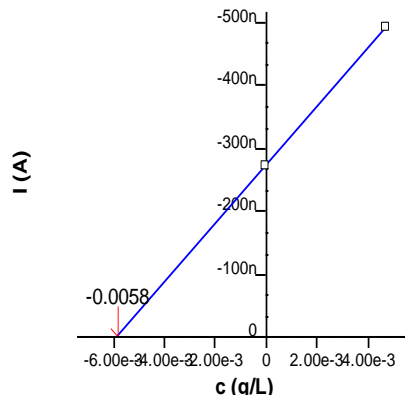
Determination of nitrate. AB 70  
50ppm rep4 run12



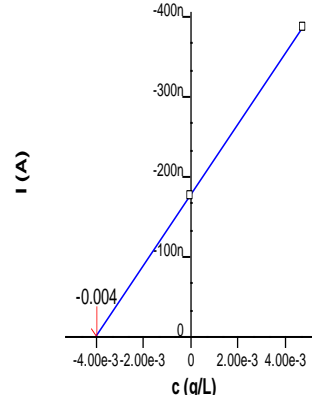
NO3  
c = 62.054 mg/L  
+/- 0.000 mg/L (0.00%)

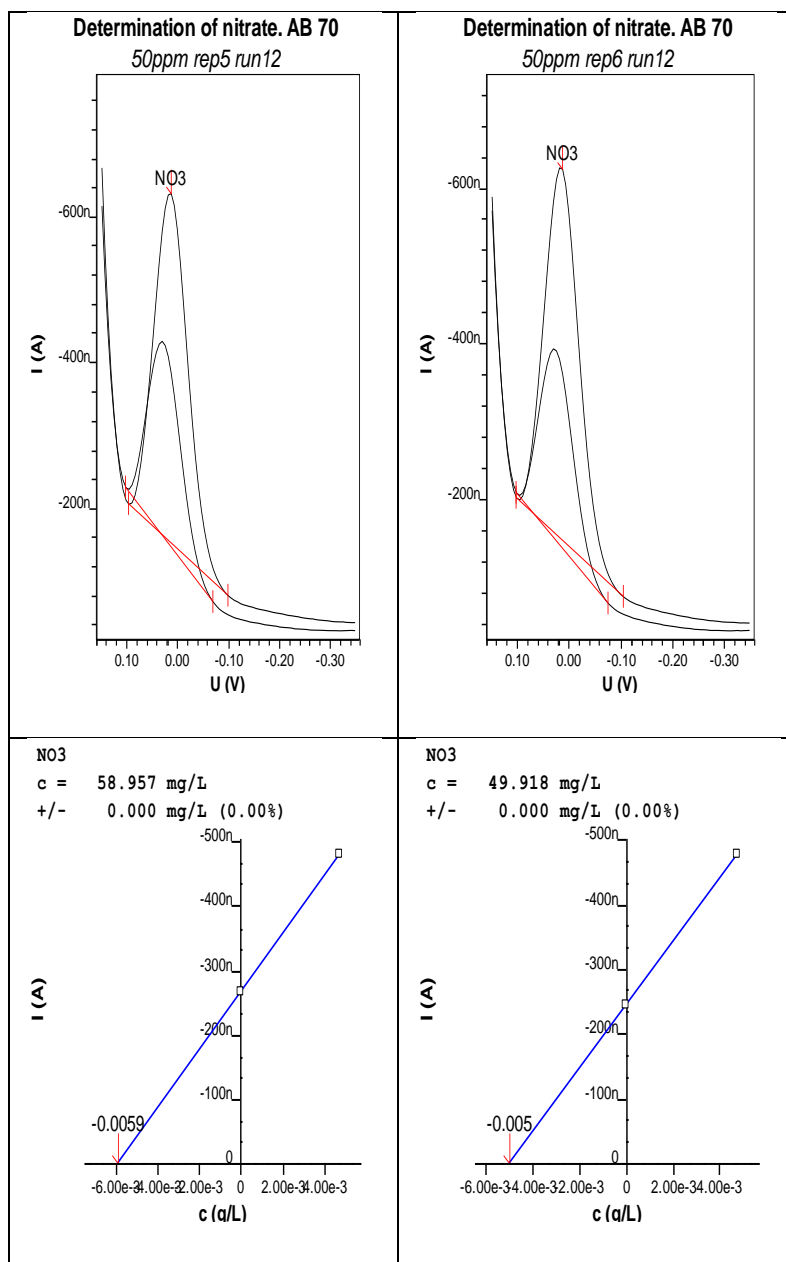


NO3  
c = 58.223 mg/L  
+/- 0.000 mg/L (0.00%)

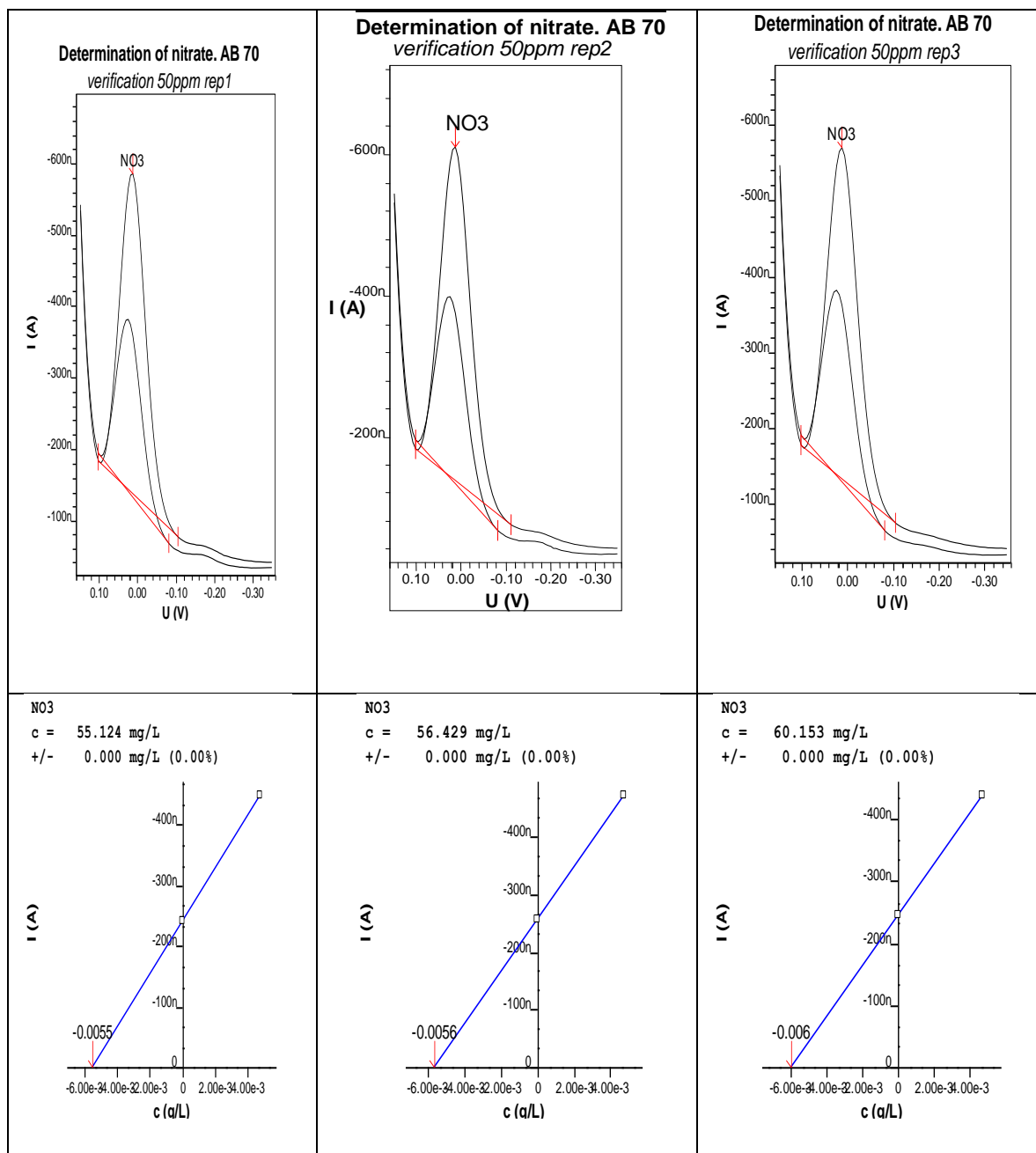


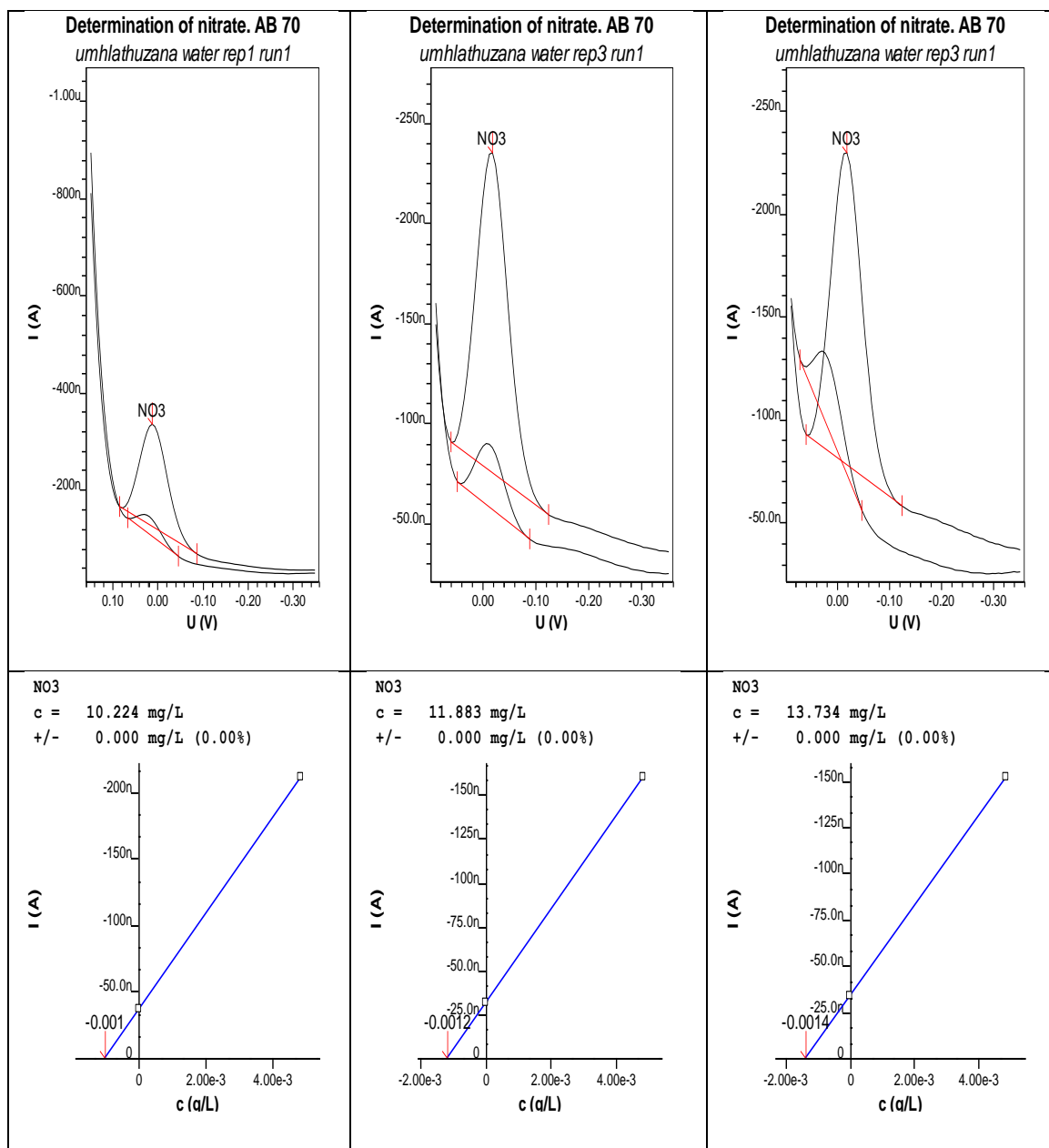
NO3  
c = 39.852 mg/L  
+/- 0.000 mg/L (0.00%)

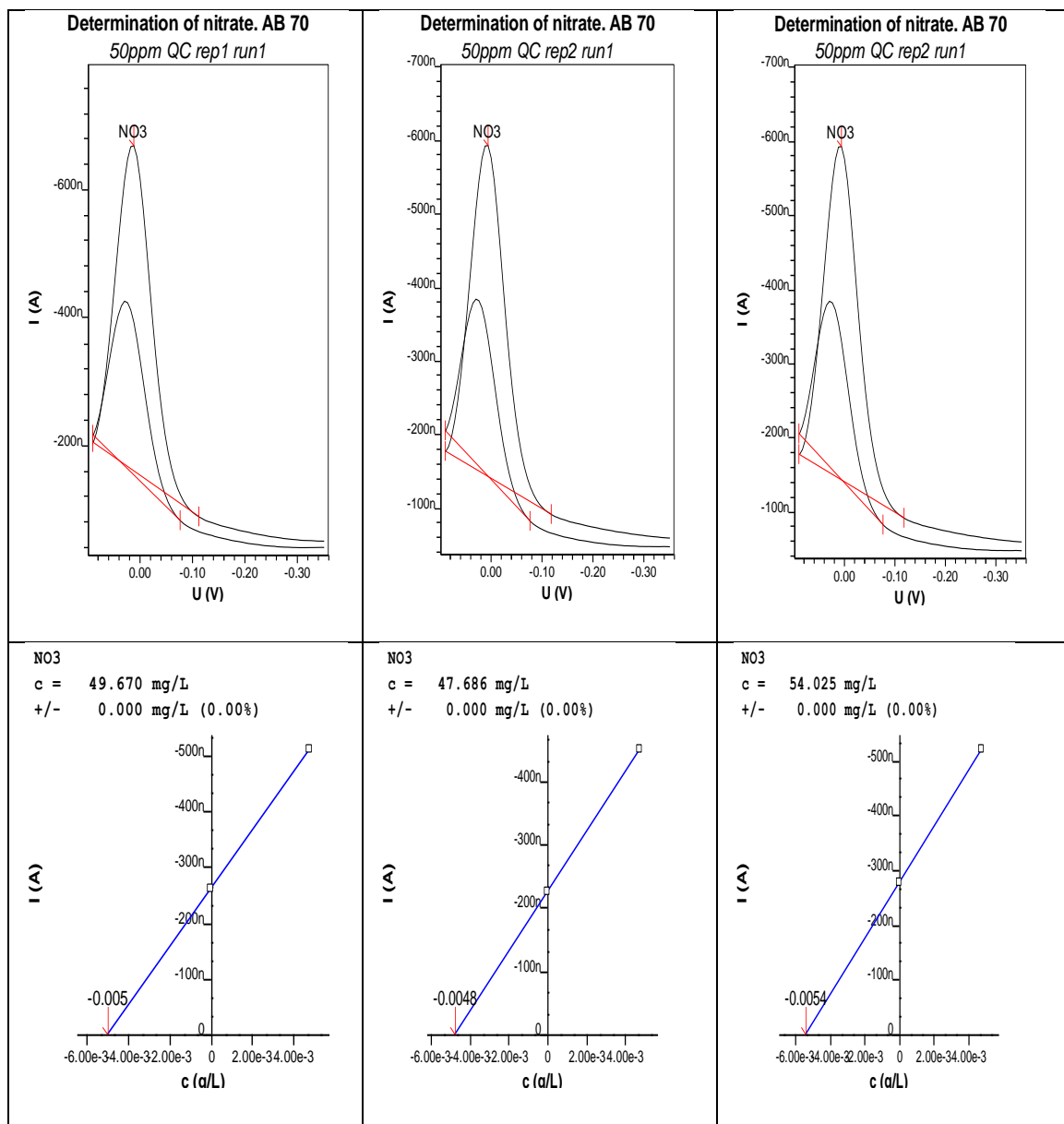


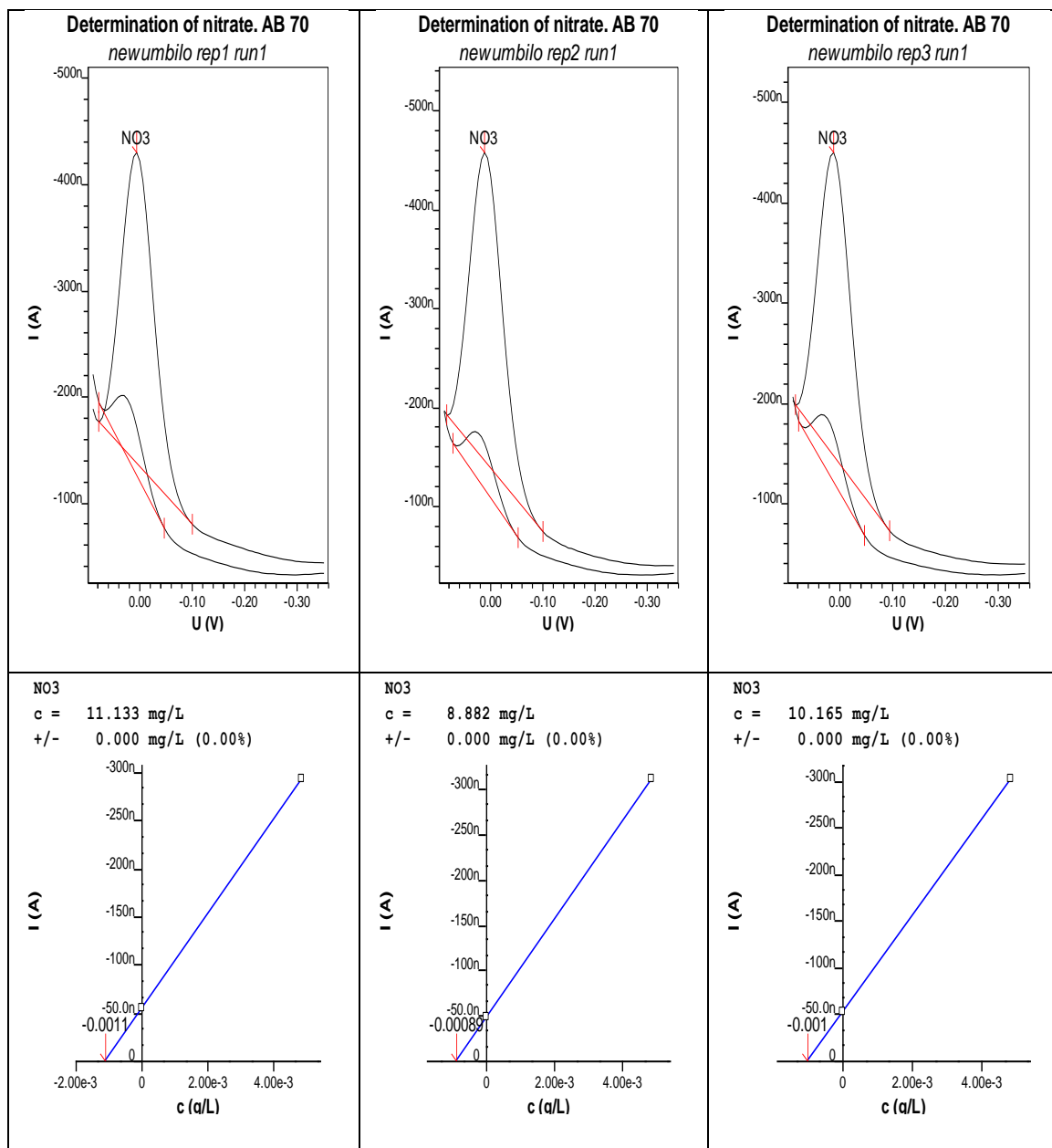


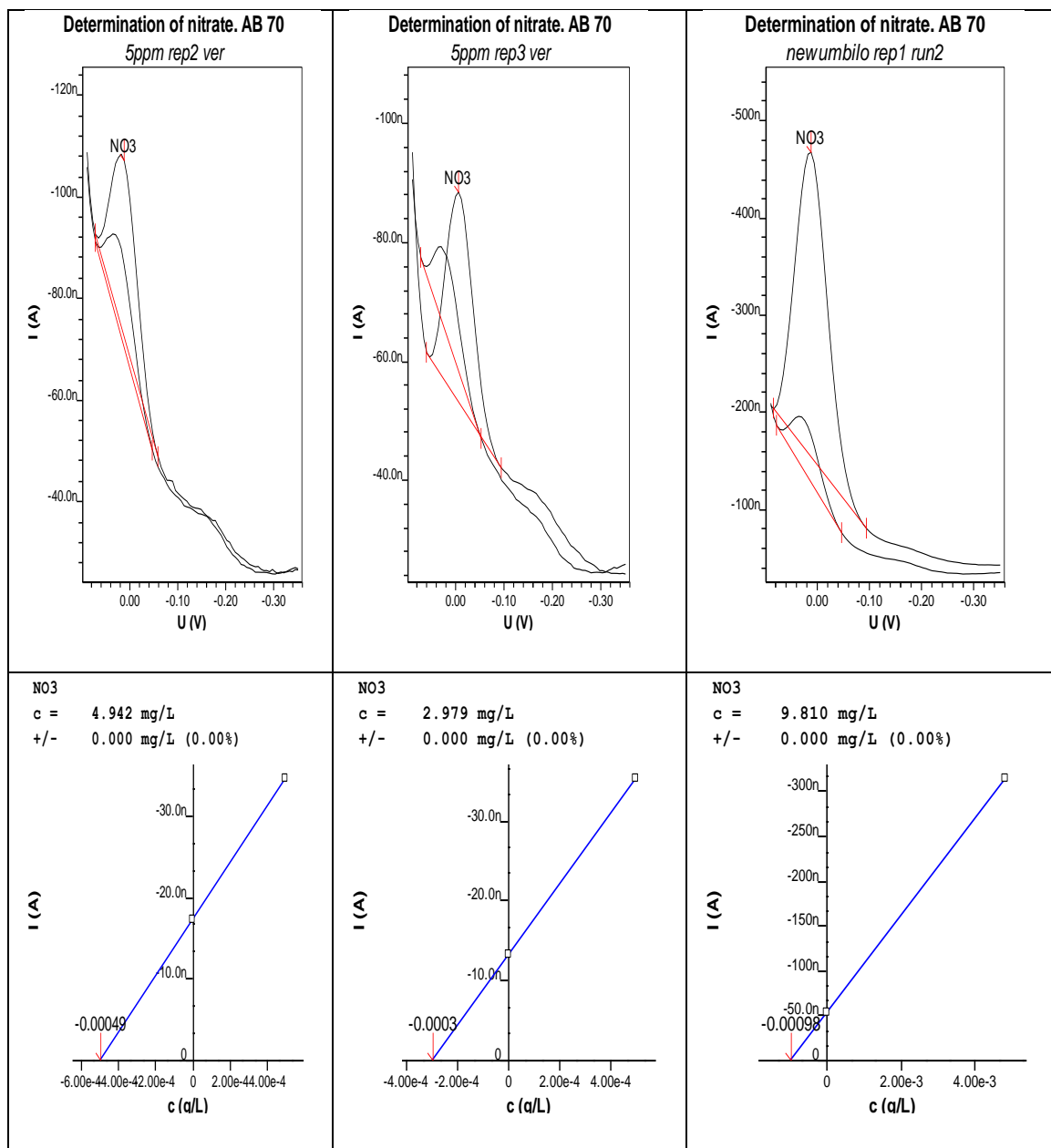
## Appendix 5 Sample analysis verification and quality control results

