



Llengües (bio)electròniques aplicades a l'anàlisi i caracterització de begudes

Tesi Doctoral

Xavier Cetó Alsedà

Universitat Autònoma de Barcelona
Març 2013



Llengües (bio)electròniques aplicades a l'anàlisi i caracterització de begudes

Memòria presentada per
Xavier Cetó Alsedà

Tesi doctoral

Programa de doctorat en Química

Directors:
Manel del Valle Zafra
Francisco Céspedes Mulero

Departament de Química
Facultat de Ciències

2013

Memòria presentada per aspirar al Grau de Doctor per Xavier Cetó Alsedà

Vist i plau

Manel del Valle Zafra

Professor titular de Química Analítica

Francisco Céspedes Mulero

Professor Agregat de Química Analítica

Bellaterra, 7 de gener de 2013

Grup de Sensors i Biosensors
Departament de Química
Universitat Autònoma de Barcelona
Edifici C, 08193 Bellaterra

El treball descrit en aquesta memòria ha estat possible gràcies a les següents fonts de finançament:

- Projecte finançat per Ministeri de Ciència i Innovació
“Estrategias en el desarrollo y aplicación de lenguas electrónicas empleando biosensores y nuevos materiales de reconocimiento”
Projecte MCINN CTQ2010-17099
Investigador principal: Manel del Valle

- Beca Predoctoral per a la Formació de Personal Investigador (FI)
Agència de Gestió d’Ajuts Universitaris i de Recerca (AGAUR)

- Beca per Recerca en l’Estranger (BE-DGR)
Agència de Gestió d’Ajuts Universitaris i de Recerca (AGAUR)

El treball dut a terme al llarg de la present tesi doctoral ha donat lloc a les següents publicacions:

- 1. Voltammetric Electronic Tongue in the analysis of cava wines**
X. Cetó, J. M. Gutiérrez, L. Moreno-Barón, S. Alegret and M. del Valle
Electroanalysis, **23**(1): 72-78
- 2. A new amperometric bienzymatic biosensor based on biocomposites for the determination of gluconic acid in wines**
X. Cetó, F. Céspedes, J. Capdevila and M. del Valle
Talanta, **85**(2): 1207-1212
- 3. Bioelectronic Tongue Employing Enzyme-Modified Sensors for the Resolution of Phenolic Antioxidant Mixtures**
X. Cetó, F. Céspedes, M.I. Pividori, J.M. Gutiérrez, Z. Haddi, B. Bouchikhi and M. del Valle
AIP Conference Proceedings, **1362**(1): 96-97
- 4. Data Fusion from Voltammetric and Potentiometric Sensors to Build a Hybrid Electronic Tongue Applied in Classification of Beers**
Z. Haddi, A. Aziz, B. Bouchikhi, J. M. Gutiérrez, X. Cetó, A. Mimendia and M. del Valle
AIP Conference Proceedings, **1362**(1): 189-190
- 5. Resolution of phenolic antioxidant mixtures employing a voltammetric bio-electronic tongue**
X. Cetó, F. Céspedes, M.I. Pividori, J.M. Gutiérrez and M. del Valle
Analyst, **137**(2): 349-356
- 6. Determination of total polyphenol index in wines employing a voltammetric electronic tongue**
X. Cetó, J.M. Gutiérrez, M. Gutiérrez, F. Céspedes, J. Capdevila, S. Mínguez, C. Jiménez-Jorquera and M. del Valle
Analytica Chimica Acta, **732**(0): 172-179
- 7. Bioelectronic Tongue for the quantification of total polyphenol content in wine**
X. Cetó, F. Céspedes and M. del Valle
Talanta, **99**(0): 554-551
- 8. The Application of an Array of Sensors based on Boronic Acid Derivative for the Quantitative Analysis of Amino Acids**
A. Kutyla-Olesiuk, M. Jańczyk, X. Cetó, M. del Valle, P. Ciosek and W. Wróblewski
Procedia Engineering, **47**(0): 522-525
- 9. Rapid Field Identification of Subjects Involved in Firearm-Related Crimes Based on Electroanalysis Coupled with Advanced Chemometric Data Treatment**
X. Cetó, A.M. O'Mahony, I.A. Samek, J.R. Windmiller, M. del Valle and J. Wang
Analytical Chemistry, **84**(23): 10306-10314
- 10. Assessment of individual polyphenol content in beer by means of a voltammetric BioElectronic Tongue**
X. Cetó, F. Céspedes and M. del Valle
Electroanalysis, **25**(1): 68-76

- 11. Chapter 6: BioElectronic Tongues: when the sensor array incorporates biosensors**
M. del Valle, X. Cetó and M. Gutiérrez
in *Multisensor systems for chemical analysis - materials and sensors* by L. Lvova, D. Kirsanov,
C. Di Natale and A. Legin
Pan Stanford Publishing Pte Ltd.
- 12. Hybrid Electronic Tongue based on Multisensor Data Fusion for Discrimination of Beers**
J.M. Gutiérrez, Z. Haddi, A. Amari, B. Bouchikhi, A. Mimendia, X. Cetó and M. del Valle
Sensors and Actuators B: Chemical, **177**(0): 989-996
- 13. Application of an Electronic Tongue towards the analysis of brandies**
X. Cetó, M. Llobet, J. Marco and M. del Valle
Analytical Methods, **5**(5): 1120-1129
- 14. Comparison of Methods for the Processing of Voltammetric Electronic Tongues Data**
X. Cetó, F. Céspedes and M. del Valle
Microchimica Acta, **180**(5-6): 319-330
- 15. Simultaneous identification and quantification of nitro containing explosives by advanced chemometric data treatment of cyclic voltammetry at screen printed electrodes**
X. Cetó, A.M. O'Mahony, M. del Valle and J. Wang
Talanta, **107**(0): 270-276
- 16. Resolution of amino acid mixtures by an array of potentiometric sensors based on boronic acid derivative in a SIA flow system**
A. Kutyla-Olesiuk, M. Jańczyk, X. Cetó, M. del Valle, P. Ciosek and W. Wróblewski
Sensors and Actuators B: Chemical, doi: 10.1016/j.snb.2013.02.089
- 17. Beer Classification by means of a Potentiometric Electronic Tongue**
X. Cetó, M. Gutiérrez, D. Calvo and M. del Valle
Food Chemistry **141**(3): 2533-2540
- 18. Voltammetric Electronic Tongue for qualitative analysis of beers**
X. Cetó, J.M. Gutiérrez, A. Mimendia, F. Céspedes and M. del Valle
Electroanalysis **25**(7): 1635-1644
- 19. Cava wine authentication employing a voltammetric Electronic Tongue**
X. Cetó, F. Capdevila, S. Mínguez and M. del Valle
submitted
- 20. Voltammetric BioElectronic Tongue for the analysis of phenolic compounds in rosé cava wines**
X. Cetó, F. Capdevila, S. Mínguez and M. del Valle
Submitted
- 21. Development and application of electronic tongue for detection and monitoring of nitrate, nitrite and ammonium levels in aqueous medium**
L. Nuñez, X. Cetó, M.I. Pividori, M. del Valle
Microchemical Journal **110**(0):273-279

A tots els que heu contribuït a fer-ho possible,
i m'heu acompanyat en aquest llarg camí.
Simplement, Gràcies!

No estudio per saber més,
sinó per ignorar menys.
~ Juana Inés de la Cruz

Resum

Resum

En els darrers anys, s'ha observat una creixent demanda de nous mètodes analítics capaços d'oferir el màxim de informació analítica amb el mínim "esforç" possible, els quals són necessaris per tal d'assolir els nous reptes en camps tals com el monitoratge ambiental, la seguretat alimentària i la salut pública. En aquest sentit, la indústria està molt interessada en el desenvolupament de mètodes ràpids i de baix cost que puguin ser utilitzats de forma rutinària com a cribratge (*screening analysis*) per tal de detectar qualsevol adulteració o contaminació del producte ja sigui durant la seva elaboració o per comprovar que compleixen els estàndards de control de qualitat.

En aquest sentit, les línies clàssiques de investigació en el camp dels sensors químics s'han centrat en el desenvolupament de dispositius cada vegada més selectius envers espècies determinades, i al mateix temps amb una major sensibilitat i menor límit de detecció. Malauradament, es disposa de molts pocs sensors químics que operin de forma òptima i sense cap interferència o efecte matriu en les condicions requerides en els casos de mostres reals. Precisament, la dificultat per obtenir sensors amb una selectivitat i sensibilitat apropiada per qualsevol analit, ha estat la que ha donat pas a l'aparició de noves estratègies com és el cas de les llengües electròniques per tal de donar resposta a aquests problemes.

Aquests sistemes analítics consisteixen en la imitació del sentit de gust dels mamífers, on uns pocs receptors són capaços de respondre a una gran varietat de substàncies, un principi que juntament amb una etapa complexa de tractament de la informació anàloga a la que té lloc en el cervell, permet la quantificació o classificació d'un gran nombre de substàncies. Aquests sistemes biomimètics, en oposició als sistemes clàssics, estan basats en la combinació de matrius de sensors amb una baixa selectivitat i/o una resposta creuada per tal d'obtenir un valor afegit en la generació de la informació analítica.

Un dels darrers avanços en el disseny de llengües electròniques ha estat la incorporació de biosensors per tal de fer front a nous camps d'aplicació o millorar els ja existents. Aquests sistemes, coneguts com a llengües bioelectròniques, només es distingeixen dels convencionals en la incorporació d'un o més biosensors en la matriu de sensors que formarà la llengua electrònica, normalment compartint el mateix principi de detecció per tal de facilitar la compatibilitat entre ells.

En aquest marc, el present treball pretén demostrar l'aplicabilitat d'aquests sistemes en l'anàlisi i caracterització de begudes, en el sector vinícola i de les begudes alcohòliques, tant per l'extracció de informació qualitativa i identificació de classes com per la quantificació de paràmetres d'interès analític, responent en ambdós casos a les necessitats corresponents en cada sector. Concretament, s'ha estudiat l'aplicació d'aquestes eines en el cava, el brandi, la cervesa i el vi; quatre dels sectors més importants en quant a begudes alcohòliques en el nostre país.

A més, donada la importància que han assolit els compostos fenòlics en els darrers anys degut a la seva activitat antioxidant, beneficiosa per a la salut, s'ha abordat la quantificació d'aquests tant a nivell global com la discriminació individual; abordant la seva detecció a partir d'una llengua electrònica clàssica i d'una llengua bioelectrònica, comparant així els beneficis derivats de la incorporació de biosensors en la matriu de la llengua.

Per últim, donada la dificultat que genera el tractament de les dades generades amb aquests sistemes, especialment en el cas dels sensors voltamperomètrics, també s'ha treballat en el desenvolupament i aplicació de noves estratègies de processament per tal de reduir-ne la complexitat i millorar els resultats obtinguts; comparant a més les diferents estratègies proposades entre elles.

Summary

Over the last years, there has been an increasing demand of new analytical methods with the high sensitivity and selectivity, and fast response needed to meet new challenges in environmental monitoring, food safety and public health. In this fashion, industry is increasingly focused on fast-response and low-cost methods, as those based on chemical sensors, that may be used for screening or detecting any adulteration or contamination of the products, either during or after its production, or to assess they guarantee quality control standards.

In this sense, classical research lines in the field of chemical sensors have focused on the development of ever more selective devices towards a particular species, and sensitive to lower concentrations at the same time. Unfortunately, there are few optimally operating chemical sensors that function without any interference or matrix effect in the required conditions when dealing with real samples analysis. Precisely, the difficulty to obtain sensors with appropriate selectivity and sensitivity for a given analyte has led to the appearance of new strategies such as electronic tongues in order to tackle these problems.

These analytical systems are inspired by the sensory ability of taste in mammals, where a few receptors can respond to a large variety of substances. This principle is coupled with complex data treatment analogous to the applied in the brain, which allows to quantify or to classify a large amount of substances. These biomimetic systems, opposed to conventional approaches, are directed towards the combination of low selectivity sensors array response (or cross response features) in order to obtain some added value in the generation of analytical information.

One of the recent advances in the design of electronic tongues has been the incorporation of biosensors, in order to tackle new application fields or to improve existing ones. These bioelectronic tongues, as they have been named, are only distinguished from conventional ones in the incorporation of one or several biosensors into the sensor array, normally sharing the same transduction principle to facilitate compatibility.

In this context, the work presented herein aims to demonstrate the applicability of these systems towards the analysis and characterization of beverages, specifically towards wine and alcoholic beverages, either for the extraction of qualitative information and its classification or the quantification of analytical parameters of

interest, responding in both cases to the needs of each sector. Concretely, its application towards cava wine, brandy, beer and wine has been studied; the most important sectors in terms of alcoholic beverages in our country.

Additionally, given the importance that phenolic compounds have achieved in the recent years due to its antioxidant activity, with huge health benefits, the quantification of these compounds has been addressed from both points of view: its global content and the individual discrimination; tackling it using either a classical electronic tongue and a bioelectronic tongue, comparing the benefits of the incorporation of biosensors in the *e*-tongue array.

Lastly, given the difficulties derived in the treatment of the data generated with such systems, specially in the case of voltammetric sensors, much attention has been paid to the development and application of novel processing strategies in order to reduce its complexity and improve the obtained results; besides comparing the different proposed strategies between each other.

Índex

Índex

| | | |
|------------|--|----|
| 1. | Introducció | 1 |
| 1.1. | Sensors i biosensors | 1 |
| 1.1.1. | Definició general..... | 1 |
| 1.1.2. | Propietats d'un sensor ideal | 2 |
| 1.1.3. | Classificació | 4 |
| 1.1.4. | Biosensors | 6 |
| 1.1.5. | Matrius de sensors i/o biosensors | 8 |
| 1.2. | Tècniques electroquímiques..... | 10 |
| 1.2.1. | Voltamperometria | 12 |
| 1.3. | Anàlisi multivariable i tractament quimiomètric | 17 |
| 1.3.1. | Equiparació de la resposta | 18 |
| 1.3.2. | Compressió del senyal | 19 |
| 1.3.3. | Anàlisi qualitativa | 21 |
| 1.3.3.1. | Anàlisi de Components Principals (PCA)..... | 21 |
| 1.3.3.2. | Anàlisi Discriminant Lineal (LDA) | 22 |
| 1.3.3.3. | Anàlisi Discriminant - Regressió Parcial per Mínims Quadrats (PLS-DA)..... | 23 |
| 1.3.3.4. | Màquines de Suport Vectorial (SVMs)..... | 23 |
| 1.3.4. | Anàlisi quantitativa | 25 |
| 1.3.4.1. | Regressió Lineal Múltiple (MLR) | 25 |
| 1.3.4.2. | Regressió de Components Principals (PCR)..... | 26 |
| 1.3.4.3. | Regressió Parcial per Mínims Quadrats (PLS) | 26 |
| 1.3.4.4. | Regressió Parcial per Mínims Quadrats Múltiple (nPLS) | 28 |
| 1.3.4.5. | Xarxa Neuronal Artificial (ANN)..... | 29 |
| 1.4. | Sistemes biomimètics..... | 30 |
| 1.4.1. | Concepte | 30 |
| 1.4.2. | Avantatges i inconvenients | 32 |
| 1.4.3. | Aproximacions i exemples..... | 34 |
| 1.4.3.1. | Llengües electròniques..... | 34 |
| 1.4.3.2. | Nassos electrònics..... | 36 |
| 1.4.3.3. | Ulls electrònics..... | 37 |
| 1.4.3.4. | Fusió de dades..... | 38 |
| 1.4.3.5. | Xarxes Neuronals Artificials..... | 39 |
| 1.4.3.5.1. | Fonament biològic | 39 |

| | | |
|--------------|--|----|
| 1.4.3.5.2. | El perceptró..... | 41 |
| 1.4.3.5.3. | Perceptró multicapa o xarxa multicapa..... | 42 |
| 1.4.3.5.4. | El procés d'aprenentatge..... | 43 |
| 1.4.3.6. | Aplicació..... | 47 |
| 1.5. | Camp d'aplicació del present treball..... | 50 |
| 1.6. | Bibliografia..... | 52 |
| 2. | Objectius..... | 57 |
| 3. | Experimental..... | 59 |
| 3.1. | Construcció dels sensors voltamperomètrics..... | 59 |
| 3.2. | Disseny experimental..... | 61 |
| 3.3. | Procediment de mesura..... | 62 |
| 3.4. | Construcció dels models de resposta..... | 63 |
| 3.4.1. | Preprocessament de les dades..... | 64 |
| 3.4.1.1. | Centrat i normalització..... | 65 |
| 3.4.1.2. | Selecció de variables (Feature selection)..... | 66 |
| 3.4.1.3. | Ajust polinomial..... | 69 |
| 3.4.1.4. | Integral de finestra seccionada..... | 70 |
| 3.4.1.5. | Anàlisi de Components Principals (PCA)..... | 71 |
| 3.4.1.6. | Transformada Ràpida de Fourier (FFT)..... | 72 |
| 3.4.1.7. | Transformada Wavelet Discreta (DWT)..... | 72 |
| 3.4.1.8. | Grau de reconstrucció del senyal..... | 73 |
| 3.4.2. | Construcció i validació del model..... | 74 |
| 3.4.2.1. | Anàlisi qualitativa..... | 76 |
| 3.4.2.1.1. | Anàlisi de Components Principals (PCA)..... | 77 |
| 3.4.2.1.2. | Anàlisi Discriminant Lineal (LDA)..... | 78 |
| 3.4.2.2. | Anàlisi quantitativa..... | 80 |
| 3.4.2.2.1. | Xarxa Neuronal Artificial (ANN)..... | 81 |
| 3.4.2.3. | Mètodes de validació..... | 85 |
| 3.4.2.3.1. | Subdivisió del conjunt de dades..... | 85 |
| 3.4.2.3.2. | Paràmetres indicadors de la bondat de l'ajust..... | 87 |
| 3.4.2.3.2.1. | Models qualitatiu..... | 88 |
| 3.4.2.3.2.2. | Models quantitatiu..... | 90 |
| 3.4.3. | Software..... | 92 |
| 3.5. | Bibliografia..... | 93 |
| 4. | Resultats i discussió..... | 95 |
| 4.1. | Anàlisi de caves..... | 97 |
| 4.1.1. | Classificació dels caves..... | 99 |

| | | |
|----------|--|-----|
| 4.1.1.1. | Quantitat de sucre afegida..... | 99 |
| 4.1.1.2. | Temps de criaça | 104 |
| 4.1.2. | Quantificació del sucre afegit | 109 |
| 4.1.2.1. | Ús del mètode de l'addició estàndard de segon ordre..... | 109 |
| 4.1.2.2. | Ajust d'un model multivariable | 111 |
| 4.1.3. | Quantificació de compostos fenòlics en caves rosats | 114 |
| 4.2. | Anàlisi de brandis | 119 |
| 4.2.1. | Destil·lats de vi | 121 |
| 4.2.2. | Correlació amb el panel de tast..... | 124 |
| 4.2.3. | Quantificació compostos indicadors de defectes en els brandis | 125 |
| 4.2.4. | Detecció de l'ús de mètodes d'envelliment alternatius..... | 128 |
| 4.2.5. | Quantificació de compostos fenòlics | 129 |
| 4.2.6. | Identificació de l'anyada..... | 131 |
| 4.3. | Anàlisi de cerveses..... | 132 |
| 4.3.1. | Classificació de cerveses..... | 133 |
| 4.3.1.1. | Tipus de cervesa..... | 137 |
| 4.3.1.2. | Procés d'elaboració..... | 138 |
| 4.3.1.3. | Discriminació simultània d'ambdós efectes | 139 |
| 4.3.2. | Quantificació de compostos fenòlics en cervesa | 140 |
| 4.4. | Anàlisi de vins..... | 149 |
| 4.4.1. | Índexs de polifenols totals | 150 |
| 4.4.1.1. | Quantificació mitjançant la Llengua Electrònica..... | 151 |
| 4.4.1.2. | Quantificació mitjançant la Llengua (Bio)Electrònica | 155 |
| 4.4.2. | Discriminació individual de compostos fenòlics | 159 |
| 4.4.2.1. | Anàlisi qualitativa mitjançant la Llengua Electrònica..... | 160 |
| 4.4.2.2. | Anàlisi qualitativa mitjançant la Llengua (Bio)Electrònica..... | 162 |
| 4.4.2.3. | Resolució de mescleres ternàries mitjançant la Llengua BioElectrònica..... | 165 |
| 4.5. | Comparativa de les diferents estratègies de tractament del senyal | 170 |
| 4.6. | Bibliografia | 179 |
| 5. | Conclusions..... | 183 |
| 6. | Publicacions que formen part del compendi de la tesi doctoral | |
| 6.1. | Voltammetric Electronic Tongue in the analysis of cava wines | |
| 6.2. | Resolution of phenolic antioxidant mixtures employing a voltammetric bio-electronic tongue | |

- 6.3. Determination of total polyphenol index in wines employing a voltammetric electronic tongue
- 6.4. Bioelectronic Tongue for the quantification of total polyphenol content in wine
- 6.5. Assessment of individual polyphenol content in beer by means of a voltammetric BioElectronic Tongue
- 6.6. Chapter 6: BioElectronic Tongues: when the sensor array incorporates biosensors

- 7. Publicacions fora del compendi que formen part de la tesi doctoral com a annex
 - 7.1. Application of an Electronic Tongue towards the analysis of brandies
 - 7.2. Comparison of methods for the processing of voltammetric Electronic Tongues data
 - 7.3. Voltammetric Electronic Tongue for qualitative analysis of beers
 - 7.4. Cava wine authentication employing a voltammetric Electronic Tongue
 - 7.5. Voltammetric BioElectronic Tongue for the analysis of phenolic compounds in rosé cava wines

Abreviatures

Abreviatures i símbols

| | |
|------------|--|
| § | Apartat |
| abv | <i>Alcohol by volume</i> Percentatge d'alcohol per volum |
| Ag | Sensor modificat amb nanopartícules de plata |
| ANN | <i>Artificial Neural Network</i> Xarxa Neuronal Artificial |
| arb. unit. | <i>Arbitrary unit</i> Unitat arbitrària |
| b | Ordenada a l'origen |
| BioET | <i>BioElectronic Tongue</i> Llengua BioElectrònica |
| BP | <i>Back-propagation</i> Retropropagació |
| CI | <i>Causal Index</i> Índex de Causalitat |
| CINVESTAV | Centre de Investigacions i Estudis Avançats del institut Politècnic Nacional de Mèxic |
| CNM | Centre Nacional de Microelectrònica |
| coefs. | Coefficients |
| CoPc | Sensor modificat amb ftalocianina de cobalt (II) |
| Cu | Sensor modificat amb nanopartícules de coure |
| DA | <i>Discriminant Analysis</i> Anàlisi Discriminant |
| DFs | <i>Discriminant Functions</i> Funcions Discriminants |
| DO | Denominació d'Origen |
| DWT | <i>Discrete Wavelet Transform</i> Transformada Wavelet Discreta |
| E | Potencial electroquímic |
| EC | <i>Enzyme Commission number</i> Número EC |

| | |
|------------------|---|
| EE | <i>Electronic Eye</i> |
| e-eye | Ull Electrònic |
| EN | <i>Electronic Nose</i> |
| e-nose | Nas Electrònic |
| ET | <i>Electronic Tongue</i> |
| e-tongue | Llengua Electrònica |
| f_c | Factor de comparació |
| FC | Folin-Ciocalteu |
| FFT | <i>Fast Fourier Transform</i> Transformada Ràpida de Fourier |
| GEC | <i>Graphite-Epoxy Composite</i> Compòsit de grafit-epoxi |
| GOX | Biosensor modificat amb glucosa oxidasa |
| GSB | Grup de Sensors i Biosensors |
| i | Intensitat de corrent |
| I ₂₈₀ | Índex de polifenols/compostos fenòlics |
| IC | Interval de Confiança |
| IMB | Institut de Microelectrònica de Barcelona |
| INCAVI | Institut Català de la Vinya i el Vi |
| Int | <i>Voltammogram windowed sliced integral</i> Integral de finestra seccionada |
| IUPAC | <i>International Union of Pure and Applied Chemistry</i> Unió Internacional de Química Pura i Aplicada |
| Lac | Biosensor modificat amb lacasa |
| LDA | <i>Linear Discriminant Analysis</i> Anàlisi Discriminant Lineal |
| m | Pendent |
| MLR | <i>Multiple Linear Regression</i> Regressió Lineal Múltiple |
| logsig | <i>Log Sigmoid transfer function</i> Funció de transferència los sigmoidea |
| nPLS | <i>N-way Partial Least Squares</i> Regressió Parcial per Mínims Quadrats Múltiple |
| NRMSE | <i>Normalized Root Mean Square Error</i> Error Quadràtic Mitjà Normalitzat |

| | |
|----------------|---|
| LOO | <i>Leave-one-out</i> |
| PCs | <i>Principal Components</i> Components Principals |
| PCA | <i>Principal Component Analysis</i> Anàlisi de Components Principals |
| PCR | <i>Principal Components Regressions</i> Regressió de Components Principals |
| PLS | <i>Partial Least Squares regression</i> Regressió Parcial per Mínims Quadrats |
| PLS-DA | <i>Partial Least Squares regression - Discriminant Analysis</i> Anàlisi Discriminant - Regressió Parcial per Mínims Quadrats |
| PANI | Sensor modificat amb polianilina |
| <i>purelin</i> | <i>Linear transfer function</i> Funció de transferència lineal |
| PVC | <i>Polyvinyl Chloride</i> Clorur de Polivinil |
| PPy | Sensor modificat amb polipirrol |
| r | Coefficient de correlació |
| R ² | Coefficient de determinació |
| RMSE | <i>Root Mean Square Error</i> Error Quadràtic Mitjà |
| <i>satlin</i> | <i>Saturating linear transfer function</i> Funció de transferència lineal de saturació |
| <i>satlins</i> | <i>Symmetric saturating linear transfer function</i> Funció de transferència lineal de saturació simètrica |
| SOSAM | <i>Second Order Standard Addition Method</i> Mètode de l'Addició Estàndard de Segon Ordre |
| SSR | <i>Split Sample Resampling</i> |
| SVM | <i>Support Vector Machine</i> Màquina de Suport Vectorial |
| tr | Subconjunt d'entrenament (<i>train</i>) |
| ts | Subconjunt d'entrenament (<i>test</i>) |
| <i>tansig</i> | <i>Hyperbolic tangent sigmoid transfer function</i> Funció de transferència |
| Tyr | Biosensor modificat amb tirossinassa |

1. Introducció

1. Introducció

1.1. *Sensors i biosensors*

1.1.1. Definició general

Un sensor es defineix com un dispositiu o element sensible a determinats estímuls que és capaç de captar-los i produir uns senyals, generalment elèctrics, que són transmesos a un altre dispositiu o a un sistema que actua d'acord amb els senyals rebuts. En el cas dels sensors químics, la IUPAC defineix aquests com un dispositiu que transforma la informació química, abastant des de la concentració d'un component específic d'una mostra a l'anàlisi de la composició total, en un senyal analíticament útil. La informació química, pot originar-se a partir d'una reacció química de l'analit o d'una propietat física del sistema investigat [1]. Per tant, en el cas analític, un sensor es tracta d'un dispositiu dissenyat per tal de realitzar mesures selectives de paràmetres físics, químics o biològics generant un senyal físic quantificable [2]. A més, idealment aquest dispositiu ha de presentar les següents característiques: selectivitat, simplicitat, dimensions reduïdes, robustesa, baix cost i capacitat de generar informació en temps real i de forma contínua; com a trets més importants.

Un sensor químic és un component essencial d'un analitzador; el qual a més del sensor, pot contenir dispositius que realitzin les funcions següents: mostreig, transport de la mostra, processament dels senyals i processament de les dades. Essent així, l'analitzador una part essencial d'un sistema automatitzat; actuant inclús com a sistema de monitoratge en el cas que aquest actuï d'acord amb un pla de mostreig que sigui funció del temps.

Els sensors químics contenen dos unitats funcionals bàsiques: la part del receptor i la part del transductor (Figura 1.1). Addicionalment, alguns sensors poden incloure un separador com pot ser, per exemple, una membrana.

En la part del receptor, el qual aporta selectivitat al sensor, la informació química es transforma en una forma d'energia que pot ser mesurada pel transductor. La part del transductor és un dispositiu capaç de transformar l'energia que conté la informació química sobre la mostra en un senyal analític útil; el qual, com a tal, no mostra cap tipus de selectivitat. Per tant, ambdues parts han d'estar connectades ja que el senyal primari

generat en la reacció de reconeixement serà convertit pel transductor en un senyal secundari, en últim terme, del domini elèctric.

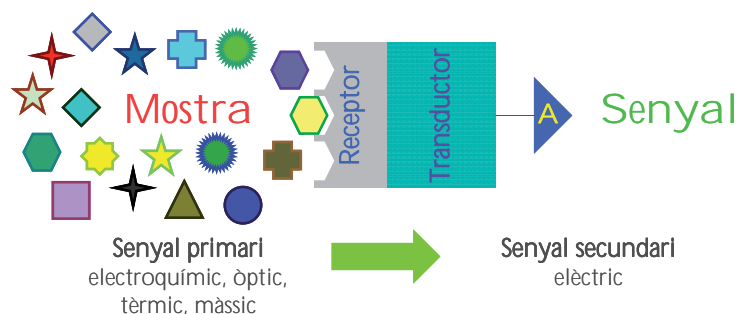


Figura 1.1. Esquema del funcionament d'un sensor químic i del procés de transducció. Només una part dels components de la mostra són reconeguts pel receptor mitjançant processos físics, químics o biològics (R); generant un senyal associat, el qual és transformat a un senyal elèctric pel transductor (T). Aquest senyal és posteriorment amplificat, condicionat, processat i presentat com a dada (A).

La part del receptor d'un sensor químic pot estar basada en principis diferents, essent així una possible font de classificació d'aquests:

- Físic, on no es du a terme cap reacció química. Un exemple d'aquests són els sensors basats en la mesura del canvi d'absorbància, índex de refracció, conductivitat, temperatura o massa.
- Químic, en els que una reacció química en la que intervé l'analit dona lloc al senyal analític.
- Bioquímic, en els que un procés bioquímic és la font del senyal analític. Exemples típics d'aquests són els sensors potenciomètrics microbians, immunosensors o els sensors enzimàtics. Tot i que aquests poden ser considerats com un subgrup dels sensors químics, són comunament diferenciats i anomenats biosensors.

No obstant, en alguns casos, no és possible decidir inequívocament si un sensor opera sota un principi químic o físic, com per exemple el cas en que el senyal és degut a un procés d'adsorció.

1.1.2. Propietats d'un sensor ideal

Així doncs, independentment del principi de transducció o del tipus de sensor, existeixen un conjunt de característiques que defineixen el que seria un sensor ideal, respecte les quals un bon sensor s'hauria d'aproximar [3].

- Senyal: ha de ser mesurable i proporcional a la concentració de l'analit, essent relacionable amb aquesta mitjançant un model matemàtic.
- Selectivitat: ha de manifestar una preferència elevada (major resposta) vers l'espècie de interès i un baix nivell de interferència cap a la resta d'analits presents en la matriu de la mostra.
- Sensibilitat: determina la capacitat del sensor de discriminar entre petites variacions en la concentració de l'analit, essent aquesta en el rang de concentració d'aplicació.
- Resolució: variació de la concentració més petita que pot ser detectada quan la concentració és variada de forma contínua.
- Relació Senyal/Soroll (S/N): aquesta relació determina el límit de detecció del dispositiu; és a dir, la mínima concentració d'analit per la qual el sensor és capaç de mostrar una resposta diferenciada.
- Temps de resposta: el temps requerit pel sensor per a detectar un canvi en la concentració de l'analit. Condiciona la possibilitat/viabilitat de construir un dispositiu que permeti dur a terme la monitorització en temps real; així doncs, aquesta ha de ser el menor possible.
- Linealitat: el rang en que la resposta del sensor mostra una relació directa a la concentració de l'analit.
- Reversibilitat: després d'haver estat exposat a la presència de l'analit, en l'absència posterior d'aquest, la resposta del sensor ha de retornar al seu nivell basal.
- Absència d'histèresi: la diferència en les característiques del sensor front a concentracions creixents o decreixents de l'analit.