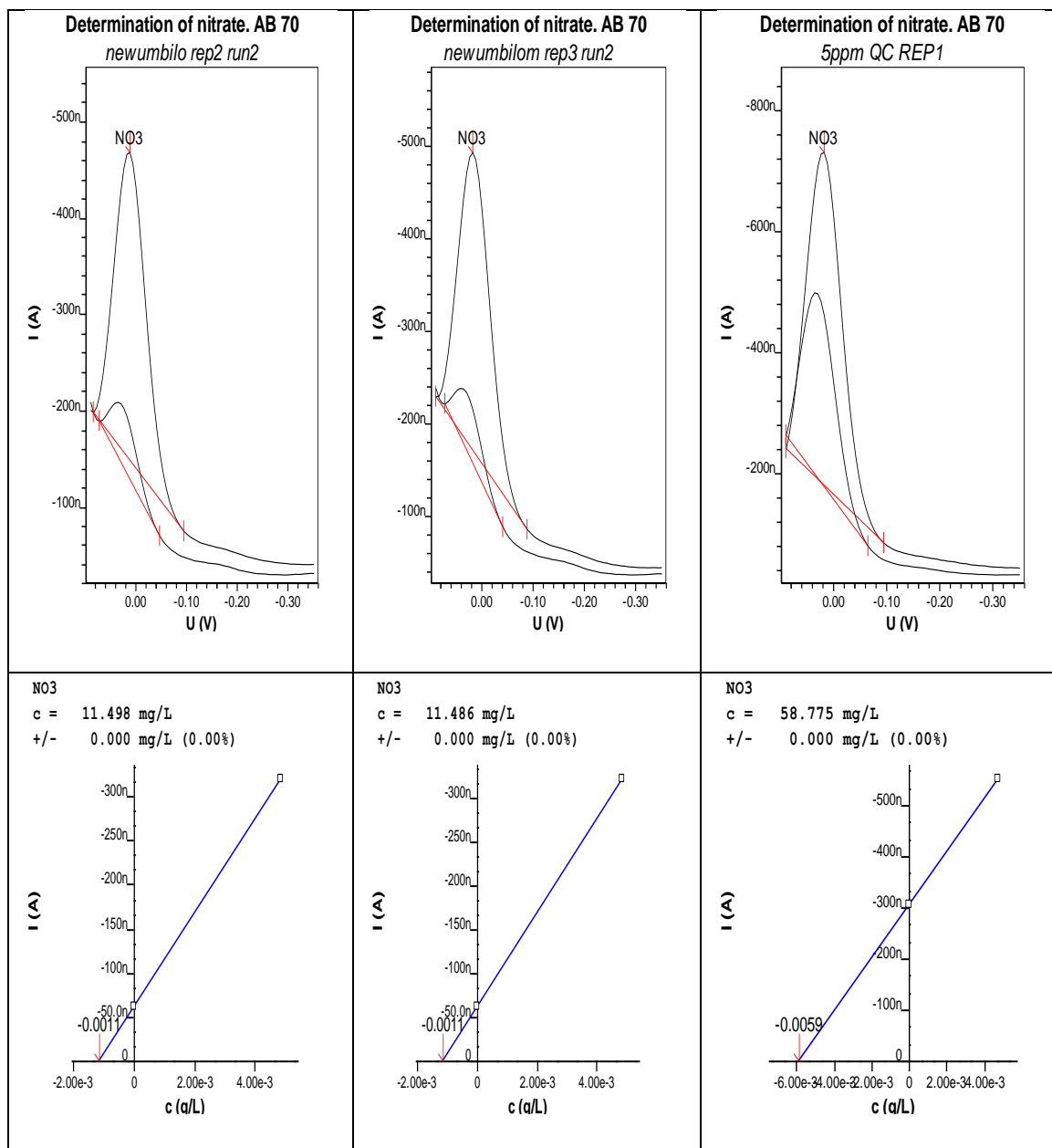




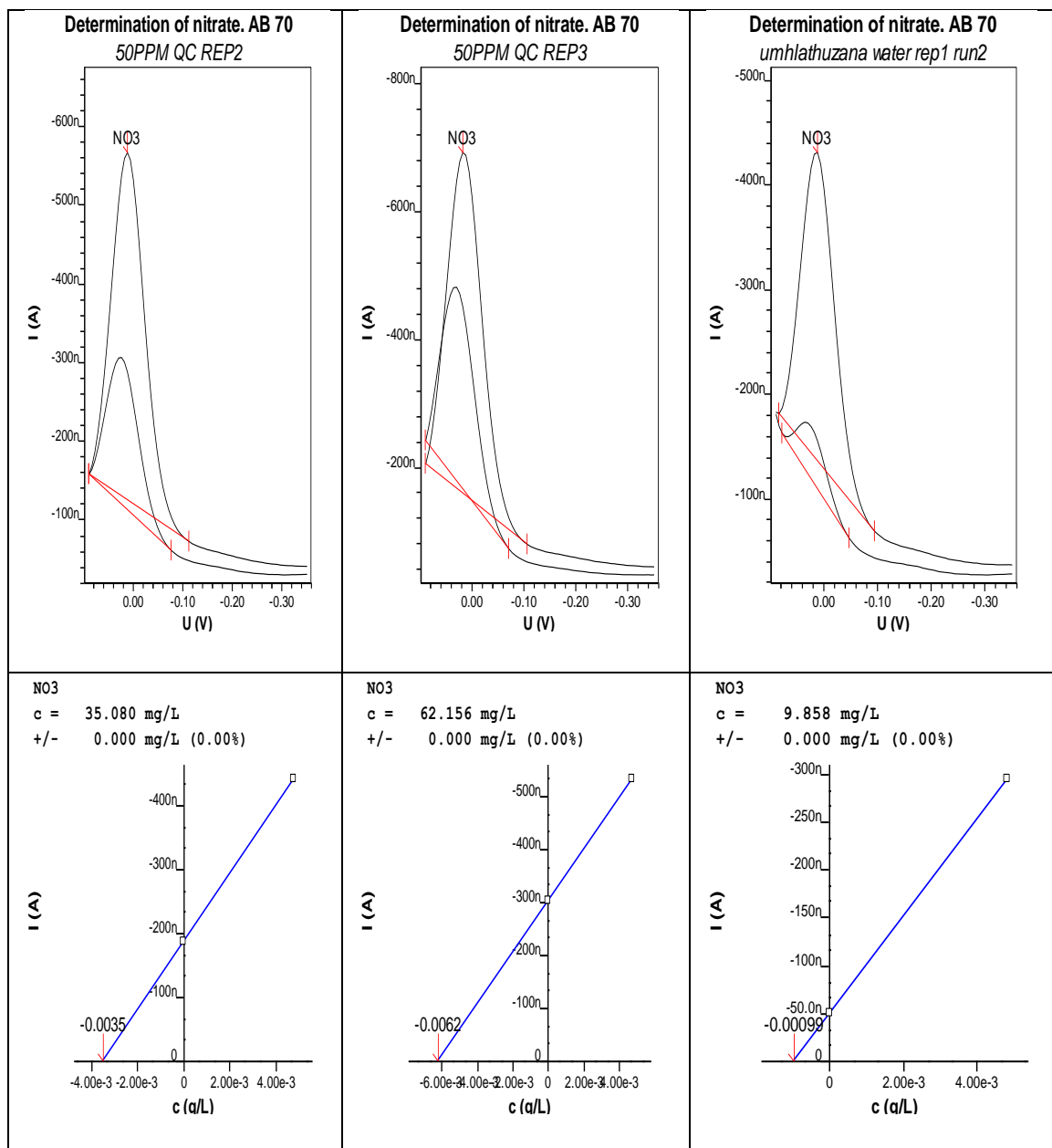


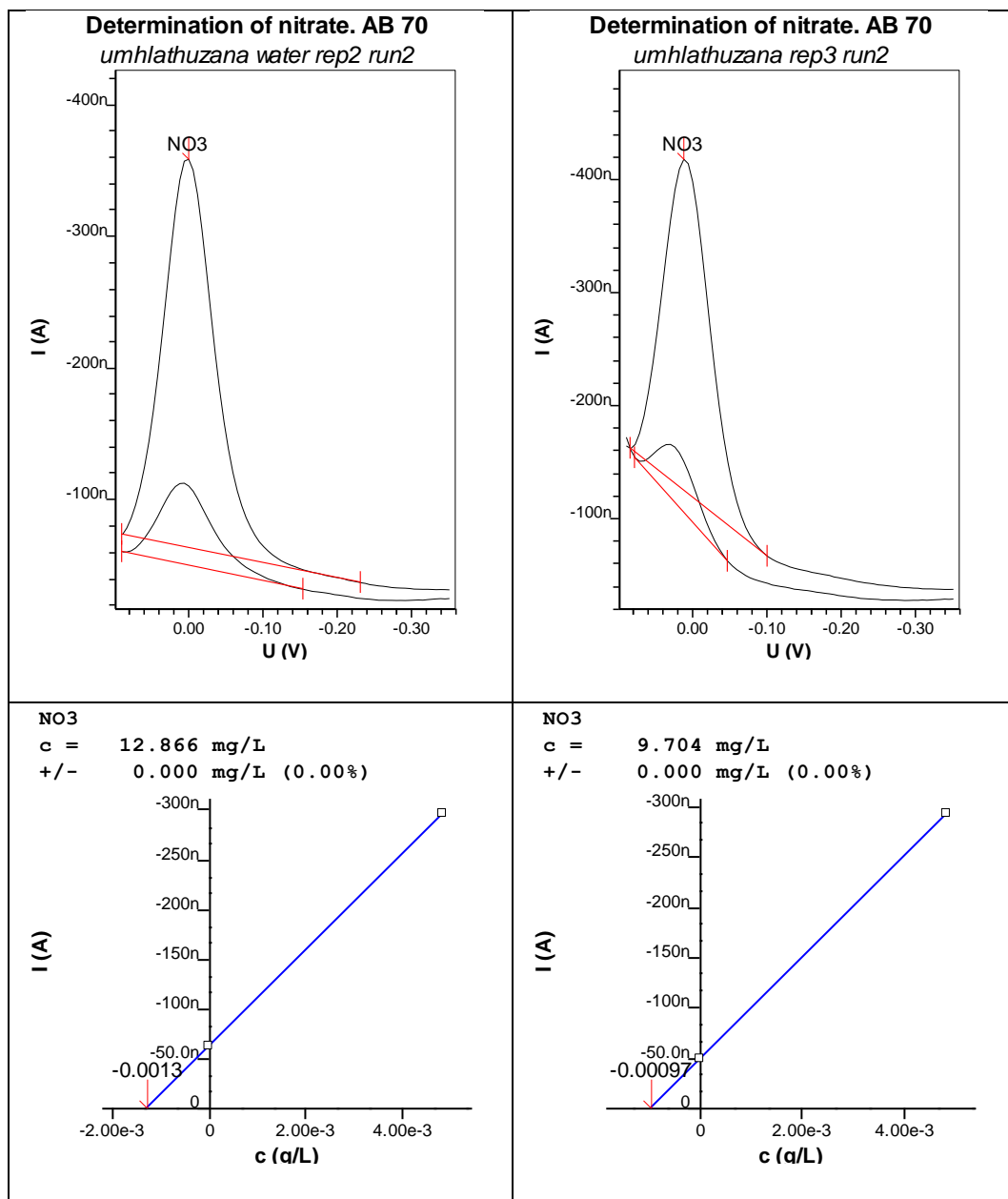




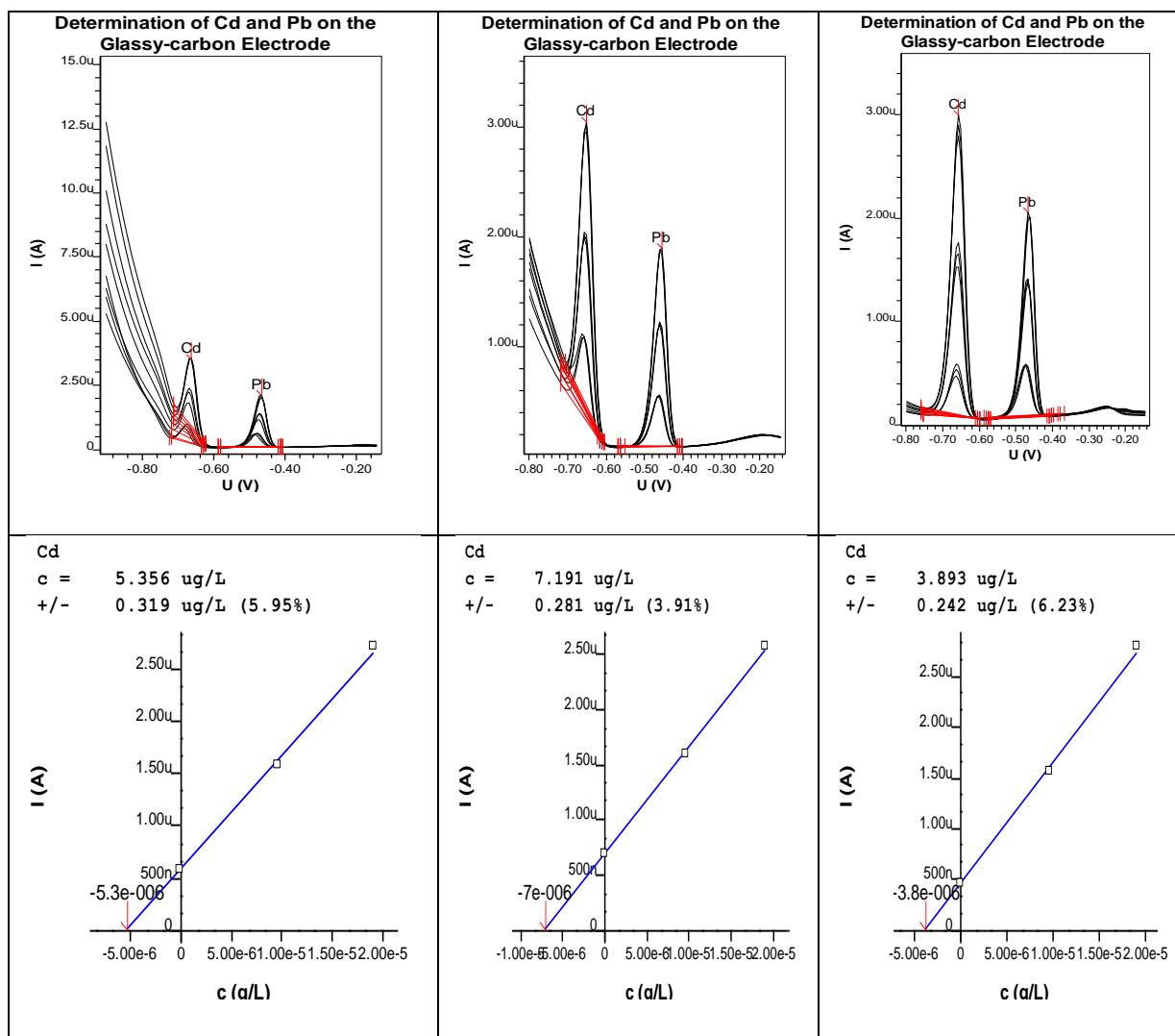


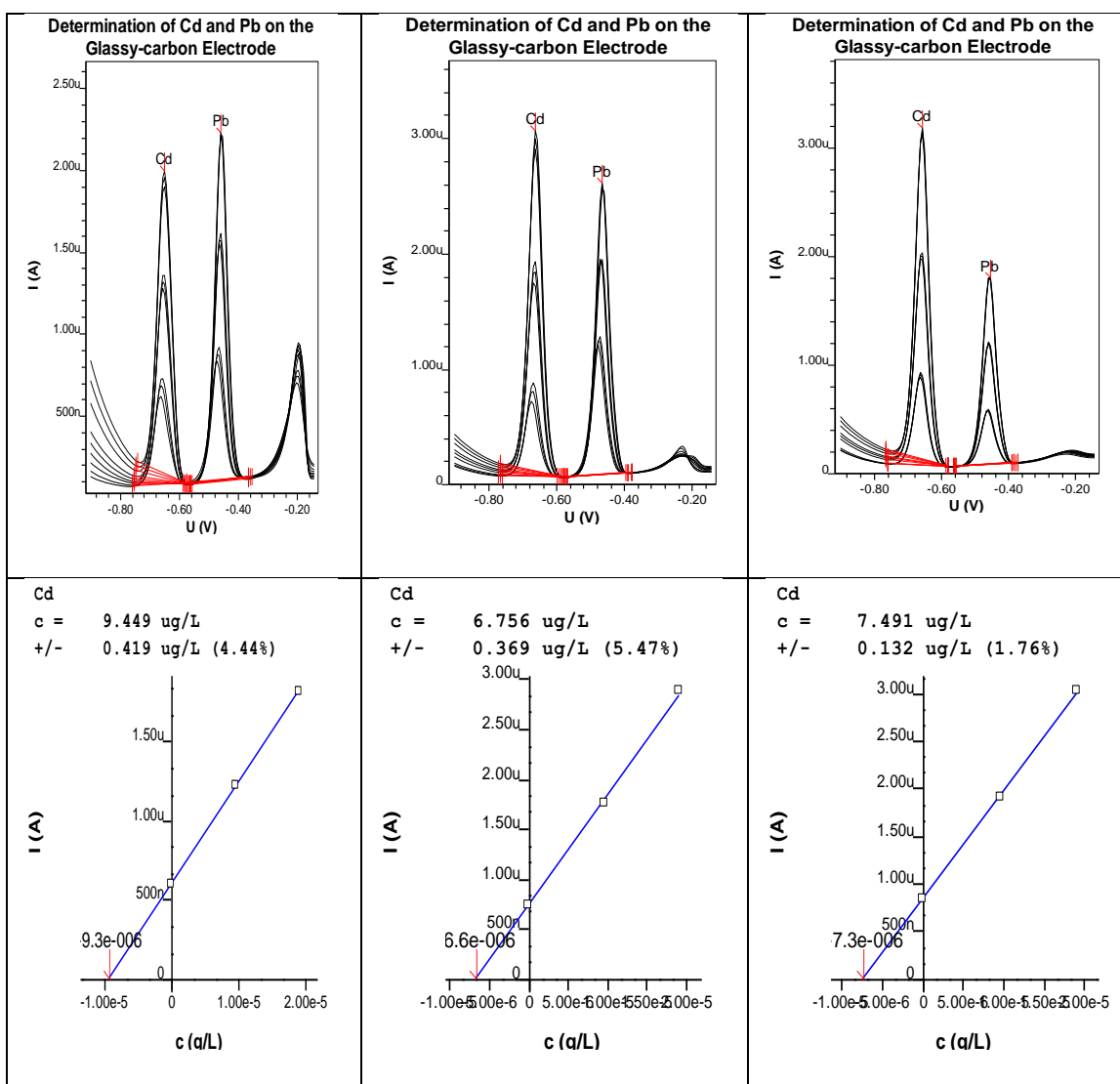
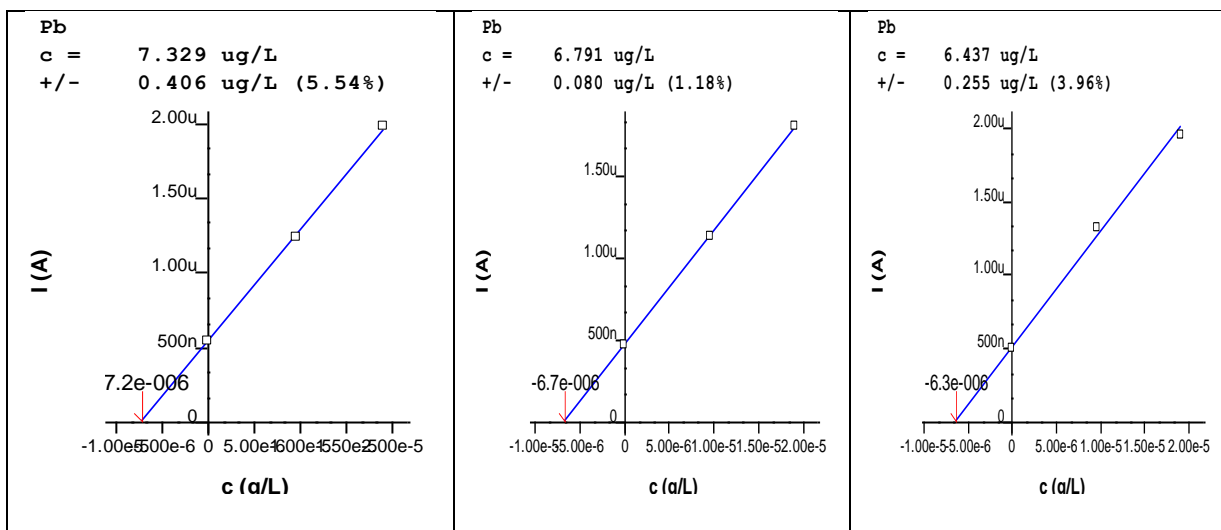


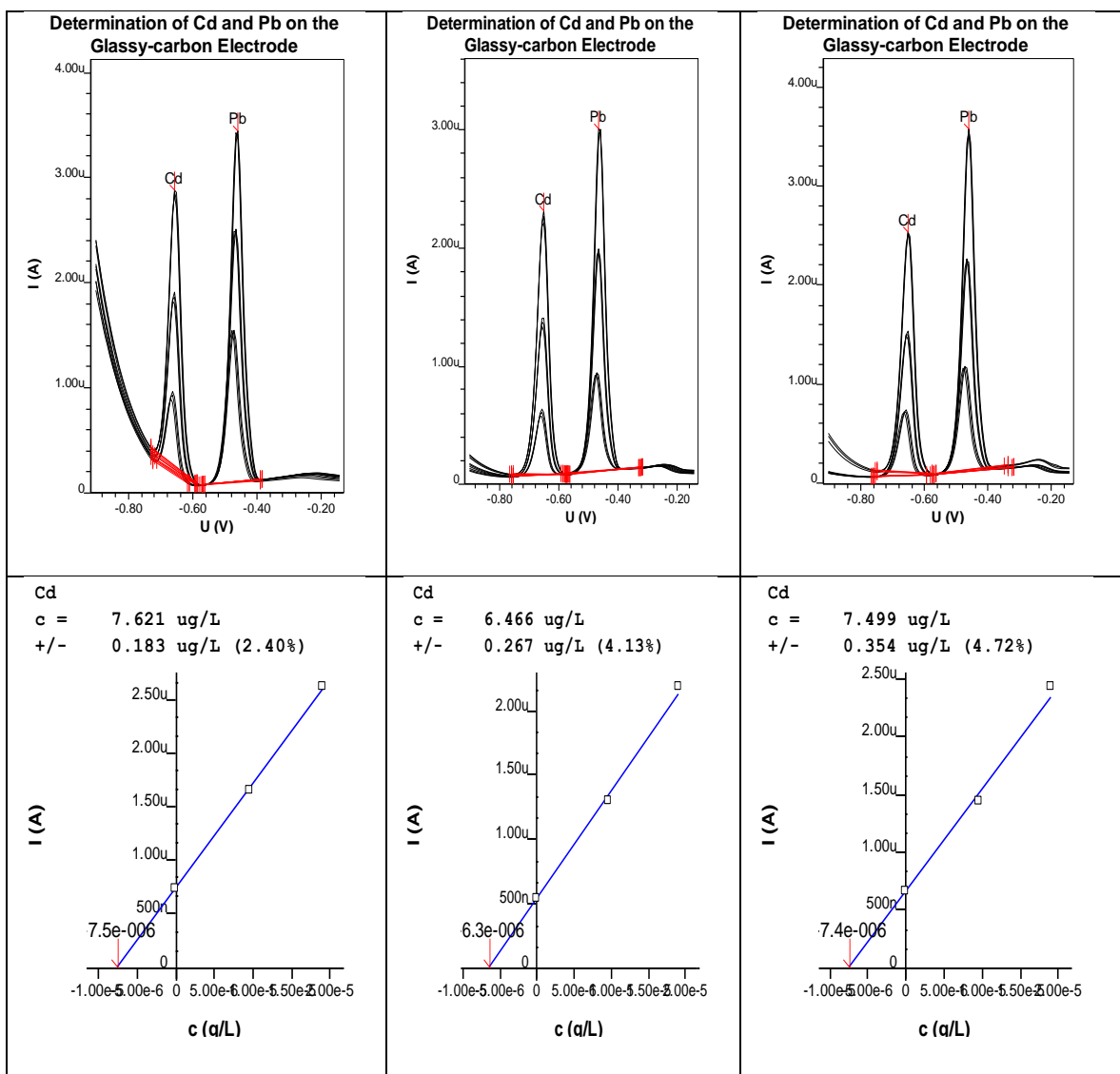
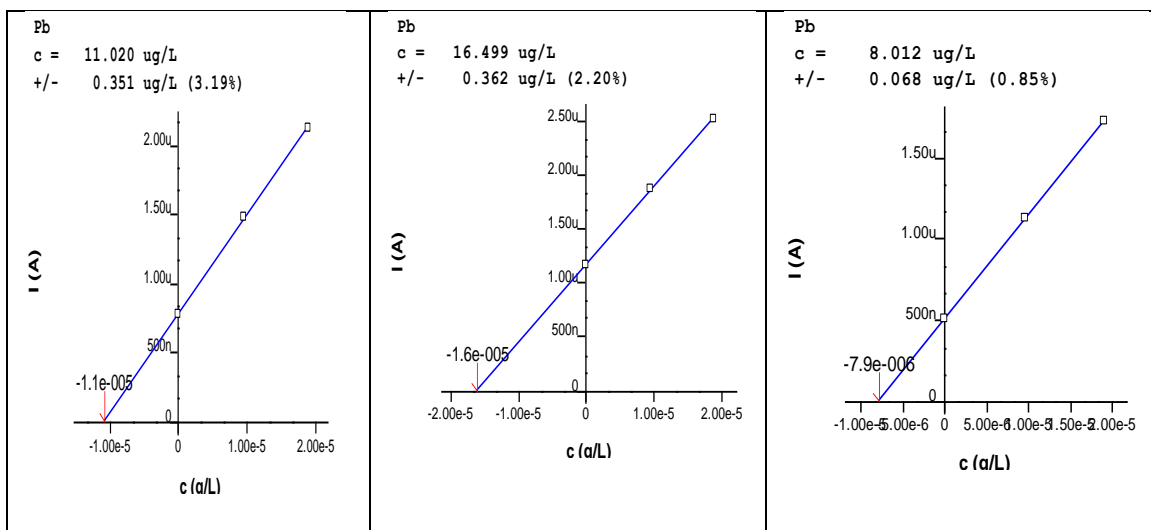


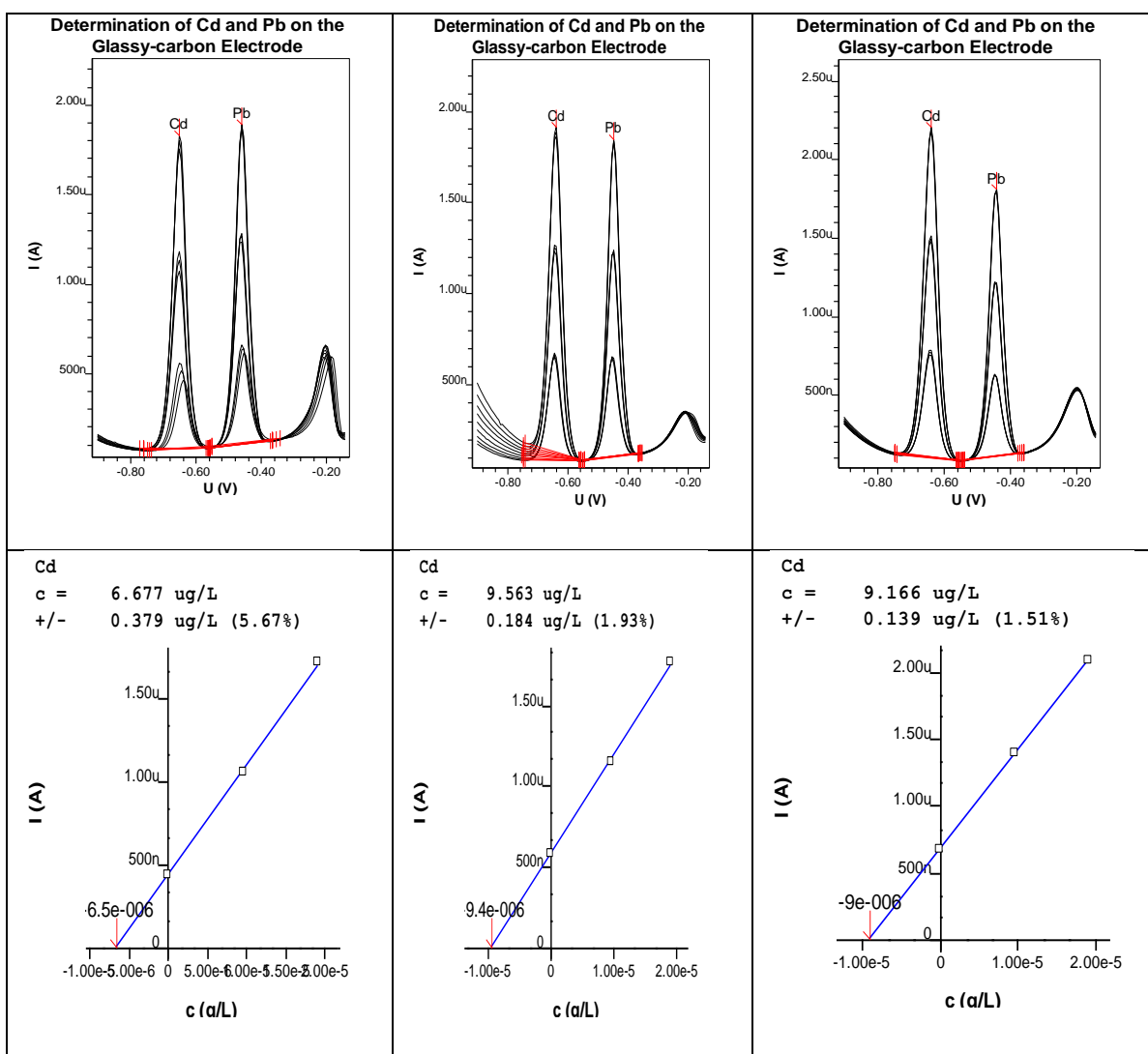
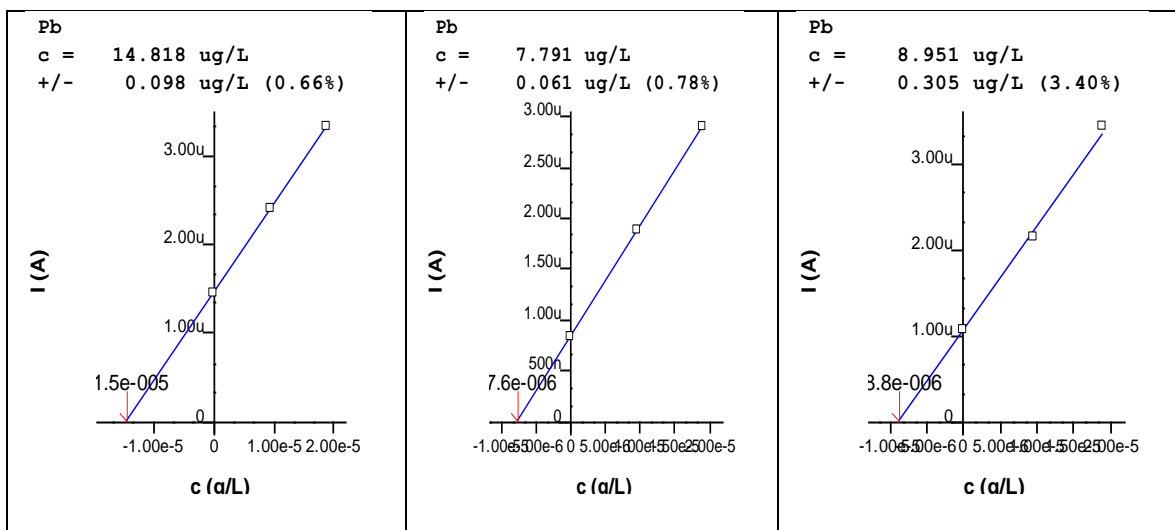


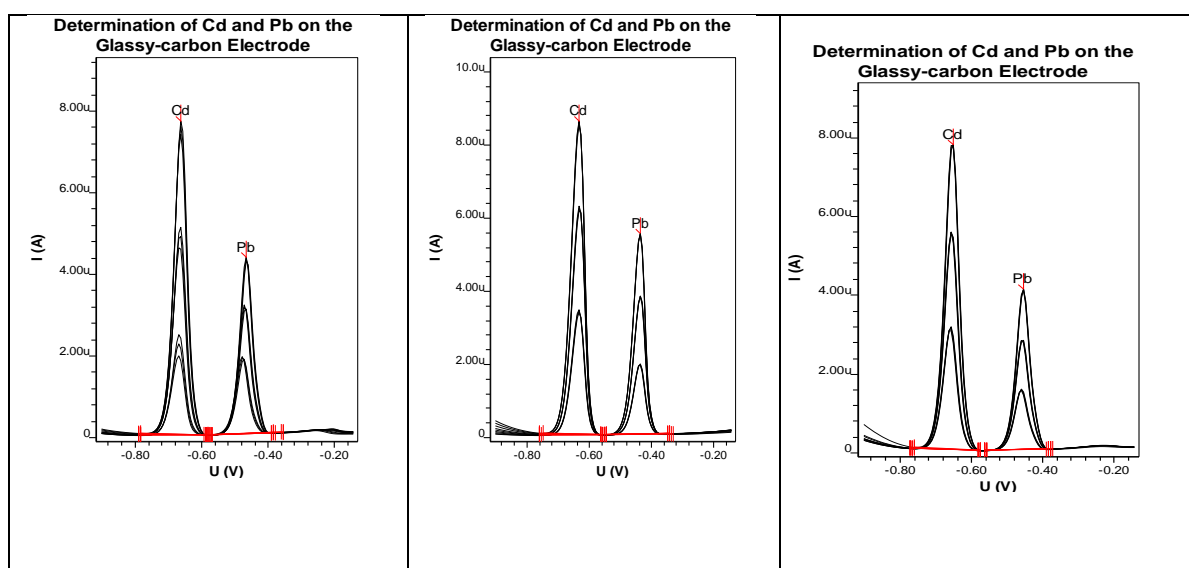
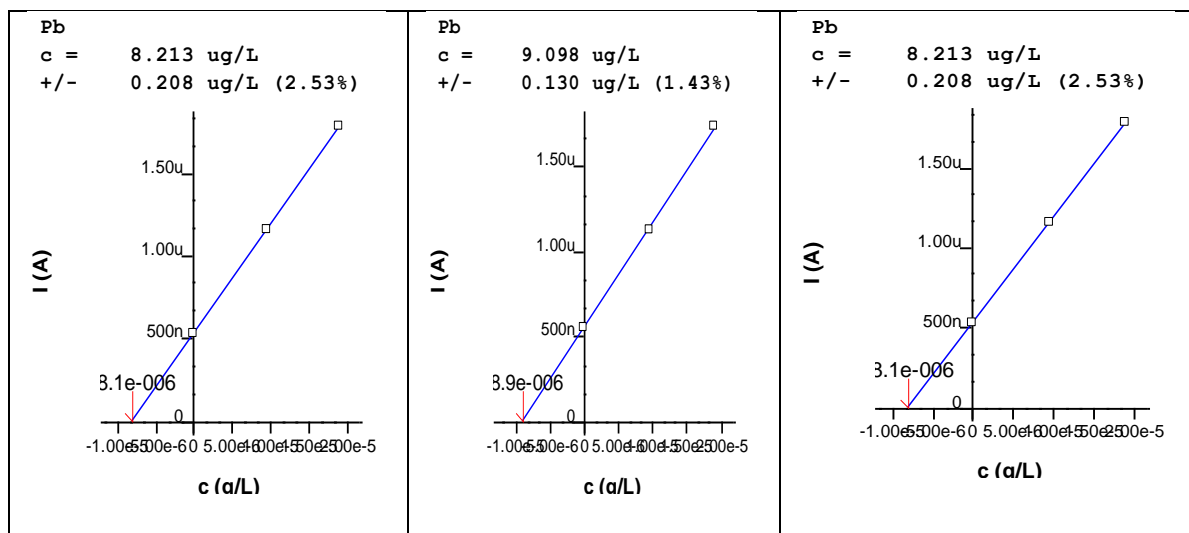
## APPENDIX 6: Results for a pre-validation study of trace metal determination in marine water