Adicionalment existeixen altres característiques que, tot i no ser indispensables en aquest, si que en contribueixen a millorar la seva resposta o prestacions. Per exemple, també interessa que el dispositiu presenti un elevat temps de vida, ampli rang lineal (major aplicabilitat evitant dilucions de les mostres), baix cost (afavoreix el seu futur ús un cop desenvolupat per part de la població o indústria), elevada portabilitat (possibilitat de dur a terme mesures *on-field*), facilitat d'ús (no requerir de personal especialitzat per la seva utilització) o dimensions adequades (relacionat amb la portabilitat d'aquest, volum de mostra requerit, etc.) entre altres.

Per últim, s'ha de considerar també els requeriments específics a l'hora de prendre la mesura, els quals dependran de l'aplicació a que aquest estigui determinat: mesures

en continu, sistemes destructius de la mostra, biocompatibilitat per a sistemes invasius, automatització del procés de mesura, etc.

En darrer terme, sempre cal adoptar un compromís entre les diferents característiques que un sensor ideal hauria de tenir i les del nostre sensor, que com el propi nom indica, només representa una situació utòpica. Per tant, a l'hora de desenvolupar el nostre dispositiu cal tenir en compte l'aplicació final d'aquest, considerant també el cost de fabricació i la complexitat del sensor, decidint quines propietats cal prioritzar respecte d'altres o cercant solucions alternatives com seria l'ús de matrius de sensors (§1.1.5 i §1.4).

1.1.3. Classificació

El desenvolupament de la instrumentació, la microelectrònica i els ordinadors ha fet possible el disseny de sensors basats en l'ús de la major part de principis físics, químics i biològics coneguts que s'utilitzen en química. En aquest sentit, els sensors químics poden ser classificats d'acord amb el principi de transducció com ja s'ha esmentat anteriorment (§1.1.1), o segons el principi de funcionament del transductor d'acord amb les següents categories:

- i. *Òptics*: dispositius que transformen els canvis de fenòmens òptics, els quals són resultat d'una interacció de l'analit amb la part del receptor. Aquest grup pot ser subdividit en funció del tipus de propietats òptiques que s'han aplicat als sensors químics en les següents categories: absorbància, reflectància, fluorescència, índex de refracció, efectes optotèrmics, *light scattering*, etc.
- ii. *Electroquímics*: dispositius que transformen l'efecte de la interacció electroquímica analit-elèctrode en un senyal útil. Aquests efectes poden ser estimulats elèctricament (e.g. amperometria) o ser el resultat d'una interacció espontània en condicions d'absència de corrent (e.g. potenciomètria). Entre aquests, es poden distingir els següents subgrups: sensors voltamperomètrics (incloent els dispositius amperomètrics), sensors potenciomètrics, transistors d'efecte de camp sensibilitzats químicament (*chemically sensitized field effect transistor*, CHEMFET) i sensors potenciomètrics elecrolítics sòlids per a gasos (*potentiometric solid electrolyte gas sensors*).

Essent en l'ús dels sensors voltamperomètrics els quals ens centrarem en la present tesi, donat que han estat aquests els sensors escollits per dur a terme el treball de recerca.

- iii. *Elèctrics*: dispositius basats en mesures on no té lloc cap procés electroquímic, però en els que el senyal prové de canvis en les propietats elèctriques causades per la interacció amb l'analit; entre els quals trobem: sensors de metalls d'òxids semiconductors (usats generalment per a la detecció de gasos), sensors semiconductors orgànics, sensors electrolítics de conductivitat i sensors de permitivitat elèctrica.
- iv. *Màssics (o sensibles a la massa)*: dispositius que transformen el canvi de massa en una superfície, especialment modificada, en un canvi d'una propietat del material de suport, essent el canvi de massa originat per l'acumulació de l'analit. E.g. dispositius piezoelèctrics o dispositius d'ona acústica de superfície (*surface acoustic wave devices, SAW*)
- v. *Magnètics*: dispositius basats en el canvi de les propietats paramagnètiques del gas analitzat; representats per certs tipus de monitors d'oxigen.
- vi. *Tèrmics*: basats en la mesura dels efectes de calor d'una reacció química específica o adsorció en la que està involucrada l'analit; podent ser aquests efectes mesurats de diferents maneres. Per exemple, en el cas dels sensors catalítics, en la mesura mitjançant d'un termistor del calor d'una reacció de combustió o enzimàtica. Alternativament, els dispositius basats en la mesura d'efectes optotèrmics també poden ser inclosos en aquest grup.
- vii. *Altres*: dispositius basats en la mesura d'altres propietats físiques tals com poden ser la radiació X, β o Γ poden formar la base d'un sensor químic en cas que s'utilitzin per determinar la composició química.

Aquesta classificació dels sensors descrita anteriorment representa només una de les possibles categoritzacions; podent ser classificats a partir d'altres característiques. En aquest sentit, els sensors també han estat classificats d'acord amb el mètode utilitzat per mesurar l'efecte, enlloc de l'efecte primari com a tal. L'exemple més clar el trobem en el cas dels *dispositius catalítics*, en els que es mesura el canvi de calor després en el procés primari pel canvi en la conductivitat d'un termistor. A més, existeixen lleugeres discrepàncies segons els autors, com per exemple en el cas dels dispositius elèctrics, els quals sovint són posats en una mateixa categoria juntament amb els dispositius electroquímics.

D'altres fonts de classificació dels sensors poden ser en funció de l'aplicació d'aquests o la determinació d'un analit concret (e.g. els sensors de pH, per ions metàl·lics, per la determinació d'oxigen o d'altres gasos, etc.). Una altra possible classificació pot ser d'acord amb el seu mode d'aplicació; per exemple sensors destinats a ser utilitzats *in vivo*, sensors pel monitoratge de processos, i així successivament.

De forma similar, és possible establir anàlogament altres tipus de classificacions diferents, sempre que aquestes es basin en principis clarament definits i disposats lògicament.

Tanmateix, els biosensors no són presentats com una classe especial (tot i que la seva distinció també pot esdevenir en una font de classificació dels sensors) donat que els processos en que aquests es basen són, en general, comuns als sensors químics. Així doncs, els biosensors poden ser diferenciats d'acord amb l'element biològic utilitzat com a receptor, entre els que podem trobar: organismes, teixits, cèl·lules, orgànuls, enzims, anticossos, etc. [4]. A més, els biosensors poden tenir diversos sistemes enzimàtics acoblats, els quals serveixen per l'amplificació del senyal [5]; o inclús ser combinats per obtenir sistemes multisensors (§1.1.5 i §1.4). Justament, aquests sistemes multisensors han donat lloc al que es coneix com a llengües bioelectròniques (§1.4.3.1).

1.1.4. Biosensors

Tal com ja s'ha dit, la principal diferència entre els sensors químics i els biosensors, radica en la incorporació de material biològic en l'estructura del dispositiu analític (ja sigui en contacte o integrat com a transductor del senyal) amb la finalitat de conferir al sistema una sensibilitat específica front a un analit determinat [4]. Combinant així el poder analític dels sensors químics amb la selectivitat proporcionada pel bioreceptor, essent aquest el seu principal avantatge.

Així doncs, un biosensor és un dispositiu analític que combina un element de reconeixement (o detecció) biològic amb un transductor per tal de produir un senyal proporcional a la concentració de l'analit [6]. Aquest senyal pot provenir del canvi en la concentració de protons, en l'estat redox, alliberament o absorció de gasos, emissió o absorció de llum, i així successivament, provocat per la reacció de l'element de reconeixement biològic front del compost desitjat. Llavors, el transductor converteix aquest senyal biològic en una resposta mesurable tal com un canvi en el corrent o el potencial mesurable a través de mitjans electroquímics o l'absorció de llum per mitjans

òptics; els quals poden ser amplificats, processats i/o emmagatzemats per a la seva posterior anàlisi.

D'aquesta manera, nombrosos biosensors han estat desenvolupats a partir de la immobilització de diferents biomolècules en el transductor tals com enzims, anticossos, orgànuls o microorganismes, així com cèl·lules d'animals i plantes o teixits; els quals han estat utilitzats com a element de reconeixement biològic [4]. Essent precisament l'ús de diferents biomolècules una de les fonts de classificació dels diferents biosensors.

Finalment, un altre aspecte clau en el desenvolupament dels biosensors, quasi tan important com l'element biològic utilitzat, és el mètode d'immobilització d'aquest. A saber, el concepte de biosensor implica que l'element de reconeixement biològic estigui en contacte íntim i permanent amb l'element transductor, requerint de la fixació d'aquest: procés d'immobilització [7]. Així doncs, els objectius claus d'aquesta etapa són aconseguir la fixació del material biològic (ja sigui en superfície o en volum) sense alterar l'activitat o selectivitat d'aquest, proporcionant-li a la vegada un entorn estable i mantenint-lo actiu. D'altra banda, al igual que qualsevol procés analític, el mètode seguit ha de ser reproduïble, simple, evitant condicions extremes i la introducció de substàncies estranyes durant el procés que puguin afectar la resposta del biosensor.

En aquest sentit, existeixen una gran varietat de metodologies per tal de dur a terme la immobilització de material biològic sobre les superfícies dels transductors, que són, de forma general: adsorció física, entrecreuament, atrapament, retenció amb membranes i unió covalent; per últim, també l'ús de compòsits/biocompòsits, els quals poden ser considerats com una subclasse dins els mètode d'atrapament.

Tot i que els avantatges i inconvenients dels diferents biosensors variaran lleugerament en funció de l'element biològic utilitzat com a element de reconeixement, si és cert que aquests comparteixen un conjunt de característiques comunes, les quals es troben resumides a la Taula 1.1.

Taula 1.1. Principals avantatges i desavantatges dels biosensors com a dispositius.

| <i>Avantatges</i> | <i>Desavantatges</i> |
|---|--|
| - Elevada selectivitat i sensibilitat | - Temps de vida (inestabilitat del compost biològic) |
| - Mecanisme d'acció ben definit i caracteritzat | - Condicions de conservació i/o funcionament més "delicades" |
| - Processos catalítics | - Requereixen processos d'immobilització |
| - Combinació de diferents elements biològics; e.g. sistemes bienzimàtics, DNA-enzim, etc. | - Increment del preu del dispositiu |
| - Possibilitat de treballar per inhibició | - Reutilització no sempre possible |

Concretament, en el present treball, tots els biosensors utilitzats han estat del tipus enzimàtic, basats en la fabricació de biocompòsits (§3.1). Aquests es basen en la preparació d'un material que integra en el si del transductor (grafit en pols), l'enzim (un únic o varis; tot i també poder ser utilitzats per a immobilitzar altres tipus de biomolècules) i els possibles mediadors, catalitzadors o cofactors, juntament amb un material aglutinant que actua com a matriu de suport [8].

Així doncs, els compòsits es basen en fases conductores disperses en matrius polimèriques que combinen les propietats elèctriques del grafit amb la facilitat de manipulació de les substàncies de naturalesa plàstica (epoxi, metacrilat, tefló, etc.); obtenint un producte resultant que presenta propietats electroquímiques, físiques, mecàniques i econòmiques molt interessants en comparació amb els conductors clàssics (or, platí, grafit, etc.).

Els principals avantatges de l'ús dels compòsits són la seva capacitat d'integrar diferents materials que permetin millorar la seva sensibilitat i selectivitat, la seva flexibilitat en quan a mida i forma que aporten en la seva fabricació, la possibilitat de polir la superfície obtenint així material fresc pels següents assaigs (obtenint resultats reproduïbles ja que la composició del "compòsit" es homogènia), una major relació S/N (senyal/soroll) en comparació amb els conductors purs, fet que permet assolir millors límits de detecció, etc.; a més a més, donat el seu baix cost i la possibilitat de construir-los un mateix, permet adaptar la seva composició a cada aplicació.

1.1.5. Matrius de sensors i/o biosensors

Tal com s'ha descrit en els apartats anteriors, en el moment de desenvolupar un sensor o biosensor, el nostre objectiu serà la recerca del "sensor ideal" (§1.1.2), és a dir, un sensor que presenti resposta específica vers l'espècie d'interès. No obstant, tal com ja s'ha explicat, aquesta representa una situació utòpica, no assolible en tots els casos.

Així doncs, els sensors són dissenyats normalment per operar en condicions ben definides vers analits específics i en certs tipus de mostres. Per tant, no sempre és necessari que un sensor respongui específicament a un analit determinat; ja que en les condicions de funcionament (acuradament controlades), el senyal de l'analit pot ser independent d'altres components de la mostra, fet que permet la determinació d'aquest sense requerir cap tractament preliminar d'aquesta.

No obstant, quan això no és possible, la solució clàssica requereix de la realització d'etapes de pretractament de la mostra prèvies a la mesura, per tal d'emascarar o eliminar les espècies interferents [9]. El problema d'aquesta estratègia és que incrementa el nombre d'etapes requerides per dur a terme l'anàlisi, una situació antagònica als principis dels sensors químic, els quals tenen com a objectiu unir totes les etapes del procés analític en un únic dispositiu. A més, presenta la limitació que sovint aquests processos són específics per cadascuna de les espècies interferents, implicant així un pretractament de la mostra molt laboriós que al seu temps suposa un increment en el cost; problemàtica que incrementa exponencialment en el cas que vulguem fer una anàlisi de diversos analits.

Per contra, una tendència en el camp dels sensors que està sorgint com a alternativa a les situacions en que no podem assolir la condició anterior, és l'ús de matrius de sensors enlloc d'un únic sensor individual. Mitjançant aquesta metodologia de treball, és possible extreure molta més informació del sistema, cosa que permet resoldre situacions i problemes no assolibles seguint la metodologia clàssica.

Per tant, els sensors inespecífics, sempre que presentin una reproductibilitat satisfactòria, poden ser utilitzats en conjunt (matrius de sensors) per a l'anàlisi multicomponent basant-se en l'ús d'un model de calibratge multivariable i el processament dels senyals (eines que es descriuran amb més detall a §1.4); desplaçant així la complexitat de l'anàlisi del camp químic al del processament.

Aquests sistemes per a l'anàlisi de diversos components es denominen matrius (o xarxes) de sensors, i segons com siguin els sensors que formen la matriu, podem classificar-los en els següents tipus [10]:

- *Matrius de sensors redundants*: basades en la combinació de sensors del mateix tipus, els quals proporcionen un replicat de la mesura en temps real. Aquesta situació que pot semblar contradictòria i poc significativa des del punt de vista clàssic, contribueix a augmentar la precisió i robustesa del sistema, ja que per un costat estem replicant la mesura, mentre que per l'altre ens proporciona un mecanisme d'autodiagnosi del funcionament del sistema. És a dir, ens permet detectar el mal funcionament d'un sensor, ja que si aquest deixa de respondre correctament la seva resposta diferirà del seu replicat; mentre que si el sistema opera inadecuadament, la resposta anormal d'ambdós sensors farà pensar en una falla en aquest.

- *Matrius de sensors selectius*: basada en la combinació de diferents sensors selectius, essent cada un d'ells específic vers una substància determinada, per tal d'assolir la multideterminació en paral·lel de diferents espècies. Permetent així la construcció d'un calibratge independent per cada sensor gràcies a l'adquisició multicanal del senyal. Així doncs, el major guany en aquest cas és des del punt de vista de processament de les mostres, permetent una reducció important en el cost d'anàlisi gràcies a la minimització del volum de mostra, de l'ús de reactius i del temps d'anàlisi.
- *Matrius de sensors amb resposta creuada*: basada en la combinació de sensors inespecífics (no correlacionats), amb baixa selectivitat i que presenten una resposta creuada amb sensibilitats lleugerament diferenciades vers a un cert nombre d'analits; proporcionant un sistema sobredeterminat que pot fer possible la determinació simultània d'una sèrie d'analits en una mescla complexa, corregint les interferències que es produeixen entre ells, mitjançant l'ús d'eines de calibratge multivariable. Amb aquesta estratègia treiem profit del que fins ara era una limitació dels sensors i desplaçem la complexitat de l'anàlisi del camp químic al del processament. Encara més, aquesta configuració ens permet afrontar anàlisis més complexes, els quals no cerquen la determinació d'una espècie o conjunt d'espècies concretes, sinó el d'una percepció o qualitat complexa com poden ser les relacionades amb el sentit del gust o l'olfacte (*taste sensor* [11]), o espècies per les quals no disposem d'un sensor (*software sensor* [10]). És precisament aquesta variant, bioinspirada en el mecanisme de reconeixement biològic, la que representa el fonament de les llengües, nassos i ulls electrònics; conceptes que es discutiran més a fons a §1.4.

1.2. Tècniques electroquímiques

La química electroanalítica compren un grup de mètodes analítics quantitius basats en les propietats elèctriques d'una dissolució d'analit quan forma part d'una cel·la electroquímica [12], les quals es poden relacionar directa o indirectament amb l'espècie d'interès, fonamentades generalment en la mesura d'un paràmetre de naturalesa elèctrica o un canvi degut a aquest.

Els mètodes electroanalítics presenten certs avantatges de caràcter general sobre altres tipus de procediments analítics tals com poden ser la seva especificitat per un estat d'oxidació particular d'un element (e.g. permeten diferenciar la concentració de cada una de les espècies en una mescla de diferents metalls tals com M^{a+}/M^{b+} , mentre que altres tècniques només proporcionen la concentració total d'aquests) o l'ús de instrumentació econòmica i fàcilment miniaturitzable la qual permet dur a terme mesures *on-field* (e.g. biosensors de glucosa). A més, certs mètodes proporcionen informació relativa a l'activitat de les espècies enlloc de la concentració, el qual pot ser considerat un avantatge o un desavantatge depenen de la situació; e.g. en estudis fisiològics, la mesura de les activitats de ions com el calci o el potassi aporten major informació.

Existeixen una gran varietat de mètodes electroanalítics, els quals en funció de les condicions experimentals i del paràmetre mesurat podem classificar d'acord amb l'esquema mostrat a la Figura 1.2. Així doncs, primerament distingirem entre mètodes interfacials (basats en fenòmens que tenen lloc en la interfase entre la superfície dels elèctrodes i la fina capa de la solució adjacent a la superfície d'aquests; essent aquests els més utilitzats i amb major aplicació) i mètodes en el si de la solució (basats en fenòmens que tenen lloc en el si de la solució, intentant evitar qualsevol efecte interfacial). Dins els primers podem distingir a més dos grans categories en funció de com opera la cel·la electroquímica: estàtics i dinàmics, i.e. en absència o presència de corrent.

Essent els mètodes estàtics basats en mesures potenciomètriques, en que es mesura un potencial electroquímic d'una solució el qual està condicionat per la seva composició, de particular importància degut a la seva rapidesa i selectivitat. D'altra banda, dins els mètodes dinàmics sorgeix una nova diferenciació en funció de si es controla el potencial aplicat i es mesura la intensitat generada arran d'aquest potencial (potencial controlat), o bé es fixa un valor de intensitat i es mesuren altres propietats com per exemple el temps necessari per fer reaccionar l'analit d'interès (intensitat constant). Generalment, els mètodes de potencial controlat són sensibles i presenten intervals dinàmics relativament amplis (varis ordres de magnitud); a més, molts d'aquests procediments poden ser aplicats amb volums de mostra de l'ordre d'uns pocs microlitres, fet que permet assolir límits de detecció encara més baixos.

Així doncs, donat que en concret la tècnica utilitzada al llarg de la present tesi ha estat la voltamperometria, ens centrarem en aquesta al llarg dels següents paràgrafs.

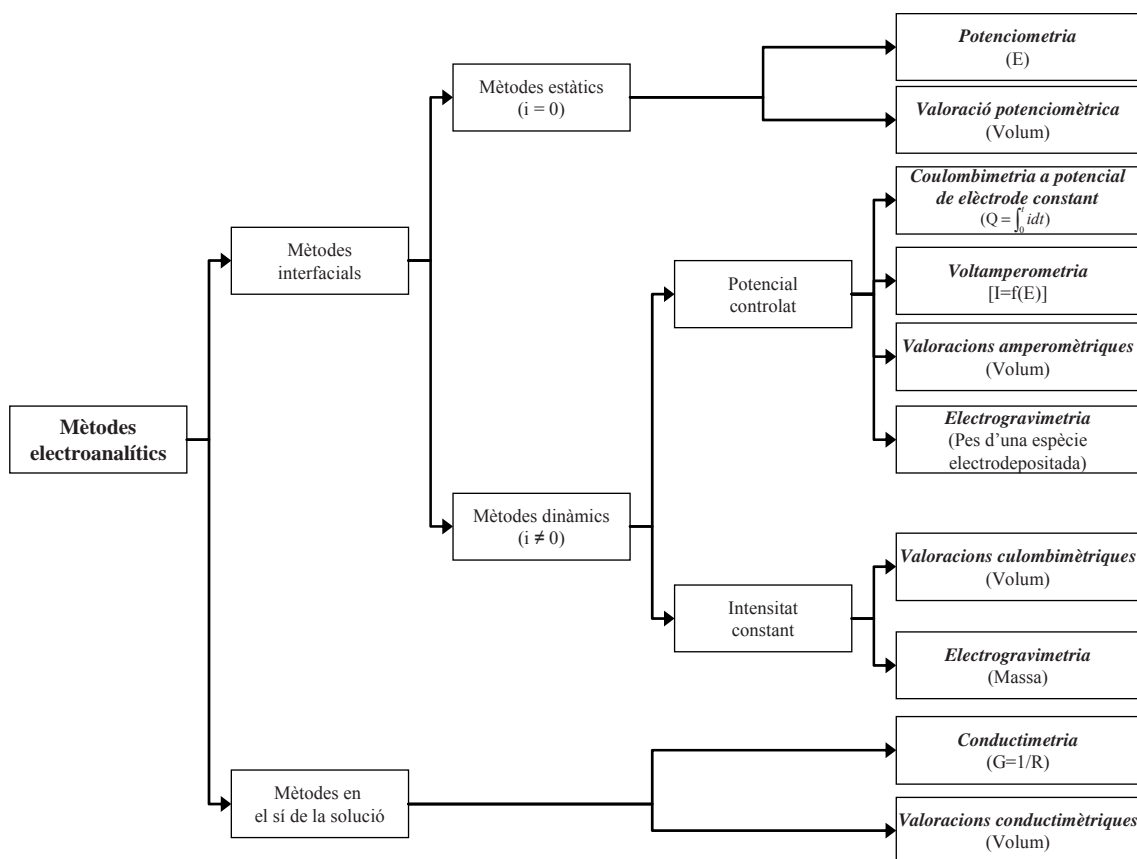


Figura 1.2. Esquema dels diferents mètodes electroquímics. El paràmetre mesurat apareix en cada cas entre parèntesi: *i*: intensitat de corrent, *E*: potencial, *R*: resistència, *G*: conductància, *Q*: quantitat d'electricitat, *t*: temps, volum: volum de la dissolució patró.

1.2.1. Voltamperometria

La voltamperometria comprèn un grup de mètodes electroanalítics en els que la informació sobre l'analit s'extreu a partir de la mesura de la intensitat de corrent en funció del potencial aplicat, en condicions en que s'afavoreixi la polarització per concentració en l'elèctrode de treball. La generació d'aquest corrent prové de l'aplicació d'un potencial extern capaç de provocar la transferència electrònica entre l'elèctrode de treball i una espècie en dissolució, a fi que aquesta pugui ésser oxidada o reduïda.

És important destacar que aquesta situació condueix a una situació en la que hi ha un consum mínim d'analit, fet que significa una mínima alteració o destrucció de la mostra.

i. Cel·la electroquímica

Amb la finalitat d'aplicar el potencial adequat mentre es mesura una intensitat de corrent és indispensable l'ús d'una configuració de tres elèctrodes en el disseny de la cel·la electroquímica. Per tant, la cel·la està formada pels següents 3 elèctrodes: un

elèctrode indicador o de treball, sobre la superfície del qual té lloc la reacció electroquímica, un elèctrode auxiliar, que serveix de contraelèctrode per tancar el circuit, i un elèctrode de referència, que assegura que el potencial aplicat és el que es vol imposar a fi que la reacció electroquímica tingui lloc.

Així doncs, el potencial de l'elèctrode de treball es manté constant mentre que l'elèctrode auxiliar serveix simplement per possibilitar el pas de corrent elèctric cap a, o des de, l'elèctrode de treball, en funció del potencial aplicat. És doncs interessant observar, que en funció del potencial aplicat el nostre elèctrode de treball es comportarà com un ànode o com un càtode, de manera que podrem utilitzar mesures voltamperomètriques per a la determinació de substàncies que presentin tant capacitats de reducció com d'oxidació.

Tanmateix, en sistemes en els que desitgem aplicar una variació de potencial al llarg del temps, realitzarem aquesta variació en el potencial de l'elèctrode de treball amb l'ajut de la instrumentació, la qual serà l'encarregada de mitjançant petits increments en el potencial aplicat (salt de potencial), assolir la situació descrita anteriorment i realitzar la mesura per a cada potencial aplicat.

ii. Senyals d'excitació – diferents mètodes voltamperomètrics

Tal com ja s'ha dit, les mesures voltamperomètriques es basen en la mesura de la intensitat generada (resposta) en funció del potencial aplicat (senyal d'excitació, variable) a una cel·la electroquímica. Per tant, en funció del tipus de senyal d'excitació aplicat, obtindrem una resposta d'intensitat de corrent característica en la qual es basa el mètode; permetent l'obtenció de resultats diferenciats en cada cas, per la qual cosa és un requisit indispensable conèixer la informació que volem extreure i el tipus de senyal que ens interessa aplicar.

En aquest sentit, la Figura 1.3 mostra els quatre tipus de senyals d'excitació comunament més utilitzats en voltamperometria:

- *escombrat lineal*: on el potencial de corrent contínua aplicat a la cel·la augmenta linealment en funció del temps; correspon al senyal clàssic en voltamperometria.
- *diferencial de polsos*: derivada de l'anterior, on el senyal d'excitació correspon a la superposició d'una sèrie de polsos regulars sobre l'escombrat lineal, mesurant la diferència de intensitat abans i després de cada impuls.

- *ona quadrada*: on s'apliquen variacions de potencial en forma d'escala (una ona quadrada es superposa sobre l'escombrat de potencial), mesurant la diferència de corrent després de cada canvi de potencial.
- *voltamperometria cíclica*: on la funció de potencial presenta forma triangular, proporcionant una variació de potencial cíclica entre dos valors, augmentant i disminuint entre aquests dos amb la mateixa velocitat.

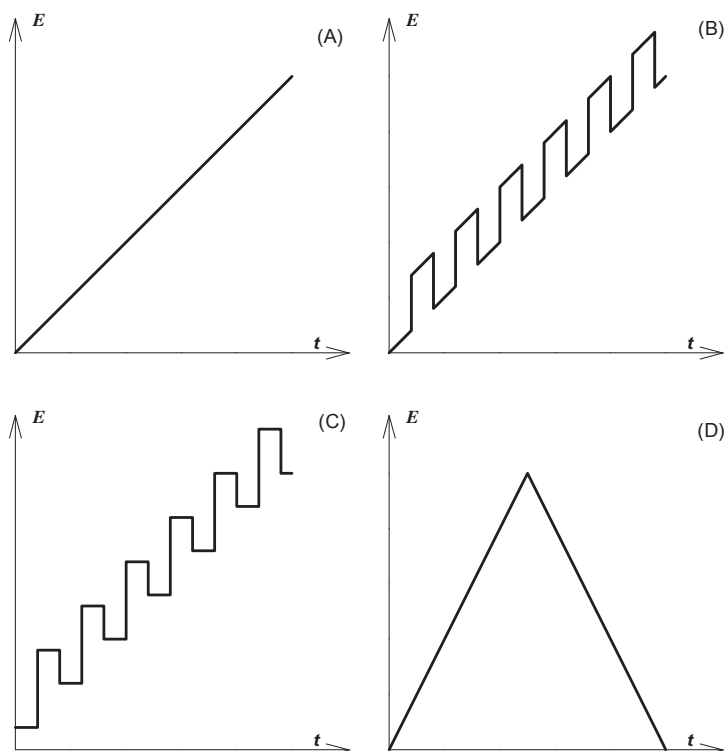


Figura 1.3. Esquema dels diferents senyals de potencial d'excitació utilitzats habitualment en voltamperometria: (A) escombrat lineal, (B) diferencial de polsos, (C) ona quadrada i (D) voltamperometria cíclica.

Evidentment, i tenint en compte la complexitat dels processos redox que es poden dur a terme dins les cel·les electroquímiques, existeixen multitud de paràmetres que podem controlar per tal d'afavorir algun procés concret. Entre ells destaquen la velocitat amb la que variem el potencial al llarg del temps (i.e. la velocitat d'escombrat), la diferència de potencial entre mesura i mesura (i.e. el salt de potencial) o l'amplitud de potencial en els casos en que s'apliquen variacions o polsos [13]. Controlant aquests paràmetres som capaços d'obtenir diferents voltamperogrames per una mateixa dissolució. És amb temps i experiència que podem, a priori, predir amb certa garantia d'èxit, unes condicions inicials tals que permetin obtenir els resultats desitjats.

Entre els mètodes descrits anteriorment, el més emprat segurament és la voltamperometria cíclica, utilitzada normalment com a eina per a la caracterització tant de substàncies redox com de les propietats del transductor utilitzat (i.e. el sensor). Aquest tipus de voltamperometria permet distingir parells redox que presentin reversibilitat en les reaccions d'oxidació i de reducció, al igual que permet trobar el potencial al que les substàncies electroactives presenten reactivitat electroquímica. Per tot això, la voltamperometria cíclica acostuma a ser el primer pas en la determinació de substàncies amb propietats redox desconegudes per l'analista o en la caracterització de nous transductors.

A més a més, aquesta és generalment la primera tècnica seleccionada en l'estudi de mecanismes i velocitats de reacció dels processos redox, en particular per sistemes orgànics i organometàl·lics, sempre que el sistema sigui tractable des del punt de vista electroquímic; sent especialment útil en la detecció d'intermedis de reacció.

Juntament amb la voltamperometria cíclica, es pot considerar la voltamperometria diferencial de polsos com al següent mètode més estès; relacionat habitualment amb metodologies per a la determinació de metalls pesats i altres substàncies que permetin la seva deposició (acumulació) a la superfície de l'elèctrode de treball. En aquest cas, es sol fer ús de la redissolució anòdica [13], abans de procedir a la realització de la mesura voltamperomètrica com a tal, per tal de rebaixar els límits de detecció gràcies a la preconcentració assolida mitjançant la seva deposició.

Així doncs la voltamperometria ens ofereix la possibilitat de determinar un gran ventall de substàncies a un preu relativament assequible. Tot i així, a dia d'avui, difícilment és una tècnica que pugui desplaçar la cromatografia o la espectrometria de masses en els sistemes d'anàlisi industrials. En aquest sentit, un dels objectius de la present tesi és demostrar l'aplicabilitat d'aquesta tècnica, especialment com a mètode de cribratge (*screening analysis*).

iii. Voltamperogrames

La voltamperometria cíclica és una tècnica útil per caracteritzar el comportament electroquímic, en un ampli interval de potencial, d'un transductor i de les espècies electroactives presents a una determinada mostra; convertint-la en una tècnica poderosa per a l'extracció de informació tant de la mostra com del transductor. Tanmateix, podent usar la variació del transductor com a font per obtenir informació addicional i complementària relativa a la mostra (§1.1.5 i §1.4.3.1).

Com ja s'ha dit, aquesta tècnica consisteix en mesurar la intensitat de corrent que s'obté de la variació lineal del potencial aplicat en funció del temps, és a dir, quan es realitza un escombrat des d'un potencial inicial a un potencial determinat i s'inverteix després el sentit de l'escombrat per tornar de nou al potencial de partida (Figura 1.3D). El resultat d'aquest registre es coneix com a voltamperograma cíclic (Figura 1.4), i consisteix en la representació de la intensitat de corrent (i) que circula entre l'elèctrode de treball i l'auxiliar, en funció del potencial aplicat en cada moment (E).