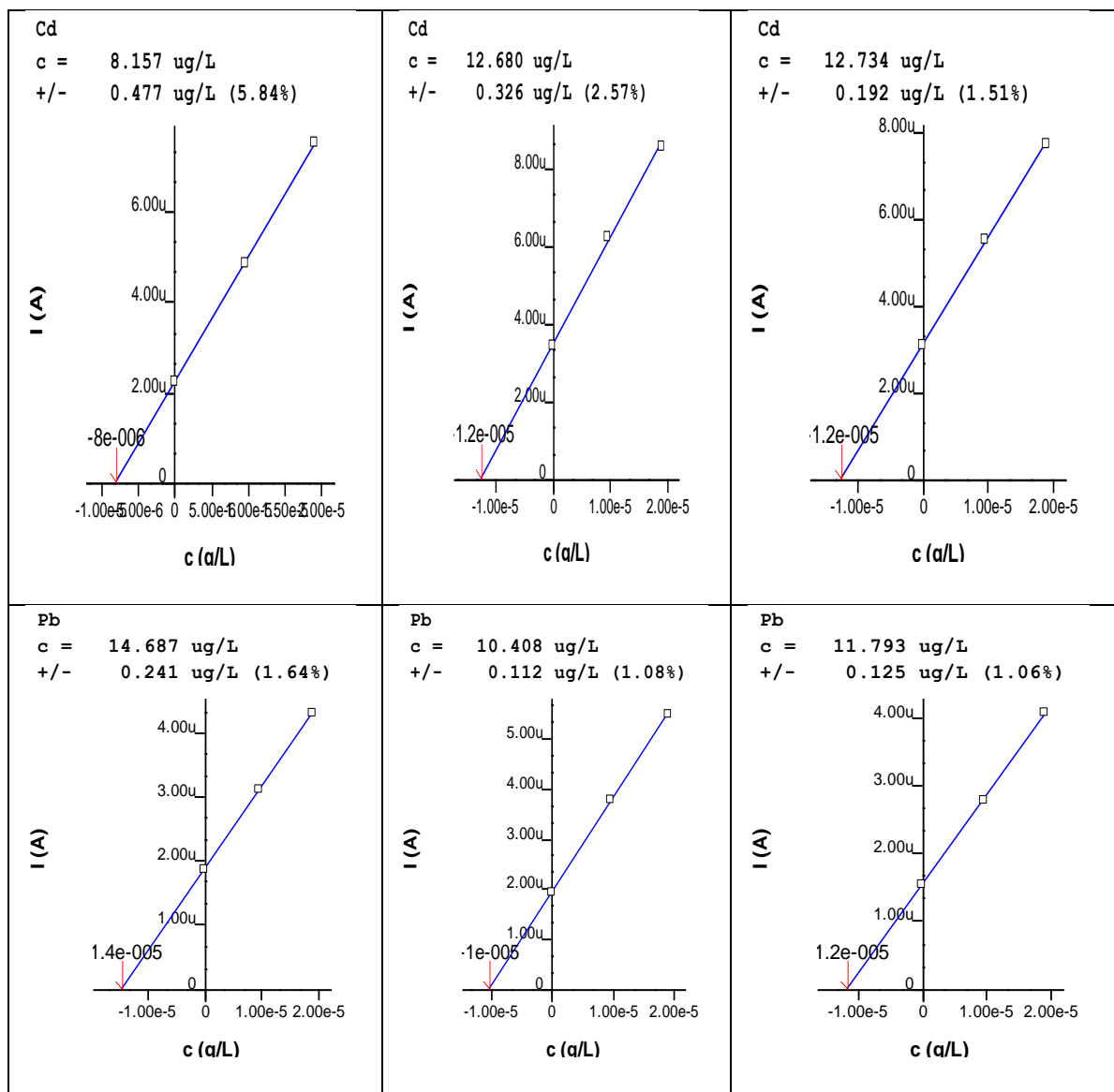




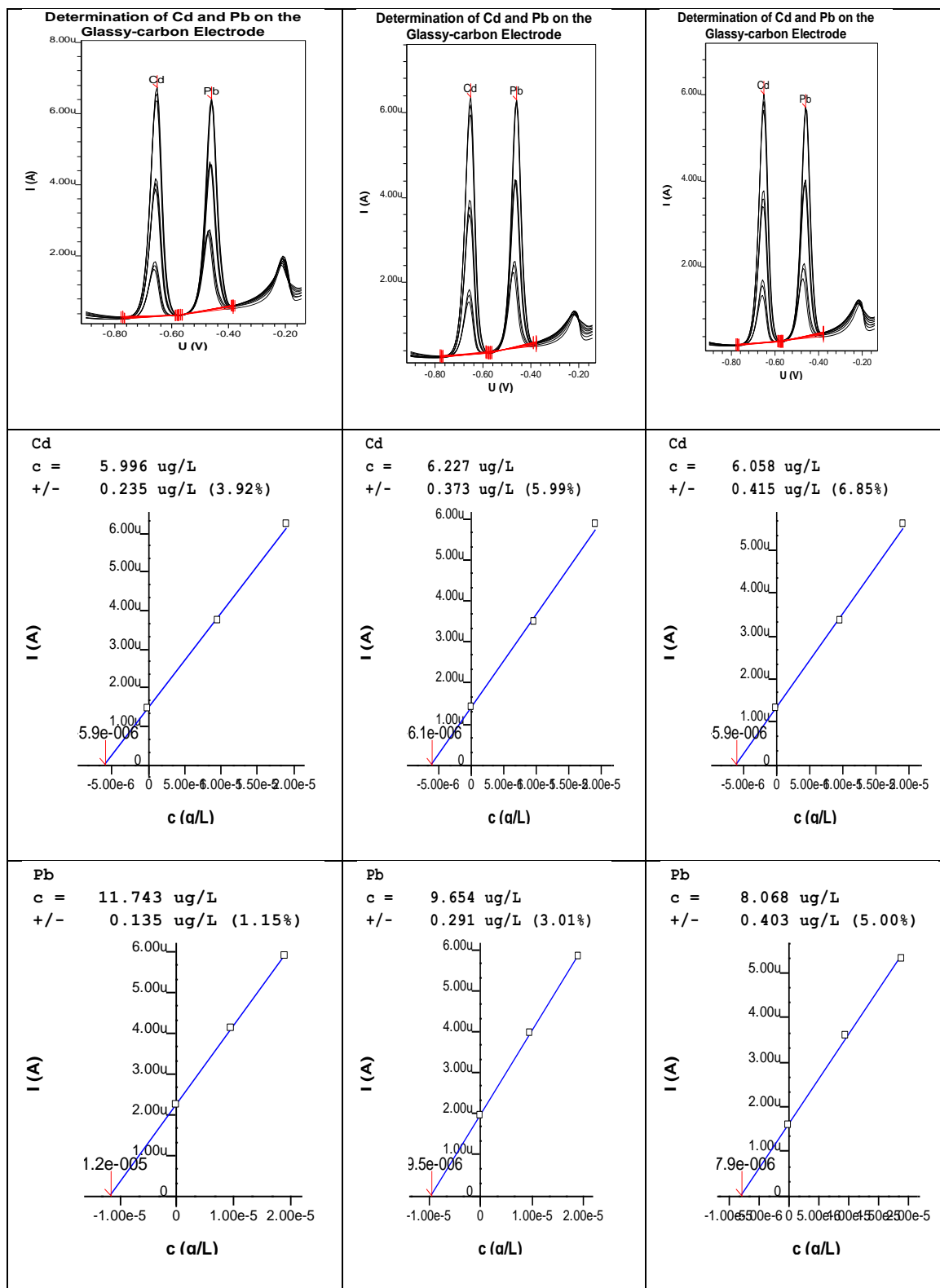


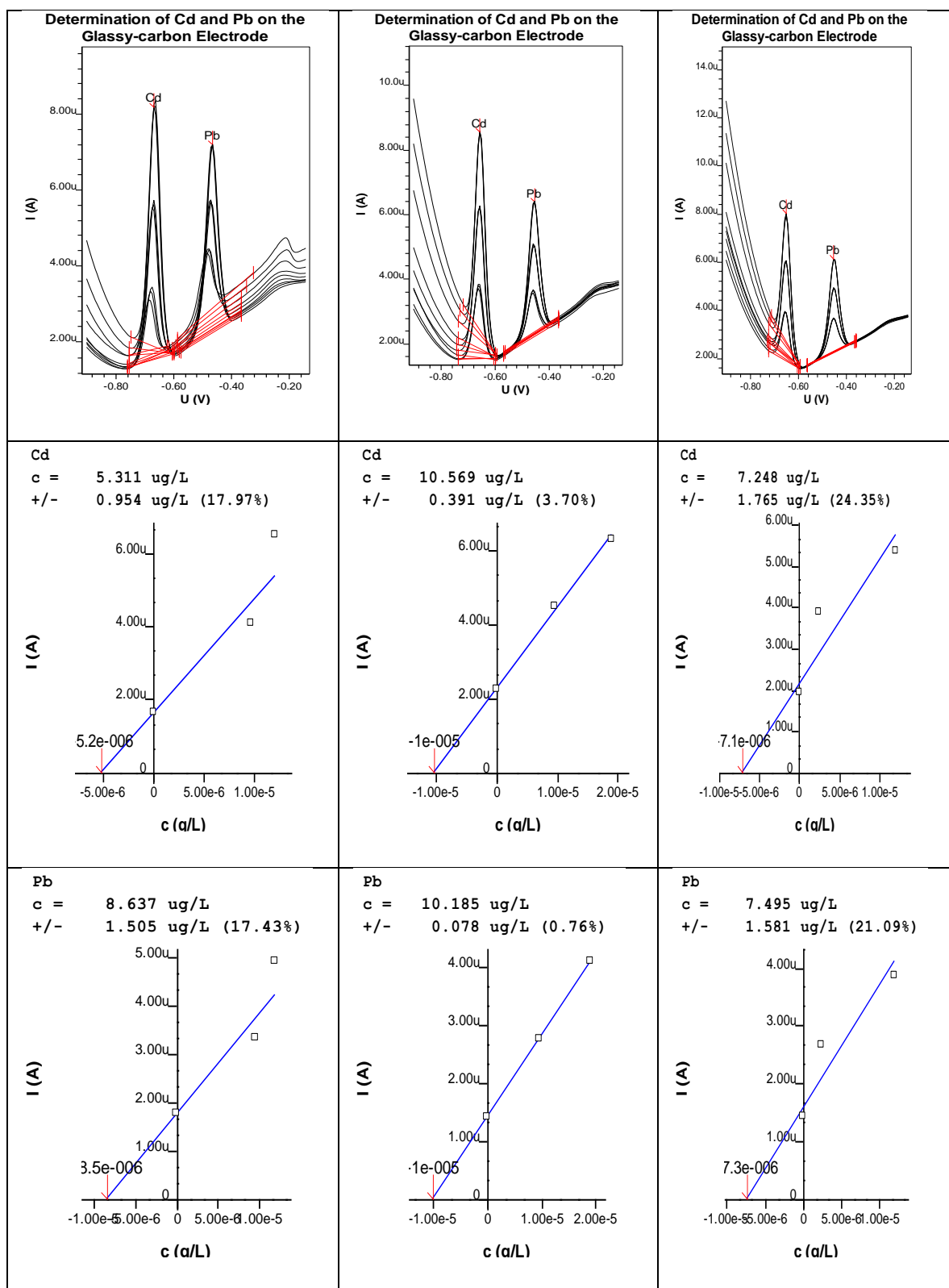




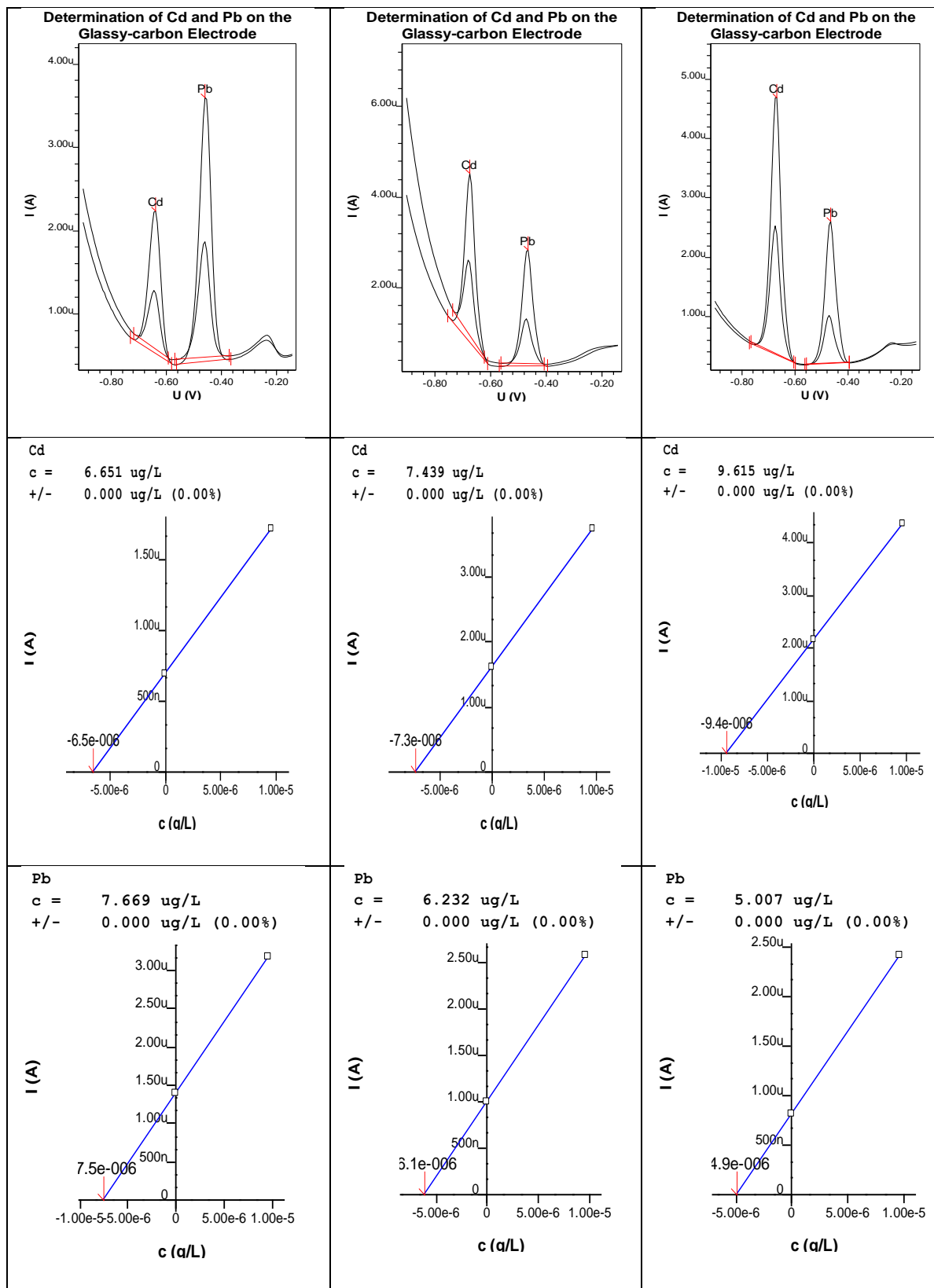


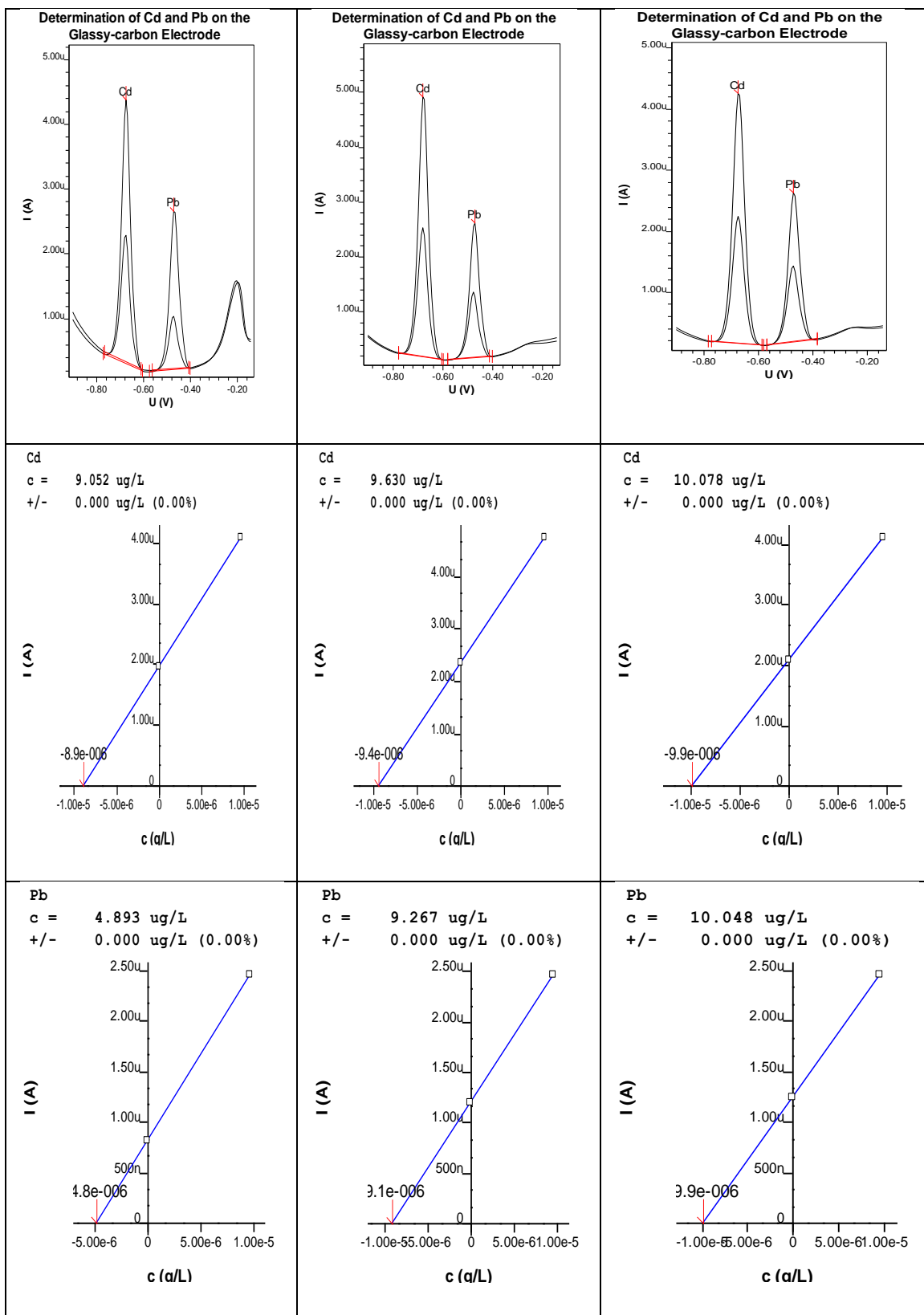


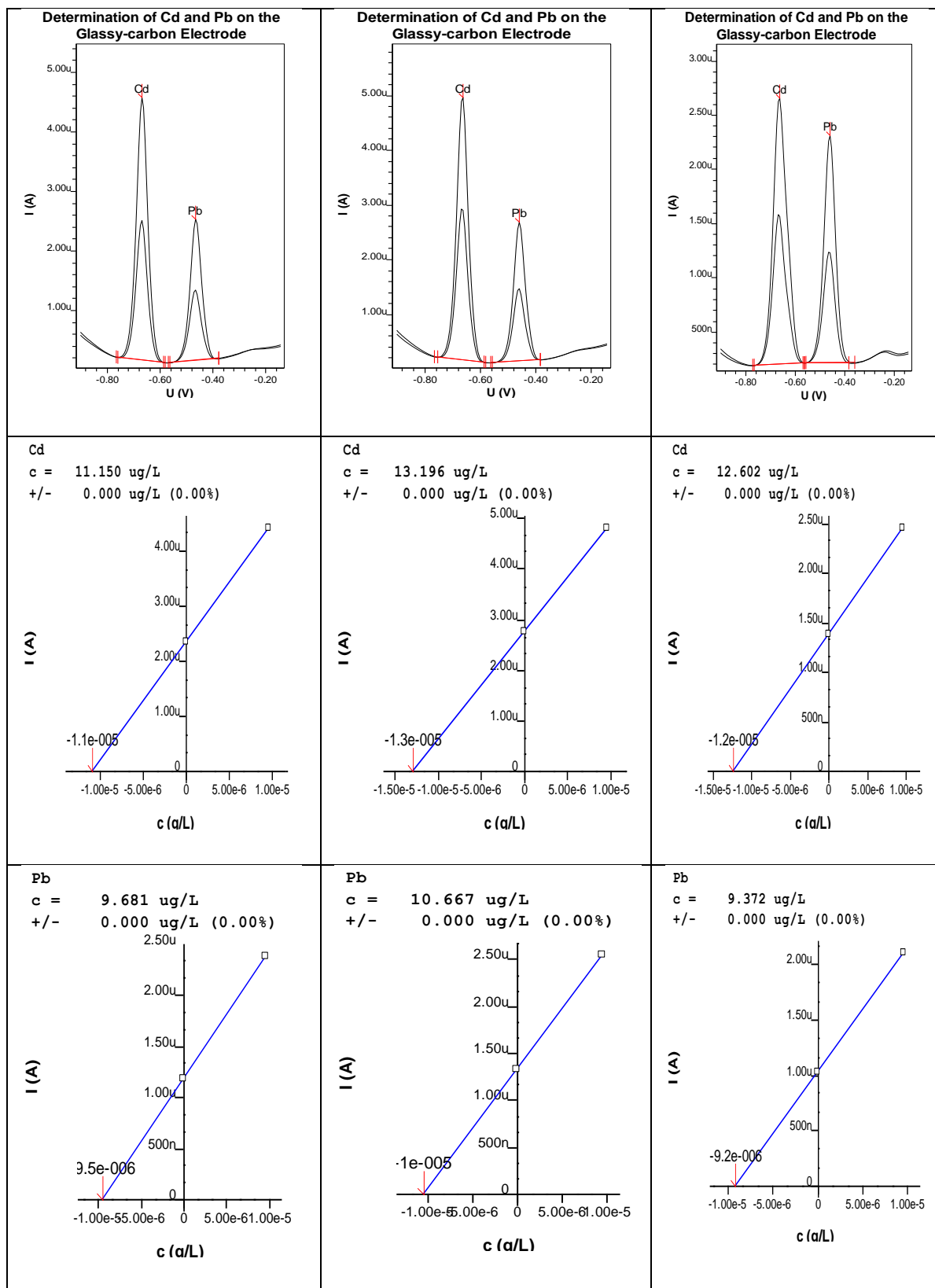


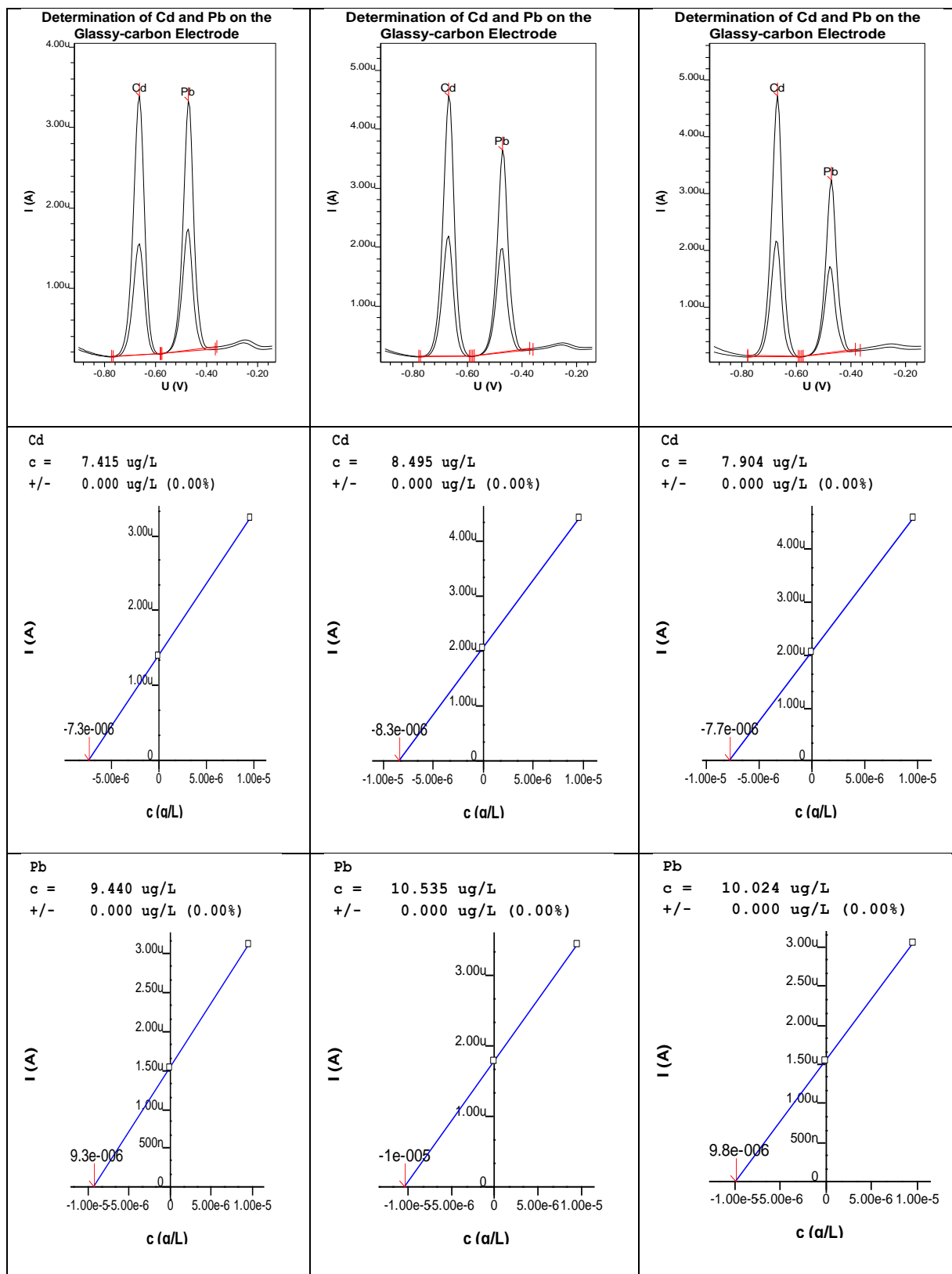


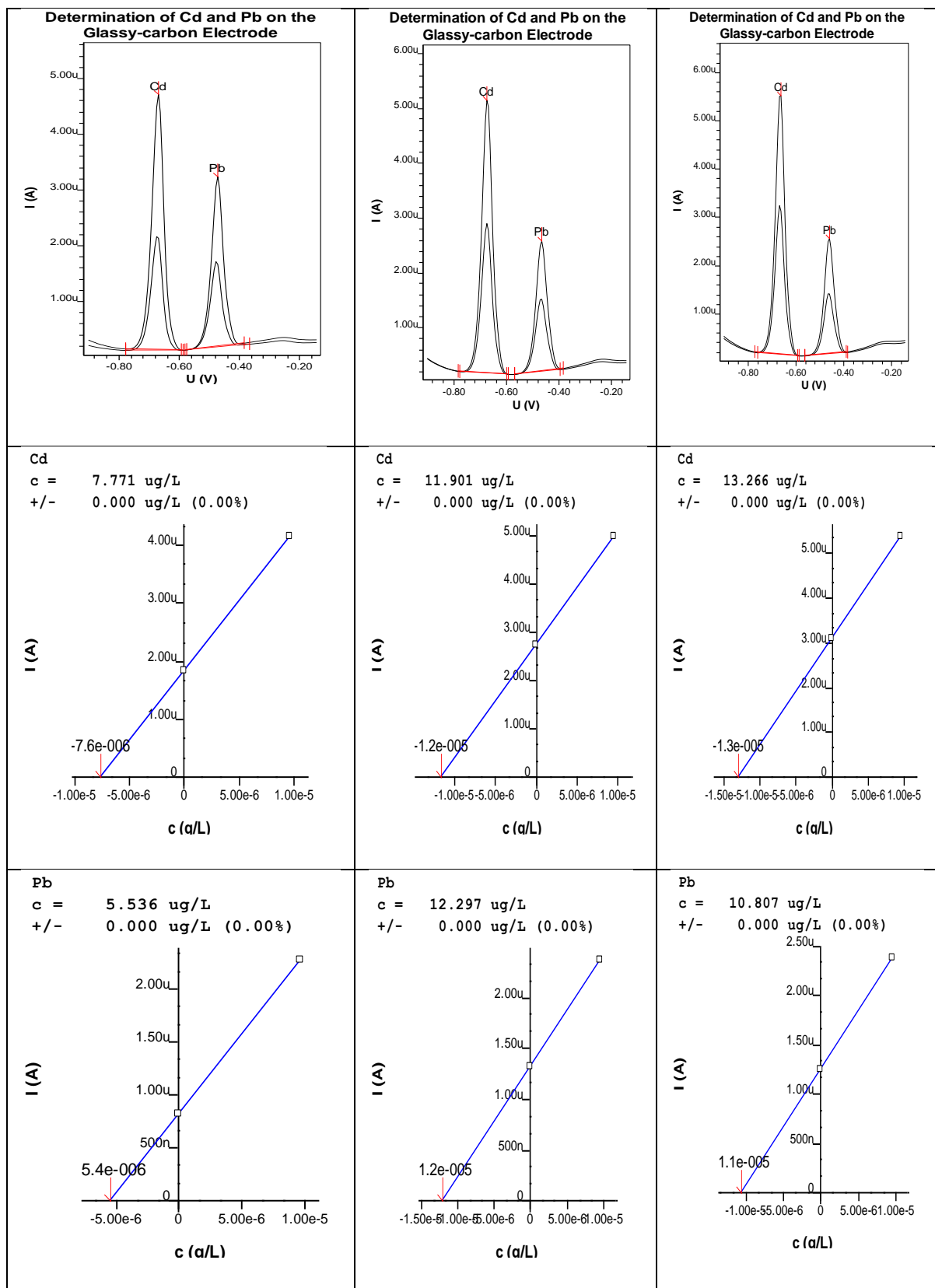
## Appendix 7: Results for 10ppb validation study for trace metal determination for case study 2