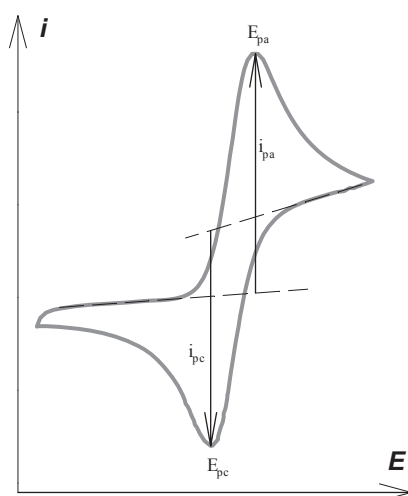


Figura 1.4. Exemple d'un voltamperograma cíclic característic d'un sistema redox reversible.

En anar cap a potencials positius, l'elèctrode es converteix en un ànode (oxidant) de manera que, si inicialment es té tota l'espècie electroactiva en la seva forma reduïda, s'afavoreix la seva oxidació sobre la superfície de l'elèctrode, la qual cosa provoca un augment de la intensitat anòdica. Atès que les mesures es fan en solucions sense agitació, aquesta intensitat deixa d'augmentar un cop s'ha consumit tota l'espècie que es troba en contacte amb l'elèctrode i únicament arriba analític per difusió (l'elèctrode es troba polaritzat), de manera que hi ha un decreixement del senyal i s'obté, com a conseqüència, un pic d'intensitat.

Seguidament, s'inverteix el sentit de l'escombrat de potencial i , en anar cap a potencials negatius, l'elèctrode esdevé un càtode (reductor) que provoca la reducció de les espècies oxidades que s'han acumulat durant el procés anterior sobre la seva superfície (en cas que la reacció sigui reversible) o d'altres espècies electroactives presents en la solució. Això ocasiona un augment de la intensitat catòdica fins a un

valor màxim, que correspon al punt en què ja s'haurà consumit tota l'espècie oxidada present a la superfície de l'elèctrode i, per tant, el senyal comença a decreïxer.

Els paràmetres més importants, per a un estudi qualitatiu d'un sistema redox concret, que s'obtenen d'un voltamperograma cíclic són els potencials i les intensitats de corrent dels pics anòdic (E_{pa} i i_{pa}) i catòdic (E_{pc} i i_{pc}), respectivament. No obstant, en el nostre cas no utilitzarem aquests paràmetres, sinó que construirem un model multivariable capaç d'extreure tant informació qualitativa com quantitativa a partir del registre voltamperomètric complet tal com s'explica en el següent apartat (§1.3).

1.3. Anàlisi multivariable i tractament quimiomètric

El treball amb sistemes voltamperomètrics, en els que es realitza un registre complet per cadascuna de les mostres i sensor, genera un gran volum de dades multidimensionals, la gestió de les quals desborda els mètodes clàssics de processament; requerint així de l'ús de mètodes estadístics i matemàtics per tal d'anàlitzar i extreure tota la informació present en aquestes: les eines quimiomètriques [14-16].

La quimiometria és la disciplina que s'encarrega de l'extracció de informació de sistemes químics a partir del control de les dades [17]. És una ciència molt interdisciplinària, la qual utilitza mètodes emprats freqüentment en disciplines de pur anàlisi de dades tals com l'estadística multivariant, les matemàtiques aplicades i les ciències de la computació (informàtica), per tal d'abordar problemes químics, bioquímics, mèdics, biològics o d'enginyeria química.

Dit d'una altra manera, la quimiometria és la disciplina que utilitza mètodes matemàtics i estadístics per a dissenyar o seleccionar procediments òptims de mesura i d'experiments, i obtenir la màxima informació química de les anàlisis; essent aquests dos els seus objectius bàsics [17]. Així doncs, aquesta intenta combinar els avantatges del processament computacional amb els senyals d'origen químic per tal d'extreure'n la màxima informació. Essent aquesta darrera vesant l'àmbit d'ús d'aquesta disciplina en la present tesi, i sobre la qual aprofundirem al llarg dels següents apartats.

L'avanç assolit en l'anàlisi multiparamètrica o multivariant ha estat possible, en gran part, gràcies a l'ús de la informàtica en l'enregistrament i tractament de les dades generades. L'actual simbiosi entre la instrumentació analítica i, el programari i maquinari actual permet tasques inimaginables fa només unes dècades.

Precisament, la creixent demanda de sistemes de tractament de dades que incorporessin la possibilitat de processar un gran volum de dades simultàniament ha permès a la Química Analítica disposar d'eines capaces de discriminar, classificar, quantificar o modelar sistemes complexos i/o mesclades fins aleshores impossibles de desxifrar.

Dins de tot el ventall d'eines, farem una classificació d'aquestes segons la informació extreta i el mecanisme d'aprenentatge. A saber, en el primer cas distingirem entre anàlisi qualitativa o quantitativa, segons el tipus de informació que ens permeti extreure l'eina de processament; mentre que en el darrer, distingirem entre mètodes supervisats o no supervisats, depenent de si la relació que s'intenta establir entre les mostres es fa a partir de dades no "etiquetades" o s'empra una part de les mesures per a la construcció d'aquesta. Per tant, categoritzant els mètodes segons la informació proporcionada al model durant el processament de les dades.

Tanmateix, i com ja s'ha dit, aquestes eines quimiomètriques solen tenir un elevat rerefons matemàtic bastant complex i elevat [18], dificultant-ne així la seva comprensió. No obstant, actualment és possible trobar paquets informàtics comercials (e.g. EasyNN, Neural Network Software; The Unscrambler, CAMO Software; Matlab, MathWorks) que ens permeten la fàcil utilització d'aquestes eines sense la necessitat de conèixer tot el seu mecanisme de funcionament intern (i.e. incorporen de forma preprogramada la part matemàtica); requerint només del coneixement conceptual i bases teòriques d'aquestes (i.e. avantatges i limitacions de cada mètode). És a dir, només necessitem conèixer el problema analític que volem resoldre, i escollir l'eina quimiomètrica més adequada per assolir la resolució d'aquest.

D'aquesta manera, l'objectiu al llarg dels següents apartats és establir de forma molt breu, les bases de les diferents eines quimiomètriques i el seu àmbit d'aplicació.

1.3.1. Equiparació de la resposta

Al igual que en els sistemes convencionals, i per tal de garantir una correcta modelització de les dades i evitar arribar a conclusions errònies, la qualitat de la informació recollida és un factor determinant. És a dir, independentment de que estiguem fent ús d'unes eines quimiomètriques molt poderoses i efectives a l'hora de resoldre situacions molt complexes, si la informació de partida no és adequada o conté errors hi haurà una manca de veracitat en la informació analítica extreta pel model. Així

doncs, hi ha una sèrie de precaucions prèvies a l'etapa de modelatge que cal tenir en compte per tal d'assegurar la fiabilitat de les mesures.

En aquest sentit, el primer paràmetre clau és el control del procés d'adquisició de dades evitant qualsevol font d'error: errors instrumentals, derives en les mesures, connexions defectuoses, degradació de la resposta dels sensors, histèresi dels sensors o fins i tot, un error per part de l'analista.

Amb el fi de detectar totes aquestes possibles fonts d'error, existeix una solució molt comuna i senzilla, consistent en analitzar el conjunt de mostres de forma aleatòria. Una altra precaució que s'ha pres consisteix en analitzar una mostra de control cada cert període de temps per tal de comprovar, mitjançant la comparació de la seva resposta, que aquesta no ha sofert variacions significatives. Per últim, una altra possible mesura seria el replicat de les mostres, donat que aquest permetria disminuir el soroll de les mesures, ja que la mitjana de senyals replicats compensa els errors aleatoris, la qual cosa permet augmentar la relació senyal/soroll (S/N).

D'altra banda, quan pretenem construir un model de resposta que englobi la resposta de diferents sensors, un altre factor que cal tenir en compte és la importància relativa de les variacions de cada variable, per tal d'equiparar a priori la contribució d'aquestes en el model final [19], assolint una correcta modelització per part de les eines quimiomètriques. Per aquest motiu, amb independència tant del mètode utilitzat en l'etapa de modelatge com de l'ús d'una etapa de compressió del senyal, sempre caldrà dur a terme una etapa de centrat i/o normalització de les dades abans de la construcció del model. Per fer-ho, algunes de les estratègies que es poden utilitzar són: el centrat, l'estandardització, l'autoescalat o la normalització de les dades (§3.4.1.1).

1.3.2. Compressió del senyal

Tot i els avenços assolits tant en el camp quimiomètric com informàtic, l'elevada dimensionalitat de les dades generades quan es treballa amb una matriu de sensors voltamperomètrics (mostres x sensors x potencials aplicats) suposa encara una limitació tant des del punt de vista de modelatge com computacional per tal d'assolir-ne el seu tractament en cru en determinats casos; requerint així d'una etapa de compactació i reducció de la dimensionalitat de les mateixes.

En aquest sentit, la solució més utilitzada i senzilla quan es treballa amb un conjunt de voltamperogrames és la utilització d'una etapa de preprocessament per a la

compressió i reducció de les dades, prèvia a l'etapa de modelatge. El principal objectiu d'aquest pas és reduir la complexitat del senyal d'entrada, preservant la informació rellevant i augmentant la seva compatibilitat amb el model; procés el qual permet a més obtenir avantatges en el temps d'entrenament, evitar la redundància en les dades d'entrada i obtenir un model amb una major capacitat de generalització [20].

Així doncs, el primer pas consisteix en el desplegament (*unfolding*) de les dades originals en una matriu de 2D (Figura 1.5) [21], seguit de l'ús de les eines quimiomètriques estàndard de calibratge lineal o no lineal com la regressió de components principals (PCR), la regressió per mínims quadrats parcials (PLS) o les xarxes neuronals artificials (ANNs). Combinant, aquest desplegament amb una eina capaç de comprimir el senyal; el qual, tot i no ser un requisit per alguns mètodes (e.g. PCR o PLS), s'ha demostrat que fins i tot en aquests casos, el seu ús en millora el comportament del model generat [22, 23].

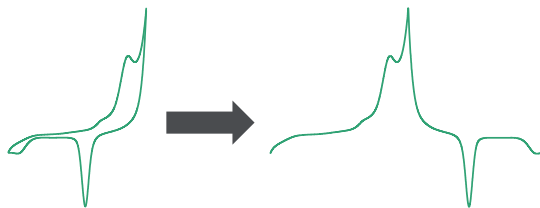


Figura 1.5. Esquema del procés d'*unfolding* d'un voltamperograma.

Per tal d'assolir aquesta etapa de compressió existeixen diferents mètodes descrits a la bibliografia tals com són: la selecció de característiques (*feature selection*) [24], l'anàlisi de components principals (PCA) [25], els polinomis de Legendre [26], les funcions "kernel" [27], la integral de finestra seccionada [28], la transformada ràpida de Fourier (FFT) [29] o la transformada Wavelet discreta (DWT) [30]. El fonament dels quals i diferències es troba descrit a §3.4.1, degut a l'àmplia utilització que se n'ha fet al llarg de la present tesi.

D'altra banda, existeix una solució directa al problema anterior, basada en l'ús de mètodes trilineals tals com els models PARAFAC o Tucker en el cas de l'anàlisi qualitativa, o la regressió parcial per mínims quadrats múltiples (nPLS) en el cas quantitatiu [31]. Aquestes eines, representen les noves tendències en l'anàlisi quimiomètrica, i són el resultat del desenvolupament de nous mètodes amb l'objectiu de fer front a les creixents necessitats d'eines més poderoses. No obstant, la complexitat d'aquestes tècniques és també un factor crític [32], el qual en dificulta la seva aplicació.

1.3.3. Anàlisi qualitativa

Aquest tipus d'anàlisi és el més difós i àmpliament utilitzat, permetent la classificació o discriminació de mostres gràcies als mapes generats mitjançant l'ús de les diferents eines quimiomètriques; entre les quals destaquen el PCA [33], el LDA [34], el PLS o PLS-DA [35] i les SVM [36], entre altres.

1.3.3.1. Anàlisi de Components Principals (PCA)

El PCA és un mètode matemàtic de reorganització de la informació d'un conjunt de dades provinent de diferents mostres, el qual busca aquelles variables de la nostra matriu multiparamètrica on existeixi la major variabilitat entre les diferents mostres (Figura 1.6).

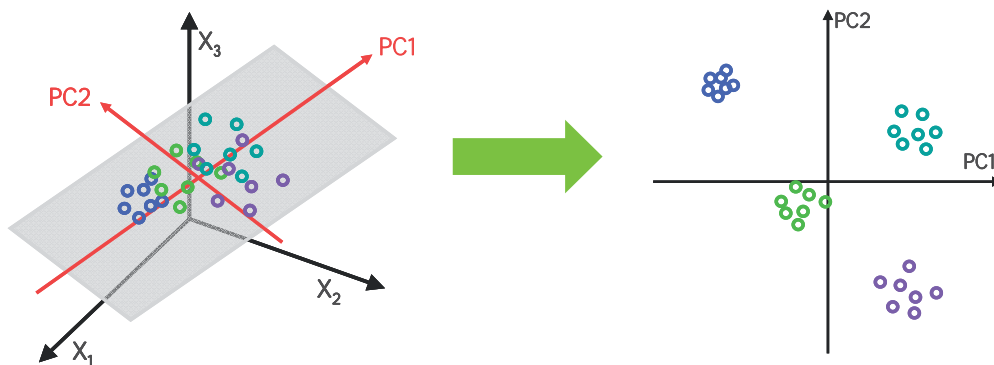


Figura 1.6. Esquema de construcció del model PCA. Representació d'un conjunt de dades originalment en tres variables i com aquest queda reduït a dos PCs amb l'aparició de diferents zones d'agrupament.

Bàsicament, aquest procediment matemàtic pot ser considerat com una transformació o projecció de la informació continguda en les variables originals a un número menor de variables adjacents ("latents") anomenades components principals (PCs) amb noves coordenades sobre aquests eixos anomenades "scores", obtinguts després de realitzar la transformació matemàtica [33]. El principal avantatge d'aquesta transformació rau en el fet que proporciona una eina capaç d'expressar la major part de la variabilitat del conjunt de dades inicial amb un nombre reduït de variables, la representació de les quals permet detectar les relacions entre les mostres i les variables; proporcionant, a partir de les variables de mesura, un nou punt de vista de com certes mostres mostren similituds entre elles, o com es diferencien les unes de les altres.

D'aquesta manera, el PCA pot ser utilitzat quan el conjunt conté informació de només unes poques variables, però esdevé més útil quan hi ha un gran nombre de

variables, com en el cas de tenir tot un registre voltamperomètric o inclús varis en el cas de treballar amb matrius de sensors.

No obstant, cal tenir en compte que aquest només presenta una millor representació visual de les dades, i que no pot ser considerat per si sol com un mètode de reconeixement de patrons o classificador, requerint d'una eina de modelatge, tal com les ANNs, per poder actuar com a tal [37, 38].

1.3.3.2. Anàlisi Discriminant Lineal (LDA)

El LDA està relacionat amb el PCA en que, al igual que el mètode anterior, aquest aplica una transformació al conjunt de dades inicials projectant aquestes en uns nous eixos (anomenats funcions discriminants, DFs) que permeten visualitzar les similituds i diferències d'una forma més senzilla i directa (Figura 1.7) [39]. No obstant, a diferència del mètode anterior, aquest es tracta d'un mètode supervisat, en el que enlloc de construir els nous eixos basant-se en les variables que proporcionin la màxima variabilitat entre les mostres, aquests són construïts a partir de combinacions lineals d'aquestes que proporcionin la millor discriminació entre els diferents grups de mostres. A més, una altra diferència és el nombre d'eixos en que la informació és projectada, ja que en aquest cas obtenim tantes funcions discriminants com classes a separar menys una ($k-1$); i encara més important, a diferència del PCA, el LDA si que pot actuar com a classificador.

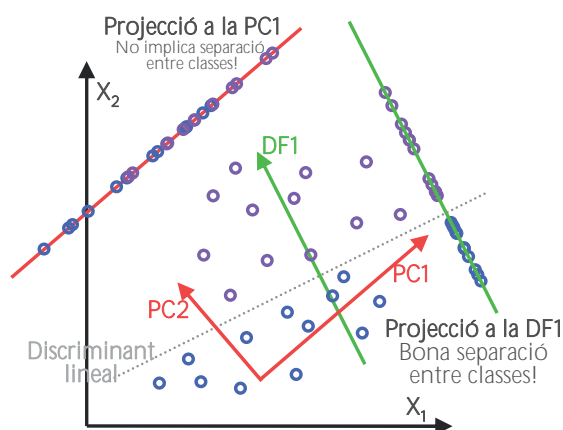


Figura 1.7. Esquema comparatiu entre el mètode PCA i LDA. Breument, el PCA busca la direcció que presenta la màxima variància entre les variables, mentre que el LDA busca la direcció que proporcioni la màxima separació entre aquestes. A més cal notar que en el cas del PCA obtenim tants nous PCs com variables, mentre que en el cas del LDA obtenim tantes com $\#grups-1$.

Per últim destacar que hi ha una gran varietat de mètodes que utilitzen el nom de LDA; així, tot i que aquests estan relacionats, existeixen petites diferències entre ells i cal tenir-ho en compte. Aquest mètode, tal i com va ser descrit originalment, es basa en l'ús de la fórmula de Bayes, i es va dissenyar com a mètode capaç de trobar una funció capaç de discriminar el conjunt de mostres en dos classes diferents; donant lloc al LDA en la seva versió "lineal" i al QDA en la "quadràtica".

Molt relacionats amb aquest, i com a extensió, trobem l'anàlisi discriminant lineal de Fisher o anàlisi de variació canònica (*Canonical Variate Analysis*, CVA), el qual pot ser considerat una extensió de l'anterior a situacions multiclasse [34]; fins el punt que es troba descrit sota el propi nom de LDA en molts casos (e.g. en el codi de programari **R**). Demostrant així, la manca d'un criteri específic i unívoc, que eviti possibles confusions; motiu pel qual, al llarg del present treball i per tal d'evitar confusions, ens referirem a aquest mètode sempre sota el nom LDA, tot i que cal tenir en compte la manca de precisió que això comporta.

1.3.3.3. Anàlisi Discriminant - Regressió Parcial per Mínims Quadrats (PLS-DA)

El PLS-DA és una variant del PLS utilitzada quan la matriu de *targets* (y) és de tipus binari, per tant, basada en els mateixos principis que aquest (§1.3.4.3) [35]. En aquest sentit, el PLS-DA mostra certes similituds tant amb el PCA com el LDA. A saber, la transformació matemàtica utilitzada es fonamenta en l'anàlisi PCA, mentre que al igual que en el cas del LDA, es tracta d'un mètode supervisat.

La diferència entre aquest i el PCA és que en aquest cas no es projecten els PCs que presenten major variabilitat, sinó que s'utilitzen aquelles direccions que presentin major capacitat de predicció del model analític presentat. Per tant, els components obtinguts no corresponen als trobats pel PCA sinó que a aquests s'ha aplicat una correcció que permeti una major capacitat predictiva, de manera que obtenim direccions on tenim les distàncies mínimes entre els valors i el component trobat.

1.3.3.4. Màquines de Suport Vectorial (SVMs)

Les SVMs són mètodes supervisats basats en algorismes d'aprenentatge estadístics iteratius per a l'anàlisi i reconeixement de patrons, utilitzats per a la classificació de

mostres (separació de les diferents classes), al igual que el LDA. Tot i que aquestes van ser desenvolupades originalment per la classificació lineal de conjunts de dades [40], també són aplicables a dades no lineals mitjançant l'ús de *kernels* (procés conegut com a *kernel trick*) [41]; essent aquest fet, el que les ha convertit en una eina quimiomètrica molt poderosa [42].

El model construït mitjançant les SVMs és un model basat en una funció que descriu un hiperplà de separació òptima entre les classes. És a dir, una representació de cada mostra com a punts en l'espai, separats de tal manera que les mostres pertanyents a diferents classes es trobin tan separades com sigui possible; classificant llavors els nous casos en funció de quin costat de l'espai apareixen (Figura 1.8).

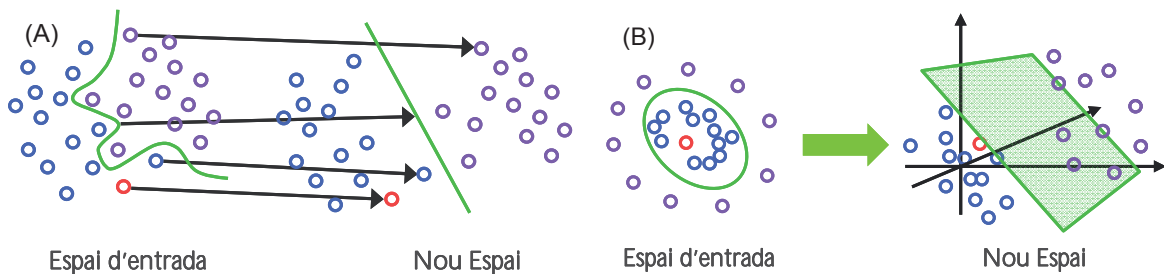


Figura 1.8. Esquema il·lustratiu del procés de transformació de les dades a un nou espai mitjançant l'ús de *kernels*, permeten així la correcta classificació de dades no lineals: (A) espai bidimensional i (B) com al augmentar la dimensió del nou espai facilita la discriminació.

Donat que no sempre l'ús d'una funció lineal és suficient per modelar la separació entre aquestes, el conjunt de dades és transportat a un nou espai dimensional; aquesta transformació és pot fer mitjançant una funció *kernel* (la qual pot ser de diferents formes) permeten així la capacitat al sistema d'assolir la classificació de casos no lineals (Figura 1.8A). D'aquesta manera, els *kernels* poden ser considerats com a una transformació de les dades no lineals a un espai dimensional superior (Figura 1.8B), proporcionant al mateix temps una drecera que permet als algorismes lineals treballar en aquestes situacions; definint els vectors de suport com el conjunt de dades d'entrenament del *kernel*.

Resumint, les SVMs presenten avantatges sobre altres mètodes de classificació, com les xarxes neuronals, ja que tenen una solució única, i una menor tendència al sobreentrenament en comparació amb altres mètodes de classificació no lineal; essent per descomptat, la validació del model l'aspecte fonamental per evitar-ne el seu sobreentrenament igual que per qualsevol altre mètode (§1.4.3.5.4ii). Les SVMs són especialment eficaces per a la modelització de dades no lineals.

1.3.4. Anàlisi quantitativa

Dins de l'anàlisi quantitativa, farem una categorització dels diferents mètodes utilitzats en base a com l'anàlisi de regressió té lloc; distingint així entre mètodes lineals i no lineals. A saber, en els cas dels models no lineals, el conjunt de dades és modelat per una funció la qual és una combinació no lineal dels paràmetres del model i depèn d'una o més variables independents, les quals són ajustades mitjançant un mètode d'aproximacions successives. Mentre que en el cas dels mètodes lineals, es tracta d'un ajust de regressió lineal, ajustat normalment a través del mètode de mínims quadrats.

Aquesta etapa de modelatge pot ser assolida mitjançant l'ús de mètodes lineals tals com el MLR [22], el PCR [43], el PLS [32] o el nPLS [32], o no lineals tals com les ANNs [44], entre altres; els quals es descriuen breument a continuació.

1.3.4.1. Regressió Lineal Múltiple (MLR)

El MLR és un mètode de regressió clàssic que combina, de forma lineal, un conjunt de predictors o variables de resposta (matriu de resposta del conjunt de sensors, "x"), que correlacionen tan estretament com és possible a una única resposta corresponent (*target*, normalment les concentracions de cadascun dels analítics, "y") (Figura 1.9) [45]. Per tant, en cas de tenir un sistema en el que volem modelar diferents predictors a la vegada, caldrà construir models de resposta separats per cadascun del analítics; aquesta mateixa limitació existeix també en el cas del PLS-1, on només una única variable de resposta pot ser modelada a la vegada. En aquest sentit, el PCR pot ser considerat com una alternativa al PLS-1 [45].

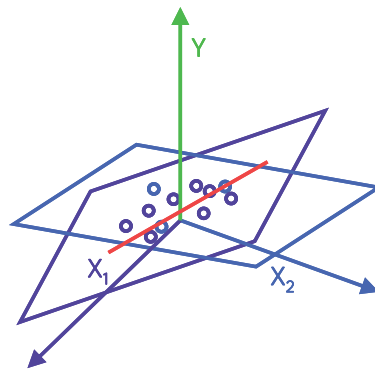


Figura 1.9. Esquema d'ajust d'un model MLR. Bàsicament, i de forma simplificada, es tracta d'una extensió de la regressió lineal univariable habitual: $Y = b_0 + b_1x_1 + b_2x_2 + \dots + b_kx_k$

1.3.4.2. Regressió de Components Principals (PCR)

El PCR és un mètode que permet relacionar la variància entre la variable objectiu (*target*, normalment les concentracions de cadascun dels analits, “y”) amb la variància del conjunt de predictors (matriu de resposta del conjunt de sensors, “x”), amb fins explicatius o predictoris. El PCR és un procediment en dues etapes en que primer es descompon la matriu de dades de resposta mitjançant l’ús del PCA, i després s’ajusta mitjançant un model de Regressió Lineal Múltiple (MLR, §1.3.4.1) utilitzant un número de *scores* dels diferents PCs enlloc de les variables de resposta originals com a predictors (Figura 1.10) [33].

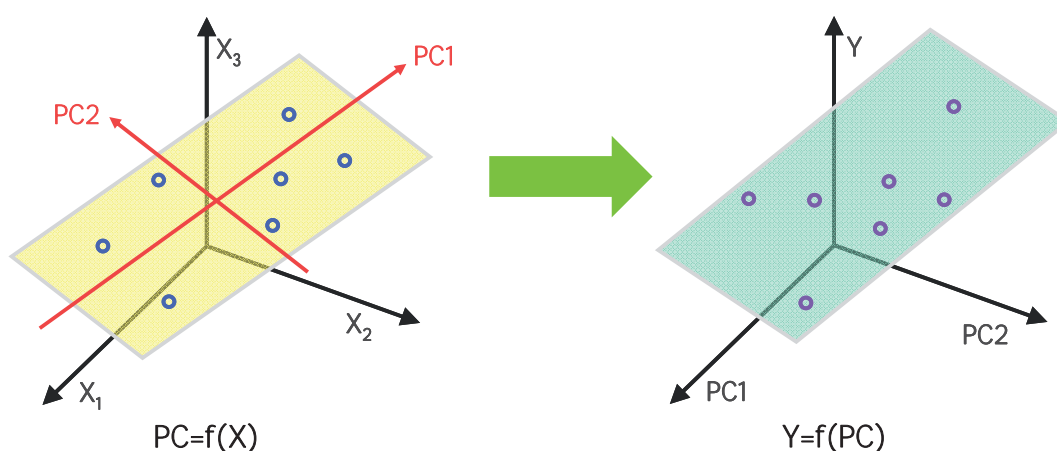


Figura 1.10. Esquema d’ajust d’un model PCR. En una primera etapa s’utilitza el PCA per trobar les components ortogonals, mentre que en una segona etapa s’utilitza el MLR per relacionar els PCs amb la resposta (Y).

Aquest mètode funciona particularment bé quan les diferents variables de resposta expressen informació comuna, és a dir, quan hi ha una gran correlació, o fins i tot col·linealitat [46]. Atès que les puntuacions són ortogonals, la solució MLR és estable i per tant, el model de PCR no pateix dels efectes de col·linealitat. Per a la seva optimització, només cal tenir en compte el nombre de PCs utilitzats per a la regressió MLR, escollint el nombre que permeti assolir un menor valor de RMSE (§3.4.2.3.2.2) pels valors predits amb el model.

1.3.4.3. Regressió Parcial per Mínims Quadrats (PLS)

El PLS és un mètode que també emprava una transformació lineal molt similar a l’anterior; no obstant, aquest modela de forma simultània tant les matrius de resposta dels sensors (*x*) com la de *targets* (*y*) per tal de trobar les variables latents (o ocultes) de

x que millor prediguin les variables latents de y (Figura 1.11). Aquests nous components PLS, anomenats variables o factors latents [47], són similars als PCs extrets en l'anàlisi PCA, però amb la diferència que en lloc de buscar hiperplans de variància màxima entre la resposta i les variables independents, aquest busca un model de regressió lineal mitjançant la projecció de les variables pronosticades i les variables observades a un nou espai. És a dir, el PLS descompon simultàniament x i y en un conjunt de variables latents que maximitza la covariància entre els dos, per després construir un model lineal que maximitzi la relació entre les noves variables. Aquest procés comporta com a avantatge, el fet d'assolir de forma habitual la convergència del sistema utilitzant un menor nombre de factors.

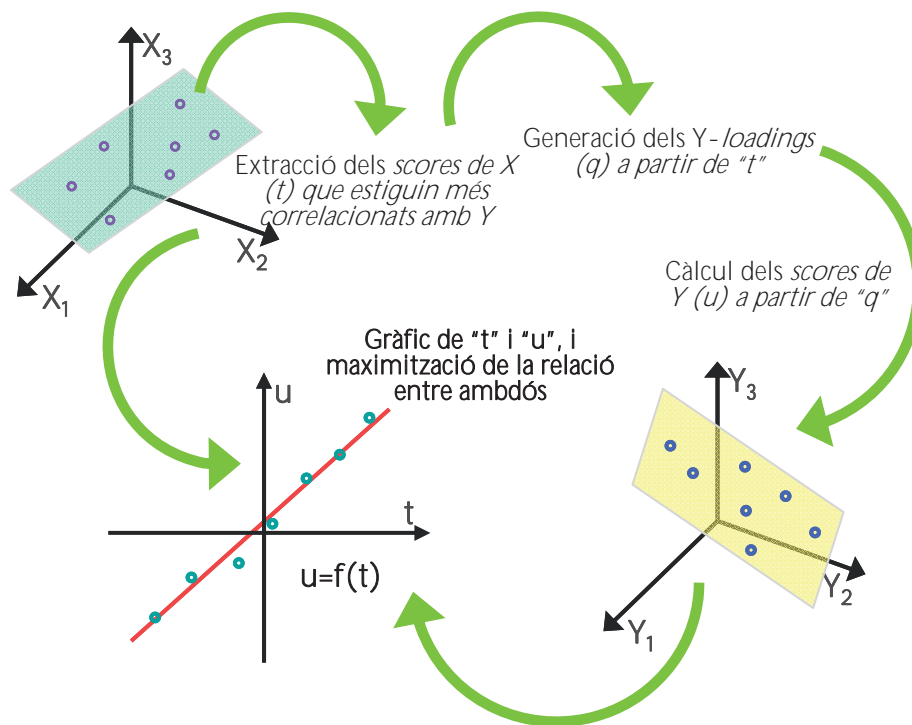


Figura 1.11. Esquema d'ajust d'un model PLS. A diferència del PCR que primer descomposava la matriu " x " i després aplicava la regressió; en aquest cas, tant " x " com " y " són descomposats a la vegada, maximitzant la covariància entre els dos.

Donat que tant les dades de x com de y són projectades a nous espais, la família dels mètodes de PLS es coneixen com a models de factors bilineals. El número afegit al final de la sigla identifica el nombre de modes de les respostes, que en el cas de PLS està restringit fins a 2 [21]. És a dir, el PLS-1 on un bloc de dades x es compon d'una sola variable y o el PLS-2, on tots dos blocs són matrius multidimensionals. En conseqüència, a diferència del PCR o el PLS-1, l'anàlisi PLS-2 pot ser dut a terme amb una o més variables y , el que significa que múltiples respostes y poden ser modelades en

un únic pas durant l'etapa de regressió. A més, en ambdós casos, al igual que en el cas anterior, l'únic paràmetre que cal optimitzar és el número de variables latents tals que permetin obtenir el menor RMSE.