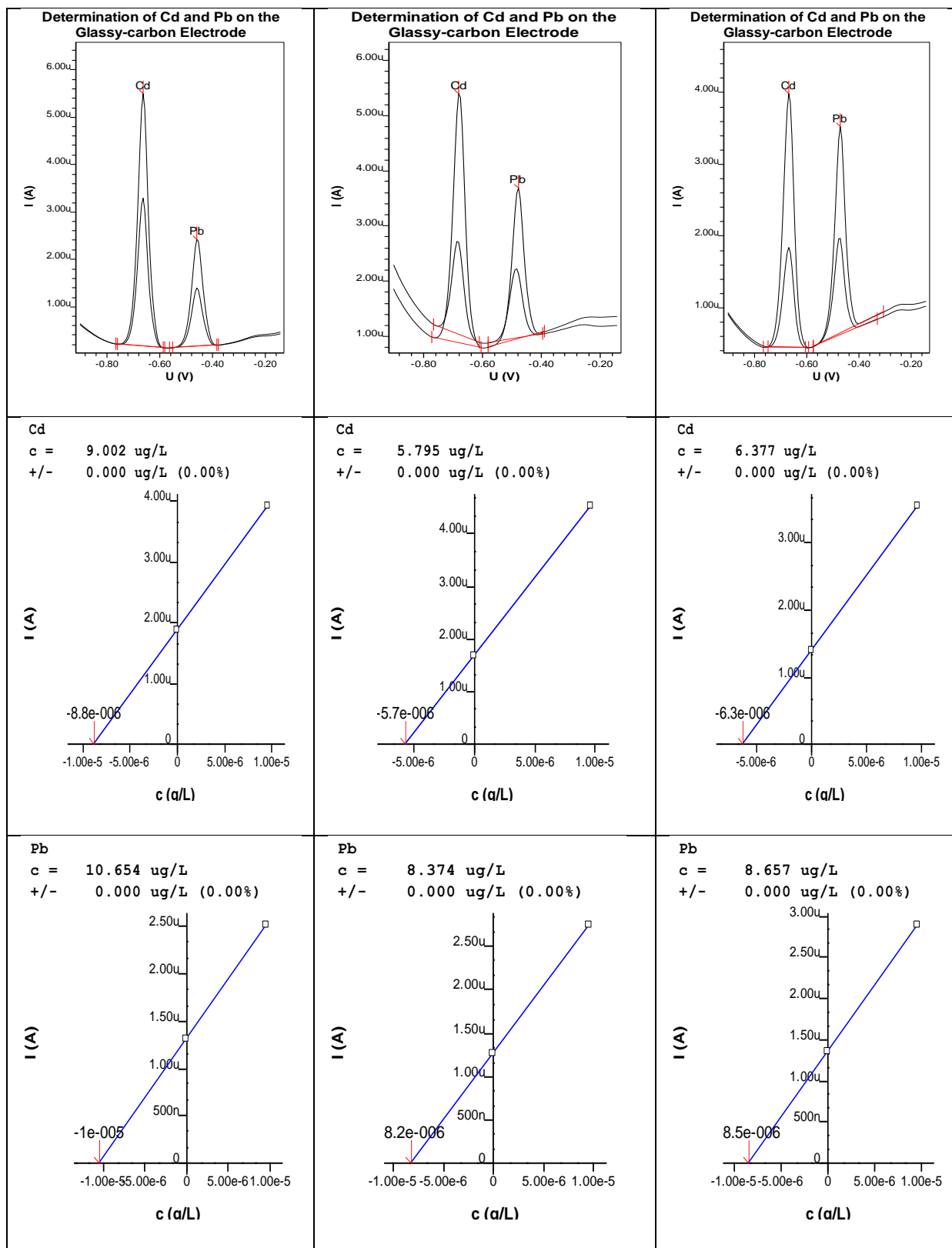




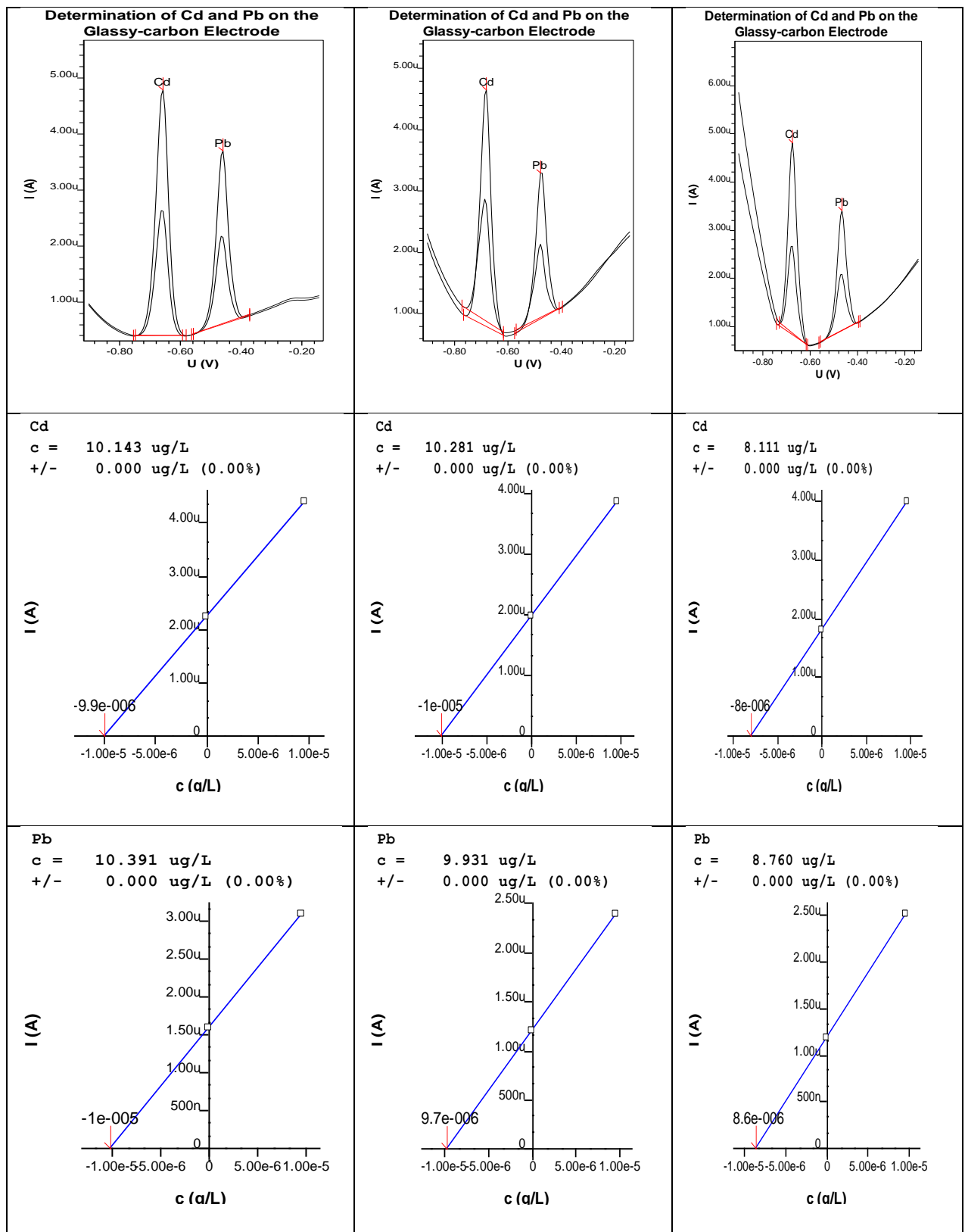


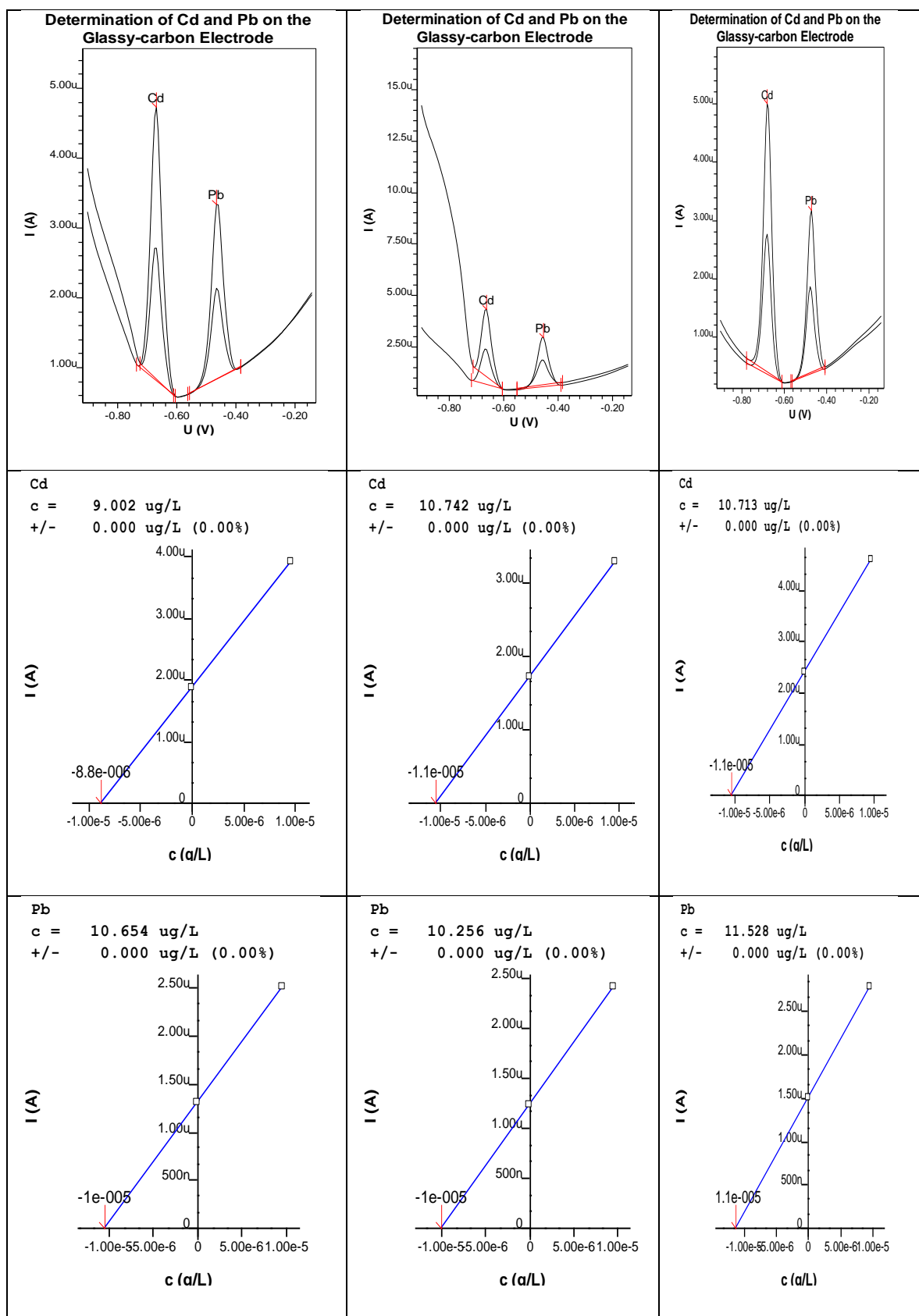


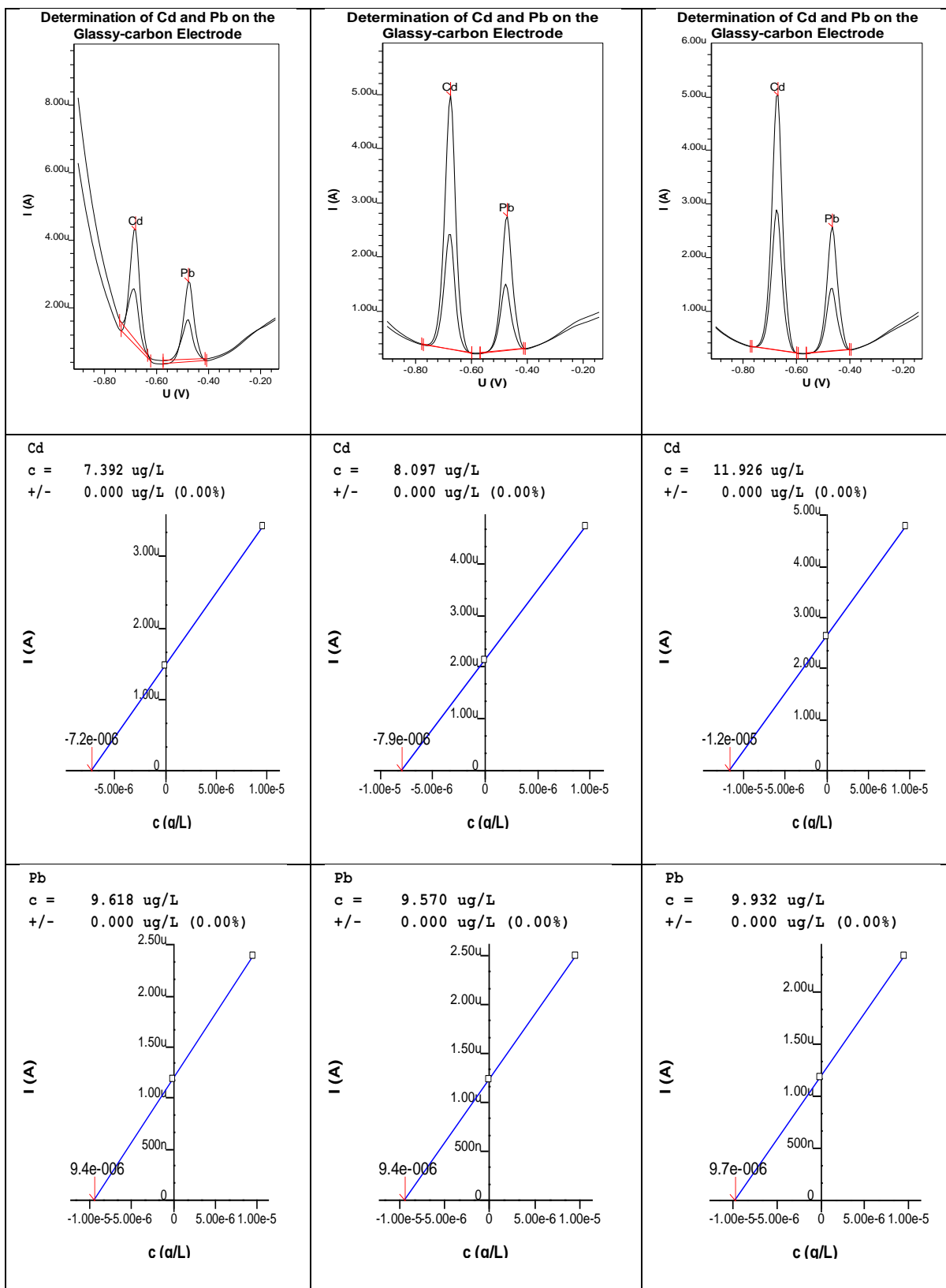


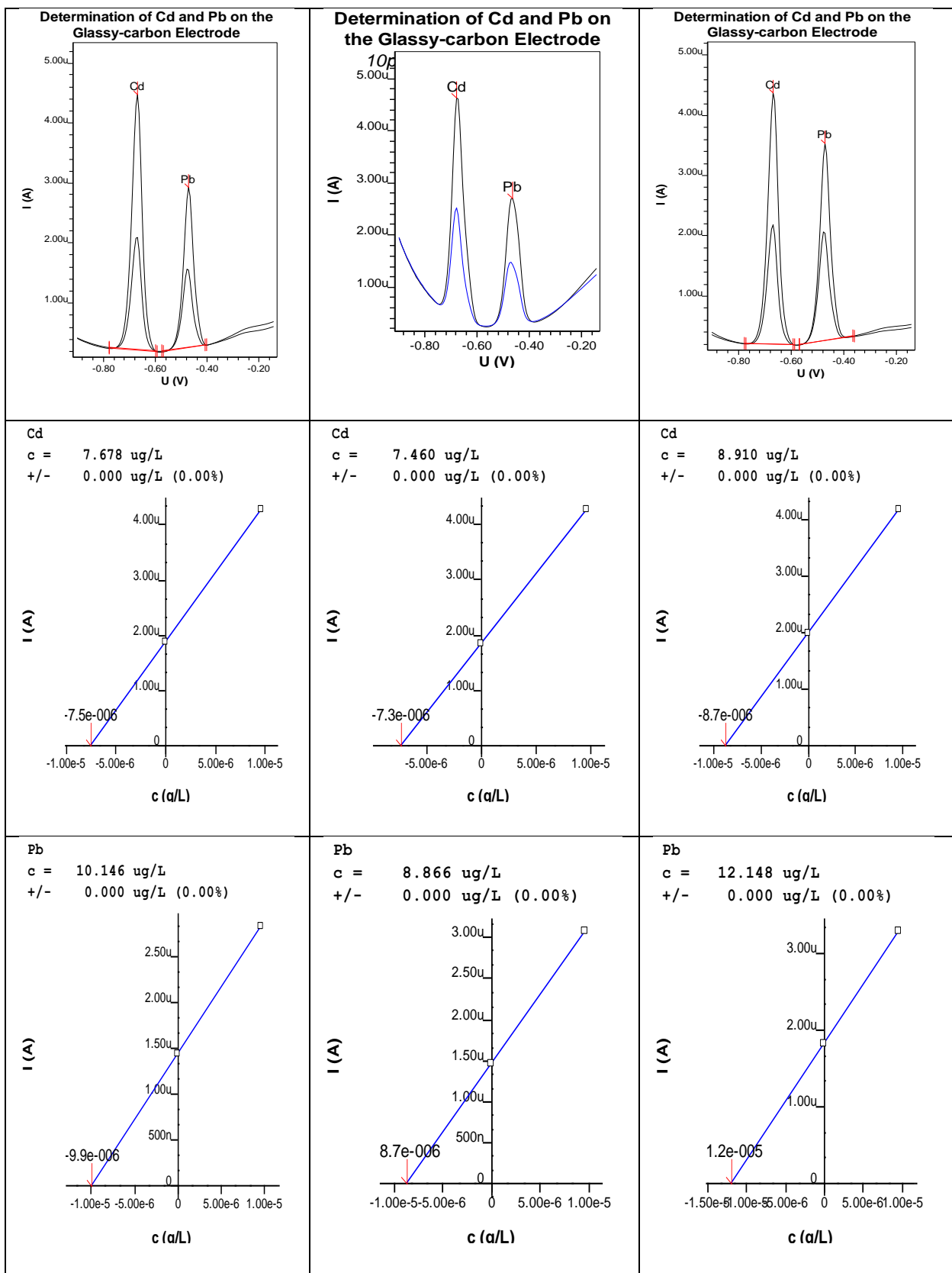


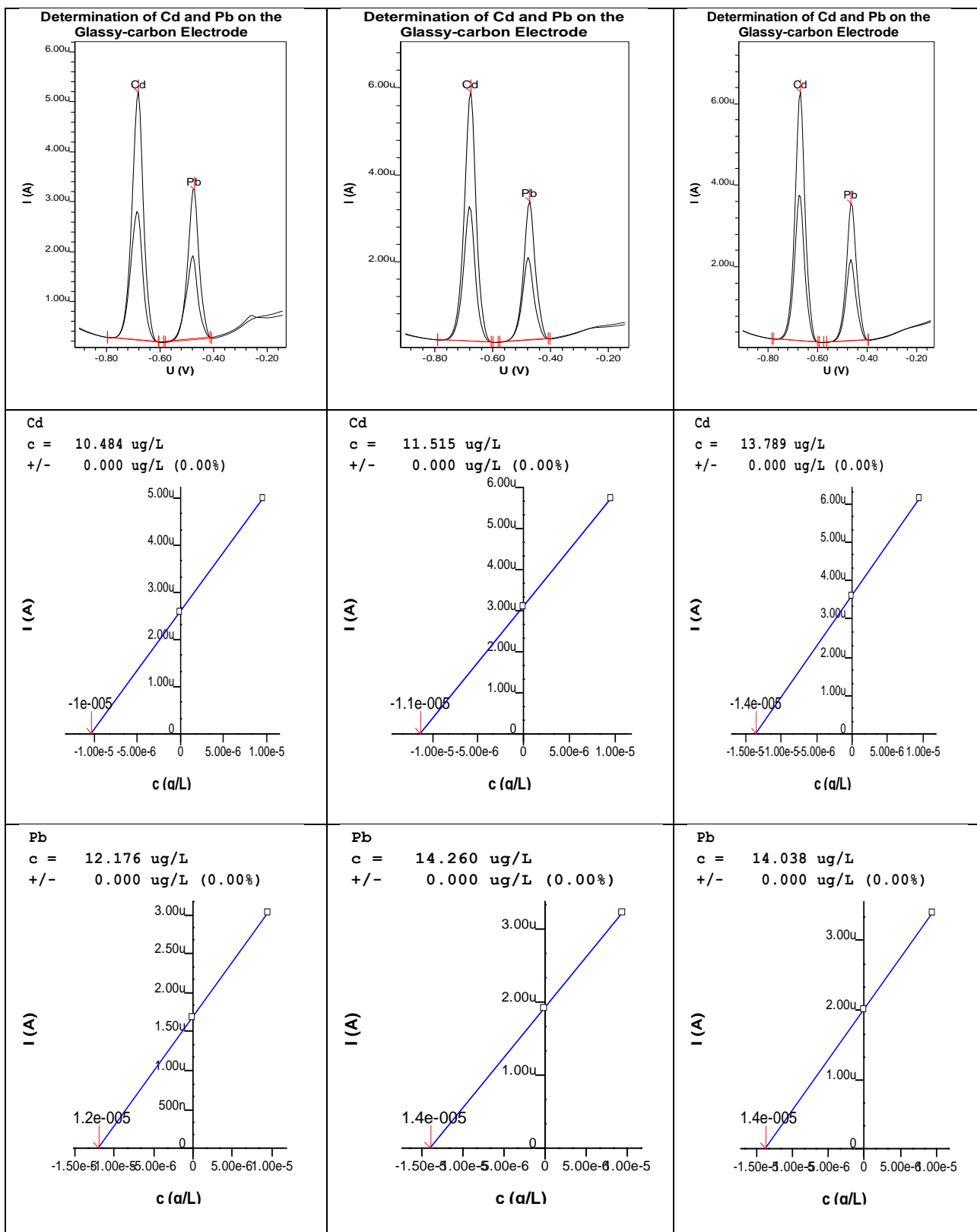


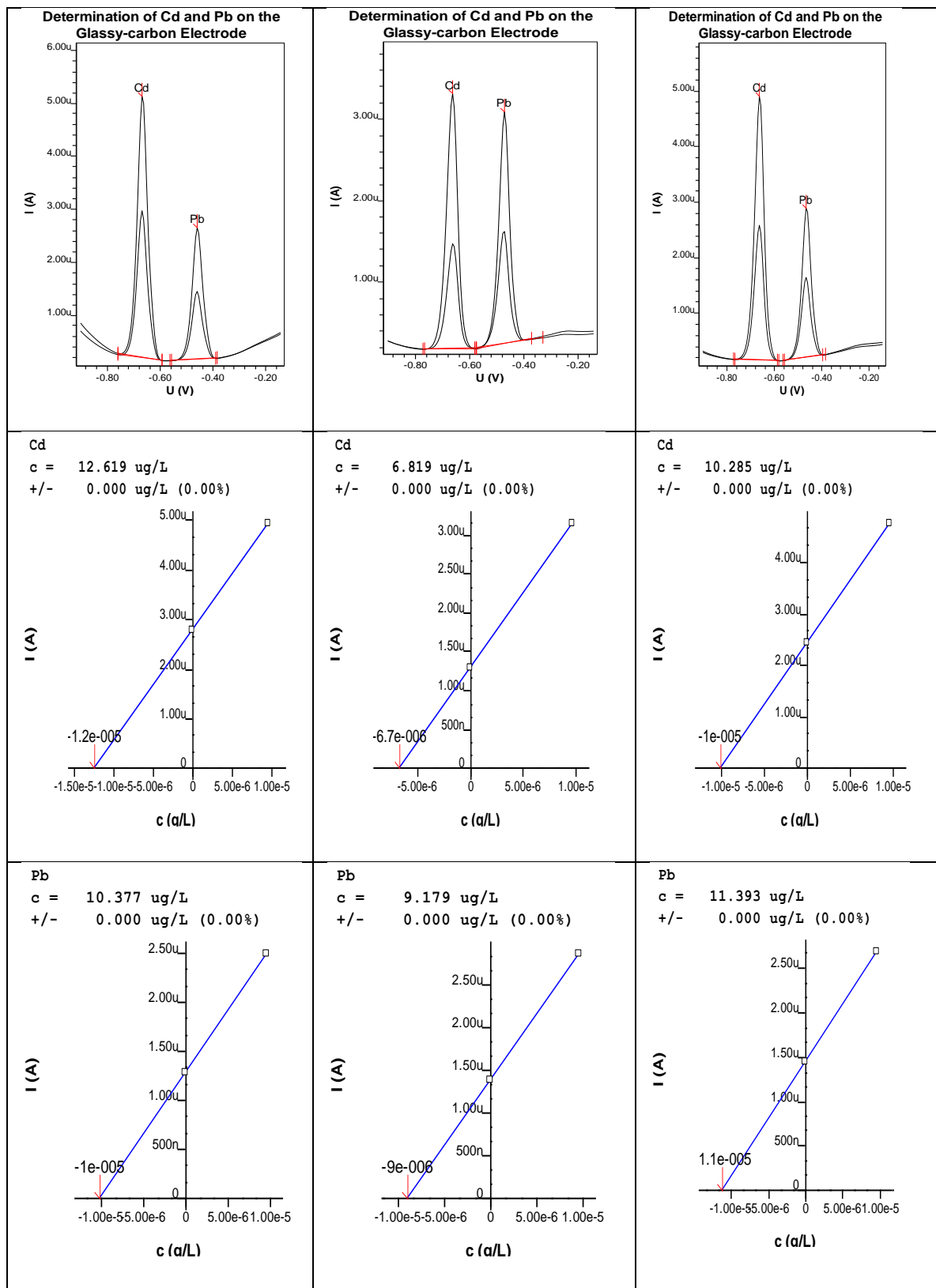


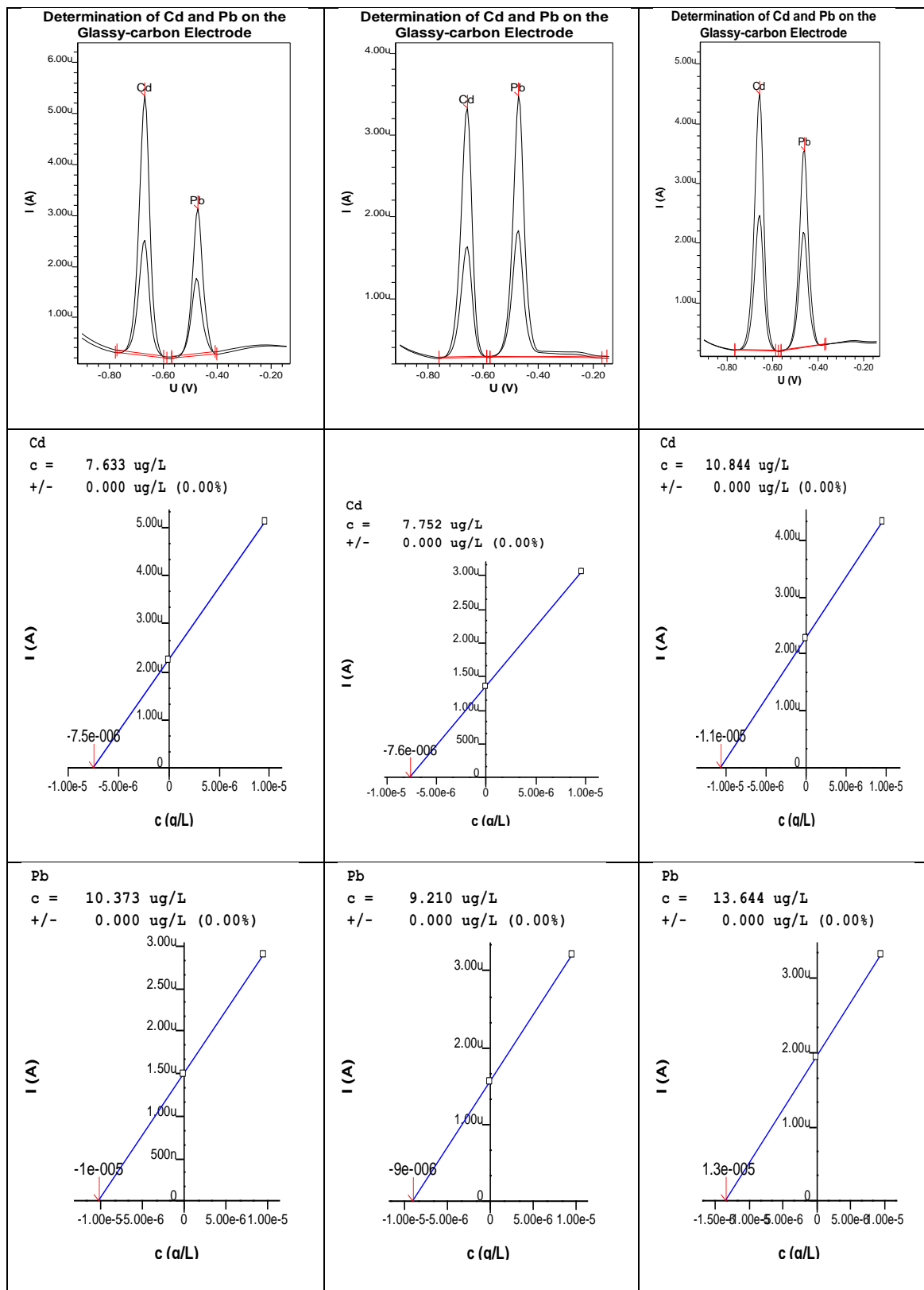


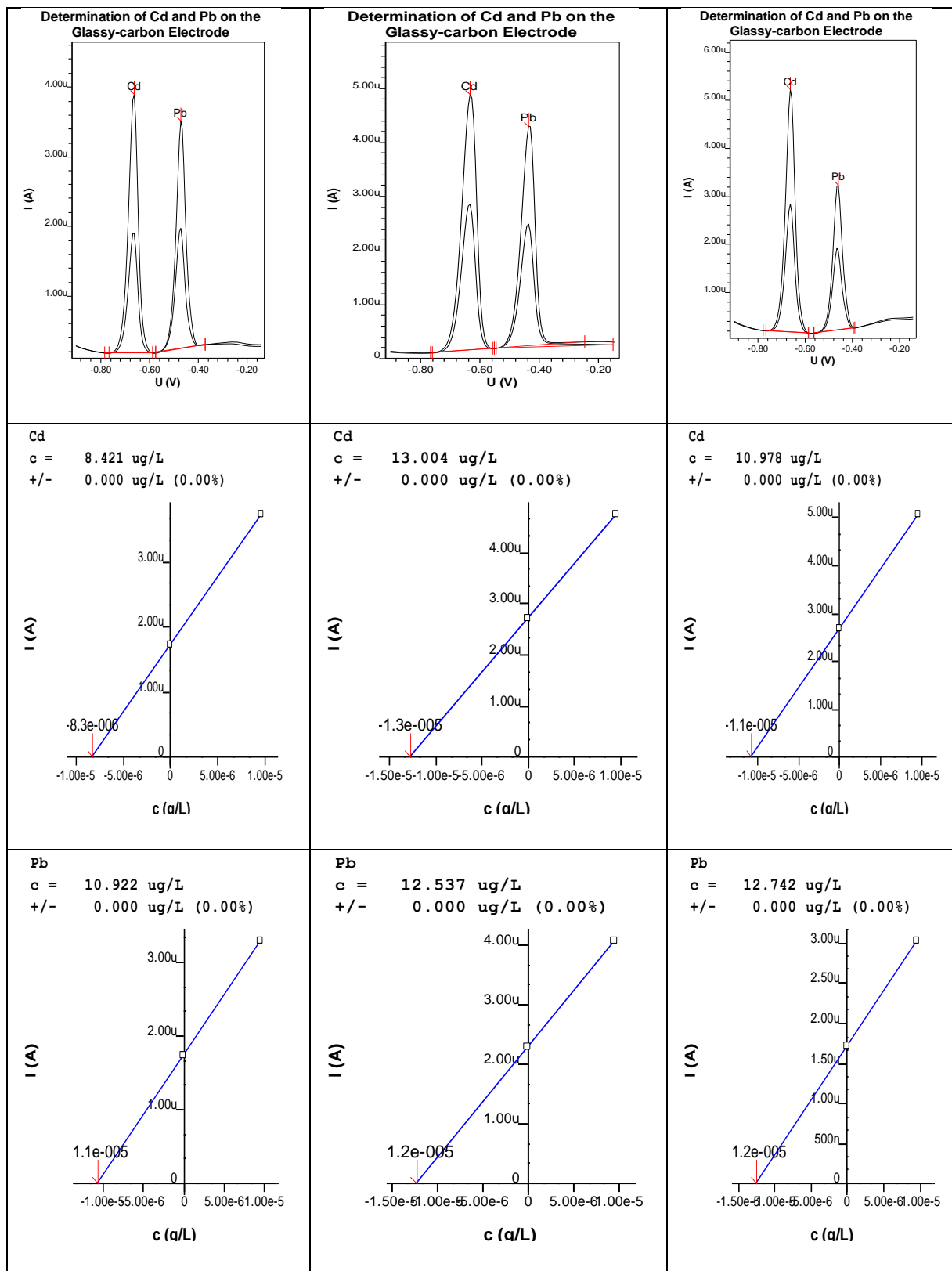




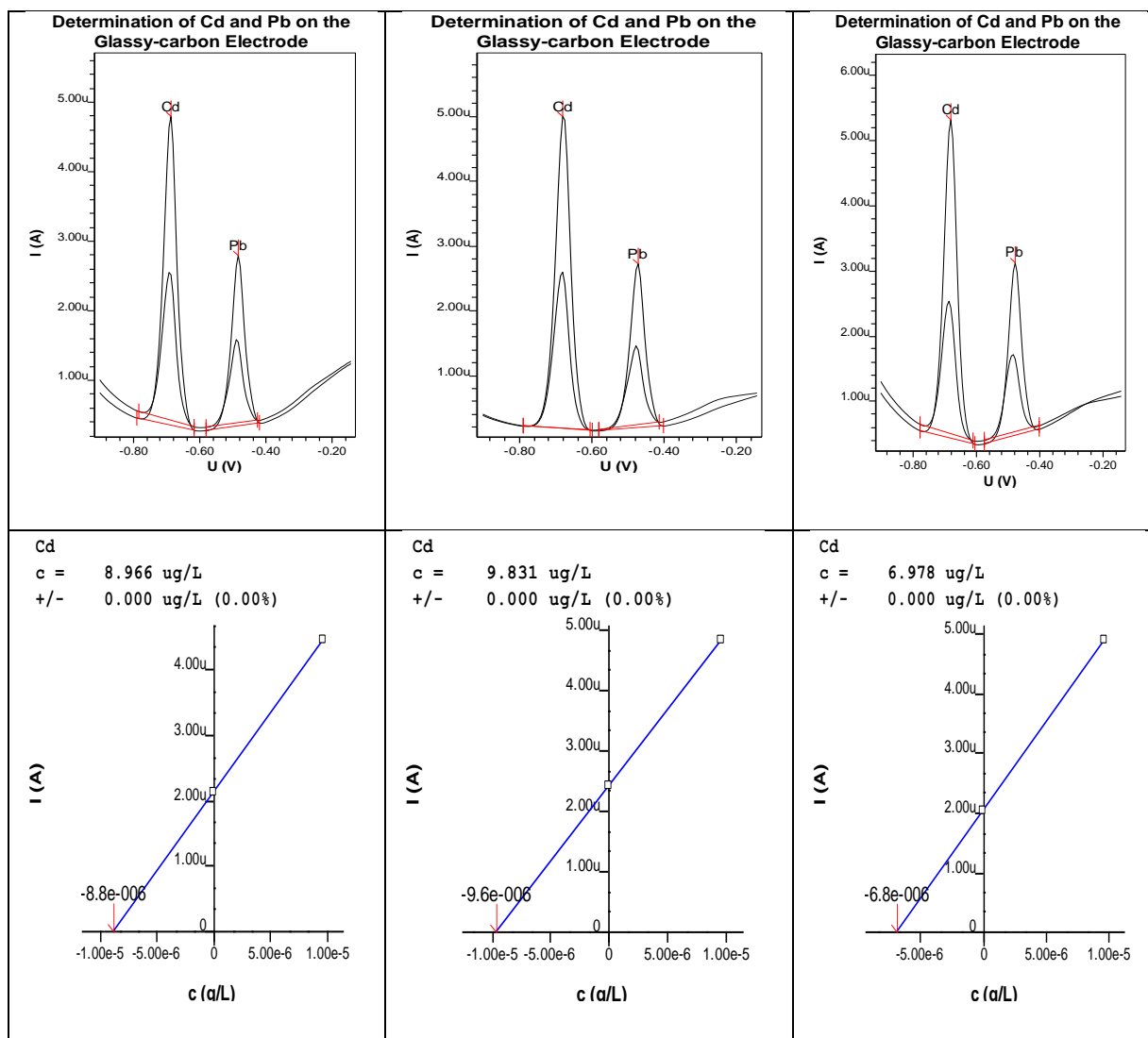


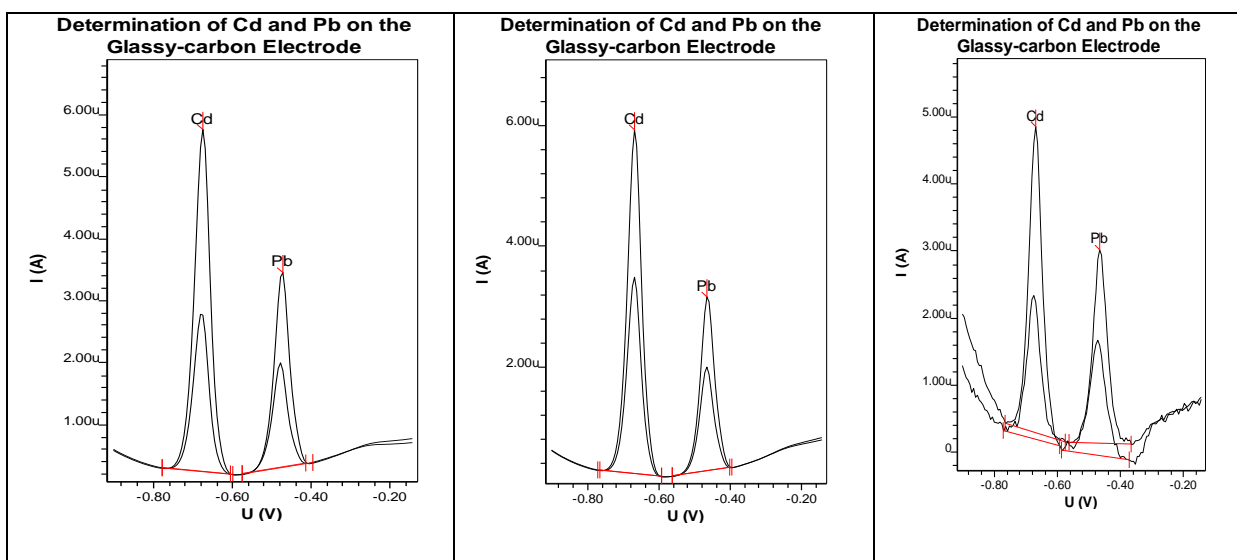
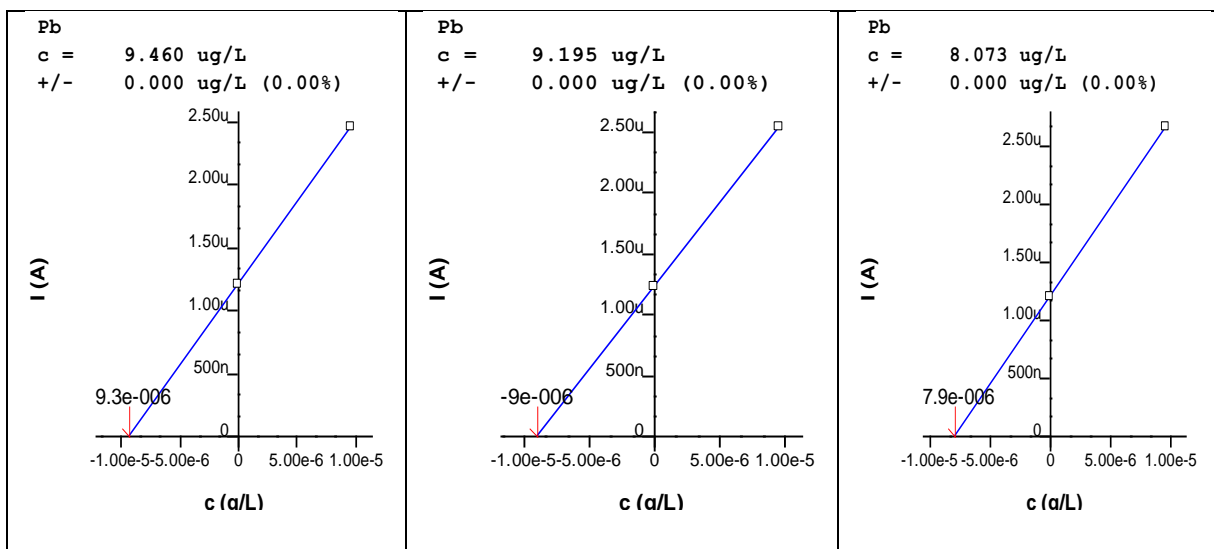


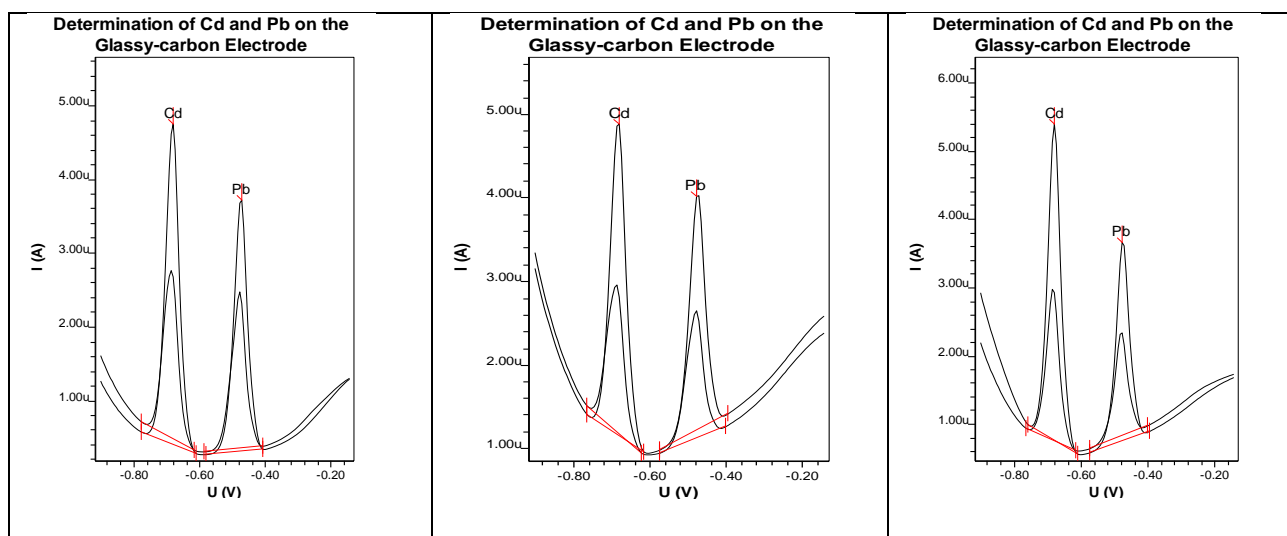
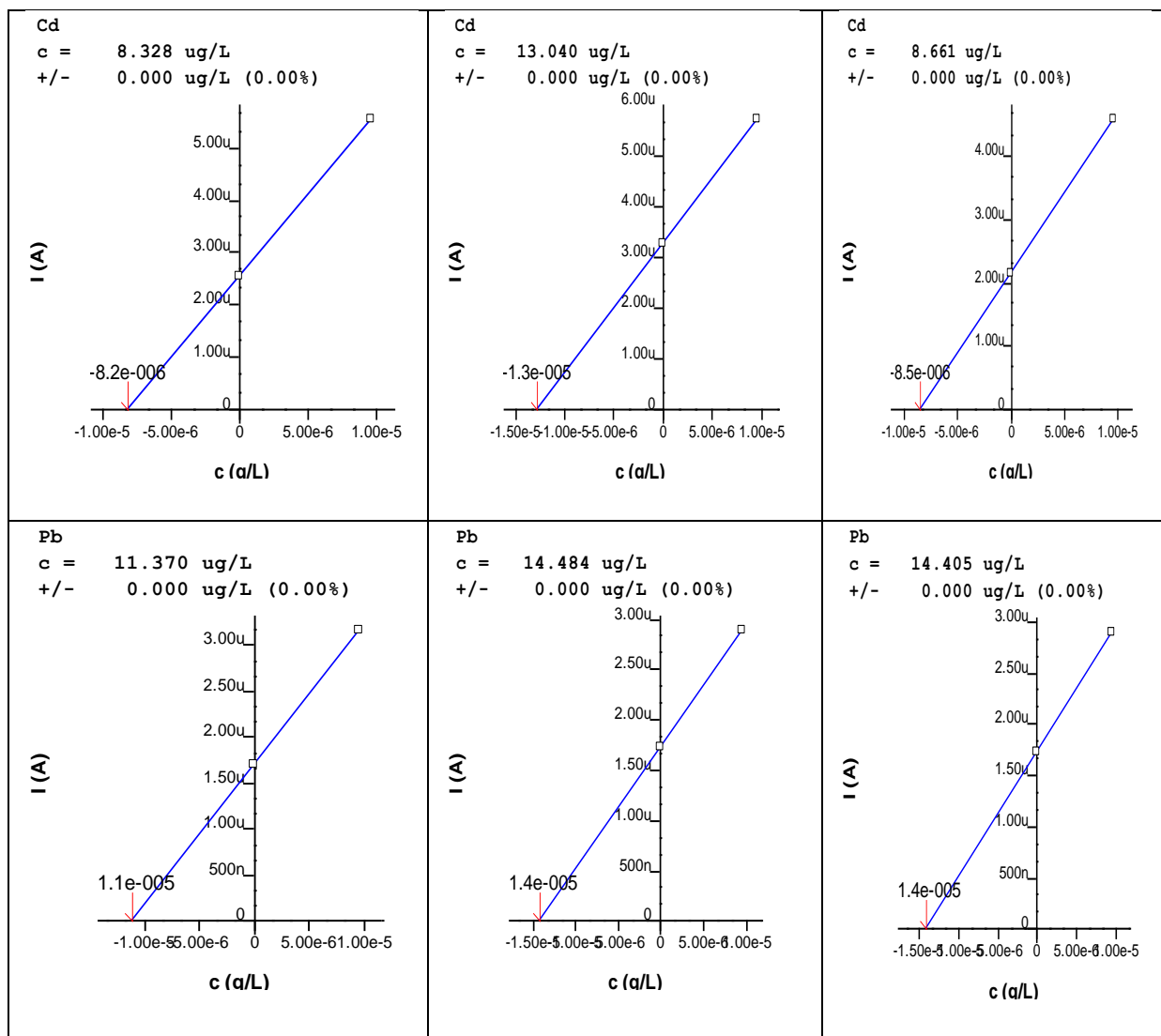