1.3.4.4. *Regressió Parcial per Mínims Quadrats Múltiple (nPLS)*

Aquest mètode és una extensió del mètode estàndard de regressió PLS (§1.3.4.3), per ser utilitzat en estructures de dades amb un major nombre de dimensions de les variables (*multiway data*; tensor) [48]. Com s'ha indicat, aquest mètode presenta l'avantatge que permet relacionar les variacions en una de les diverses variables de resposta (y) disposades en una taula 2D amb les variacions de diversos predictors disposats en una taula 3D (variables “ x ” primària i secundària), amb propòsits explicatius o predictius. Per tant, superant els inconvenients causats durant el procés de *unfolding*, i.e. les discontinuïtats creades d'un enregistrament al següent (e.g. la unió de dos voltamperogrames provinents de dos sensors diferents, Figura 3.7), el qual causa la pèrdua de la relació inicial i crea limitacions per al calibratge de primer ordre [49].

Com a extensió a la regressió PLS de dues vies (PLS-2), la regressió nPLS construeix un model de calibratge que incorpora una relació entre els conjunts de predictors i de respostes basada en l'estructura *multiway* de les matrius. La principal diferència del nPLS respecte el PLS-2 és l'ús de matrius de *loadings* de major dimensionalitat per a la descomposició dels arranjaments multidimensionals de dades (tensors). L'objectiu d'aquesta metodologia és ajustar simultàniament models multilineals formats pels vectors de *scores* i *loadings* tant dels arranjaments de predictors com de resposta, juntament amb un model de regressió que relaciona els dos models de descomposició. D'acord amb la teoria subjacent de la regressió PLS, el model és ajustat de tal manera que els vectors de *scores* tenen màxima covariància amb la part no explicada de la variable dependent [21].

A més, els models obtinguts mitjançant l'aplicació de mètodes de primer ordre a conjunts de dades, les quals la seva estructura *multiway* no es té en compte, poden ser menys robustos i difícils d'interpretar, a diferència dels models basats en mètodes *multiway*, els quals a més solen donar millors prediccions i majors possibilitats per a l'exploració de les dades [48, 50].

En aquest sentit, les dos principals limitacions d'aquest mètode són la seva major complexitat i el fet que pressuposa linearitat (i.e. l'ús d'un mètode lineal, enlloc de

mètodes no lineals tals com les ANNs); no obstant, presenta com a principals avantatges la seva major capacitat de modelització i l'anomena't "avantatge de segon ordre" (*second-order advantage*) [51]. A saber, la capacitat de predir de forma correcta les concentracions d'anàlits d'interès en presència d'interferents, els quals no han estat modelats pel sistema [52].

1.3.4.5. Xarxa Neuronal Artificial (ANN)

Les ANNs van ser originalment dissenyades per tal de mimetitzar el funcionament del cervell humà i aplicades tant a l'anàlisi qualitativa com quantitativa durant la darrera dècada (§1.4.3.5) [53]. Aquestes consisteixen en un número d'unitats de processament simples (o neurones) unides a través de interconnexions ponderades modificables (Figura 1.12) [54].

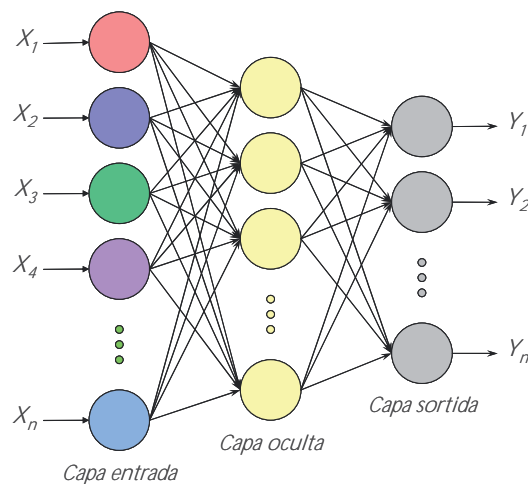


Figura 1.12. Esquema d'una xarxa neuronal artificial (ANN).

En comparació amb els mètodes presentats anteriorment, les ANNs es caracteritzen per ser un mètode de modelització més flexible donat que poden utilitzar tant funcions lineals com no lineals (o inclús combinacions d'aquestes) en les connexions entre les neurones. Així, el seu ús és especialment convenient en el cas de treballar amb sensors que presentin una resposta no-lineal. Això permet també establir relacions més complexes entre un espai de descriptors d'elevada dimensionalitat i les variables a modelar, fet que pot comportar una millor capacitat de predicció del model ANN resultant en comparació amb altres mètodes lineals.

Tanmateix, el principal desavantatge de les ANNs està també directament relacionat

amb la complexa infraestructura del model, el qual requereix de certs coneixements previs per a la selecció de la topologia o arquitectura de la xarxa, i d'un esforç significatiu en l'optimització d'aquesta donada la dificultat de predir-ne la configuració òptima per endavant [55]. Es tracta doncs d'un procés d'assaig-error en el que varis paràmetres (algorismes d'entrenament, número de neurones en la capa oculta, funcions de transferència, etc.) són ajustats per tal de trobar la millor configuració que optimitzi l'eficàcia del model neuronal. D'altra banda, tal com ja s'ha esmentat, aquestes requereixen generalment d'una etapa de preprocessament per tal de reduir la dimensionalitat de les dades d'entrada abans de construir el model ANN.

1.4. Sistemes biomimètics

La constant demanda de nous mètodes analítics capaços d'oferir la màxima informació relativa a una mostra amb el mínim temps i cost possible, sumada a la dificultat d'obtenir sensors químics, envers qualsevol analit, que operin de forma òptima i sense cap interferència o efecte matriu en les condicions requerides en els casos de mostres reals, ha donat lloc a l'aparició de noves estratègies per tal de donar resposta a aquests problemes.

És en aquest context, que ha tingut lloc l'aparició dels sistemes biomimètics [56]. Bàsicament, aquests sistemes analítics consisteixen en la imitació dels diferents sentits dels mamífers, on uns pocs receptors són capaços de respondre a una gran varietat de substàncies, un principi que juntament amb una etapa complexa de tractament de la informació en el cervell, permet la quantificació o qualificació d'un gran nombre de substàncies. Aquests sistemes biomimètics, en oposició als sistemes clàssics, estan basats en la combinació de matrius de sensors amb una baixa selectivitat i/o una resposta creuada per tal d'obtenir un valor afegit en la generació d'informació analítica.

1.4.1. Concepte

Així doncs, aquests sistemes usen la naturalesa com a font d'inspiració per tal de resoldre problemes ja resolts en aquesta. En el cas de la Química Analítica, i prenent com a referència els mamífers, existeix la tendència en la biometització dels sentits de l'olfacte, del gust i de la vista; donant lloc a dispositius coneguts com a nassos, llengües i ulls electrònics, respectivament, en clara referència als sistemes que pretenen imitar.

El fonament bàsic d'aquests sistemes és l'ús de matrius de sensors que presentin resposta davant un gran nombre dels components presents en la mostra (i.e. baixa selectivitat), amb l'objectiu d'obtenir resposta entrecreuada entre ells, sent aquesta la característica que aportarà major valor a la informació recollida (§1.1.5); per posteriorment, realitzar el tractament de la informació obtinguda normalment amb potents eines matemàtiques, associades sovint a la branca de la intel·ligència artificial (§1.3). Així doncs, podem dividir aquests sistemes en dos parts ben diferenciades, per una banda els sensors encarregats d'imitar els diferents sentits i percepcions que aquests ens aporten, i per l'altra, l'eina quimiomètrica, encarregada de processar la informació recollida i obtenir una resposta tal com ho faria el cervell.

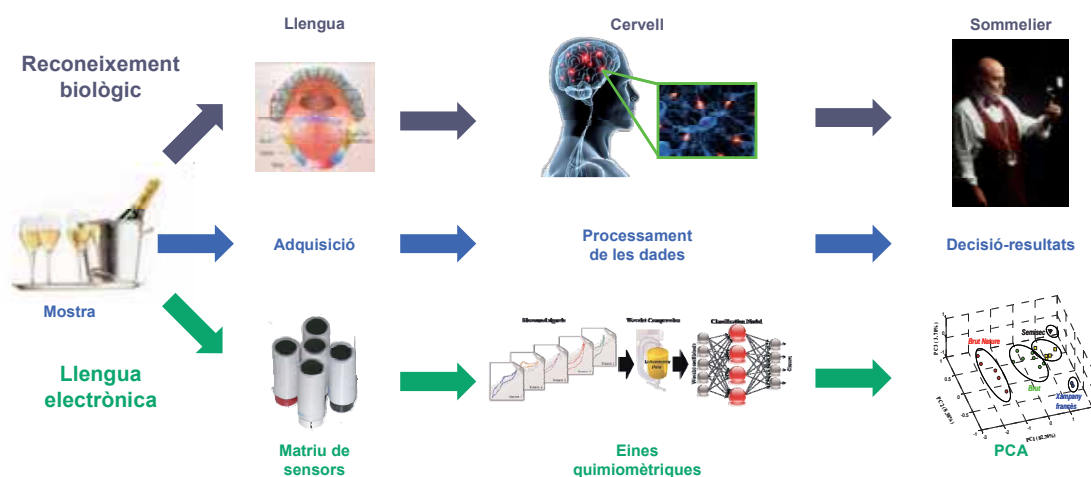


Figura 1.13. Comparació del procés de reconeixement d'una mostra d'acord amb el mecanisme biològic (part superior) i un sistema biomimètic (part inferior).

D'aquesta manera, la Figura 1.13 il·lustra la comparació entre el procés biològic i el sistema biomimètic en el cas d'una llengua electrònica. Inicialment, una matriu de sensors (amb una resposta diferenciada, ampla i complementària) és utilitzada per obtenir informació química de la mostra en el que és l'equivalent dels nostres sentits; en aquest cas la llengua, on cada sensor seria l'equivalent a les diferents papil·les gustatives. Després, en el cas de la percepció humana, la percepció dels sentits és transduïda pels nervis fins al cervell mitjançant senyals elèctrics; de forma similar, en el cas de les llengües electròniques, es genera un senyal elèctric com pot ser una variació de potencial o intensitat de corrent. Finalment, la percepció del gust i el reconeixement es basa en la construcció o reconeixement dels patrons sensorials dels nervis activats pel cervell i en la empremta que aquests han deixat; aquest darrer pas és aconseguit en el cas de les llengües electròniques gràcies a la utilització de software estadístic que

interpreta les dades obtingudes per la matriu de sensors. Al igual que en el mecanisme biològic, fa falta un procés d'aprenentatge o educació de la llengua electrònica per tal que aquesta sigui capaç de reconèixer les propietats en les que estem interessats. És per tot això que diem que es tracta d'un sistema biomimètic, perquè intenta imitar el sistema de reconeixement biològic a partir de mesures experimentals.

Tanmateix, tot i que els termes de “nas electrònic” o “llengua electrònica” neixen basats en la inspiració dels respectius sentits, podent ser aplicats a la detecció d'olors o gustos des d'un punt de vista instrumental, aquests dispositius biomimètics no estan sempre enfocats a la resolució d'aquests problemes. Tal com proposa Mielle et al. [57], *aquests sistemes analítics són evidentment electrònics però no un nas o una llengua*. És a dir, presenten analogies en el medi de detecció i el principi de processament de la informació: com el sistema dels mamífers “mesuren” gasos o líquids i envien la informació a un “òrgan” de reconeixement; però el principi d'operació, el número de “sensors” i la seva selectivitat o sensibilitat són molt diferents. Per tant, aplicant aquests sistemes a altres camps i no restringint-los als casos de naturalesa biològica.

Així doncs, el sistema pot ser utilitzat tant per l'extracció de informació qualitativa com quantitativa de la mostra en funció dels requeriments de l'anàlisi i, conseqüentment, de l'eina de quimiomètrica escollida (§1.3). En aquest sentit, és en l'anàlisi qualitativa (presentant normalment resultats en forma d'agrupacions per la classificació de mostres) on aquests sistemes han tingut una major aplicabilitat, possiblement gràcies a la seva idoneïtat, com ja s'ha discutit, i la menor complexitat en el tractament de la informació. D'altra banda, aquests sistemes també estan emergint amb força en aplicacions quantitatives, gràcies a la seva capacitat de dur a terme la discriminació i quantificació simultània de diferents substàncies o propietats físico-químiques relatives a la mostra; essent aquesta molt interessant en problemes analítics amb presència d'interferents, on una etapa prèvia de separació és freqüent, o fins i tot, indispensable en el cas de l'anàlisi clàssica, però no per aquests sistemes.

1.4.2. Avantatges i inconvenients

El principal avantatge d'aquests sistemes deriva del seu fonament teòric. A saber, a diferència de les tècniques clàssiques d'anàlisi en que la determinació d'una propietat de la mostra implica la identificació de l'analít responsable d'aquella característica i el desenvolupament d'un sensor específic per aquesta; aquests sistemes biomimètics

permeten la identificació de les propietats d'aquesta, inclús en els casos en que els compostos responsables no han estat identificats. Això és gràcies a que aquests sistemes combinen tota la informació química extreta de la matriu de sensors per generar un model tal com el cervell humà faria, evitant així centrar-se en quins són els compostos responsables d'aquestes característiques, però reconeixent-ne al mateix temps la seva empremta i permetent-ne la seva identificació o quantificació; tal com s'il·lustra més endavant amb diferents exemples (§4.1, §0).

En aquest sentit, aquesta aproximació és especialment útil en l'anàlisi qualitativa de mostres, sobretot a nivell organolèptic, quan es pretén imitar des d'un punt de vista analític i objectiu la classificació d'aquestes de forma anàloga a com ho faria un panel de tast; representant a més, una solució directa a l'absència de coneixement dels analits concrets responsables d'aquestes característiques gràcies al seu mecanisme biomimètic.

A més, existeix un altre avantatge derivat que és el fet que al emprar sensors no específics és possible utilitzar les llengües i els nassos per a diferents problemes analítics sense variar cap paràmetre instrumental [58, 59]. Això és possible ja que també permeten modelar sistemes amb més d'un analit i/o interferent que presenten solapament entre els senyals de les mostres, poden presentar inclús relacions no lineals.

Justament, en aquest sentit, un avantatge derivat és la capacitat del sistema per actuar com a *software sensor* [10]. A saber, quan mitjançant l'ús del dispositiu és possible predir de forma quantitativa una propietat de la mostra, la qual no és directament mesurada per cap dels sensors utilitzats, però, tot i així, essent les eines quimiomètriques capaces d'extreure'n aquesta a partir de les matrius de resposta dels sensors. Aquesta aproximació s'utilitza quan una determinada variable difícil de mesurar, és predita utilitzant mesures secundàries, les quals són més simples i/o barates, mitjançant l'acoblament amb algun algorisme matemàtic. Per tant, d'aquesta manera, la informació recollida a partir d'un nombre determinat de sensors pot ser utilitzada per extreure'n informació d'alguna altra espècie, la qual el seu sensor no es troba disponible, correlacionant la informació disponible amb la concentració real; essent aquesta metodologia indirecta especialment útil i adequada per a determinacions analítiques complexes. Alguns exemples serien la determinació del contingut de suc de taronja natural a partir d'una ET potenciomètrica [60]; o la predicció de fòsfor (fosfat) [61] o amoníac [62] en aigües residuals a partir de paràmetres fisicoquímics tals com el pH o la demanda química d'oxigen.

Per tant, cal destacar l'elevada projecció d'aquests sistemes en problemes de

classificació de mostres de naturalesa semblant, on els nassos i les llengües electrònics s'utilitzen com a eines de discriminació, possibilitant la creació de mapes d'agrupació.

No obstant, a pesar dels seus avantatges, aquest tipus de sistemes no gaudeix encara d'una gran aplicació industrial, essent utilitzats només en alguns casos com a instruments complementaris a altres mètodes d'anàlisi. Aquesta complementarietat situa a aquests sistemes normalment com a eines de cribratge (*screening analysis*), procedint posteriorment a l'anàlisi de les mostres mitjançant mètodes clàssics com són l'ús de tècniques com la cromatografia o la espectrometria de masses en el moment en que es vol conèixer a fons la composició d'aquestes.

D'altra banda, el gran avantatge d'aquests sistemes és a la vegada el seu major inconvenient; doncs, com s'ha dit, aquests sistemes requereixen d'un procés d'entrenament o aprenentatge. Per tant, a l'hora de desenvolupar el nostre sistema requerirem d'un elevat nombre de mostres patró per tal de construir el model, més un altre subconjunt per tal de validar aquest; mostres que han de ser representatives del problema analític (tant en nombre com en composició) i de les que hem de conèixer perfectament els paràmetres a determinar. Per tant, necessitem tenir un coneixement absolut del problema analític que volem resoldre, ja que en el moment de fer el disseny del conjunt de mostres que utilitzarem per construir el model cal tenir en compte el rang de treball o la presència d'altres components que presentin sensibilitat analítica, entre altres, per tal d'assegurar la fiabilitat del model construït i dels resultats que n'extraïem d'aquest. Per tant, és molt important doncs acotar perfectament el disseny experimental del model per tal que, en cap cas, el nostre sistema hagi d'extrapolar.

Un altre factor que crea reticències al sector industrial, és el fet d'assegurar que la reproductibilitat del sistema es manté constant al llarg del temps. On en el cas d'aparició de derives o passivacions dels sensors, caldria assegurar-se de corregir-les d'alguna forma [63, 64]; intentant evitar així el reentrenament del sistema donat l'elevat nombre de mostres requerit.

1.4.3. Aproximacions i exemples

1.4.3.1. Llengües electròniques

D'acord amb la definició de la IUPAC [65], una llengua electrònica (ET) és “un sistema multisensor, el qual consisteix en un nombre de sensors amb baixa selectivitat i

que utilitza procediments matemàtics avançats per dur a terme el processament dels senyals, basats en reconeixement de patrons (PARC) i/o anàlisi de dades multivariants [xarxes neuronals artificials (ANNs), anàlisi de components principals (PCA), etc.]". Aquests sistemes van aparèixer posteriorment al desenvolupament del concepte de nas electrònic, inspirats en l'aplicació dels mateixos conceptes, però aquest cop en mostres líquides [66].

Per tant, les ETs són sistemes analítics aplicats a l'anàlisi de líquids formats per una matriu de sensors capaç de generar una resposta multidimensional, acoblada amb una eina quimiomètrica de processament capaç d'extreure informació significativa a partir d'aquesta, relacionant-la amb el seu significat analític [10]. A més, l'etapa de processament pot compensar qualsevol efecte matriu o interferència de la pròpia mostra, o permetre la resolució de les interferències, derives o no linealitat obtingudes amb els sensors [64]. Així, amb aquesta metodologia, és possible aconseguir una determinació en paral·lel d'un gran nombre d'espècies diferents, mentre que les possibles interferències són resoltes mitjançant la compensació de la seva presència [52].

En el disseny d'una ET, hi ha dos paràmetres clau que determinen la informació obtinguda, que són: l'eina quimiomètrica (§1.1.5) i el tipus de sensors utilitzats (§1.2 i §1.3). En aquest sentit, entre les diferents famílies de sensors químics i biosensors que poden formar la ET, els sensors potenciomètrics, voltamperomètrics i impedimètrics són les principals variants utilitzades [10, 67]. Aquesta classificació respon a la natura del senyal primari del dispositiu, el qual després d'una interacció amb l'analit respon ja sigui amb la generació d'un potencial, d'un corrent electroquímic o d'un canvi de resistència/capacitància.

En el cas de la variant potenciomètrica, aquesta pot contemplar xarxes de sensors utilitzant ISEs (*Ion-Selective Electrodes*) i/o ISFETs (*Ion-Sensitive Field-Effect Transistors*). D'altra banda, la variant voltamperomètrica (o amperomètrica) implica la utilització de sensors voltamperomètrics, podent generar la multidimensionalitat en la resposta de diferents maneres; per exemple, mitjançant un nombre d'elèctrodes polaritzats a diferents potencials fixos, o realitzant un escombrat voltamperomètric complet per cadascun dels sensors utilitzats. Finalment, aquesta variabilitat també es pot trobar en el cas dels sensors impedimètrics, donat que podem utilitzar diferents sensors en una freqüència de corrent altern (AC) fixa, o realitzar el registre de tot un espectre d'alta freqüència pel sensor o sensors utilitzats.

Fins ara, les ETs han proporcionat alternatives a molts mètodes clàssics,

especialment en el camp de l'anàlisi d'aliments [10, 38, 68]. Tal com ja s'ha mencionat, el tipus d'aplicació d'aquests sistemes pot ser tan qualitativa (e.g. identificació de determinades espècies, classificació de mostres, identificació d'episodis d'adulteració, etc.) com quantitativa (normalment, la multideterminació d'un conjunt d'espècies químiques o índexs).

D'altra banda, existeix també una tendència en l'aplicació d'aquestes com a eina de tast artificial [11], un concepte més bioinspirat, amb l'objectiu de dur a terme la percepció del gust de forma automatitzada; una aplicació amb especial interès en l'àmbit industrial. Aquest concepte va ser introduït pel professor Toko, i consisteix en el desenvolupament d'una matriu de sensors inspirats en les papil·les gustatives de la llengua i/o que responguin als tipus de sabors bàsics (dolç, àcid, salat i amarg; més el concepte oriental "d'*umami*", el qual significa deliciós) amb l'objectiu final d'imitar el sentit del gust humà per a la realització d'una avaluació anàloga en els casos en què aquesta no és una opció possible, com és en el control automàtic, control de mostres verinoses, etc.

Finalment, destacar l'aparició de les llengües bioelectròniques (BioET) com a nou concepte emergent dins d'aquest camp, el qual respon a la variant més recent de les ETs. Aquestes estan basades en la incorporació d'un o varis biosensors (§1.1.4) en la matriu de sensors que forma la llengua electrònica; és a dir, l'ús d'un element biològic com a element de reconeixement en lloc de diferents sensors basats en l'ús de diferents catalitzadors electroquímics com a modificadors, o diferents materials transductors.

La diferència entre ambdues respon simplement al tipus de sensors utilitzats, però amb un gran avantatge respecte les primeres, que és la seva resposta millorada. És a dir, per una banda tenim l'especificitat i selectivitat que ens aporten els biosensors; mentre que per l'altra, l'ús d'eines quimiomètriques tals com les ANNs que permeten extreure informació significativa a partir de dades complexes o inespecífiques. Així doncs, aquesta variant coneguda com a BioET, pot ser postulada com una eina que combina la quimiometria per tal de solucionar els problemes d'interferències dels biosensors, i els biosensors com l'eina que soluciona els problemes de selectivitat de les ETs.

1.4.3.2. *Nassos electrònics*

El concepte de nas electrònic (EN) va ser proposat el 1982, pels professors Persaud i Dodd, com un dispositiu que pretenia imitar la discriminació dels olors del sistema olfactiu dels mamífers [69]; essent així el primer dels sistemes biomimètics basats en

sensors que pretenia imitar algun dels sentits dels animals, i que va permetre el ràpid desenvolupament d'aplicacions interessants en diferents camps [70].

No obstant, cal tenir en compte que deguda a la gran ambició de la proposta i la complexitat que comporta, aquesta fita encara no ha estat avui assolida; derivant així el concepte inicial també cap a altres aplicacions no relacionades amb el casos biològics tal com ja s'ha discutit anteriorment.

Així doncs, els professors Gardner i Bartlett han definit el nas electrònic com “un instrument, que comprèn un conjunt de sensors químics electrònics amb especificitat parcial i un sistema de reconeixement de patrons apropiat, capaç de reconèixer olors simples o complexos” [71]. Cal denotar que aquesta definició només contempla el reconeixement d'olors, sense incloure la detecció o determinació de components d'una mescla gasosa; no considerant aquesta fins a la dècada dels 90s [71].

Dins la gran varietat de famílies de sensors que presenten característiques interessants pel desenvolupament de nassos electrònics trobem els semiconductors d'òxids metàl·lics, polímers conductors, sensors d'ona acústica, sensors piezoelèctrics, sensors d'efecte de camp, sensors electroquímics o sensors de fibra òptica [59].

1.4.3.3. *Ulls electrònics*

Per últim, destacar també l'aparició del concepte d'ull electrònic (EE); el qual, tot i diferir lleugerament dels dos anteriors, també es basat en l'ús d'una tècnica analítica per tal d'obtenir informació relativa a la mostra i el posterior processament d'aquesta mitjançant una eina quimiomètrica.

En aquest cas doncs, el desenvolupament de sistemes d'EE es basa sobretot en mesures espectromètriques, principalment en la zona de l'espectre visible (380-750 nm), tot i també poder fer ús d'altres zones de l'espectre electromagnètic; essent l'objectiu la captació del “color” de la mostra. D'aquesta manera, al igual que en el cas de les llengües electròniques en que podíem utilitzar com a font de multidimensionalitat el registre complet de tot un voltamperograma, en aquest cas podem realitzar un escombrat a diferents regions de l'espectre electromagnètic.

Encara més, és possible també la utilització d'un colorímetre basat en una càmera CCD (*charge-coupled device*, dispositiu de càrrega acoblada) que realitzi l'adquisició d'imatges en alta qualitat, adquirint així no tan sols informació relativa al “color” de la mostra, sinó també a la forma d'aquesta. Extraient així major informació relativa a la

mostra, la qual és processada de forma conjunta en l'etapa de processament, millorant la resposta del model generat.

1.4.3.4. Fusió de dades

Anant un pas més enllà, fins i tot és possible la combinació de les respostes obtingudes de famílies de sensors diferents, una estratègia coneguda com a fusió de dades (*data fusion*). En un principi, podríem pensar en aquesta com la combinació de dades provinents de sensors dins d'un mateix sistema, e.g combinació mesures provinents de sensors potenciomètrics i voltamteromètrics en el cas d'una llengua electrònica [72], encara que no existeix tal limitació [73].

En aquest sentit, existeixen també dins la bibliografia treballs d'integració de dos o més sistemes, mostrant per exemple una clara complementarietat entre els nassos i les llengües electròniques [74, 75]. Aquest fet no és d'estranyar doncs, si tenim en compte que es tracta de sistemes biomimètics en els quals pretenem imitar el sistema de percepció dels animals; en els que l'obtenció d'una resposta per part del cervell deriva de la combinació d'estímuls provinents dels diferents sentits.

Un bon exemple n'és el procés de tast ("cata") d'un vi, en el qual combinem la informació que ens aporta el sentit de la vista relatiu al seu aspecte (i.e. color, limpidesa, fluïdesa, efervescència), l'olfacte respecte la seva olor (aroma, *bouquet*, aroma de boca via retronassal) i gust relatiu al seu sabor (gust pròpiament dit, astringència, causticitat, bombolleig, consistència, fluïdesa, untuositat, temperatura) [76]. Justament és aquest procés el que ha inspirat l'aparició del que s'anomena el *panel sensorial electrònic*, el qual es basa en la combinació de la resposta obtinguda a partir d'un nas electrònic, una llengua electrònica i un ull electrònic [73, 77].

A pesar de l'evidència conceptual de l'aproximació, s'ha tenir en compte la complexitat d'aquesta; doncs degut a la natura diferenciada de les diferents respostes provinents de cada sensor, cal tenir en compte la importància relativa de cada variable i de cada tècnica.

La fusió de dades és doncs una estratègia que combina i analitza les dades provinents de múltiples fonts amb el fi d'aprofitar les seves característiques i millorar la representació de la informació per construir un model de predicció o decisió [78]; per tant, és una eina que permet compatibilitzar les mesures de natura diferenciada. Malauradament, la recerca d'aquesta en química es troba encara en una etapa de

desenvolupament, tot i que ja hi ha alguns treballs que en descriuen els seus beneficis en aquest àmbit [79]. Així, depenent de l'etapa de processament en què té lloc la fusió de dades, aquesta és classificada sovint en els següents nivells [80]:

- baix (*low-level fusion*): concatenació de les matrius de resposta. Les dades obtingudes mitjançant diferents fonts (i.e. les diferents famílies de sensors o tècniques) són combinades en brut per obtenir un nou conjunt de dades, el qual serà considerat com un nou conjunt únic de respostes. E.g. [81]
- intermedi (*mid-level fusion*): concatenació de les matrius de resposta després del seu preprocessament. Les dades obtingudes de les diferents fonts són preprocessades per separat i combinades posteriorment. E.g. [77]
- alt (*high-level fusion*): combinació de les sortides de diferents models. Les dades obtingudes són processades per separat, i les prediccions de dos o més models són llavors combinades per generar una resposta "fusionada". E.g. [82]

1.4.3.5. Xarxes Neuronals Artificials

Una xarxa neuronal artificial (ANN) és un model matemàtic inspirat en el funcionament del sistema nerviós i del cervell dels animals, utilitzat com a eina quimiomètrica en l'anàlisi multivariant.

Aquestes no responen a un model algebraic explícit, sinó que consisteixen en un conjunt d'entitats lògiques artificials interconnectades anomenades neurones, processant la informació mitjançant la modificació d'aquestes connexions; mecanisme que pretén imitar el funcionament del sistema nerviós [83]. Així doncs, es tracta d'un sistema adaptatiu que varia la seva estructura durant una fase d'aprenentatge, i que s'utilitza per modelar relacions complexes entre entrades i sortides, o per trobar patrons en les dades.

Per tant, una definició força real del que són les ANNs és: "programes informàtics basats en un model simplificat del cervell, que no pretenen copiar els detalls fins del seu funcionament, però que procuren reproduir el seu funcionament lògic utilitzant unes unitats de processament semblants a les neurones" [84].

1.4.3.5.1. Fonament biològic

Tal com s'ha dit, les ANNs són també un sistema biomimètic, que pretén imitar el sistema nerviós dels animals. Aquest és l'encarregat de recollir la informació procedent

dels sentits, emmagatzemar-la, processar-la i/o transmetre-la a altres parts encarregades d'interpretar-la per dur a terme alguna funció. Aquesta etapa de processament té lloc en el cervell, el funcionament del qual està basat en les unitats fonamentals que el formen: les neurones (Figura 1.14)

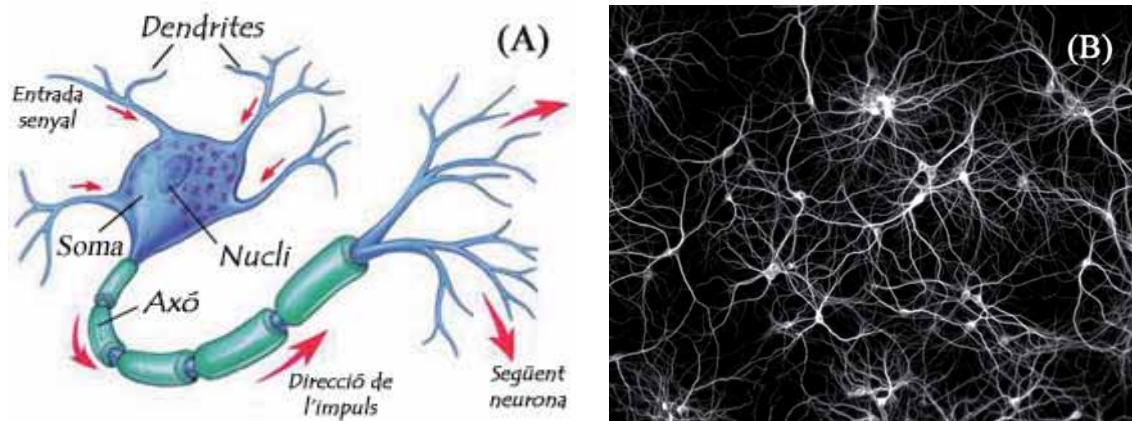


Figura 1.14. (A) Infografia de les parts d'una neurona biològica i (B) xarxa neuronal biològica.

Una neurona és una cèl·lula del teixit nerviós, que es troba formada per les següents parts: soma (o cos cel·lular), dendrites i axó (Figura 1.14A), i que es troben a l'encèfal, la medul·la espinal i els ganglis nerviosos [85].

La informació arriba a la neurona gràcies a uns canals anomenats dendrites fins al cos cel·lular (soma) on és processen els senyals rebuts per tal de generar una resposta de sortida, que es distribueix a través de l'axó a altres cèl·lules. L'axó es ramifica per propagar la informació de sortida a diferents neurones i s'uneix a les dendrites d'aquestes neurones mitjançant una unió anomenada *sinapsi* (Figura 1.14B).

Els impulsos elèctrics del cervell, generats pels estímuls rebuts per tots i cadascuns dels sensors repartits pel cos, flueixen constantment al llarg de tot l'entramat de neurones; els quals són processats per aquestes; processament que comporta la combinació de tots els senyals rebuts per les dendrites. Si aquest supera un valor límit, es dona l'activació de la neurona i aquesta envia un senyal de sortida a través del seu axó, retornant al seu estat inicial; propagant d'aquesta manera la informació a través del teixit neuronal.

Gràcies a l'ajust de la conductivitat de les diferents sinapsis, que fa que els senyals es transmetin en menor o major intensitat, s'aconsegueix que la xarxa sigui capaç d'aprendre la resposta adequada a una sèrie d'estímuls. Si un estímulo es reproduïx de forma continuada, aconseguirem una resposta apropiada de forma automàtica. És

important destacar la independència del tipus d'impuls rebut, essent simplement rellevant la intensitat en que aquest es presenta.

El coneixement s'associa doncs a la repetició dels estímuls al llarg d'unes vies preferents, i són aquestes connexions les que al llarg del temps es reforcen. Aquest processament tan simple adquireix la categoria de processament intel·ligent quan la xarxa neuronal és prou gran. En aquest sentit, el cervell humà s'estima que conté ca. 10^{11} neurones, cadascuna de les quals té una mitjana de 7000 connexions sinàptiques amb altres neurones. Així doncs, es calcula que el cervell d'un nen de 3 anys té ca. 10^{15} sinapsis; nombre que disminueix amb l'edat, fins a estabilitzar-se en l'edat adulta entre 10^{14} i $5 \cdot 10^{14}$.

La imitació del sistema biològic descrit, ha donat lloc a les xarxes neuronals artificials com a sistema empleat en intel·ligència artificial [86]. Aquestes últimes es basen en la combinació de diferents neurones, les quals reben una sèrie d'entrades a través de les interconnexions d'aquestes i emeten una sortida. Aquesta sortida ve determinada per tres funcions diferents: una funció de propagació, que generalment consisteix en el sumatori de cada entrada multiplicada pel pes de la seva interconnexió; una funció d'activació, que modifica l'anterior i que pot inclús no existir, essent en aquest cas la sortida la pròpia funció de propagació; i una funció de transferència, que s'aplica al valor retornat per la funció d'activació. Aquesta s'utilitza per acotar la sortida de la neurona i generalment ve donada per la interpretació que volem donar-li a l'esmentada sortida.

1.4.3.5.2. *El perceptró*

Les ANNs tenen com a unitat bàsica la *neurona artificial* o *perceptró*; el qual és doncs l'element fonamental de processament computacional que s'assimila a les neurones [55, 83]. Aquest és una unitat de presa de decisió (Figura 1.15), format per un conjunt d'entrades (x_i) i generalment una única sortida (a_k).

Al igual que en la neurona biològica, aquest rep els diferents senyals i procedeix a fer la suma ponderada dels senyals rebuts, mitjançant l'aplicació d'un valor d'importància anomenat *pes* (w_{ki}) i la suma també d'un valor b corresponent al *biaix* que dona com a resultat un cert valor y . Igual que en la neurona biològica, si aquest valor y no supera un cert valor llindar (t_{ki}), el perceptró roman inactiu; mentre que si el supera, aquest envia una senyal de sortida (a_k , resultat d'aplicar una funció de transferència (F))

al valor y) i retorna a l'estat inicial. Evidentment, les respostes cerebrals no són de tipus binari, sinó que modulen la sortida en funció de les entrades. És per això que els perceptrons utilitzen diferents funcions de transferència per tal d'imitar aquesta modulació, i oferir senyals de sortida no només de tipus binari.

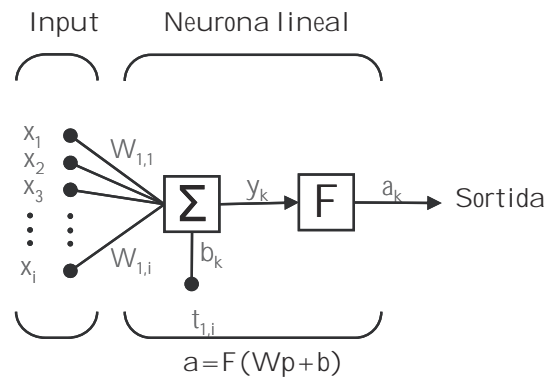


Figura 1.15. Esquema del perceptró o neurona artificial.

Gracies a l'ús de funcions de transferència contínues podem fer que les neurones responguin amb una sortida gradual i no limitar-les a la resposta binària d'uns i zeros. En el fons, qualsevol tipus de funció matemàtica podria aplicar-se com a funció de transferència, sempre que compleixi les següents condicions mínimes que són: ser monòtona creixent acotada i tenir una derivada continua en l'interval d'aplicació. Tot i així, existeixen un conjunt de funcions de transferència que són les que més s'utilitzen habitualment en la construcció de models amb ANNs, a saber: *tansig*, *logsig*, *purelin*, *satlin* i *satlins* (Figura 3.8).

Per tant, el procés d'aprenentatge per part del perceptró consistirà en l'ajust dels valors dels pesos aplicats a cada senyal i el biaix, el valor llindar que ha de superar per transferir la informació i la funció de transferència adequada per tal que aquest arribi a produir l'estímul desitjat. Aquests paràmetres són ajustats de forma que minimitzin una funció d'error, la qual compara els valors obtinguts a la sortida amb els valors esperats per una sèrie de patrons coneguts. Aquests valors s'emmagatzemen llavors en una matriu de pesos, ja que seran els que posteriorment utilitzarà el perceptró per tal de processar nova informació.

1.4.3.5.3. Perceptró multicapa o xarxa multicapa

Al igual que en el mecanisme biològic i basant-nos en l'estructura multicapa del cervell humà, per tal de fer front a les limitacions i incapacitat del perceptró per resoldre

certs tipus de problemes relativament senzills, es va recórrer a la connexió de diverses capes successives de perceptrons [54, 55]. Aquesta nova estratègia va donar lloc al naixement de les ANNs, en les quals un elevat número de perceptrons es col·loquen en capes que propaguen la informació al llarg de l'entramat de perceptrons (Figura 1.12).

A més, donada la disposició de les connexions sempre en un mateix sentit i entre capes diferents, aquest tipus de xarxes s'anomenen també xarxes unidireccionals (*feed-forward*). Concretament, aquest model multicapa es troba format per tres tipus diferents de capes:

- Capa d'entrada: és la primera de les capes, és l'encarregada d'introduir la informació a l'interior de la xarxa i conté tantes neurones com *inputs*. Aquesta rep el conjunt d'entrades (x_i) i s'encarrega de distribuir-les a la següent capa de perceptrons sense alterar-les.
- Capa oculta: habitualment consistent en una única capa ja que aquesta sol ser suficient, tot i que també és possible utilitzar dos o més capes de neurones connectades successivament en aquesta. Conté un nombre variable de perceptrons, i és en aquesta en les que té lloc el processament de la informació.
- Capa de sortida: rep els senyals processats en les capes anteriors i és la que proporciona la informació útil a l'usuari.

D'altra banda, al igual que en el cas del perceptró, durant la generació del model mitjançant la xarxa neuronal, cal cercar la millor combinació dels pesos de les connexions sinàptiques mitjançant iteracions successives amb les dades d'una sèrie de patrons destinats a l'entrenament d'aquesta; així com també d'aplicar una funció de transferència.

1.4.3.5.4. El procés d'aprenentatge

La generació del model neuronal requereix però d'un procés d'aprenentatge, al igual que en el cas biològic, amb independència de la complexitat d'aquest; és a dir, tant si treballem amb un únic perceptró com si tenim una xarxa multicapa. Durant aquest, té lloc la cerca de la millor combinació de pesos de les connexions sinàptiques mitjançant interaccions successives amb les dades d'una sèrie de patrons destinats a l'entrenament de la xarxa. Un cop trobat el model òptim, aquest es pot emprar per a la predicció de mostres desconegudes.

Per tant, seguint amb l'analogia amb el procés biològic, al igual que qualsevol nen,

la xarxa neuronal s'ha de sotmetre a un procés d'ensenyament en el que realitza l'aprenentatge, facilitant a aquesta tant el conjunt de paràmetres corresponents a les entrades com a les sortides, les quals utilitzarà per construir el model. Per posteriorment ser avaluada, entregant només els paràmetres corresponents a les entrades i avaluant la seva capacitat per “predir” correctament les sortides.

i. ANNs amb retropropagació (back-propagation)

Aquest procés d'entrenament és necessari ja que la xarxa (i.e. cada perceptró) inicialment es genera amb total ignorància respecte el procés que ha d'aprendre; a saber, valors dels pesos inicials definits aleatòriament. Per tant, durant el procés d'aprenentatge cal ajustar aquests per tal que la sortida sigui igual al valor esperat, ja que evidentment és molt poc probable que això passi amb els valors aleatoris fixats inicialment.

Es tracta doncs d'un procés iteratiu, en que es van variant aquests valors de forma que cada cop els valors de les sortides s'assemblin més als valors reals [87]. Així, en cada iteració, es va comparant el resultat predit per la xarxa amb el resultat esperat, reajustant els pesos en cada iteració per tal de millorar la resposta en la sortida, fins a assolir que els valors de sortida de la xarxa siguin el més semblants possibles a la sortida desitjada.

Cal tenir en compte que el procés d'aprenentatge en les xarxes neuronals és més complex que en el cas del perceptró, ja que s'ha d'avaluar com els canvis entre els pesos sinàptics d'una capa fan que variïn els de la següent (xarxes de propagació unidireccional, §1.4.3.5.3). Això s'aconsegueix gràcies a algorismes d'entrenament supervisats que controlen que la modificació dels pesos condueixi a la minimització de l'error en la capa de sortida; essent els algorismes de propagació unidireccional amb aprenentatge per retropropagació d'errors (*back-propagation*) els més emprats [54].

Amb aquest procediment, es calcula l'error com la semisuma de la diferència al quadrat entre la sortida esperada (s_{nj}) i l'obtinguda (a_{nj}) per cada neurona (n) i per cada patró (p) (Equació 1.1). Llavors, s'utilitza aquest error per tal d'efectuar variacions en els pesos tals que en minimitzin el seu valor. Cada cicle complet de propagació a través de la xarxa i la conseqüent reassignació de pesos és conegut com a *època*.

$$E_p = \frac{1}{2} \sum_j^n (s_{nj} - a_{nj})^2 \quad \text{Equació 1.1}$$

El procés d'aprenentatge tendirà doncs a aconseguir minimitzar l'error, mitjançant un procés iteratiu on s'assignen els pesos de manera que l'error disminueixi el més ràpidament possible. No obstant, existeixen diferents algorismes d'entrenament en funció de com té lloc la modificació dels pesos en cada època durant la construcció del model; essent així, la retropropagació amb descens de gradient [88] o amb l'algorisme Levenberg-Marquard [89], dos dels més comuns. El primer es basa en la modificació dels pesos en la capa de sortida i la retropropagació d'aquest ajust a la capa anterior, i així successivament fins a arribar a la capa d'entrada; mentre que el segon utilitza optimitzacions numèriques de nivell avançat que ofereixen aprenentatges molt més acurats de forma més ràpida i eficient, on s'apliquen optimitzacions de les matrius d'error a cadascuna de les èpoques durant el procés d'entrenament.

ii. Sobreentrenament (overfitting)

A l'hora d'ajustar un model cal tenir en compte els avantatges i inconvenients de les eines que estem utilitzant. A saber, si bé és cert que estem treballant amb eines de càlcul molt potents, capaces de resoldre situacions i problemes molt complexos, també és cert que aquesta situació pot donar lloc a un sobreajust del model generat; situació coneguda com a *sobreentrenament (overfitting)*.

Aquest fenomen té lloc quan enlloc de construir una xarxa que modeli de forma generalitzada el nostre problema, generem un model sobreajustat. És a dir, la reducció de l'error en l'etapa d'aprenentatge ha tingut lloc de tal forma que s'assoleix un error associat al subconjunt de mostres d'entrenament tan petit, que quan s'introdueix a la xarxa el conjunt de mesures pertinents a una mostra que no ha format part del subconjunt d'entrenament s'obté un error associat molt gran.

Tanmateix, fent el símil una altra vegada, podríem parlar que la xarxa enlloc d'estar aprenent, està memoritzant la sortida que pertany a cada conjunt d'entrades. Per tant, no sent capaç de reproduir la tasca que se li ha intentat ensenyar.

Per tal d'il·lustrar aquest concepte, la Figura 1.16 mostra com exemple un cas de sobreajust d'un conjunt de dades resultat d'augmentar el grau polinòmic de la funció utilitzada. Així doncs, es veu clarament com en augmentar aquest, l'ajust per les mostres d'entrenament millora, però a expenses de la resta de mostres. És a dir, el model sobreentrenat presentarà errors molt petits pel subconjunt d'entrenament, mentre que els errors pel subconjunt de test seran enormes ja que aquestes mostres no han intervingut durant l'ajust del model.

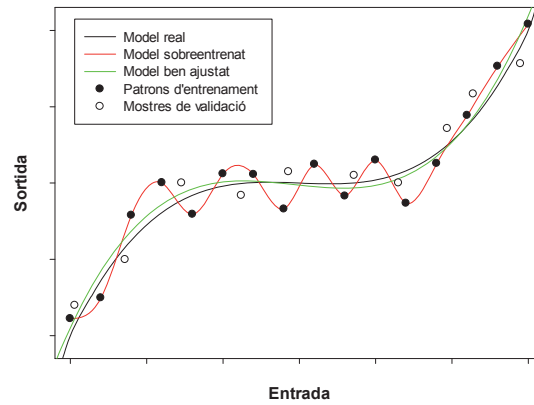


Figura 1.16. Representació esquemàtica d'un model sobreentrenat i un model ben ajustat.

Per tant, queda demostrada la importància d'utilitzar un conjunt de validació per tal d'evitar situacions de sobreentrenament. Amb aquesta finalitat, podem fer ús de dos estratègies diferents:

- *Aturada en el temps* [90]: consisteix en utilitzar un tercer subconjunt de mostres per a controlar l'ajust del model per aquest, aturant el procés d'aprenentatge en cas d'aparició de sobreentrenament per part de model. Així doncs, en aquest cas el conjunt de mostres inicial el dividirem en tres subconjunts diferents: entrenament (utilitzat per a generar el model), validació interna (utilitzat per mesurar el sobreentrenament durant el procés d'aprenentatge) i validació externa o test (utilitzat per avaluar la capacitat de predicció del model un cop generat).

Durant el procés d'aprenentatge, al llarg de cada època, els pesos de la xarxa són ajustats amb el subconjunt d'entrenament tal com ja s'ha explicat; calculant a la vegada també l'error pel conjunt de validació interna. Així doncs, inicialment, a mesura que el model va aprenent, els errors per ambdós subconjunts decauen i el procés d'aprenentatge continua; per contra, si el model tendeix a una situació de sobreentrenament, aquesta és detectada gràcies a la disminució en l'error pel subconjunt d'entrenament i l'augment d'aquest pel subconjunt de validació interna ja que aquest no intervé en el procés d'aprenentatge. Aturant així el procés si això succeeix per un nombre successiu d'èpoques, i retornant els valors dels pesos d'interaccions als valors anteriors. Finalment, al igual que abans, la capacitat de predicció del model és avaluada amb el subconjunt de test un cop ajustat el model.

- *Regularització Bayesiana* [91, 92]: presenta com a principal avantatge la cerca de la xarxa neuronal més simple que ajusti millor el sistema a modelar, oferint

millors resultats per a les mostres que no han participat en el procés d'entrenament sense necessitat de fer ús d'un conjunt de validació interna. En aquest sentit, aquest algorisme no tant sols minimitza l'error global de la xarxa, tal com feia l'algorisme del descens del gradient, sinó que també considera la importància dels pesos, intentant minimitzar-los de manera que el model generat esdevingui menys dependent d'alguns pesos en concret.

Obtenint així models més simples, amb una xarxa "suavitzada", que aconsegueix evitar el sobreentrenament sense necessitat d'un conjunt de validació interna, presentant generalment resultats molt més acurats.

1.4.3.6. Aplicació

Tot i que l'eclosió de la majoria d'aquests dispositius ha tingut lloc al llarg de les darreres dos dècades i que com a tals encara es troben en un estadi bastant primari i en procés de desenvolupament, cal destacar el fet que ja és possible trobar al mercat dispositius complets venuts sota el nom de nassos, llengües o ulls electrònics. Tot i així, cal tenir present que aquest tipus d'instrumentació és encara incapaç de competir amb la instrumentació pesada emprada en els laboratoris, relegant aquests dispositius com a instrument d'anàlisi complementari, especialment indicat com eina de cribratge (*screening analysis*).

i. Dispositius comercials

Tot i el menor mercat que han assolit les llengües electròniques, cal destacar la gran variabilitat de dispositius presents en el cas dels nassos electrònics, en els quals podem comptar més d'una vintena d'empreses amb aquest tipus de dispositius en el mercat [59]. Alguns exemples serien els equips comercialitzats per la marca *Alpha Mos* (França), la qual disposa de quatre nassos, una llengua i un ull electrònic en el seu catàleg; assolint inclús la monitorització *online* en temps real en el cas d'algun dels dispositius. Justament aquesta podria ser considerada l'empresa més coneguda i amb un major assortit d'aquests dispositius. O les empreses *Insent (Intelligent Sensor Technology Inc.)*, (Japó) i *McScience* (Corea), les quals comercialitzen ambdues diferents llengües electròniques. En aquest aspecte cal destacar la primera ja que aquesta porta des del 1993 comercialitzat llengües electròniques. A pesar de que només disposa generalment d'una única llengua electrònica en el seu catàleg, aquesta és el resultat de

l'evolució de les versions anteriors: tres dispositius descatalogats actualment i que han donat lloc a la quarta versió disponible des del 2007.

ii. *Camp d'aplicació i exemples*

Donada l'ambigüitat en la definició d'aquests dispositius (a saber i de forma resumida, la unió d'un conjunt de sensors amb una resposta diferenciada i una eina quimiomètrica capaç d'extreure informació significativa d'aquests), ens trobem davant d'un sistema que presenta un gran ventall de possibilitats de del punt de vista de la seva arquitectura. No obstant, tots amb l'objectiu comú d'obtenir una eina d'anàlisi ràpida i poderosa que permeti la resolució de problemes complexos en el camp analític.

En aquest sentit, aquests dispositius han estat àmpliament aplicats a nombroses situacions de naturalesa diferenciada [10, 38, 59, 68, 93-95], però amb especial atenció al sector alimentari i de les begudes. A continuació es pretén il·lustrar breument aquest fet amb alguns exemples d'aplicacions descrites a la bibliografia segons el sector:

- *Anàlisi agroalimentaria*: control de qualitat durant la producció i emmagatzematge de diferents productes tals com aigua, vi, cafè, llet, etc.; detecció de frauds; mesura de la frescor o deteriorament d'aliments; optimització de bioreactors; control del procés d'envelliment; tast artificial automàtic; etc.
- *Medicina*: diagnòstic no invasiu (orina, suor, etc.); monitoratge clínic *in vivo*; anàlisi de compostos farmacèutics; etc.
- *Seguretat*: detecció d'explosius i/o armes biològiques o químiques; detecció de drogues; detecció de *GSR* (*gunshot residues*); etc.
- *Monitoratge ambiental*: control d'indicadors mediambientals (tant en aire com aigua); control d'agents contaminants (metalls pesants, detergents, etc.); identificació de substàncies tòxiques; etc.
- *Cosmètica*: anàlisi de fragàncies tant en el control com en el desenvolupament de perfums i colònies; etc.
- *Indústria química*: puresa dels productes; detecció grups funcionals; control d'embalatges i paquets; etc.
- *Tast artificial*: aportar des d'un punt de vista objectiu una eina capaç de proporcionar una resposta equivalent als sentits en que es troben inspirades.

Dins de les diferents aplicacions reportades a la bibliografia, podem distingir aquestes en dos grans blocs en funció del tipus d'informació que volem extreure: anàlisi

qualitativa o quantitativa; generalment implicant un major grau de complexitat en el darrer cas. No obstant, en ambdós casos, cal generar un model de resposta a partir d'una etapa d'entrenament amb mostres estàndard, les quals han de representar la variabilitat de les concentracions de les espècies d'interès i els possibles interferents amb que es pot trobar l'eina en l'aplicació final. Posteriorment, un cop generat el model aquest és aplicat per a la predicció de les variables per les que el model ha estat entrenat, obtenint així una poderosa eina d'anàlisi amb gran aplicabilitat en camps molt diferenciats.

iii. Anàlisi qualitativa

L'objectiu en aquests casos és la classificació, discriminació o identificació d'una mostra o conjunt d'aquestes; essent el camp d'aplicació majoritari dels diferents sistemes biomimètics descrits anteriorment. Aquesta predilecció o major ús, pot ser deguda per una banda a la menor dificultat requerida generalment en aquests casos, com també al fet que la pròpia naturalesa d'aquests sistemes els converteix en eines idònies per assolir aquestes tasques (§1.4.2).

Amb aquesta finalitat, l'eina quimiomètrica més utilitzada és el PCA (§1.3.3.1), el qual permet la representació del conjunt de mostres analitzat en un nou espai bi- o tri-dimensional, visualitzant d'una forma més simple les relacions (i.e. similituds o discrepàncies) entre les diferents mostres. No obstant, cal recordar que és tracta només d'una eina de visualització, requerint d'una eina de modelatge per poder actuar com a classificador.

iv. Anàlisi quantitativa

D'altra banda, i encara menys freqüent que l'anàlisi qualitativa, l'ús d'aquests sistemes està irrompent amb força també com a eina d'anàlisi quantitativa per assolir la determinació simultània de diferents substàncies o paràmetres d'interès en mostres complexes. El seus avantatges deriven de l'ús de sensors amb resposta entrecruada front els diferents analits, combinats amb poderoses eines quimiomètriques capaces d'extreure informació rellevant d'aquesta, sobrepasant problemes tals com el solapament de la resposta o la presència d'interferents. Amb aquest fi, les eines quimiomètriques més utilitzades són les ANNs (§1.3.4.5) i el PLS (§1.3.4.3).

1.5. Camp d'aplicació del present treball

El nexa d'unió entre totes les aplicacions desenvolupades a la present tesi ha estat la resolució de problemes en el sector vinícola i de les begudes alcohòliques, mitjançant l'ús de les llengües electròniques, tant per l'extracció de informació qualitativa i la identificació de classes com per la quantificació de paràmetres d'interès analític.

Així doncs, en cada cas cal tenir en compte les diferents necessitats corresponents a cada sector a l'hora de desenvolupar els diferents sistemes, així com les característiques de les mostres des del punt de vista tan físic com químic. És a dir, el fet que des del punt de vista analític, es tracta de mostres reals complexes formades per la barreja d'una gran varietat de substàncies, que influeixen a les diferents propietats organolèptiques i sensorials d'aquestes; les propietats de les quals, a més, evolucionen en el temps. D'altra banda, factors com l'elevat contingut alcohòlic, la presència de CO₂ que es desprèn durant la mesura o la manca de coneixement dels compostos responsables de certes característiques, entre altres propietats de les mostres són les que també en dificulten la seva anàlisi.

Precisament, és per aquest motiu que la classificació de la majoria de mostres en aquests sectors es realitza a través d'un panel sensorial expert (panel de tast) [76], donades les dificultats per dur a terme aquesta avaluació utilitzant tècniques clàssiques d'anàlisi. Això es principalment degut a que els principals compostos responsables de les diferents classes no han estat encara identificats, fet que al seu torn dificulta l'ús dels mètodes clàssics, l'ús dels quals requeriria identificar els compostos responsables d'aquestes sensacions. Per contra, les ETs combinen tota la informació química extreta de la matriu de sensors per generar un model de forma anàloga al procés dut a terme en el cervell humà, evitant focalitzar-se en quins compostos específics són els responsables dels diferents sabors o gustos, però tot i així extraient una empremta digital específica, que en permet també la seva identificació o quantificació.

En aquest sentit, un cop hem entrenat correctament la ET a partir de les notes de tast del panel sensorial, aquesta pot ser utilitzada per tal de reproduir el funcionament d'aquest darrer. Per tant, permetent l'obtenció d'un mètode automatitzat que permetria reduir significativament la degustació de les mostres per part del panel, tot i no arribant a substituir-lo, però si augmentant el nombre de mostres que es poden analitzar o la freqüència de mostreig. D'aquesta manera, les ETs poden proporcionar una eina útil per ser utilitzada com a mètode de cribratge (*screening analysis*) per tal de detectar

qualsevol adulteració o contaminació durant el procés de producció o garantir que les mostres compleixen amb els estàndards de qualitat.

D'aquesta manera, les ETs representen una alternativa als mètodes tradicionals d'anàlisi per la caracterització a nivell qualitatiu de les mostres, amb avantatges sobre aquests com poden ser el seu baix temps de resposta o baix cost, essent especialment adequats per l'anàlisi de cribratge (*screening analysis*). A més, degut al seu perfil biomimètic, presenten una solució senzilla i directa a l'absència de coneixement sobre quins compostos són els principals responsables de l'estil de les mostres, superant els problemes trobats amb les tècniques analítiques convencionals com ara la cromatografia i l'espectroscòpia, l'aplicació de les quals requereix de la identificació dels compostos concrets.

Un altre punt d'unió entre els diferents tipus de mostres analitzades en la present tesi ha estat l'avaluació del seu contingut en compostos fenòlics o polifenols [96]. L'interès per la determinació d'aquest tipus de compostos ha incrementat en els darrers anys degut a les seues efectes beneficiosos per la salut, derivats de les seves propietats com a antioxidants naturals; proporcionant així un efecte protector contra les patologies degudes a l'edat tals com les malalties cardiovasculars o mutacions cancerígenes.

No obstant, cal tenir en compte que aquests dos termes no són sinònims, doncs no tots els compostos fenòlics són polifenols. La diferència entre ambdós radica en la seva "grandària", estan els darrers formats per la presència de nombroses unitats fenol en la seva estructura, mentre que els primers són molècules més petites que els anteriors. Malauradament, i degut a la inespecificitat derivada del mètode de referència utilitzat (mètode Folin-Ciocalteu) [97], el qual únicament proporciona una mesura del contingut total del conjunt d'aquests compostos (estrictament, el que proporciona és una mesura de la capacitat antioxidant de la mostra), és molt comú troba a la literatura l'ús indiscriminat d'aquests dos termes. Per tant, caldrà tenir en compte que quan es parla del contingut total, es fa referència al conjunt tant de polifenols com de compostos fenòlics d'estructura més simple; i que en molts casos, el terme polifenol és usat erròniament per la comunitat analítica i pel sector agroalimentari. Així, mantenint la terminologia utilitzada dins el sector, s'ha fet ús d'ambdós termes, tot i ser concients que no és la forma estrictament correcta; procurant fer ús del terme compostos fenòlics en el cas que estrictament es tracta de molècules senzilles i mantenint el terme polifenols en el cas del valor global.

1.6. Bibliografia

- [1] A. Hulanicki, S. Glab, F. Ingman, Chemical sensors definitions and classification, *Pure Appl. Chem.* **1991**, *63*, 1247.
- [2] J. Janata, A. Bezech, Chemical sensors, *Anal. Chem.* **1988**, *60*, 62R.
- [3] L. Senesac, T.G. Thundat, Nanosensors for trace explosive detection, *Mater. Today* **2008**, *11*, 28.
- [4] D.R. Thévenot, K. Toth, R.A. Durst, G.S. Wilson, Electrochemical biosensors: Recommended definitions and classification - (Technical Report), *Pure Appl. Chem.* **1999**, *71*, 2333.
- [5] S. Pérez, E. Fàbregas, Amperometric bienzymatic biosensor for l-lactate analysis in wine and beer samples, *Analyst* **2012**, *137*, 3854.
- [6] A.P.F. Turner, I. Karube, G.S. Wilson, Biosensors: fundamentals and applications, Oxford University Press, Oxford, **1987**.
- [7] S. Alegret, M. del Valle, A. Merkoçi, Sensores electroquímicos: introducció a los quimiosensores y biosensores, Universitat Autònoma de Barcelona, Bellaterra, **2004**.
- [8] F. Céspedes, E. Martínez-Fàbregas, S. Alegret, New materials for electrochemical sensing I. Rigid conducting composites, *TrAC-Trend. Anal. Chem.* **1996**, *15*, 296.
- [9] J. Workman, K.E. Creasy, S. Doherty, L. Bond, M. Koch, A. Ullman, D.J. Veltkamp, Process Analytical Chemistry, *Anal. Chem.* **2001**, *73*, 2705.
- [10] M. del Valle, Electronic Tongues Employing Electrochemical Sensors, *Electroanal.* **2010**, *22*, 1539.
- [11] K. Toko, A taste sensor, *Meas. Sci. Technol.* **1998**, *9*, 1919.
- [12] D.A. Skoog, F.J. Holler, T. Nieman, Principios de análisis instrumental, 5ª ed., McGraw-Hill, Madrid, Spain, **2001**.
- [13] A.J. Bard, L.R. Faulkner, Electrochemical methods: fundamentals and applications, Wiley, New York, US, **2001**.
- [14] V. Pravdová, M. Pravda, G.G. Guilbault, Role of chemometrics for electrochemical sensors, *Anal. Lett.* **2002**, *35*, 2389.
- [15] M. Esteban, C. Ariño, J.M. Díaz-Cruz, Chemometrics for the analysis of voltammetric data, *TrAC-Trend. Anal. Chem.* **2006**, *25*, 86.
- [16] Y. Ni, S. Kokot, Does chemometrics enhance the performance of electroanalysis?, *Anal. Chim. Acta* **2008**, *626*, 130.
- [17] S.D. Brown, S.T. Sum, F. Despagne, B.K. Lavine, Chemometrics, *Anal. Chem.* **1996**, *68*, 21.
- [18] E. Richards, C. Bessant, S. Saini, Multivariate Data Analysis in Electroanalytical Chemistry, *Electroanal.* **2002**, *14*, 1533.
- [19] S.M. Scott, D. James, Z. Ali, Data analysis for electronic nose systems, *Microchim. Acta* **2006**, *156*, 183.
- [20] Z. Boger, Selection of quasi-optimal inputs in chemometrics modeling by artificial neural network analysis, *Anal. Chim. Acta* **2003**, *490*, 31.
- [21] R. Cartas, A. Mimendia, A. Legin, M. del Valle, Multiway Processing of Data Generated with a Potentiometric Electronic Tongue in a SIA System, *Electroanal.* **2011**, *23*, 953.
- [22] J.M. Palacios-Santander, L.M. Cubillana-Aguilera, M. Cocchi, A. Ulrici, I. Naranjo-Rodríguez, R. Seeber, J.L. Hidalgo-Hidalgo de Cisneros, Multicomponent analysis in the wavelet domain of highly overlapped electrochemical signals: Resolution of quaternary mixtures of chlorophenols using a peg-modified Sonogel-Carbon electrode, *Chemometr. Intell. Lab.* **2008**, *91*, 110.