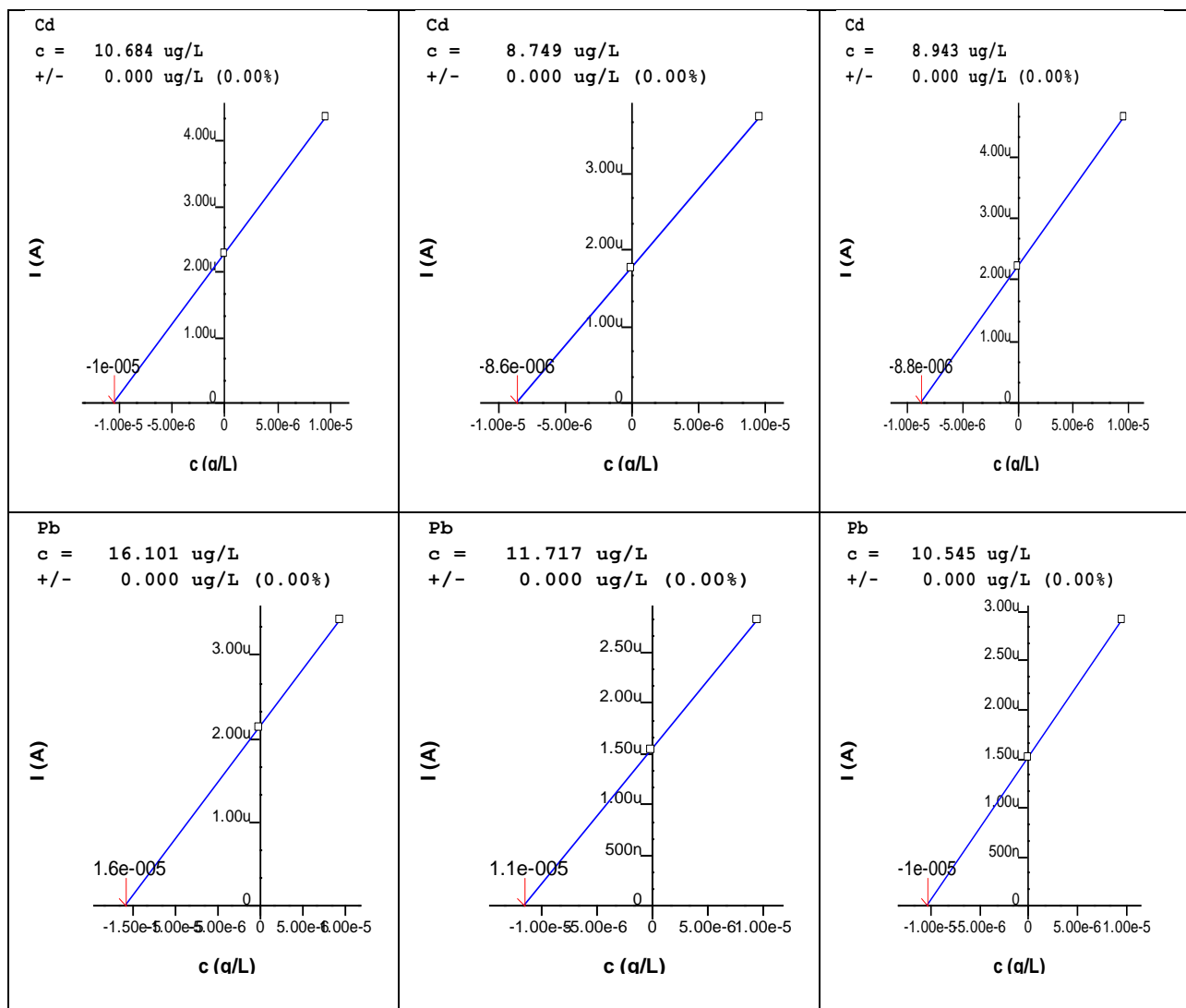


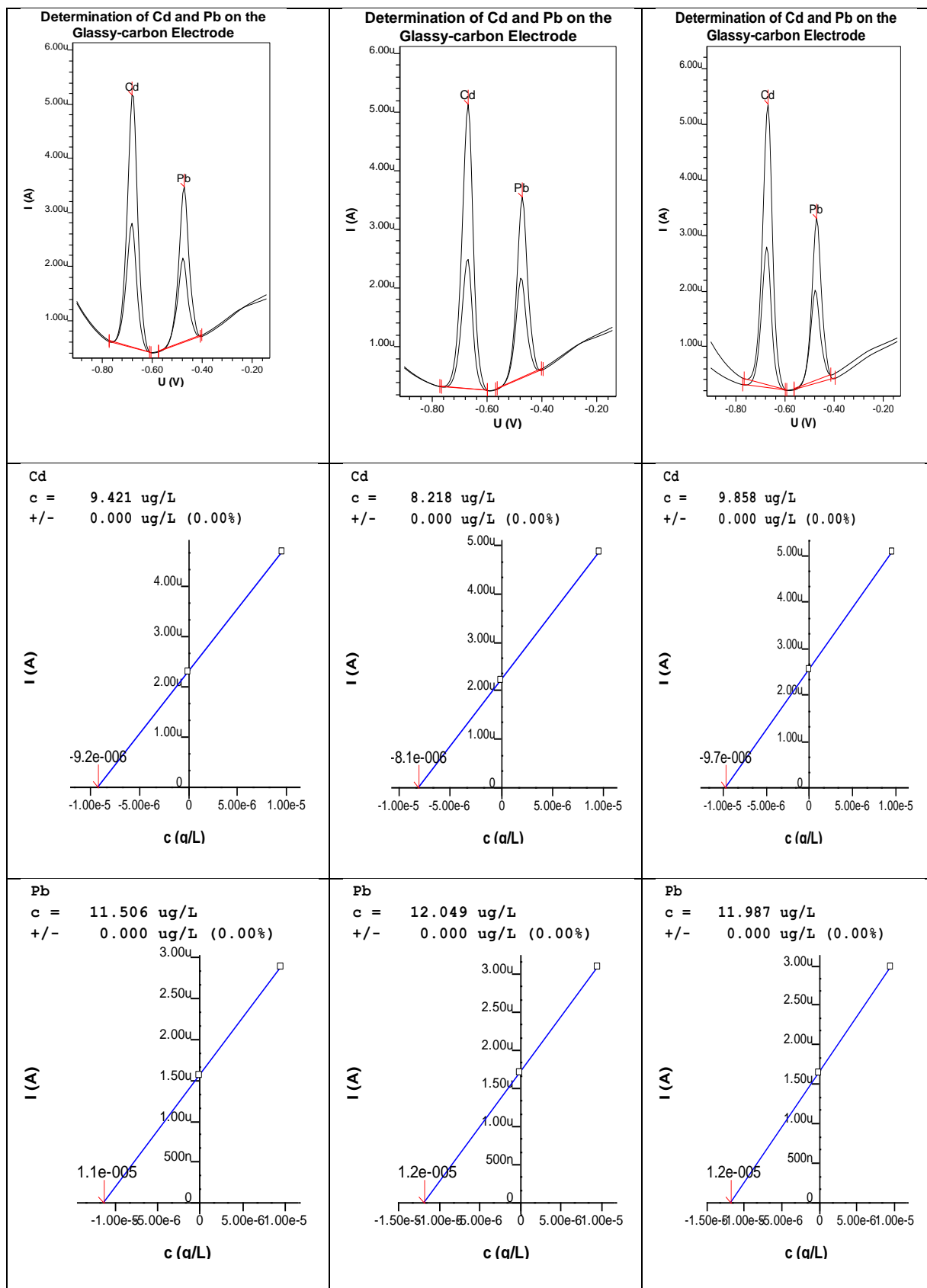


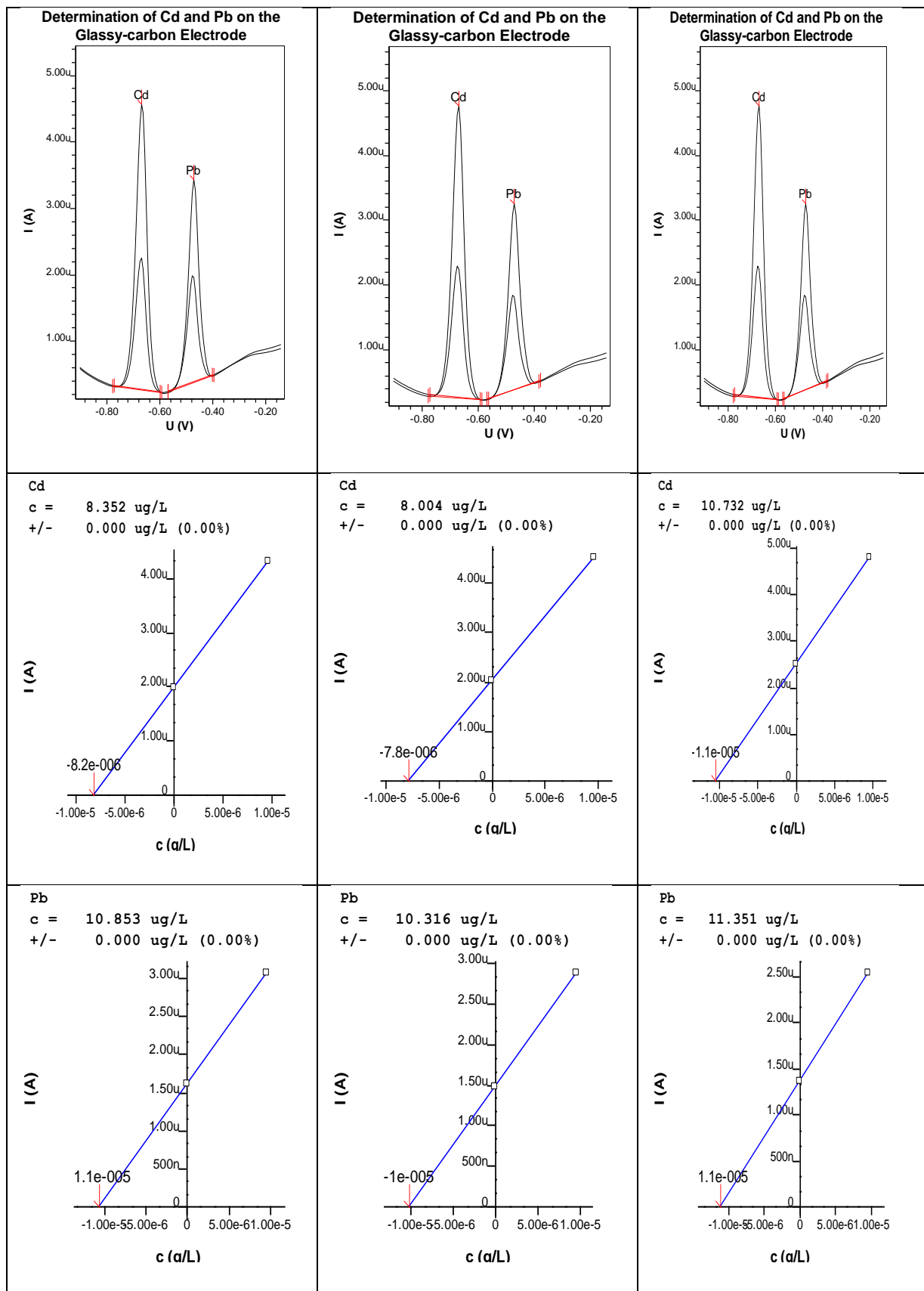


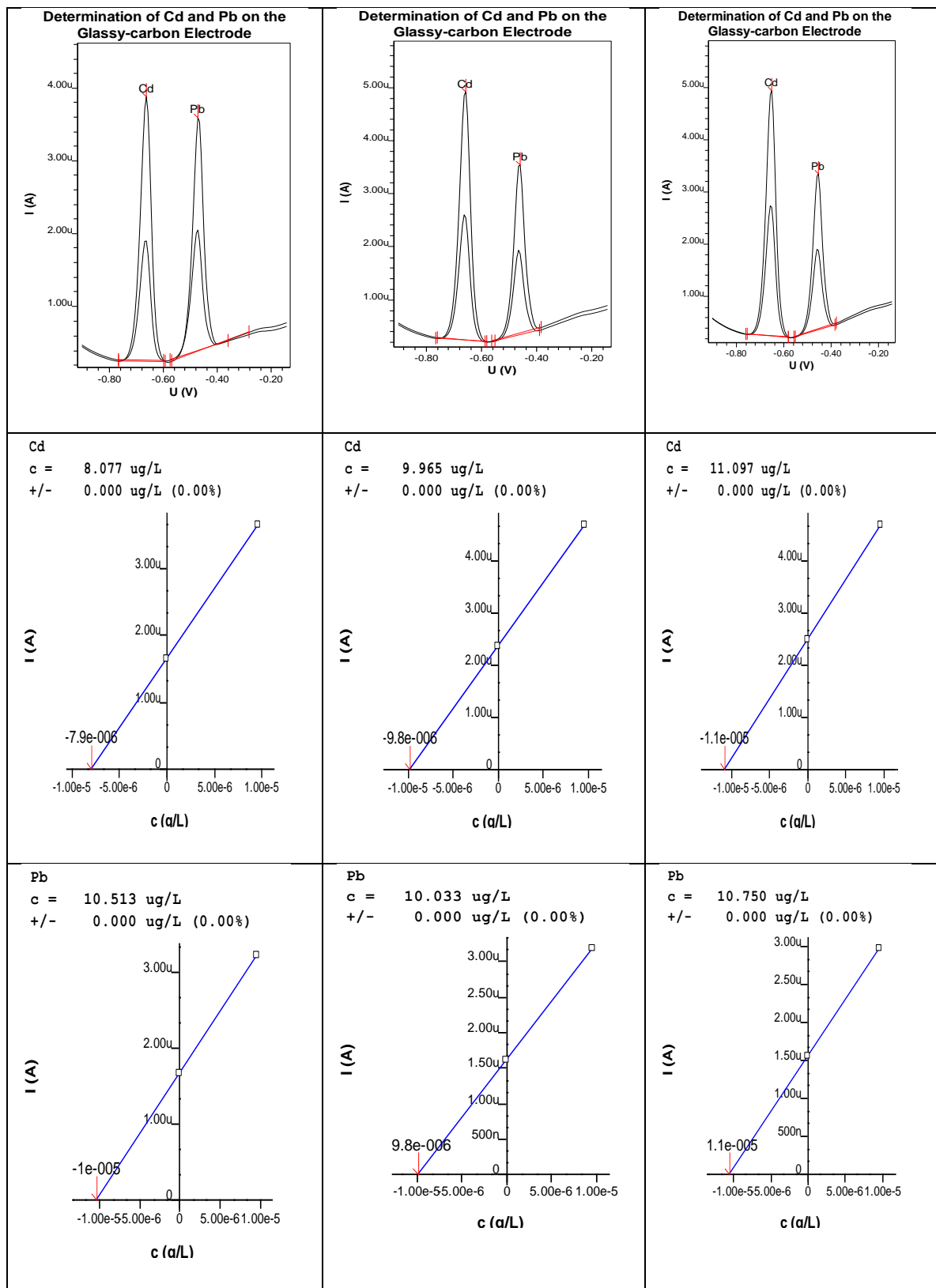


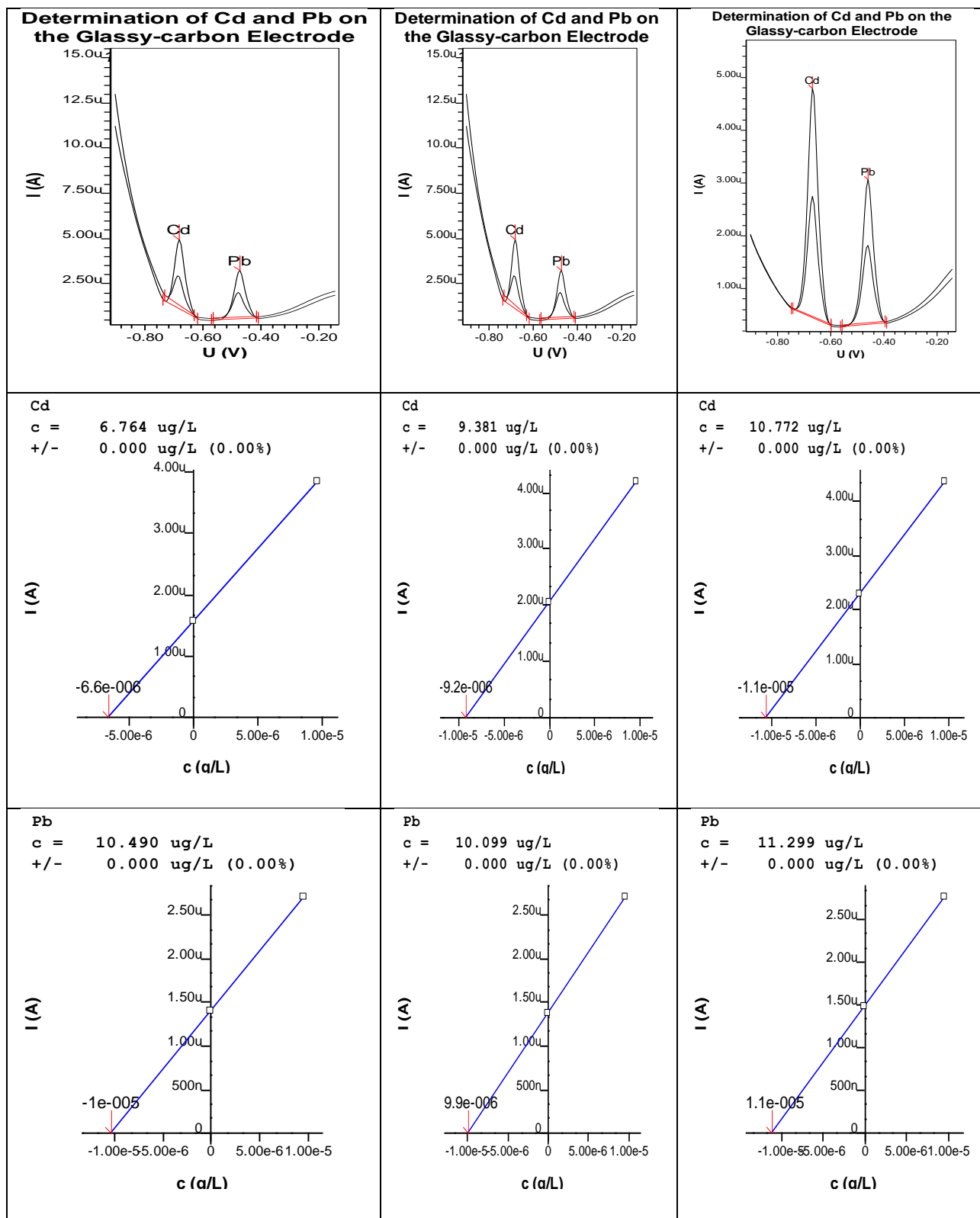




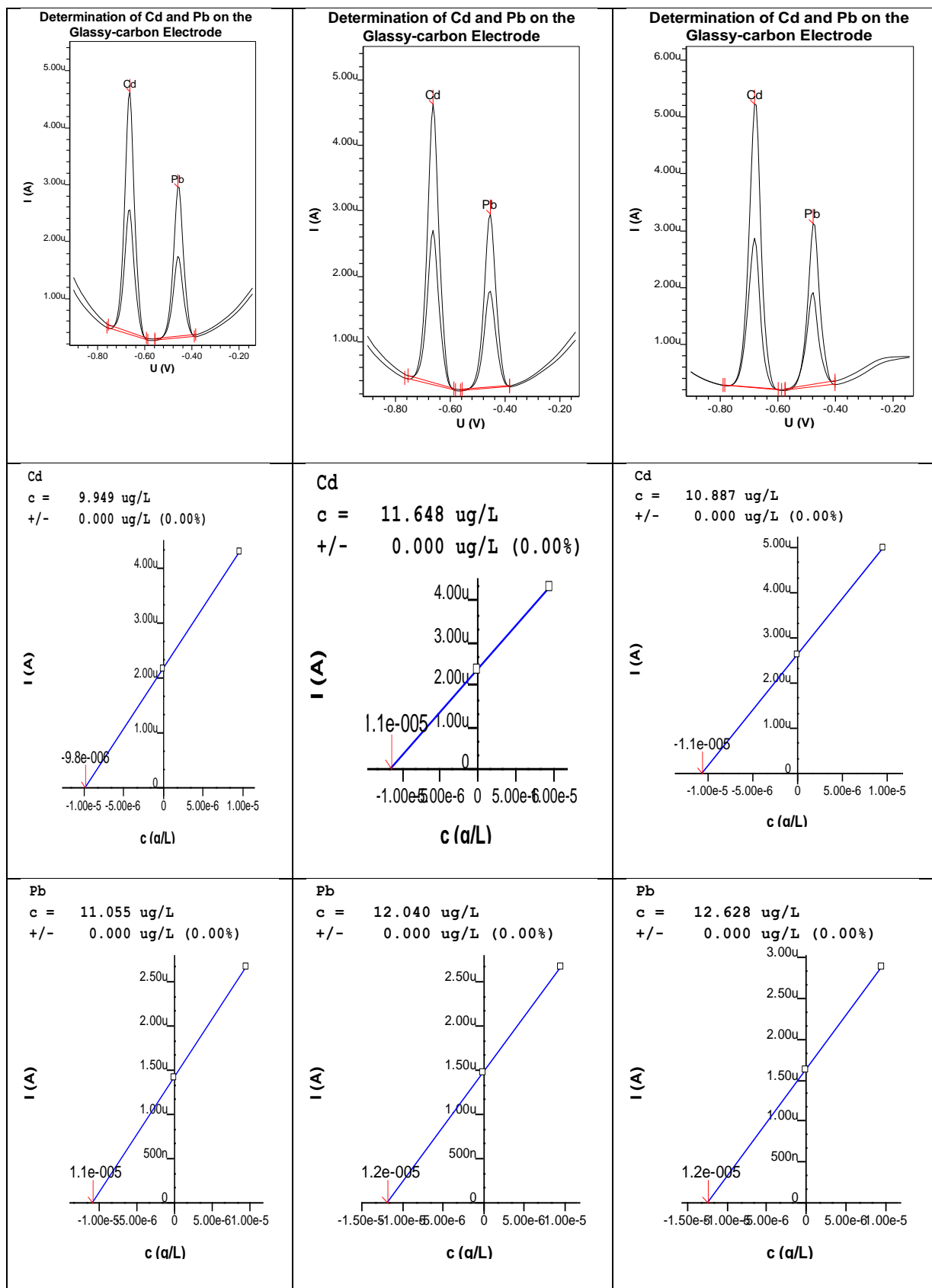


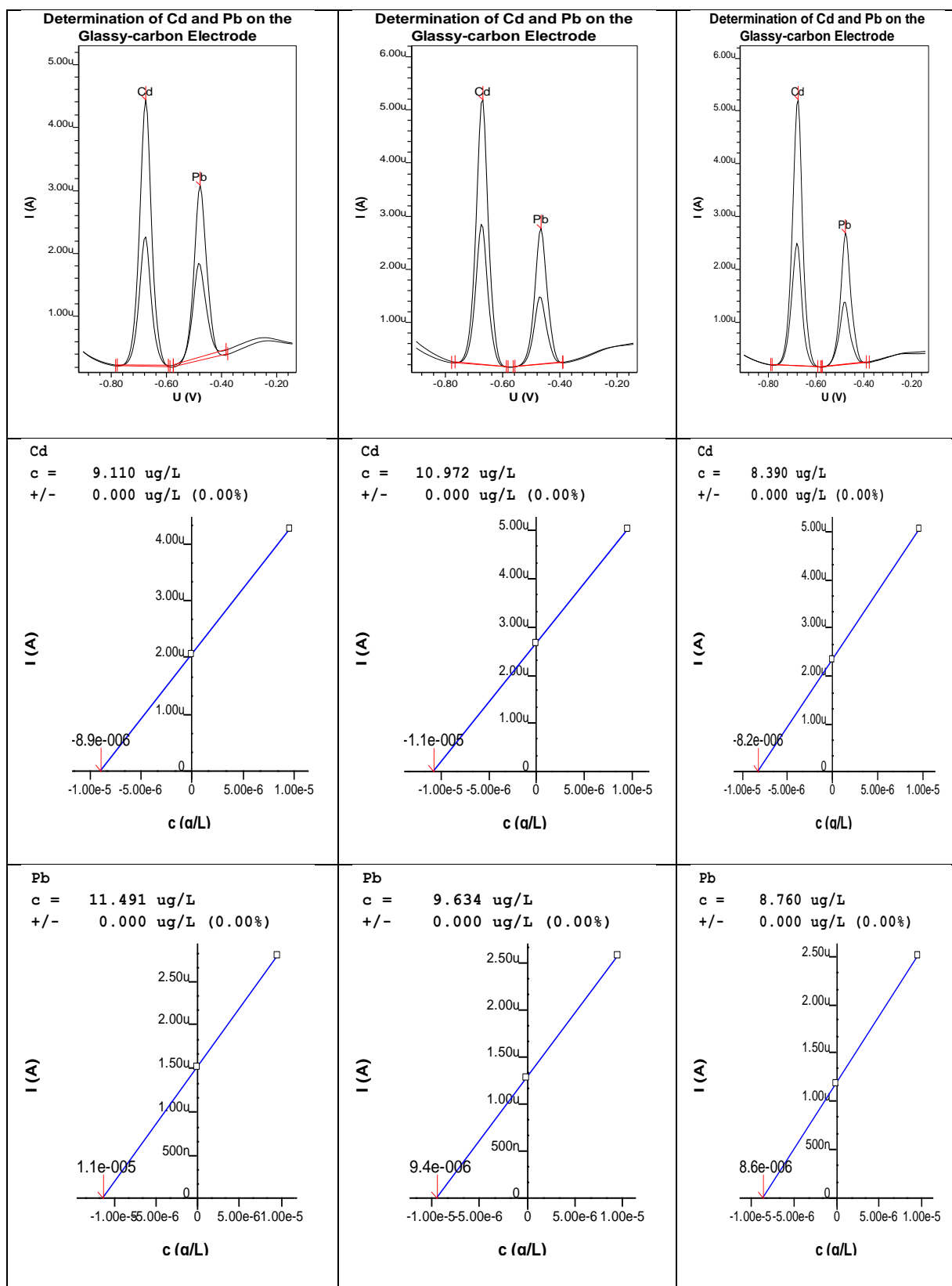


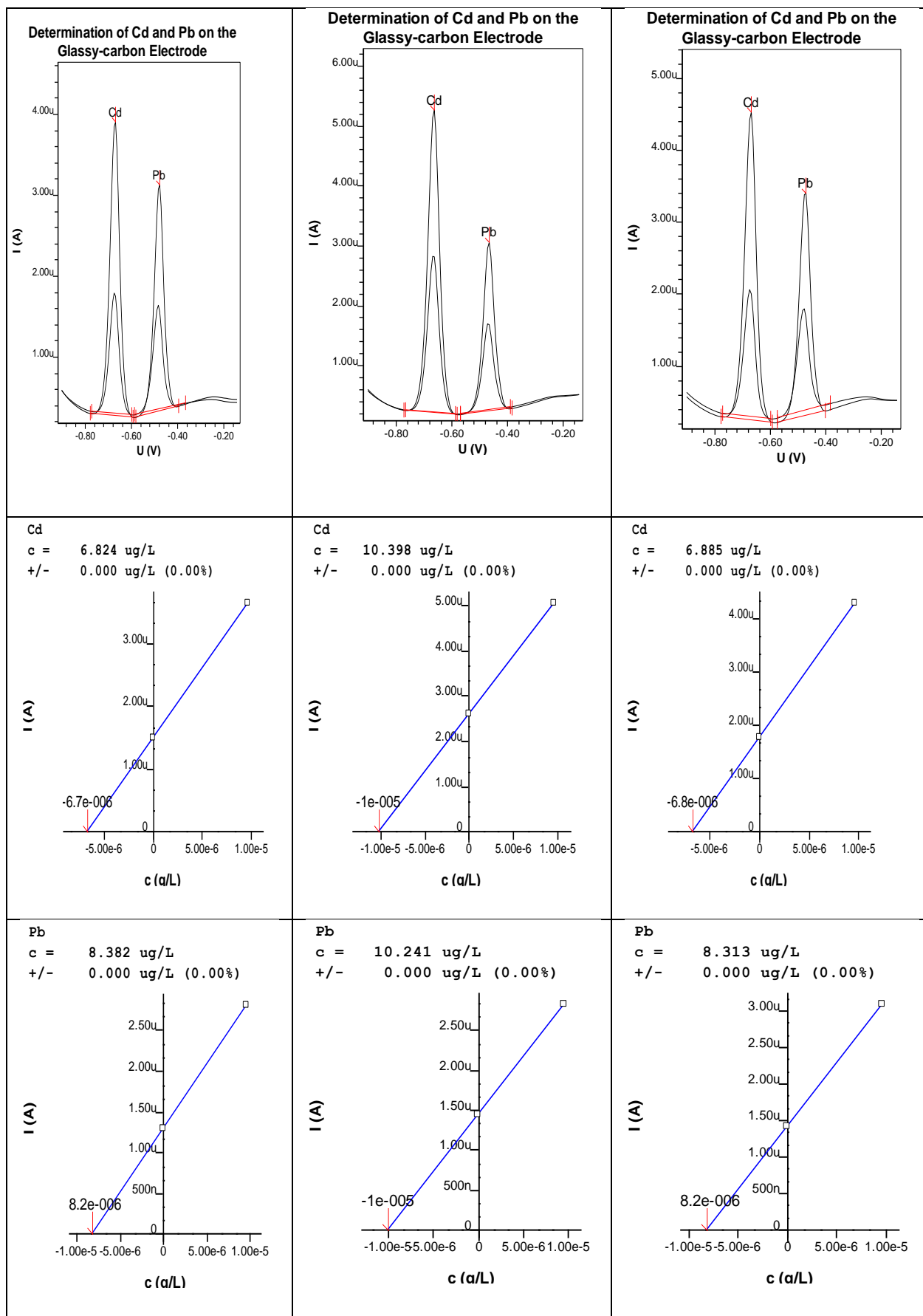












## APPENDIX 8: Marine water and QC synthetic sample results for case study 2

QC std's