- [23] X. Cetó, J.M. Gutiérrez, M. Gutiérrez, F. Céspedes, J. Capdevila, S. Mínguez, C. Jiménez-Jorquera, M. del Valle, Determination of total polyphenol index in wines employing a voltammetric electronic tongue, *Anal. Chim. Acta* **2012**, 732, 172.
- [24] X. Cetó, F. Céspedes, M.I. Pividori, J.M. Gutiérrez, M. del Valle, Resolution of phenolic antioxidant mixtures employing a voltammetric bio-electronic tongue, *Analyst* **2012**, 137, 349.
- [25] R.M. de Carvalho, C. Mello, L.T. Kubota, Simultaneous determination of phenol isomers in binary mixtures by differential pulse voltammetry using carbon fibre electrode and neural network with pruning as a multivariate calibration tool, *Anal. Chim. Acta* **2000**, 420, 109.
- [26] J. Simons, M. Bos, W.E. van der Linden, Data processing for amperometric signals, *Analyst* **1995**, 120, 1009.
- [27] R. Gutiérrez-Osuna, H.T. Nagle, A method for evaluating data-preprocessing techniques for odour classification with an array of gas sensors, *IEEE T. Syst. Man Cy. B* **1999**, 29, 626.
- [28] X. Cetó, F. Céspedes, M. del Valle, Assessment of Individual Polyphenol Content in Beer by Means of a Voltammetric BioElectronic Tongue, *Electroanal.* **2012**, 25, 68.
- [29] X. Cetó, F. Céspedes, M. del Valle, BioElectronic Tongue for the quantification of total polyphenol content in wine, *Talanta* **2012**, 99, 544.
- [30] L. Moreno-Barón, R. Cartas, A. Merkoçi, S. Alegret, M. del Valle, L. Leija, P.R. Hernandez, R. Muñoz, Application of the wavelet transform coupled with artificial neural networks for quantification purposes in a voltammetric electronic tongue, *Sensor Actuat. B-Chem.* **2006**, 113, 487.
- [31] R. Bro, Review on multiway analysis in chemistry - 2000-2005, *Crit. Rev. Anal. Chem.* **2006**, 36, 279.
- [32] D. Ebrahimi, E. Chow, J.J. Gooding, D.B. Hibbert, Multi-analyte sensing: a chemometrics approach to understanding the merits of electrode arrays versus single electrodes, *Analyst* **2008**, 133, 1090.
- [33] I.T. Jolliffe, *Principal Component Analysis*, Springer New York, New York, US, **2002**.
- [34] P. Mitteroecker, F. Bookstein, Linear Discrimination, Ordination, and the Visualization of Selection Gradients in Modern Morphometrics, *Evol. Biol.* **2011**, 38, 100.
- [35] *Handbook of Partial Least Squares: Concepts, Methods and Applications*, Springer, Berlin, DE, **2010**.
- [36] T. Heberer, H.J. Stan, Detection of more than 50 substituted phenols as their t-butyltrimethylsilyl derivatives using gas chromatography mass spectrometry, *Anal. Chim. Acta* **1997**, 341, 21.
- [37] M. del Valle, Sensor Arrays and Electronic Tongue Systems, *Int. J. Electrochem.* **2012**, 2012,
- [38] P. Ciosek, W. Wroblewski, Sensor arrays for liquid sensing - electronic tongue systems, *Analyst* **2007**, 132, 963.
- [39] R.A. Johnson, D.W. Wichain, *Applied multivariate statistical analysis*, Pearson Education, Harlow, GB, **2007**.
- [40] C. Cortes, V. Vapnik, Support-vector networks, *Mach Learn* **1995**, 20, 273.
- [41] B.E. Boser, I.M. Guyon, V.N. Vapnik, in: *Proceedings of the fifth annual workshop on Computational learning theory*, ACM, Pittsburgh, Pennsylvania, United States, **1992**, pp. 144.

- [42] D. Meyer, F. Leisch, K. Hornik, The support vector machine under test, *Neurocomputing* **2003**, 55, 169.
- [43] L. Lvova, S.S. Kim, A. Legin, Y. Vlasov, J.S. Yang, G.S. Cha, H. Nam, All-solid-state electronic tongue and its application for beverage analysis, *Anal. Chim. Acta* **2002**, 468, 303.
- [44] A. Gutés, A.B. Ibáñez, F. Céspedes, S. Alegret, M. del Valle, Simultaneous determination of phenolic compounds by means of an automated voltammetric "electronic tongue", *Anal. Bioanal. Chem.* **2005**, 382, 471.
- [45] H. Martens, T. Naes, Multivariate calibration, John Wiley & Sons, Inc., Chichester, **1991**.
- [46] K.H. Esbensen, Multivariate Data Analysis - in practice, CAMO Sosftware, Oslo, **2002**.
- [47] H. Abdi, Partial least squares regression and projection on latent structure regression (PLS Regression), *WIREs. Comput. Stat.* **2010**, 2, 97.
- [48] R. Bro, Multi-way analysis in the food industry. Models, algorithms and applications, Ph.D. Thesis, *University of Amsterdam*, **1998**
- [49] N.M. Faber, J. Ferre, R. Boque, J.H. Kalivas, Second-order bilinear calibration: the effects of vectorising the data matrices of the calibration set, *Chemometr. Intell. Lab.* **2002**, 63, 107.
- [50] A.K. Smilde, R. Bro, P. Geladi, Multi-way analysis: applications in the chemical sciences, John Wiley & Sons Ltd., Chichester (UK), **2004**.
- [51] V. Gómez, M.P. Callao, Analytical applications of second-order calibration methods, *Anal. Chim. Acta* **2008**, 627, 169.
- [52] A. Mimendia, J.M. Gutiérrez, L.J. Opalski, P. Ciosek, W. Wróblewski, M. del Valle, SIA system employing the transient response from a potentiometric sensor array—Correction of a saline matrix effect, *Talanta* **2010**, 82, 931.
- [53] J. Zupan, J. Gasteiger, Neural networks: A new method for solving chemical problems or just a passing phase?, *Anal. Chim. Acta* **1991**, 248, 1.
- [54] F. Despagne, D. Luc Massart, Neural networks in multivariate calibration, *Analyst* **1998**, 123, 157R.
- [55] D.R. Hush, B.G. Horne, Progress in supervised neural networks, *IEEE Signal Processing Magazine* **1993**, 10, 8.
- [56] K. Toko, Biomimetic Sensor Technology, *Meas. Sci. Technol.* **2001**, 12, 221.
- [57] P. Mielle, B. Hivert, G. Mauvais, Are gas sensors suitable for on-line monitoring and quantification of volatile compounds?, in: P. Etiévant, P. Schreier (Eds.) Proceedings of Bioflavour 95, INRA éditions, Dijon, **1995**, pp. 81.
- [58] F. Winquist, P. Wide, I. Lundström, An electronic tongue based on voltammetry, *Anal. Chim. Acta* **1997**, 357, 21.
- [59] A. Wilson, M. Baietto, Applications and Advances in Electronic-Nose Technologies, *Sensors* **2009**, 9, 5099.
- [60] J. Gallardo, S. Alegret, M. del Valle, Application of a potentiometric electronic tongue as a classification tool in food analysis, *Talanta* **2005**, 66, 1303.
- [61] Å. Jansson, J. Röttorp, M. Rahmberg, Development of a software sensor for phosphorus in municipal wastewater, *J. Chemometr.* **2002**, 16, 542.
- [62] M.H. Masson, S. Canu, Y. Grandvalet, A. Lynggaard-Jensen, Software sensor design based on empirical data, *Ecol. Model.* **1999**, 120, 131.
- [63] R. Gutierrez-Osuna, in: J.W. Gardner, K.C. Persaud (Eds.) Proceedings of the 7th International Symp. On Olfaction and Electronic Nose, Taylor & Francis, Brighton, UK, **2000**, pp. 147.

- [64] F. Hossein-Babaei, V. Ghafarinia, Compensation for the drift-like terms caused by environmental fluctuations in the responses of chemoresistive gas sensors, *Sensor Actuat. B-Chem.* **2010**, *143*, 641.
- [65] Y. Vlasov, A. Legin, A. Rudnitskaya, C. Di Natale, A. D'Amico, Nonspecific sensor arrays ("electronic tongue") for chemical analysis of liquids (IUPAC Technical Report), *Pure Appl. Chem.* **2005**, *77*, 1965.
- [66] Y. Vlasov, A. Legin, Non-selective chemical sensors in analytical chemistry: from "electronic nose" to "electronic tongue", *Fresenius J. Anal. Chem.* **1998**, *361*, 255.
- [67] M. del Valle, Electrochemical micro(bio)sensor arrays, *Microchim. Acta* **2008**, *163*, 1.
- [68] J. Zeravik, A. Hlavacek, K. Lacina, P. Skladal, State of the Art in the Field of Electronic and Bioelectronic Tongues - Towards the Analysis of Wines, *Electroanal.* **2009**, *21*, 2509.
- [69] K. Persaud, G. Dodd, Analysis of discrimination mechanisms in the mammalian olfactory system using a model nose, *Nature* **1982**, *299*, 352.
- [70] J.W. Gardner, P.N. Bartlett, Electronic noses: principles and applications, Oxford University Press, Oxford, **1999**.
- [71] J.W. Gardner, P.N. Bartlett, A brief history of electronic noses, *Sensor Actuat. B-Chem.* **1994**, *18*, 210.
- [72] F. Winquist, S. Holmin, C. Krantz-Rülcker, P. Wide, I. Lundström, A hybrid electronic tongue, *Anal. Chim. Acta* **2000**, *406*, 147.
- [73] M.L. Rodriguez-Mendez, C. Apetrei, I. Apetrei, S. Villanueva, I.J.A. de Saja, I. Nevares, M. del Alamo, IEEE International Symposium on Industrial Electronics, **2007**, pp. 2782.
- [74] C. Di Natale, R. Paolesse, A. Macagnano, A. Mantini, A. D'Amico, A. Legin, L. Lvova, A. Rudnitskaya, Y. Vlasov, Electronic nose and electronic tongue integration for improved classification of clinical and food samples, *Sensor Actuat. B-Chem.* **2000**, *64*, 15.
- [75] S. Buratti, S. Benedetti, M. Scampicchio, E.C. Pangerod, Characterization and classification of Italian Barbera wines by using an electronic nose and an amperometric electronic tongue, *Anal. Chim. Acta* **2004**, *525*, 133.
- [76] R.S. Jackson, Wine tasting: a professional handbook, Academic Press, London, **2002**.
- [77] L. Vera, L. Aceña, J. Guasch, R. Boqué, M. Mestres, O. Busto, Discrimination and sensory description of beers through data fusion, *Talanta* **2011**, *87*, 136.
- [78] K.H. Ruhm, Sensor fusion and data fusion – Mapping and reconstruction, *Measurement* **2007**, *40*, 145.
- [79] Z. Ahmad, J. Zhang, Combination of multiple neural networks using data fusion techniques for enhanced nonlinear process modelling, *Comput. Chem. Eng.* **2005**, *30*, 295.
- [80] L.A. Klein, Sensor and Data Fusion: A Tool for Information Assessment and Decision Making, Society of Photo Optical, **2004**.
- [81] C. Söderström, A. Rudnitskaya, A. Legin, C. Krantz-Rülcker, Differentiation of four *Aspergillus* species and one *Zygosaccharomyces* with two electronic tongues based on different measurement techniques, *J. Biotechnol.* **2005**, *119*, 300.
- [82] T.G. Doeswijk, A.K. Smilde, J.A. Hageman, J.A. Westerhuis, F.A. van Eeuwijk, On the increase of predictive performance with high-level data fusion, *Anal. Chim. Acta* **2011**, *705*, 41.
- [83] F. Rosenblatt, The perceptron: A probabilistic model for information storage and organization in the brain, *Psychol. Rev.* **1958**, *65*, 386.

- [84] M. Peris, L. Escuder-Gilabert, A 21st century technique for food control: Electronic noses, *Anal. Chim. Acta* **2009**, 638, 1.
- [85] R. Paniagua Gómez-Álvarez, *Citología e histología vegetal y animal*, McGraw-Hill/Interamericana, Madrid, **2007**.
- [86] R. Lippmann, An introduction to computing with neural nets, *IEEE ASSP Mag.* **1987**, 4, 4.
- [87] A.P. Plumb, R.C. Rowe, P. York, M. Brown, Optimisation of the predictive ability of artificial neural network (ANN) models: A comparison of three ANN programs and four classes of training algorithm, *Eur. J. Pharm. Sci.* **2005**, 25, 395.
- [88] M.T. Hagan, H.B. Demuth, M.H. Beale, *Neural Network Design*, Pws Pub., Boston, **1996**.
- [89] M.T. Hagan, M.B. Menhaj, Training feedforward networks with the Marquardt algorithm, *IEEE T. Neural Networ.* **1994**, 5, 989.
- [90] M.H. Beale, M.T. Hagan, H.B. Demuth, in, MathWorks, **2012**.
- [91] D.J.C. MacKay, Bayesian Interpolation, *Neural Comput.* **1992**, 4, 415.
- [92] F. Dan Foresee, M.T. Hagan, International Conference on Neural Networks, **1997**, pp. 1930.
- [93] E. Schaller, J.O. Bosset, F. Escher, 'Electronic Noses' and Their Application to Food, *LWT-Food Sci. Technol.* **1998**, 31, 305.
- [94] L. Escuder-Gilabert, M. Peris, Review: Highlights in recent applications of electronic tongues in food analysis, *Anal. Chim. Acta* **2010**, 665, 15.
- [95] M. Scampicchio, D. Ballabio, A. Arecchi, S.M. Cosio, S. Mannino, Amperometric electronic tongue for food analysis, *Microchim. Acta* **2008**, 163, 11.
- [96] J. Garrido, F. Borges, Wine and grape polyphenols—A chemical perspective, *Food Res. Int.* **2011**, 44, 3134.
- [97] V.L. Singleton, J.A. Rossi, Colorimetry of total phenolics in grapes and wine with phosphomolybdic -phosphotungstic acid reagents, *Am. J. Enol. Viticult.* **1965**, 16, 144.

2. Objectius

2. Objectius

En els darrers anys, s'ha observat una creixent demanda de nous mètodes analítics capaços d'oferir el màxim de informació analítica amb el mínim "esforç" possible, els quals són necessaris per tal d'assolir els nous reptes en camps tals com el monitoratge ambiental, la seguretat alimentària i la salut pública. En aquest sentit, la indústria està molt interessada en el desenvolupament de mètodes ràpids i de baix cost que puguin ser utilitzats de forma rutinària com a cribratge (*screening analysis*) per tal de detectar qualsevol adulteració o contaminació del producte ja sigui durant la seva elaboració o per comprovar que compleixen els estàndards de control de qualitat.

En aquest sentit, les línies clàssiques de investigació en el camp dels sensors químics s'han centrat en el desenvolupament de dispositius cada vegada més selectius envers espècies determinades, i al mateix temps amb una major sensibilitat i menor límit de detecció. Malauradament, es disposa de molts pocs sensors químics que operin de forma òptima i sense cap interferència o efecte matriu en les condicions requerides en els casos de mostres reals. Precisament, la dificultat per obtenir sensors amb una selectivitat i sensibilitat apropiada per qualsevol analit, ha estat la que ha donat pas a l'aparició de noves estratègies com és el cas de les llengües electròniques per tal de donar resposta a aquests problemes.

Aquests sistemes analítics consisteixen en la imitació del sentit de gust dels mamífers, on uns pocs receptors són capaços de respondre a una gran varietat de substàncies, un principi que juntament amb una etapa complexa de tractament de la informació anàloga a la que té lloc en el cervell, permet la quantificació o classificació d'un gran nombre de substàncies. Aquests sistemes biomimètics, en oposició als sistemes clàssics, estan basats en la combinació de matrius de sensors amb una baixa selectivitat i/o una resposta creuada per tal d'obtenir un valor afegit en la generació de la informació analítica.

Un dels darrers avanços en el disseny de llengües electròniques ha estat la incorporació de biosensors per tal de fer front a nous camps d'aplicació o millorar els ja existents. Aquests sistemes, coneguts com a llengües bioelectròniques, només es distingeixen dels convencionals en la incorporació d'un o més biosensors en la matriu de sensors que formarà la llengua electrònica, normalment compartint el mateix principi de detecció per tal de facilitar la compatibilitat entre ells. En aquest sentit, un dels objectius de la present tesi serà el la incorporació de biosensors en les diferents llengües

electròniques per tal de demostrar les seves millors prestacions, sense menystenir l'ús de les llengües electròniques convencionals, i comparar els resultats d'ambdues.

Així doncs, l'objectiu general de la present tesi és demostrar l'aplicabilitat de les llengües electròniques en el sector vinícola i de les begudes alcohòliques, tant per l'extracció de informació qualitativa i identificació de classes com per la quantificació de paràmetres d'interès analític, responent en ambdós casos a les necessitats corresponents en cada sector i amb un nivell de complexitat creixent. Per tal d'assolir aquest objectiu s'han seguit les següents directrius:

- Plantejar les diferents necessitat o aplicacions d'interès en cada sector, definint quins serien els compostos a analitzar o les classes a discriminar
- Construir i avaluar els diferents sensors i biosensors voltamperomètrics que permetin la determinació dels diferents conjunts d'analits d'interès, i que formaran la matriu de sensors de la llengua (bio)electrònica
- Preparació del disseny experimental i dels diferents conjunts de mostres, així com del protocol de mesura d'aquestes
- Aplicació de les diferents eines quimiomètriques que permetin extreure la informació desitjada, seleccionant en cada cas el tractament més adequat. Així, en aquest sentit, es pretén estudiar i avaluar l'ús de diferents estratègies de preprocessament i processament del senyal

D'aquesta manera, seguint les directrius descrites anteriorment, és va plantejar abordar la classificació de diferents begudes tals com el cava, els brandis i les cerveses, amb l'objectiu de desenvolupar una eina analítica objectiva que permeti la categorització d'aquestes i/o la identificació de frau en la seva producció.

Paral·lelament, i pensant especialment en el cas dels vins, es va plantejar com a objectiu abordar la detecció de compostos fenòlics en aquests. A més, degut a que en aquest cas es busca la detecció d'una família de compostos concrets, es va voler utilitzar aquest com a cas d'estudi per avaluar les diferències en la resposta de la llengua electrònica i la llengua (bio)electrònica.

3. Experimental

3. Experimental

A continuació es descriuen de forma concisa els aspectes més significatius referents al procediment experimental i metodologia seguida en la realització d'aquest treball, incloent una breu introducció i descripció dels conceptes utilitzats habitualment en el camp de les llengües electròniques.

Així doncs, en els següents apartats es descriurà la construcció dels sensors, el disseny experimental, el procediment de mesura i els aspectes bàsics del processament de les dades. Posant major èmfasi en aquest darrer apartat, i en l'ús de les xarxes neuronals artificials (ANNs), les quals han estat el mètode més bastament utilitzat al llarg d'aquest treball, relegant la resta de mètodes a fins comparatius.

3.1. *Construcció dels sensors voltamperomètrics*

En tots els casos s'ha emprat elèctrodes basats en compòsits rígids grafit-epoxi, llargament usats dins el Grup de Sensors i Biosensors (GSB). Els compòsits són materials formats per mesclades de dos o més components amb propietats diferents a les mostrades pels predecessors [1]. En aquest cas es combina una resina epoxi, que realitzarà les funcions d'aglutinant, amb el grafit, que ofereix la conductivitat elèctrica com a propietat principal. De la unió d'aquestes dues substàncies, i una vegada endurit, s'aconsegueix un nou material robust, un compòsit, que permet la regeneració de la seva superfície mitjançant un lleuger poliment físic amb paper abrasiu.

Per a la construcció dels compòsits s'ha utilitzat la configuració convencional (o cilíndrica) [2]. Així doncs, els passos seguits per l'elaboració dels diferents sensors són els següents, tal com s'il·lustra a la Figura 3.1:

- El primer pas és la soldadura amb estany d'un disc de coure de 5mm de diàmetre a un connector elèctric. Aquesta làmina s'ha rentat prèviament amb àcid clorhídric a fi d'eliminar la capa d'òxid que es forma sobre la superfície d'aquest metall.

- Posteriorment s'insereix el connector dins d'un tub de PVC de 20mm de llarg i 6mm de diàmetre intern. Amb això s'aconsegueix generar un petit espai buit, d'uns 4mm d'alçada on es dipositarà la pasta de grafit-epoxi preparada.

- El següent pas és la preparació de la mescla grafit-epoxi. La diferència entre els diferents sensors i biosensors es dona en aquest pas, ja que podem variar la composició d'aquesta i inclús incorporar diferents catalitzadors electroquímics o enzims durant la

seva preparació. El primer pas consisteix en pesar els 2 components de la resina (Epotek H77, part A i B) en la proporció 20:3 (w/w) i la seva homogeneïtzació. Seguidament, s'afegeix la quantitat pertinent de grafit en pols (mida de partícula 50 μ m), i dels modificadors en cas que sigui necessari, donant lloc a una segona etapa d'homogeneïtzació.

- A continuació es diposita la barreja en la cavitat, donant pas al procés de curat, el qual consisteix en deixar l'elèctrode a l'estufa per tal que la resina polimeritzi i s'endureixi. Generalment aquest procés té lloc a 80°C durant 2-3 dies, excepte en el cas d'incorporar material biològic (biocompòsits) que per tal d'evitar la desnaturalització tèrmica d'aquest, el procés té lloc a 40°C durant una setmana.

- Finalment, un cop endurida la resina, cal polir l'excés de material amb paper abrasiu de diàmetre de gra decreixent i paper d'alúmina fins a obtenir una superfície fina i especular. L'àrea final de l'elèctrode és de 28mm².

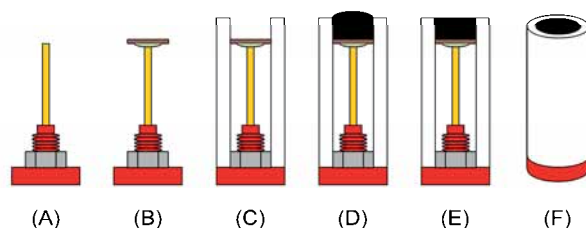


Figura 3.1. Esquema del procés de construcció dels elèctrodes. (A) Connector elèctric, (B) sobre el qual es solda un disc de coure, (C) i s'introdueix en un tub de PVC. (D) Preparació de la barreja grafit-epoxi, incorporació dels modificadors i deposició en la cavitat formada pel disc i el tub. (E i F) Aspecte final del compòsit un cop curat i polit.

Taula 3.1 Composició en % en pes dels diferents elèctrodes voltamperomètrics que formen part de les diferents llengües electròniques utilitzades.

| | <i>Sensor</i> | <i>Epotek H77*</i> | <i>Grafit</i> | <i>Modificador</i> |
|-------|--------------------|--------------------|---------------|--|
| ET | Grafit-epoxi (GEC) | 85% | 15% | - |
| | CoPc | 83% | 15% | 2% ftalocianina de Cobalt(II) |
| | Cu | 83% | 15% | 2% nanopartícules de coure |
| | Pt | 83% | 15% | 2% nanopartícules de platí |
| | Ag | 83% | 15% | 2% nanopartícules de plata |
| | PANI o Pol | 83% | 15% | 2% polianilina |
| | PPy o Pyr | 83% | 15% | 2% polipirrol |
| - | GOX | 83% | 15% | 2% Glucosa oxidasa d'Aspergillus niger (EC 1.1.3.4, 174.4 U·mg ⁻¹) |
| BioET | Grafit-epoxi (GEC) | 85% | 15% | - |
| | Tyr | 83% | 15% | 2% Tirossinassa de bolet (EC 1.14.18.1, 4276 U·mg ⁻¹) |
| | Lac | 83% | 15% | 2% Lacasa de Trametes versicolor (EC 1.10.3.2, 21 U·mg ⁻¹) |
| | Cu | 83% | 15% | 2% nanopartícules de coure (50nm) |

*Prèviament es barreja la resina (part A) amb el corresponent enduridor (part B) en la proporció 20:3 (w/w)

Aquest procés és va repetir, variant el modificador que s'incorporava en cada cas, per cadascun dels sensors que formen part tant de la llengua electrònica (ET) com la llengua bioelectrònica (BioET). Així doncs, la Taula 3.1 resumeix la composició dels diferents elèctrodes utilitzats al llarg del present treball.

3.2. Disseny experimental

Un dels principals inconvenients de les llengües electròniques i dels models quimiomètrics aplicats és l'elevat volum de mostres necessari a l'hora de construir un model i dur a terme l'avaluació i validació d'aquest. Això és degut a que per tal de construir un model robust i amb una bona capacitat de predicció, aquest haurà de contenir totes les possibles interaccions entre els diversos analits i els sensors. Per aquest motiu, pren especial rellevància el disseny experimental escollit en el moment de preparar i escollir les mostres.

A l'hora de dissenyar un experiment el primer que haurem de fer és definir el domini experimental, és a dir, el rang de concentracions per cadascun dels analits d'interès. A continuació, caldrà escollir el tipus de disseny a utilitzar, fet que delimitarà el número de mostres requerides per la construcció del nostre model i la seva distribució espacial. En aquest treball, i basant-se en resultats anteriors obtinguts en el nostre grup [3], s'ha optat per utilitzar un disseny factorial complet (3^3) modificat (Figura 3.2). Aquesta modificació, consisteix en realitzar una lleugera rotació sobre cada un dels eixos del cub obtingut amb el disseny factorial complet clàssic.

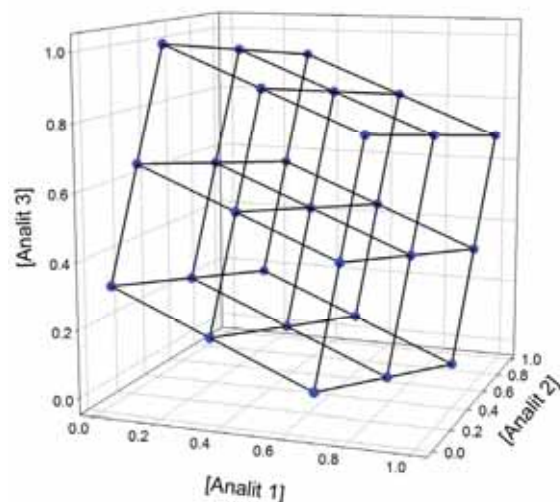


Figura 3.2. Representació espacial de les concentracions de cada analit utilitzant un disseny factorial complet modificat amb 3 nivells a 3 analits.

Un cop definit el nostre domini experimental i el número de mostres requerides per construir el nostre model (conjunt d'entrenament o *train*), caldrà preparar un altre grup de mostres les quals seran utilitzades per avaluar la capacitat de predicció d'aquest (conjunt de prova o *test*). Aquestes últimes es troben distribuïdes de forma aleatòria dins del nostre domini experimental, i el seu número sol variar entre 2/3 i 1/4 del número de mostres utilitzades en el conjunt d'entrenament. És a dir, mantenint una relació entre els dos conjunts de 60:40 a 80:20 per a entrenament i validació, respectivament.

3.3. Procediment de mesura

La tècnica utilitzada per dur a terme les diferents mesures ha estat la voltamperometria cíclica (§1.2.1). Aquesta és una tècnica que es basa en la mesura de la intensitat de corrent generada en funció del potencial elèctric aplicat, de forma que s'afavoreixi la polarització per concentració de l'elèctrode de treball. És interessant observar que en funció del potencial aplicat, el nostre elèctrode de treball es comportarà com un ànode o com un càtode, de manera que podrem utilitzar les mesures voltamperomètriques per a la determinació de substàncies que presentin tant capacitat de reducció com d'oxidació.

La cel·la electroquímica estava formada per la matriu de diferents sensors de la llengua electrònica els quals actuaven com a elèctrodes de treball, un elèctrode de referència de doble unió Ag/AgCl (Thermo Orion 900200) i un elèctrode auxiliar de platí (Crison 52-67). En l'elèctrode de referència s'utilitzava com a solució interna una solució de KNO₃ saturada amb AgCl subministrada per la mateixa firma (Orion 900002), que actua com a solució de referència interna a fi de proporcionar un potencial constant, igual al d'un elèctrode de calomelans saturat. Com a solució de referència externa s'utilitzava la mateixa solució usada com a medi electrolític (habitualment tampó fosfat 0.1M i KCl 0.1M, a diferents pH segons els requisits experimentals) o una solució de KNO₃ al 10% subministrada per la mateixa firma (Orion 900003).

Per dur a terme les mesures voltamperomètriques es va utilitzar un potenciostat Autolab PGSTAT20 en una configuració de multielèctrode, utilitzant el programa *GPES Multichannel 4.7* pel control de l'equip i l'adquisició de les dades, els quals permetien la possibilitat de mesurar fins a 6 canals de forma simultània (Figura 3.3). Les condicions de mesura es varen variar lleugerament en cada cas per tal d'obtenir la millor resposta, adaptant tant els rangs de l'escombrat de potencial com el salt de

potencial al llarg d'aquest, mantenint en gairebé tots els casos una velocitat d'escombrat de $100 \text{ mV} \cdot \text{s}^{-1}$.

Durant la mesura de les mostres de cadascun dels diferents conjunts, totes les mesures es van dur a terme sense realitzar cap tipus de regeneració física dels elèctrodes de treball. En aquest sentit, per tal de prevenir l'acumulació d'impureses en la superfície d'aquests (*fouling*), es realitzava una etapa de neteja electroquímica entre cada mostra. Aquesta consistia en l'aplicació d'un potencial lleugerament superior al potencial aplicat en l'escombrat durant un període de 40s després de cada mesura en un vas que contenia aigua destil·lada.

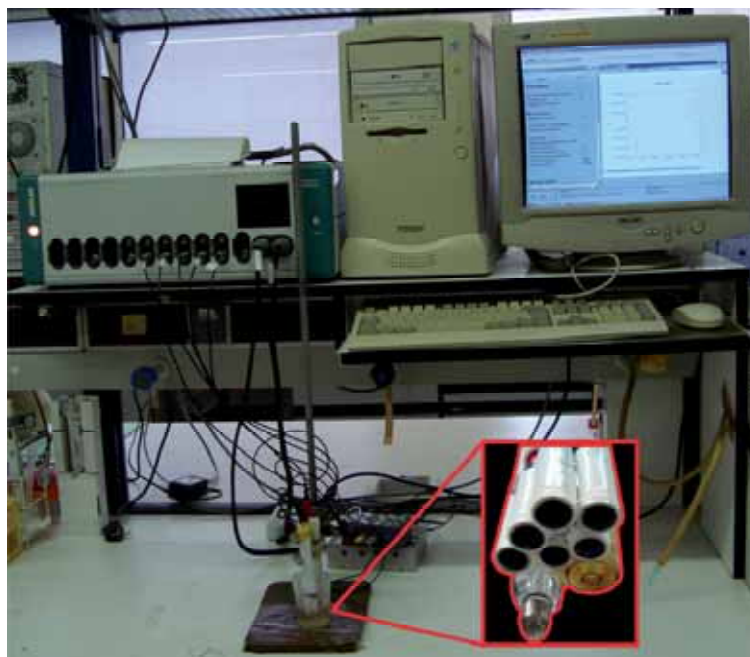


Figura 3.3. Fotografia de l'equip de mesura i detall de la cel·la electroquímica utilitzada.

3.4. Construcció dels models de resposta

Com ja s'ha explicat, per tal de poder extreure el màxim de informació de les dades obtingudes en el procés de mesura, es requereix de l'ús d'eines quimiomètriques que permetin la construcció d'un model de resposta multivariable (Figura 3.4). Aquests mètodes consisteixen en l'ajust d'un model matemàtic, que donat un conjunt de predictors (resposta dels sensors) sigui capaç de predir els paràmetres d'interès de la nostra mostra; en el que seria l'equivalent multidimensional de la construcció d'una recta de calibrat.

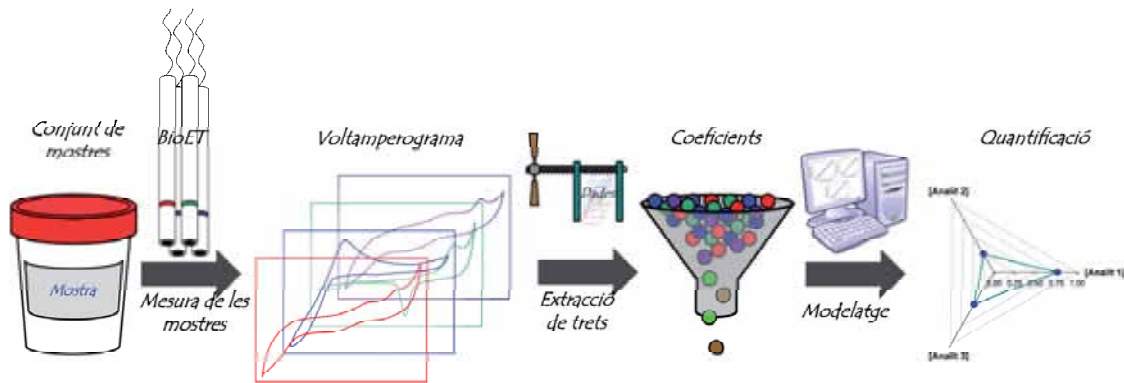


Figura 3.4. Esquema dels passos seguits per la construcció del model.

El procés de construcció del model de resposta el dividirem en tres etapes: preprocessament de les dades, construcció del model i validació d'aquest. L'objectiu de la primera etapa és reduir l'elevada dimensionalitat del conjunt de dades per tal d'obtenir avantatges en l'etapa de modelatge; tot i que aquesta etapa no és imprescindible per a tots els mètodes de modelatge (e.g. mètodes lineals tals com el PLS o el nPLS), s'ha demostrat que és possible obtenir millors resultats en realitzar el preprocessament de les dades [4, 5]. El següent pas és la construcció del model, per això s'utilitza com a entrada en el model les respostes dels elèctrodes (ja sigui la resposta dels (bio)sensors directament o els coeficients obtinguts en l'etapa de preprocessament) i les variables a predir (ja sigui les concentracions dels diferents analits o la classe a la qual pertany cada mostra) per cada mostra del subconjunt d'entrenament. Aquestes dos matrius són utilitzades per construir i ajustar un model que relacioni el conjunt de respostes amb el de *targets*. Finalment, un cop construït el model, cal avaluar la seva capacitat predictiva. Per dur a terme aquesta tasca, s'introdueix en el model només la matriu de resposta per les mostres del subconjunt de test, obtenint els valors predits per aquestes. Tot seguit, cal comparar i verificar si els valors esperats i predits coincideixen i el grau en que ho fan, la qual cosa determinarà l'eficàcia del model.

A continuació, i al llarg dels següents apartats es descriuran els diferents procediments emprats al llarg de la present tesi doctoral per a la realització d'aquestes tres etapes, acompanyats d'una breu descripció dels fonaments teòrics per tal de facilitar-ne la seva comprensió en certs casos.

3.4.1. Preprocessament de les dades

Un dels principals impediments en el cas de treballar amb les llengües electròniques voltamperomètriques és l'elevada dimensionalitat dels senyals obtinguts (mostres x

sensors x potencials aplicats; és a dir, quan per cada un dels sensors de la matriu fem el registre complet de tot el voltamperograma). Aquest fet dificulta el seu tractament en l'etapa de construcció del model, especialment en els cas de les xarxes neuronals artificials (ANNs), requerint d'una etapa intermèdia de preprocessament del senyal entre el procés de mesura i de modelatge. Així doncs, el principal objectiu d'aquesta etapa és la reducció de la complexitat del senyal d'entrada, preservant al mateix temps la informació significativa d'aquest, reduint per tant el volum de dades emprades per la construcció del model.

A pesar de que en alguns casos es pugui observar diferències en termes d'especificitat des del punt de vista amperomètric convencional (mesura de la intensitat de corrent generada al aplicar un potencial concret), la utilització de tot el registre voltamperomètric representa una millora notable en els resultats obtinguts respecte la utilització de només els valors corresponents als pics del voltamperograma. Això és degut a que cada sensor proporciona una resposta voltamperomètrica complexa que conté informació global sobre la mostra i els compostos analitzats al llarg de tot el voltamperograma. Fet que ha donat lloc a l'aparició de diferents estratègies per tal d'assolir l'extracció d'aquesta informació en uns pocs coeficients.

D'altra banda, quan treballem amb matrius de sensors tenim que prendre certes precaucions, com ara la importància relativa de les variacions de cada variable, per tal d'equiparar a priori la contribució d'aquestes en el model final. Per aquest motiu, amb independència tant del mètode utilitzat en l'etapa de modelatge com de l'ús d'una etapa de compressió del senyal, sempre caldrà dur a terme una etapa de centrat i normalització de les dades abans de la construcció del model.

3.4.1.1. Centrat i normalització

Quan treballem amb matrius de sensors, on cadascun dels quals presenta una resposta característica, cal tenir en compte la importància relativa de les variacions en cadascuna de les variables. És a dir, que a priori la resposta de tots els sensors tingui la mateixa importància. En aquest sentit, algunes estratègies que es duen a terme per tal d'equiparar el conjunt de lectures dels diversos sensors són: el centrat, l'estandardització, l'autoescalat o la normalització de les dades [6].

- Centrat: consisteix en referenciar totes les dades a la mitjana aritmètica d'aquestes, de manera que totes queden centrades a 0 (mitjana nul·la o igual a 0, $\bar{x}=0$).

$$x_i^* = x_i - \bar{x} = x_i - \frac{1}{n} \sum_{i=1}^n x_i \quad \text{Equació 3.1}$$

- Estandardització: consisteix a equiparar la dispersió de les dades provinents de diferents sensors, de manera que totes tinguin desviació estàndard igual a la unitat ($s=1$).

$$x_i^* = \frac{x_i}{\sqrt{s_i}} = \frac{x_i}{\sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}} \quad \text{Equació 3.2}$$

- Autoescalat: correspon a la realització conjunta dels dos processos anteriors. És a dir, dades centrades a 0 i desviació estàndard 1 ($\bar{x}=0$ i $s=1$).

$$x_i^* = \frac{x_i - \bar{x}}{\sqrt{s_i}} \quad \text{Equació 3.3}$$

- Normalització: correspon a l'escalat de les dades entre dos valors concrets, habitualment entre 0 i 1 (Equació 3.4) o entre -1 i 1 (Equació 3.5), de manera que el conjunt de dades de cada sensor presenti el mateix rang de variació i la mateixa escala.

$$x_i^* = \left(\frac{x_i - \min(x)}{\max(x) - \min(x)} \right) \quad \text{Equació 3.4}$$

$$x_i^* = 2 \cdot \left(\frac{x_i - \min(x)}{\max(x) - \min(x)} \right) - 1 \quad \text{Equació 3.5}$$

3.4.1.2. Selecció de variables (Feature selection)

La solució més simple, per assolir la reducció en aquests casos, és la mera selecció dels punts del conjunt de dades que més contribueixen al senyal dels analits d'interès (en aquest cas, generalment els corresponents als diferents pics dels voltamperogrames) i la seva inclusió en l'etapa de modelatge.

Tot i que l'ús de tot el registre voltamperomètric sempre aporta major informació al sistema, hi ha grans avantatges en identificar i eliminar els *inputs* menys significatius

fins i tot en el cas que puguem entrenar un bon model amb tot el conjunt: major precisió en el modelatge, obtenció d'un model més robust, major comprensió del sistema pel fet d'identificar els components més significatius (ergo model més fàcil de comprendre), major proporció d'*inputs*/pesos del model (per tant, una millor capacitat de generalització) [7]. Conseqüentment, obtenint millors models a partir del conjunt d'*inputs* reduït.

En aquest sentit, a continuació es presenten diferents estratègies (de menor a major complexitat) seguides per dur a terme aquesta selecció, les quals intenten capturar els punts més representatius del conjunt de dades.

i. Coneixements previs

La primera opció és basa en l'ús de l'experiència acumulada per part de l'analista, el qual pot identificar a l'observar el voltamperograma quines són les zones més significatives d'aquest, i per tant, els potencials als que apareixen les respostes voltamperomètriques més significatives. D'aquesta manera, suposant que la informació més rellevant es troba en aquestes zones i incloent només aquests punts al model, és possible dur a terme la reducció en la dimensió de les dades.

Aquesta opció és especialment útil quan es treballa amb biosensors, sensors modificats que presenten pics molt concrets en el voltamperograma o quan prèviament hem identificat i caracteritzat amb solucions estàndard els nostres sensors i coneixem les zones del voltamperograma en les que apareix la resposta voltamperomètrica.

No obstant, a pesar dels bons resultats que poden ser obtinguts d'aquesta manera, aquest procediment no és la millor de les opcions ja que requereix de molts més coneixements previs del sistema, la selecció dels *inputs* és subjectiva, no es pot assegurar que tota la informació present en les dades de partida sigui inclosa en el model i no sempre els punts més rellevants són fàcilment identificables.

ii. Canvi en la freqüència de mostreig

Tot i que de per si no és tracta de cap mètode en si, una opció és ajustar el salt de potencial a l'hora de fer les mesures (o simplement descartar 1 de cada x punts a posteriori) o ajustar el rang de l'escombrat de potencial per tal de tenir els mínims punts experimentals de partida [8]. Una altra opció, utilitzada prèviament amb el nostre grup, es basa en la manipulació matemàtica de les diferents senyals, e.g. la suma dels diferents registres voltamperomètrics pels diferents sensors [9].

iii. Paràmetres estadístics/mètodes d'inclusió per passos

En el mateix sentit, aquesta selecció dels punts que cal utilitzar com a *inputs* en l'etapa de modelatge pot ser realitzada de manera més objectiva a partir de l'anàlisi estadística. Alguns exemples de mètodes estadístics que poden ser utilitzats són l'anàlisi de correspondència (CA), l'anàlisi de factors (FA) [10] o els Algorismes Genètics (GA) [11], entre els més comuns.

D'aquesta manera, podem simplement avaluar la correlació lineal entre els diferents punts de la matriu de resposta i eliminar els que presenten major correlació entre ells; utilitzar el FA per crear un número menor de noves variables no observades (anomenades factors) a partir de les combinacions lineals de les variables de partida; o utilitzar els algorismes genètics, els quals són mètodes de cerca heurística que imiten els mètodes de selecció natural per tal de trobar solucions a problemes d'optimització.

iv. Índex de Causalitat (CI)

Finalment, una altra possibilitat és l'ús de les ANNs (descrites amb més detall a §3.4.2.2.1, Figura 1.12) com a eina de selecció de variables enlloc de utilitzar aquestes com a eina de quantificació [12]. Així enlloc de fixar-nos en els resultats obtinguts per cada model, ens fixarem en el pesos que la xarxa atorga a cadascun dels *inputs*.

La utilització de les ANNs per la selecció de variables consisteix en la cerca d'un conjunt òptim d'*inputs* que permeti classificar o predir els *outputs* (ja siguin valors qualitius o quantitius) de manera satisfactòria. Es tracta d'una xarxa de propagació unidireccional (*feed-forward*) que combina un algorisme amb aprenentatge per retropropagació d'errors (*back-propagation*), el qual és utilitzat per entrenar la xarxa d'acord amb una regla d'aprenentatge [13]. Els passos a seguir per tal de dur a terme la reducció de la dimensionalitat i el procés de selecció de variables amb aquest mètode es detallen a continuació.

El primer pas és la definició de l'arquitectura de la ANN, la qual ha de ser escollida a priori, ja que en aquest cas aquesta no serà optimitzada. Per tant, el número de neurones en la capa oculta i les funcions de transferència entre les diferents capes són escollides a priori per tal de representar la dimensionalitat intrínseca del conjunt de dades. Cal destacar que aquest serà un pas clau en l'etapa de selecció de variables ja que tindrà un alt efecte en els resultats obtinguts. Per tant, basant-nos en l'experiència prèvia i els diferents articles, el nombre de neurones escollit per dur a terme la reducció (i també les funcions de transferència) ha de ser proper al valor òptim del model final,

essent en general 5 neurones un bon nombre. En aquest sentit, la major part de l'esforç està dirigit a reduir el nombre d'entrades de la xarxa donat que el nombre de sortides està fixat pels requisits del model (i.e. les concentracions o classes a predir).

El següent pas és determinar la rellevància dels diferents *inputs* de la xarxa. Un cop hem entrenat el nostre model ANN, procedim a l'anàlisi dels pesos que el model ha assignat a cada neurona, el qual permetrà identificar fàcilment els *inputs* més significatius. Llavors, es calcula les contribucions relatives de cada un dels *inputs* a la variància de la capa de sortida en el model construït [7]. D'aquesta manera, els *inputs* amb menor contribució sobre la variació en la capa de sortida poden ser eliminats; procés conegut com "podat d'entrades" o *pruning*. Aquests baixos valors indiquen que l'*input* no canvia significativament o que durant l'etapa d'entrenament el model li ha assignat un baix pes. Mitjançant la repetició d'aquest procés utilitzant com a punt de partida el conjunt de dades reduït i seleccionant cada vegada els *inputs* més significatius, és possible la identificació d'un conjunt de *inputs* reduït proper a l'òptim; repetint aquest procés fins que l'eficàcia del model és degradada. El principal inconvenient d'aquest mètode és el fet que els conjunts reduïts d'*inputs* obtinguts no són únics, i poden variar en funció de l'arquitectura de la ANN escollida, d'aquí la importància de la seva elecció.

3.4.1.3. Ajust polinomial

Una altra opció és l'ajust del senyal registrat a alguna funció basada en una distribució polinomial de grau n , gaussiana, exponencial, potencial, equacions d'ona (e.g. sinus o cosinus), etc. [14]; utilitzant d'aquesta manera els coeficients obtinguts de la funció ajustada com a inputs.

En aquest mateix sentit, per tal d'assolir una reducció substancial de les dades i fer un filtrat del soroll a la vegada, el senyal obtingut pot ser ajustat mitjançant l'ús dels polinomis de Legendre [14]. Encara que aquest procediment pot ser vist com un simple ajust, és de fet una eina utilitzada per trobar una representació més eficient dels punts de dades [15, 16].

L'equació general dels polinomis de Legendre (Equació 3.6) pot resoldre's utilitzant el mètode de sèrie de potències (sempre que $|x| < 1$ i que n sigui un número enter no negatiu), formant les solucions una família de polinomis ortogonals anomenats polinomis de Legendre ($P_n(x)$), on cada polinomi $P_n(x)$ és un polinomi de grau n . A

més, per tal de simplificar-ne la seva expressió és possible utilitzar la fórmula de Rodrigues (Equació 3.7)

$$P_n(x) = \frac{1}{2^n n!} \frac{d^n}{dx^n} \left[(x^2 - 1)^n \right] \quad \text{Equació 3.6}$$

$$P_n(x) = \frac{1}{2^n} \sum_{k=0}^n \binom{n}{k}^2 (x+1)^{n-k} (x-1)^k \quad \text{Equació 3.7}$$

3.4.1.4. Integral de finestra seccionada

Aquest mètode consisteix en considerar l'àrea sota la corba voltamperomètrica, la qual també és pot considerar com un promig escalat dels diferents punts de la corba [17]. La idea d'aquest mètode és capturar d'alguna manera informació relativa a la forma del voltamperograma; mentre que la seva efectivitat radica en el fet que es tracta d'una aproximació de la forma de treballar tradicional en voltamperometria, en la qual s'identifica els pics corresponents a cada analit d'interès i es calcula l'alçada de pic (corrent màxima) o l'àrea d'aquest (corrent total), paràmetres els quals són proporcionals a la concentració de l'analit. Així d'aquesta manera, degut que en el cas de les llengües electròniques molts cops tenim solapament entre els senyals obtinguts per cada analit, no és possible realitzar una recta de calibrat de forma convencional i per tant, substituïm l'àrea de cada pic per l'àrea dins de la corba voltamperomètrica calculada per diferents seccions (Figura 3.5).

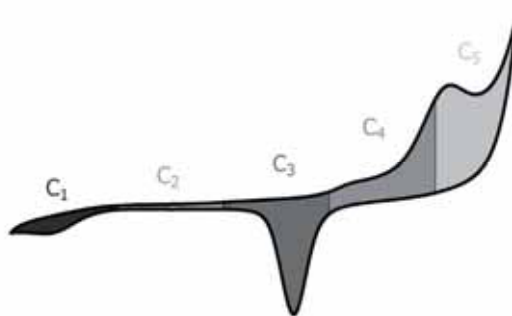


Figura 3.5. Esquema de preprocessament del senyal voltamperomètric utilitzant el mètode de la integral de finestra seccionada; en aquest cas utilitzant un valor de $k=5$.

Per tant, tal com s'il·lustra a la Figura 3.5 per tal de calcular els diferents coeficients, es divideix el voltamperograma en k seccions i es calcula l'àrea sota cada secció, la qual serà utilitzada com a *input* en l'etapa de modelatge.

Un dels paràmetres claus a l'hora d'aplicar aquest mètode és el número de seccions

(k) en que és dividit el voltamperograma, el qual és un paràmetre arbitrari i que dependrà del perfil de resposta de les nostres dades. Per una banda, al augmentar el valor de k , s'augmenta el grau de fidelitat o reproducció del senyal original, però a costa del grau de compressió. Quan el valor de k és molt proper al número de punts del senyal original, els coeficients obtinguts corresponen simplement a un valor escalat del valor dels punts d'aquest. És a dir, el senyal original multiplicat per un factor constant. D'altra banda, si el valor de k és molt petit, hi ha el perill de no extreure suficient informació del perfil de resposta i obtenir simplement un valor promig del corrents obtinguts al llarg de tot el registre. Per tant, cal escollir un valor de k tal que representi un compromís entre el grau de compressió i la fidelitat respecte el senyal original. Degut a la dificultat d'escollir aquest valor a priori, ja que només l'experiència prèvia ens pot ajudar a escollir un valor idoni, la millor opció és realitzar una optimització d'aquest de forma conjunta a la topologia del model durant el procés de modelatge.

3.4.1.5. Anàlisi de Components Principals (PCA)

Una altra opció és l'ús del PCA com a mètode d'extracció de característiques. Tot i que aquest mètode s'utilitza majoritàriament com a eina exploratòria de conjunts de dades (tal com s'ha descrit a §1.3.3.1, i amb més detall a §3.4.2.1.1) també pot ser utilitzat com a procediment per tal de reduir la dimensionalitat de grans conjunts de dades, en aquest cas els diferents voltamperogrames enregistrats per cada sensor que forma part de la matriu de la llengua electrònica; podent ser utilitzat fins i tot per trobar senyals en conjunts de dades sorolloses.

En aquest sentit, el que fa el PCA és “descobrir” noves variables, combinacions de lineals de les variables originals, anomenades components principals (PCs), que representen la major part de la variabilitat de les dades; permetent així descriure aquesta informació amb un nombre considerablement menor de variables de les que eren originalment presents. Així, en el cas dels registres voltamperomètrics, cada voltamperograma sol estar format per entre 250 i 525 punts, quedant continguda gairebé tota la informació present en aquests amb només els primers 5-10 PCs. Precisament, gràcies a aquest fet i preservant només els coeficients més significatius, descartant els demés, és el que ens permet utilitzar el PCA com a mètode de compressió del senyal.

Al igual que en el cas anterior, cal prendre un compromís entre el número de coeficients extrets (número de PCs en aquest cas) que utilitzarem com a *inputs* en

l'etapa de modelatge i la representativitat d'aquests respecte el senyal original. Justament en aquesta direcció, un altre dels avantatges que ens aporta aquest mètode és que podem utilitzar la variància acumulada en els diferents PCs com a mesura del grau de reconstrucció del senyal, obtenint un indicador del número de PCs que cal utilitzar.

3.4.1.6. *Transformada Ràpida de Fourier (FFT)*

La transformada ràpida de Fourier és un algorisme més eficient per tal de calcular la transformada discreta de Fourier (DFT); això és degut a que el càlcul a partir de la definició és massa lent en la majoria de casos per a resultar pràctic. D'aquesta manera, utilitzant la FFT, la diferència de velocitat pot ser substancial, especialment per grans conjunts de dades (a la pràctica, el temps de càlcul pot ser reduït en varis ordres de magnitud). La DFT és una eina molt important en el processament digital de senyals, la qual descompon una seqüència de valors en components de diferents freqüències.

El processament dels senyals utilitzant la transformada de Fourier es pot considerar com la descomposició d'aquest utilitzant la parella de funcions sinus/cosinus a diferents freqüències, calculant un coeficient per cada una tenint en compte la seva contribució al senyal original. D'aquesta manera, al aplicar la transformada de Fourier s'obtenen tants coeficients com punts en el senyal original, on cada contribució representa la informació de les bandes de freqüència contingudes en un segment específic dels senyals analitzats. Al igual que en el cas del PCA, la informació més rellevant es troba reflectida en els primers coeficients, i per tant és possible aconseguir un elevat grau de reconstrucció preservant només els primers coeficients i descartant la resta. A més a més, un avantatge addicional d'aquest mètode és la facilitat per eliminar el soroll present en el senyal mitjançant l'eliminació de les contribucions de freqüència elevada [18]. Finalment, per tal de seleccionar el número òptim de coeficients que cal preservar, és pot utilitzar la inversa de la transformada de Fourier, la qual permet avaluar el grau de reconstrucció respecte el senyal original.

3.4.1.7. *Transformada Wavelet Discreta (DWT)*

La transformada Wavelet (WT) és una tècnica de processament de senyals derivada de la transformada de Fourier, apareguda a finals de la dècada dels 1980s [19]. La diferència entre aquestes dos transformades és que enlloc de projectar el senyal sobre

una funció sinusoidal, la WT projecta el senyal sobre una funció *wavelet*; amb un avantatge clau respecte la FT que és la seva resolució temporal: captura tant la freqüència com la localització (temporal) de la informació.

En l'última dècada, s'han descrit un gran nombre d'aplicacions de la transformada Wavelet en l'anàlisi química, tant en electroquímica com cromatografia, tals com l'eliminació del soroll i el filtrat de senyals, la compressió de dades, la correcció de la línia base, la resolució del solapament de senyals, etc. [20].

La WT s'utilitza per la descomposició d'un senyal en un conjunt de funcions base. Aquestes funcions s'obtenen a partir de dilatacions i translacions d'una única funció anomenada funció mare *wavelet* [21]; sent la família de *wavelets Daubechies* la més utilitzada del conjunt de DWT [18]. D'aquesta manera, la transformada Wavelet descompon el senyal en contribucions marcades i localitzades per una escala i un paràmetre de posició. Aquesta és implementada utilitzant l'algorisme de la piràmide de Mallat's [19], el qual actua sobre un senyal discret de longitud M descomponent aquest en subespais ortogonals de longitud ca. $M/2$ en cada pas [22]. D'aquesta manera, incrementant el grau de descomposició a un nivell k (el qual significa repetir el procés de descomposició k vegades), el grau de compressió pot ser incrementat. No obstant, aquest increment del grau de compressió és a costa del grau de reconstrucció, donat que la fidelitat en la reproducció del senyal és degradada en cada pas.

Al igual que en el cas de la FT, la WT també presenta com a avantatge el poder ser utilitzada per eliminar el soroll present en els senyals. A més, la WT pot ser utilitzada com a mètode de compressió de dades si només utilitzem els coeficients més significatius obtinguts en la descomposició del senyal. Per tant, es tracta d'una eina molt eficient en aquestes situacions donat que ofereix l'avantatge de realitzar la compressió de les dades, l'extracció de característiques i la reducció del soroll en una sola etapa [22]. Al igual que en el cas anterior, la transformació pot ser revertida i avaluar d'aquesta manera el grau de reconstrucció, fet que permet l'elecció del número òptim de coeficients a utilitzar.

3.4.1.8. Grau de reconstrucció del senyal

En els casos que és possible desfer la transformació aplicada al senyal i reconstruir aquest a partir dels coeficients seleccionats, es va utilitzar el seu grau de reconstrucció per tal de determinar el número òptim de coeficients que calia emprar. Per tal de

quantificar el grau de reconstrucció, es va utilitzar dos paràmetres diferents: el coeficient de determinació (R^2) entre el senyal reconstruït i el senyal original, i un factor de comparació anomenat f_c que considera l'àrea sota els dos voltamperogrames quan aquests dos són superposats [22].

En ambdós casos, abans de calcular els dos paràmetres es realitza un desplegament (*unfolding*) del voltamperograma, sense tenir en compte el potencial del qual prové el corrent mesurat (Figura 1.5). El primer paràmetre (R^2) prové de realitzar la regressió lineal punt a punt i calcula la correlació entre ambdós senyals. El factor f_c es defineix com la relació entre l'àrea de la intersecció entre les dos corbes respecte l'àrea total sota les dos corbes (Figura 3.6). Des del punt de vista teòric, essent A i B l'àrea sota cadascuna de les dos corbes, el factor f_c es pot expressar d'acord amb l'Equació 3.8. Al igual que R^2 , aquest factor varia entre 0 i 1 depenent de la similitud entre els dos senyals; essent 0 quan els aquests dos no tenen res en comú i incrementant el seu valor al mateix temps que la similitud entre ambdós ho fa. En aquest sentit, el factor f_c mesura la semblança entre els dos senyals d'una forma similar al coeficient R^2 , però amb l'avantatge de ser més sensible a les petites diferències.

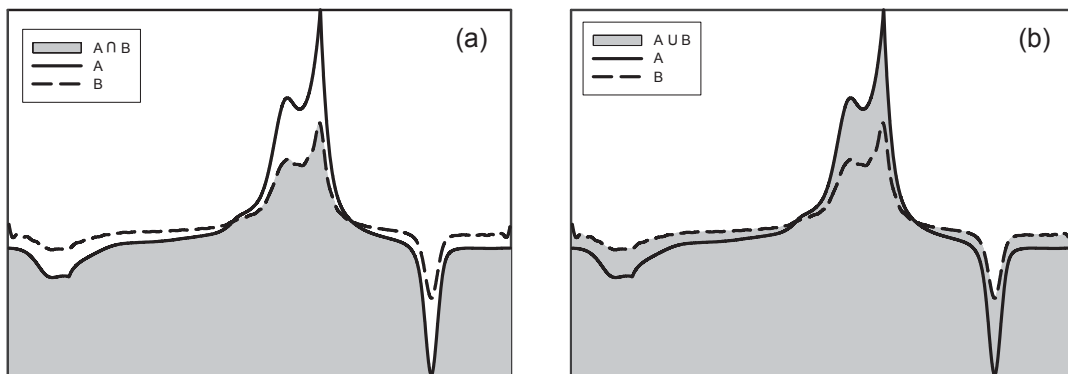


Figura 3.6. Esquema del procés de càlcul del factor f_c : (a) $A \cap B$ i (b) $A \cup B$.

$$f_c = \frac{A \cap B}{A \cup B} = \frac{\sum_{i=1}^M (\max\langle a_i, b_i \rangle - |a_i - b_i|)}{\sum_{i=1}^M (\min\langle a_i, b_i \rangle + |a_i - b_i|)} \quad \text{Equació 3.8}$$

3.4.2. Construcció i validació del model

Un cop definit el conjunt de coeficients o variables que seran utilitzats com a *inputs* en el nostre model, el següent pas és la modelització de les variables *target* a partir

d'aquests. No obstant, el primer factor que cal tenir en compte abans de la construcció del model és el tipus d'informació que es vol extreure de les mostres. És a dir, si la informació que pretenem extreure mitjançant la llengua electrònica és del tipus qualitatiu (classificació o distinció de diferents mostres) o quantitatiu (quantificació d'un o més compostos o índexs a partir de models multivariable), ja que aquest fet determinarà l'eina quimiomètrica a utilitzar a l'hora de construir el nostre model de resposta (§1.3).

En tots els casos, el total de punts experimentals és dividit en dos subconjunts abans de procedir a la construcció del model. El primer normalment conté ca. 2/3 parts del total de punts experimentals i és el que s'utilitza per al modelatge. Habitualment s'anomena aquest subconjunt com a entrenament (*train*), donat que és amb aquestes dades que l'eina quimiomètrica intenta modelar el sistema, tot ajustant l'equació matemàtica. El segon subconjunt conté el terç restant de punts experimentals, i correspon a una validació del modelatge (*test*). Aquest subconjunt de validació no participa en el procés d'entrenament, i per tant suposa una validació externa de punts no emprats anteriorment. És amb la interpolació d'aquest subconjunt de validació on es manifesta la capacitat de predicció presentada pel model creat. Una bona correlació amb els valors reals d'aquest subconjunt indica que s'ha assolit un model capaç de predir situacions contingudes dins els límits del subconjunt d'entrenament, però que no són pròpiament els valors emprats pel modelatge. D'altra banda, una mala correlació d'aquest posa de manifest la falta de capacitat del model per descriure correctament el nostre sistema, ja sigui per la falta de relació entre les mesures i les variables a predir o per un sobreentrenament d'aquest (§1.4.3.5.4ii).

Tanmateix, amb altres algorismes d'entrenament és típic utilitzar un tercer subconjunt de dades, el conjunt de validació interna, utilitzat per detectar la situació de sobreentrenament i aturar l'afinament del model en cas que aquesta es produeixi. No obstant, en el cas de les ANNs, cal dir que aquesta variant no cal utilitzar-la quan l'algorisme d'entrenament emprat és la Regularització Bayesiana, tal i com s'ha explicat a la introducció (§1.4.3.5.4), ja que aquesta implícitament ja assegura evitar el sobreentrenament [23].

En concret, al llarg dels següents apartats es descriurà l'ús del PCA, el LDA i les ANNs els quals constitueixen les eines de modelatge més utilitzades al llarg de la present tesi; havent estat ja proporcionada una breu descripció dels fonaments de la resta de mètodes a la introducció (§1.3.4). A més, també es presentaran les diferents

estratègies utilitzades en el procés de validació dels models generats, així com els paràmetres utilitzats com a indicadors de la qualitat d'aquests.

3.4.2.1. Anàlisi qualitativa

Per dur a terme l'anàlisi qualitativa de les diferents mostres s'han emprat dos mètodes diferents: l'anàlisi de components principals (PCA) i l'anàlisi discriminant lineal (LDA). El LDA està relacionat amb el PCA en el fet que tots dos mètodes busquen combinacions lineals de les variables que millor expliquin les dades [24]. La principal diferència entre aquestes dos tècniques radica en la tasca d'aprenentatge: en el cas del PCA és tracta d'un mètode de reconeixement de patrons no supervisat (i.e. no hi ha procés d'aprenentatge, fet pel qual el PCA per si sol només actua com una eina de visualització; requerint d'una eina de modelatge per a poder actuar com a classificador, e.g. PCA-ANN); mentre que en el cas del LDA, es tracta d'un mètode supervisat (i.e. existeix una etapa de modelatge d'acord amb un algorisme d'aprenentatge). És a dir, el LDA intenta explícitament construir un model que permeti explicar la diferència entre les diferents classes del conjunt de mostres, mentre que el PCA no té en compte qualsevol diferència basada en la classe.

Aquesta classificació dels mètodes és basa en la manera en que les funcions que modelen les dades són construïdes. D'una banda, en els mètodes supervisats les dades d'entrenament consisteixen en la matriu de resposta pel subconjunt d'entrenament (una fracció del conjunt de casos que són utilitzats per generar el model) més la sortida desitjada per a aquests casos. Així, el model és construït tenint en compte els paràmetres que millor prediuen la sortida desitjada. Posteriorment, una vegada generat el model, la seva resposta és avaluada emprant els casos restants no utilitzats en l'etapa d'entrenament (subconjunt de test). D'altra banda, en el cas dels mètodes no supervisats només la matriu de respostes (en la seva totalitat, i sense fer la divisió entre entrenament i test) és proporcionada al model (*learner*), sense cap tipus d'etiqueta. Presentant així només una millor representació visual de les relacions entre les mostres i les variables, i proporcionant, a partir de les variables de mesura, un nou punt de vista de com certes mostres mostren similituds entre elles, o com es diferencien les unes de les altres. Per aquesta raó, el PCA s'utilitza normalment només com a eina de visualització que permet comprovar si les mostres s'agrupen conjuntament en funció de les classes, i no pot ser considerat per si sol com un mètode de reconeixement de patrons o classificador.

3.4.2.1.1. Anàlisi de Components Principals (PCA)

L'anàlisi de components principals (PCA) és un procediment matemàtic que utilitza una transformació ortogonal per a convertir un conjunt d'observacions de variables possiblement correlacionades en un conjunt de valors de variables linealment no correlacionades, anomenades components principals (PCs) [25]. Així, la informació continguda en les variables originals és projectada en un menor número de variables adjacents ("latents"), els PCs, amb noves coordenades sobre aquest eixos anomenades *scores*, obtinguts després de realitzar la transformació matemàtica (Figura 1.6).

Tot i que la transformació no implica la reducció en el nombre de variables obtingudes, sinó només un canvi de sistema de coordenades, ja que obtenim un vector de n *scores* amb una coordenada per cadascun dels PCs; l'avantatge és que aquests nous eixos es generen de tal manera que cobreixin la major variància del conjunt de dades original. És a dir, el primer PC (PC1) és la direcció a través de les dades que explica/cobreix la major variabilitat en les dades possible (i.e. el PC1 té màxima variància); el segon component (PC2) és ortogonal (i.e. no relacionat) al PC anterior i cobreix tota la variabilitat restant que sigui possible (considerant la restricció d'ortogonalitat), i així successivament.

Un cop coneixem les direccions dels PCs és simple geometria la que ens permet expressar els valors de les mostres individuals en termes dels PCs com a sumes lineals de les dades originals, multiplicant per un coeficient que descriu el PC. Aquests nous valors se'ls coneix com a *scores*, i cada mostra té un *score* per a cada PC (tal com tenia un valor per a cadascuna de les variables de potencial en les dades originals).

Per tant, amb aquest procediment aconseguim expressar la major part de la variabilitat del conjunt de dades inicial amb un nombre reduït de variables utilitzant només els primers PCs. És per això, que en estadística, el PCA és una tècnica utilitzada per reduir la dimensionalitat d'un conjunt de dades; permetent a més, detectar les fonts de variabilitat d'aquestes i ordenar-les segons la seva importància.

No obstant, tot i la seva simplicitat teòrica, cal tenir en compte la complexitat de les eines matemàtiques necessàries per fer-ne el càlcul, les quals són relativament complexes (al igual que en la resta d'eines quimiomètriques); motiu pel qual no s'entrarà en l'aspecte matemàtic en cap dels mètodes descrit.

Finalment, a l'hora d'aplicar el PCA cal tenir en compte que aquest és sensible a l'escala relativa de les variables originals, i que només es garantirà la condició de

independència dels PCs si el conjunt de dades esta distribuït normalment de forma conjunta.

3.4.2.1.2. *Anàlisi Discriminant Lineal (LDA)*

El LDA és un mètode de classificació supervisat basat en la fórmula de Bayes que construeix un model predictiu per avaluar la pertinença al grup d'una mostra. El model està format per $k-1$ funcions discriminants lineals (essent k el nombre de grups, i generant un nou eix per a cada funció) basades en combinacions lineals de les variables predictorres que proporcionin la millor discriminació entre els grups.

En essència, enlloc de generar nous eixos basats en les direccions de màxima variància de la matriu de resposta, tal com fa el PCA, el LDA genera els nous eixos basats en la màxima discriminació entre els grups de mostres (Figura 1.7). D'aquesta manera, les mostres són agrupades tenint en compte la distància entre les observacions i els centroides del grup, la qual pot ser mesurada utilitzant la distància de Mahalanobis.

La distància de Mahalanobis s'utilitza per a determinar la similitud d'una mostra a un conjunt ja conegut donat que a diferència de la distància euclidiana, aquesta té en compte la covariància entre les variables en el càlcul de distàncies [26]. Aquest fet permet solucionar els problemes d'escala i la correlació inherent a la distància euclidiana, corregint les respectives escales de les diferents variables i tenint en compte la correlació entre les variables. En aquest sentit, la distància de Mahalanobis proporciona una mesura de la distància equivalent a l'ús de la desviació estàndard per mesures univariables; tenint en compte el fet que les desviacions en cada direcció són diferents i la covariància entre les variables.

Per tant, per la construcció del model, les funcions són generades a partir d'una mostra de casos (subconjunt d'entrenament) per als quals es coneix la pertinença al grup. Per posteriorment, aplicar les funcions obtingudes a nous casos dels quals coneixem les variables de predicció (i.e. la resposta obtinguda amb els sensors), però dels quals desconexim el grup de pertinença.

A més, una altra possibilitat en el moment de realitzar l'anàlisi LDA, és la utilització d'un mètode d'inclusió per passos [27], el qual és molt útil per tal de seleccionar i eliminar les variables que no contribueixen en absolut (o ho fan en menor grau) en l'èxit de la predicció. Així, tenint una llista de variables independents, algunes de les quals poden ser predictors útils, però algunes de les quals són segurament gairebé inútils,

l'objectiu és trobar el millor subconjunt que permeti completar la tasca de predicció, tan bé com sigui possible, amb el menor nombre de variables possible.

L'elecció de les variables es basa en criteris estadístics, entre els quals cal destacar l'ús de l'estadístic lambda de Wilk's (Λ) o la distància de Mahalanobis [27, 28]. El primer és utilitzat en anàlisi discriminant (DA) com a mesura de la separació entre els centroides de les diferents classes. Així, aquest mètode és utilitzat en LDA per seleccionar les variables que seran introduïdes en l'equació generada pel model basant-se en la seva contribució a disminuir el valor de Λ . D'aquesta manera, en cada pas la variable que minimitza la Λ global és introduïda fins que s'assoleix una eficàcia (*performance*) òptima (millor separació entre les classes). De forma anàloga, la distància de Mahalanobis proporciona una mesura de quant difereix de la mitjana per tots els casos els valors en les variables independents de un cas concret. Una distància de Mahalanobis gran identifica un cas que tingui valors extrems per una o més de les variables independents. D'aquesta manera, en cada etapa, les variables que maximitzen la distància de Mahalanobis entre dos grups propers són introduïdes fins que s'assoleix una eficàcia òptima.

La lambda de Wilk's es defineix com la proporció de les inèrcies del grup que no són explicades per la variable de resposta (la qual identifica les classes) en l'esquema clàssic de la descomposició de la variància [27, 28]. Aquesta és per tant, la relació de la inèrcia intra-classe i la inèrcia total; conseqüentment, essent el seu valor restringit en l'interval entre 0 i 1. Si només una petita fracció de la inèrcia total no s'explica per l'existència de grups, llavors aquests grups estan ben separats, i les seves mitjanes són significativament diferents. Per tant, un valor petit (proper a 0) de Λ significa que els grups estan ben separats, mentre que, un valor gran (proper a 1) de Λ significa que els grups no estan ben separats.

Recapitulant, i de forma resumida, la principal diferència entre ambdós models (PCA i LDA) és que en lloc de generar un nou sistema d'eixos basat en les direccions de màxima variància de la matriu de resposta com fa el PCA, el LDA genera el nou sistema d'eixos basant-se en la màxima discriminació entre els grups de la mostra (Figura 1.7); és a dir, l'ús d'un mètode no supervisat en front d'un supervisat. Causa de la qual se'n deriva la principal diferència, com és el fet que el PCA no pot actuar com a classificador per si sol, mentre que el LDA sí.

D'altra banda, una altra diferència radica en el nombre d'eixos generats. És a dir, en el cas del PCA es generen tants nous "eixos" (components principals) com a variables

originals; mentre que en el cas del LDA el model estarà format per només $k-1$ funcions discriminants (els nous eixos, sent k el nombre de grups), on aquestes funcions corresponen a combinacions lineals de les variables predictores que proporcionen la millor discriminació entre els grups.

3.4.2.2. Anàlisi quantitativa

Anàlogament a l'anàlisi qualitativa, l'objectiu en aquest cas serà la modelització de les variables *target* a partir dels coeficients extrets per tal d'obtenir informació quantitativa de les mostres. La principal diferència radica en la possibilitat de construir models multi-resposta; és a dir, la predicció d'un conjunt de variables *target* en disposició matricial a partir d'un únic model, mentre que en el cas qualitatiu cada un d'ells quedava restringit a un conjunt de *targets* vectorial. Aquesta etapa de modelatge pot ser assolida mitjançant l'ús de diferents mètodes lineals tals com el MLR, el PCR, el PLS o el nPLS o mètodes no lineals com les ANN, entre altres (vegeu §1.3.4).

Aquesta classificació dels mètodes es basa en la manera en que l'anàlisi de regressió té lloc. A saber, en els cas dels models no lineals, el conjunt de dades és modelat per una funció la qual és una combinació no lineal dels paràmetres del model i depèn d'una o més variables independents, les quals són ajustades mitjançant un mètode d'aproximacions successives. Mentre que en el cas dels mètodes lineals, es tracta d'un ajust de regressió lineal, ajustat normalment a través del mètode de mínims quadrats.

Donada la llarga experiència del nostre grup en la utilització de les ANNs, aquestes van ser escollides com el mètode utilitzat per defecte per dur a terme el processament de les dades (construcció dels models de resposta); utilitzant únicament els altres mètodes amb fins comparatius en ocasions puntuals. Per tant, en aquest apartat ens centrarem en la discussió de l'ús de les ANNs com a eina de modelització. No obstant, una explicació més profunda, així com detalls del procediment matemàtics dels diferents mètodes es pot trobar a [6, 29-31]. Altrament, cal tenir en compte que aquest és el mètode més "complex" des del punt de vista d'aplicabilitat, ja que requereix d'una etapa d'optimització del mateix molt més extensa que en el cas dels mètodes lineals, en els que només cal optimitzar el nombre de components principals o variables latents utilitzades per construir el model.

3.4.2.2.1. Xarxa Neuronal Artificial (ANN)

Tal com ja s'ha explicat, les ANNs representen una eina de modelització molt poderosa, amb grans avantatges degut a la seva flexibilitat i capacitat d'adaptació (i.e. utilització tant de funcions lineals com no lineals, capacitat de variació de la seva arquitectura per tal de mimetitzar millor el procés d'aprenentatge, adaptació tant dels pesos com del biaix entre les connexions de les diferents unitats de processament, etc.); però aquest avantatge suposa a la vegada el seu major desavantatge. És a dir, la complexa infraestructura del model, requereix de certs coneixements previs per tal de seleccionar la topologia o arquitectura de la xarxa i d'un esforç significatiu en l'optimització d'aquesta, donada la dificultat de predir la configuració òptima de la xarxa per endavant [32, 33].

Una altra limitació d'aquests sistemes és la dimensionalitat de la matriu d'inputs, la qual permet treballar només amb dades monodimensionals. Per tant, en el nostre cas en el que la matriu de dades obtinguda presenta una major dimensionalitat (potencial aplicat x sensors x mostres), es requereix d'una etapa prèvia per tal de reduir la dimensió de les mateixes.

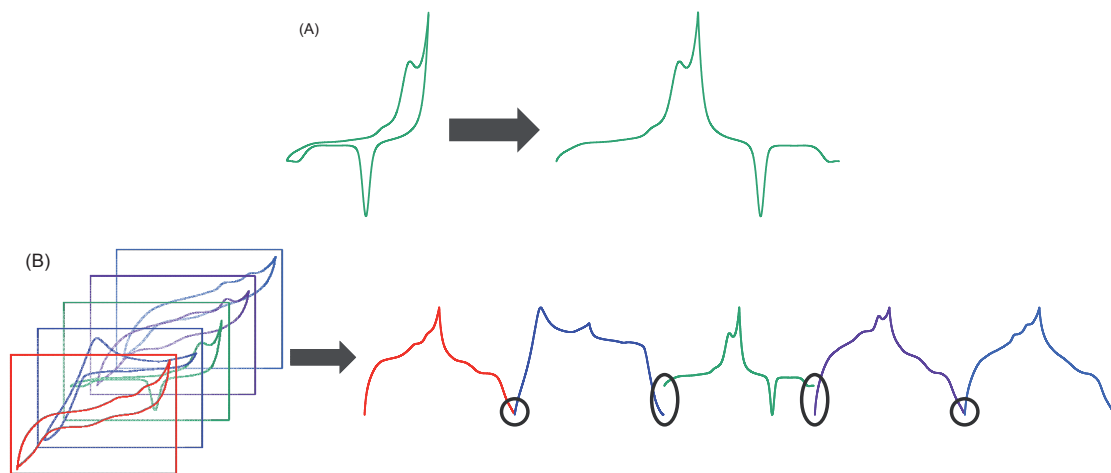


Figura 3.7. Esquema del procés (A) d'*unfolding* d'un voltamperograma i (B) d'*unfolding* dels voltamperogrames de diferents sensors; on es pot observar certes discontinuïtats generades entre la resposta de cada sensor.

El procés més habitual en aquests casos és el desplegament de les dades (*unfolding*) per passar d'una matriu tridimensional (3D) a una bidimensional (2D); aquest procés simplement consisteix en ajuntar les dades de la coordenada de sensors en una mateixa coordenada (Figura 3.7). De la mateixa manera, cada un dels voltamperogrames és "obert" o "desplegat" (*unfolded*), menyspreant el potencial del qual prové el corrent

mesurat i tenint en compte només la posició d'aquest. Evidentment, no es tracta d'una situació idònia, ja que aquest procés pot generar discontinuïtats d'un enregistrament al següent (e.g. la unió de dos voltamperogrames provinents de dos sensors diferents), el qual pot causar la pèrdua de la relació inicial i crear limitacions per al calibratge de primer ordre. En aquest sentit, certs mètodes trilineals com el nPLS intenten sobrepassar aquestes limitacions; no obstant això, tal com queda demostrat al llarg de la present tesi i concretament en §4.5 la capacitat de les ANNs és suficient per sobrepassar aquestes limitacions.

A més a més, cal tenir en compte que per tal d'obtenir una bona capacitat de generalització en el model generat i no excedir-se en la quantitat de càlculs requerits per tal d'arribar al punt òptim (ergo en el temps de processament), cal fer ús de mètodes que permetin reduir el nombre de pesos que cal ajustar (connexions), mantenint al mateix temps la capacitat de resoldre el problema. Això implica, fer ús de mètodes de preprocessament tals com els descrits a §3.4.1 per tal de reduir el nombre d'*inputs* i no excedir-se en el nombre de neurones a la capa oculta.

Altrament, abans de l'obtenció del model final, caldrà dur a terme una etapa d'optimització de l'arquitectura de la xarxa neuronal; es tracta d'un procés d'assaig-error en el que els diferents paràmetres de la topologia d'aquesta (algorismes d'entrenament, número de neurones, funcions de transferència, etc.) són ajustats per tal de trobar la millor configuració que optimitzi les prestacions del model neuronal.

En el nostre cas, en tots els models construïts, s'ha fet ús d'una única capa oculta ja que aquesta ofereix resultats prou satisfactoris i per tant no és necessari l'ús de sistemes més complexos amb més d'una capa oculta; tot i que, cal tenir present que aquesta és una configuració susceptible de ser utilitzada en casos d'elevada complexitat de les dades d'entrada. Així doncs, el número de neurones en la capa oculta serà un paràmetre que caldrà avaluar i optimitzar durant el procés de construcció del model, ja que aquest és un paràmetre que afecta substancialment els resultats finals. D'altra banda, el número de neurones en la capa sortida vindrà determinat pel nombre de paràmetres que volem avaluar mitjançant el model, mentre que el nombre de neurones en la capa d'entrada vindrà determinat pel número de coeficients o variables que decidim utilitzar com a *inputs*; quedant així ambdós prefixats, excepte en determinats casos en els que també podem utilitzar la ANN com a eina d'optimització del número de coeficients o variables que cal emprar (§3.4.1.2iv).

Paral·lelament, un altre dels paràmetres que cal avaluar és la selecció de les funcions de transferència en la capa oculta i en la capa de sortida. Entre aquestes hi ha una gran varietat de possibilitats, però basant-nos en l'experiència prèvia al grup i obviant certes funcions de tipus estadístic o de distribució de poblacions, s'han utilitzat només les funcions que presentaven una major avinença amb la naturalesa de les dades. A continuació s'il·lustren les diferents funcions de transferència utilitzades (Figura 3.8).

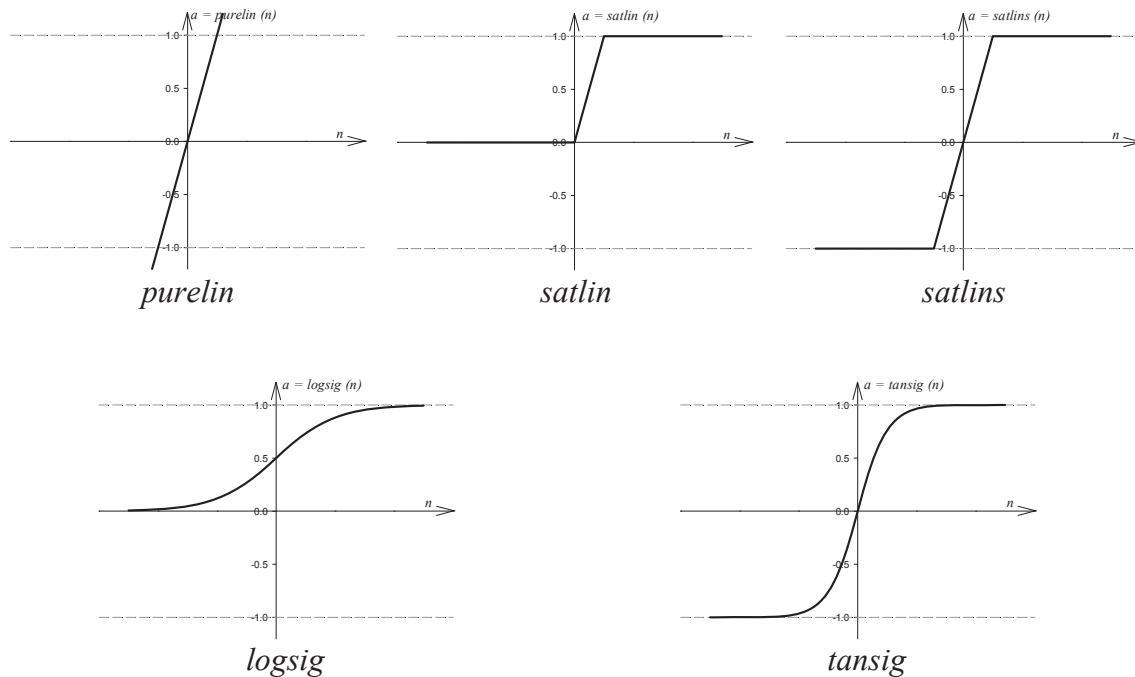


Figura 3.8. Representació de les diferents funcions de transferència. Les 3 primeres corresponen a funcions lineals (part superior), mentre que les 2 restants a funcions sigmoïdes (part inferior).

Finalment, altres paràmetres susceptibles de ser variats, van ser predefinits a partir de l'experiència prèvia en el nostre grup. Entre aquests, es troba el tipus de xarxa (propagació unidireccional (*feed-forward*) que combina un algorisme d'aprenentatge per retropropagació d'errors (*back-propagation*) amb múltiples sortides), la utilització d'una única capa oculta, la normalització de les dades, l'algorisme d'entrenament (regularització Bayesiana), la taxa d'aprenentatge ($\alpha=0.01$), el moment ($\beta=0.4$) o l'error a assolir (0.3) [34].

El conjunt de tots aquests paràmetres són els que determinen com té lloc el procés d'aprenentatge del model neuronal, és a dir, la cerca dels pesos i biaixos per cada una de les neurones que conduiran a l'ajust de l'equació final. Tot i que la seva variació pot influir en els resultats finals, aquesta no ha estat avaluada ja que es tracta d'un procés molt complex i amb molta menys influència sobre els resultats finals, degut a que

aquests han estat prèviament optimitzats per oferir els millors resultats i la seva variació només induiria a la construcció d'un major nombre de models (i.e. degut a l'increment exponencial derivat de la combinació dels diferents factors). D'altra banda, la regularització Bayesiana va ser escollit com l'algorisme d'entrenament ja que aquest presenta l'avantatge intrínsec d'evitar el sobreentrenament del model sense la necessitat de controlar la bondat de l'ajust mitjançant l'ús d'un subconjunt de validació interna, i per tant sense requerir d'aquesta precaució [23].

Així doncs, la resta de paràmetres van ser variats sistemàticament provant totes les combinacions possibles entre ells per tal d'aconseguir la major eficàcia (*performance*) en el model final. Cal tenir en compte l'elevat nombre de combinacions que es poden generar, i que l'optimització de l'arquitectura final de la xarxa passa doncs per l'estudi exhaustiu de tots aquests possibles casos. Resumint, els següents valors i combinacions d'aquests paràmetres són els que van ser considerats per dur a terme l'optimització per cada un dels diferents models ajustats:

- Número de neurones en la capa d'entrada: dependrà del número de coeficients o variables preseleccionats; en alguns casos concrets, es va utilitzar la ANN per dur a terme l'optimització també d'aquests
- Número de neurones en la capa oculta: es va variar entre 1 i 12
- Número de neurones en la capa de sortida: dependrà exclusivament de l'aplicació concreta que duem a terme, i el seu valor no variarà per un mateix cas
- Funcions de transferència en la capa oculta: *tansig*, *logsig* i *purelin*
- Funcions de transferència en la capa de sortida: *tansig*, *logsig*, *purelin* i *satlins*

Per dur a terme l'elecció de l'arquitectura final, s'ajusta un model amb les dades d'entrenament per cada una de les possibles estructures de la ANN i s'avalua el comportament d'aquest per les dades del subconjunt de test, les quals no han intervingut en cap moment en la generació del model i proporcionen una mesura objectiva de la capacitat de predicció d'aquest. Així doncs, a partir dels *inputs* introduïts per aquest subconjunt i fent ús del model generat, la ANN calcula les concentracions o índexs dels analítics que el model considera que corresponen a aquells inputs. Llavors, com nosaltres si que coneixem les concentracions reals (tot i que aquestes no han estat emprades pel model), podem construir a posteriori un gràfic comparatiu de concentracions predites vs. concentracions reals (§3.4.2.3.2.2). D'aquesta manera, a partir d'aquests, podem calcular certs paràmetres d'avaluació de la capacitat predictiva del model (§3.4.2.3.2.2),

els quals ens permetran determinar quina és la configuració òptima, és a dir, aquella que presenta una millor capacitat de predicció o ajust respecte les dades esperades.

Tot i que aquests paràmetres també són calculats pel subconjunt d'entrenament, cal evitar utilitzar aquest per la tria del model ja que en casos determinats en que augmentem molt la complexitat del model, podem arribar en situacions de sobreentrenament d'aquest que condueixin a la "memorització" de les dades, més que a l'ajust d'un model amb capacitat predictiva (§1.4.3.5.4ii). No obstant, cal comprovar que una bona tendència és obtinguda pels dos subconjunts, essent habitual l'obtenció de millors resultats pel conjunt d'entrenament.

Així doncs, aquesta elecció s'ha basat en la tria del model que ofereixi un menor RMSE (tant per cada analit concret com pel valor global), però que ofereixi també una bona tendència en els gràfics de comparació de concentracions obtingudes *vs.* esperades (valors propers al comportament ideal: ordenada a l'origen propera a 0, valors de pendent i coeficients de correlació propers a 1; i.e. concentracions esperades iguals a les concentracions predites). Adoptant així una solució de compromís, ja que és freqüent trobar ajustos amb errors menors però amb ajustos més pobres, i viceversa. Aquests casos solen aparèixer dins una mateixa configuració de funcions de transferència amb diferent nombre de neurones a la capa oculta.

3.4.2.3. Mètodes de validació

Un cop construït el model, cal avaluar la capacitat de predicció d'aquest. El procediment emprat es basa en comparar els valors obtinguts mitjançant el model amb els valors reals esperats, ja siguin valors quantitatius o qualitius; tot i que, els paràmetres que farem anar en cadascun dels dos casos diferiran lleugerament.

3.4.2.3.1. Subdivisió del conjunt de dades

Així doncs, en ambdós casos, i com ja s'ha comentat, el primer que cal fer és fer una divisió del conjunt de mostres en 2 subconjunts: entrenament (*train*) i validació (*test*). El primer conjunt serà utilitzat per construir el nostre model, aportant al sistema tant la matriu de resposta dels sensors (x) com la de *targets* (y), les quals s'utilitzaran per ajustar el model de resposta optimitzant els diferents paràmetres de cada mètode. El segon conjunt, servirà per avaluar la capacitat d'aquest i validar-lo, ja que només

proporcionarem al sistema la informació relativa a la matriu de resposta, calculant la “matriu de *targets*” amb el model ja construït. Posteriorment, i donat que nosaltres disposem de la matriu de *targets* reals, compararem els valors d’ambdues, el que permetrà avaluar la bondat de l’ajust del model de resposta.

Cal tenir en compte la importància del conjunt de validació, el qual no ha participat en cap moment en la generació del model de resposta, sent utilitzat com a subconjunt de mostres cegues per aquest, i que permetran la detecció de sistemes sobreentrenats (§1.4.3.5.4ii). És a dir, sistemes amb una modelització pel conjunt d’entrenament perfecta però amb una capacitat de predicció pèssima pel conjunt de validació. Aquesta situació és bastant comuna especialment en el cas de les ANN, sobretot quan es construeixen models amb una arquitectura molt complexa.

Un altre factor clau a l’hora de realitzar la validació del model és la forma en que es realitza la divisió de dades entre els dos subconjunts, així de forma resumida els mètodes emprats en la present tesi han estat els següents:

- Divisió *train/test*: consisteix en realitzar una única divisió de les mostres entre els dos subconjunts amb una proporció entre 60:40 i 80:20 pels subconjunts d’entrenament i validació respectivament. Aquest mètode és especialment útil en els casos en que treballem a partir d’un disseny experimental, en el que les mostres que formaran part del conjunt d’entrenament queden prèviament definides pel propi disseny; en cas de no ser així, la divisió és pot fer de forma aleatòria.
- *Leave-X-out* o *k-fold*: el conjunt de mostres originals és dividit de forma aleatòria en k subconjunts; tot i tractar-se d’una divisió aleatòria, s’intenta mantenir una proporció constant entre els diferents tipus de mostres per cadascun dels subconjunts generats. De tots els subconjunts generats, només un és utilitzat com a validació quedant fora del procés de modelització, utilitzant la resta de $k-1$ subconjunts per a l’entrenament o construcció del model. Llavors, aquest procés és repetit k vegades (d’aquí el nom), utilitzant cadascun dels subconjunts una única vegada com a conjunt de validació. Finalment, els resultats dels diferents *folds* són agrupats i promitjats per tal d’obtenir la matriu de predictors estimada pel model.

L’avantatge d’aquest mètode respecte una repetició aleatòria en la divisió dels subconjunts és que permet assegurar que totes les mostres intervenen tant en el

procés d'entrenament com en el de validació i que, a més, totes les mostres són utilitzades una única vegada en el procés de validació.

- *Leave-one-out*: una variació del cas anterior (on $k=1$) i possiblement la més estesa, en la qual es deixa només una mostra fora del procés de modelització cada vegada. D'aquesta manera, cada mostra és quantificada o classificada a partir del model generat per la resta de mostres, repetint el procés tantes vegades com mostres considerades en l'estudi.
- *Repeated random sub-sampling validation*: aquest mètode va ser suggerit originalment en estadística com una aproximació general per a la prova d'hipòtesis i el càlcul dels intervals de confiança en situacions on aparentment millors mètodes no podien ser utilitzats [35]. Amb aquest procediment, els errors estàndards són calculats a partir de la repetició de l'etapa de modelatge basada en un remostratge amb distribució aleatòria diferent cada vegada. D'aquesta manera, el mètode pot ser aplicat deixant tan sols una mostra fora, o inclús deixant varies mostres fora cada vegada [36].

Aquest mètode té l'avantatge que evita la dependència en les prediccions de la divisió específica emprada de les dades (*train/test*) i que el nombre de mostres en cada subconjunt és independent del nombre de *folds*, donat que aquest procés es repetit k vegades de forma aleatòria, avaluant la capacitat de resposta del model per cada un dels conjunts de test cada vegada [37]; posteriorment, les diferents respostes per cada un dels subconjunts són agrupades tal com es feia en el cas del *k-fold*, obtenint el nostre model de resposta, el qual permet a més calcular els intervals de confiança per a les prediccions [38].

D'altra banda, i tot i que no han estat utilitzats en la present tesi, altres mètodes de divisió de mostres utilitzats comunament són: el *bootstrapping*, el *jack-knife* i el *montecarlo*; tots ells presentant com a avantatge comú, el fet de permetre l'estimació dels intervals de confiança de les prediccions.

3.4.2.3.2. Paràmetres indicadors de la bondat de l'ajust

Finalment, per tal d'avaluar numèricament la capacitat de resposta del models s'ha realitzat una distinció segons el tipus de resposta que ens aporta el model, i.e. informació qualitativa o quantitativa.

3.4.2.3.2.1. *Models qualitius*

En el primer cas, per tal d'avaluar la capacitat d'aquest, el primer que fem és construir la matriu de confusió (Taula 3.2). Es tracta d'una eina de visualització que s'utilitza en l'aprenentatge supervisat, on cada columna representa els casos predits per aquella classe, mentre que cada fila correspon als casos reals d'aquella classe. S'anomena així, ja que permet identificar de forma ràpida i simple si el sistema està distingint o confonen alguna de les classes, és a dir, identificar quines mostres estan sent mal classificades i amb quin grup estan sent confoses. Un altra peculiaritat/avantatge d'aquesta, és el fet que totes les mostres ben classificades es troben en la seva diagonal, facilitant la identificació d'errors en la classificació ja que aquests estaran caracteritzats per presentar valors diferents de 0 fora de la diagonal.

Taula 3.2. Exemple de dos matrius de confusió diferents per un sistema hipotètic format per 100 mostres repartides en 3 classes: A, B i C. Addicionalment, es presenta els valors calculats pels paràmetres de grau d'encert, sensibilitat i especificitat.

| <i>Cas 1</i> | | | | | | | <i>Cas 2</i> | | | |
|----------------|----------|---------------|----------|----------|----------------------|-------------|---------------|----------|----------|--|
| | | <i>Predit</i> | | | | | <i>Predit</i> | | | |
| | | <i>A</i> | <i>B</i> | <i>C</i> | | | <i>A</i> | <i>B</i> | <i>C</i> | |
| <i>Esperat</i> | <i>A</i> | 34 | 0 | 0 | | | 90 | 0 | 0 | |
| | <i>B</i> | 0 | 30 | 3 | | | 5 | 0 | 0 | |
| | <i>C</i> | 1 | 6 | 26 | | | 5 | 0 | 0 | |
| | | | | | <i>Cas1</i> | <i>Cas2</i> | | | | |
| | | | | | <i>Grau d'encert</i> | 90.0% | 90.0% | | | |
| | | | | | <i>Sensibilitat</i> | 89.9% | 33.3% | | | |
| | | | | | <i>Especificitat</i> | 95.0% | 66.7% | | | |

Per tal de quantificar la bondat de l'ajust dels models qualitius, a banda de presentar la matriu de confusió, s'han utilitzat tres paràmetres addicionals els quals es calculen a partir d'aquesta, que són: grau d'encert (o percentatge de classificació correcta), sensibilitat i especificitat. El primer paràmetre és el que ens ofereix una dada sobre les prestacions generals del sistema, ja que ens proporciona una mesura directa de la proporció de mostres correctament classificades. Però no ens aporta gaire informació per si sola, ja que cal tenir en compte que aquesta dada pot ser enganyosa degut a que no té en compte el nombre d'objectes de cada classe. És a dir, si estem treballant en una situació en que cada grup tingui un nombre significativament diferent d'objectes, la bondat en la classificació dels grups majoritaris pot ocultar la manca d'encert en els grups minoritaris (e.g. Taula 3.2). Així doncs, per tal de definir millor la bondat d'ajust del model, cal complimentar el percentatge d'encert tenint també en compte el nombre de falsos negatius (*FN*) i falsos positius (*FP*), aquesta informació és proporcionada mitjançant el valor de la sensibilitat i especificitat del mètode. A continuació es detalla la definició i el càlcul detallat dels esmentat paràmetres:

- Taula de confusió, amb dos possibles casos: positiu (P) i negatiu (N). Aquesta simplificació és sempre realitzada per tal de calcular els diferents paràmetres, considerant com a positiu la classe d'interès i tota la resta de classes com a negatiu; extraient cada vegada els valors de grau d'encert, sensibilitat i especificitat per cada classe.

Taula 3.3. Taula (o matriu) de confusió per un sistema amb dos possibles cassos.

| | | <i>Predit</i> | |
|----------------|----------------|------------------------------|------------------------------|
| | | <i>Positiu</i> | <i>Negatiu</i> |
| <i>Esperat</i> | <i>Positiu</i> | Positiu verdader (TP) | Fals negatiu (FN) |
| | <i>Negatiu</i> | Fals positiu (FP) | Negatiu verdader (TN) |

- Grau d'encert: proporció total d'objectes que són classificats correctament per cada una de les classes (Equació 3.9). Indica simplement la proporció de casos correctament identificats sense tenir en compte els casos mal classificats ni la dimensió de cada grup.

$$\frac{\sum_{i=1}^n TP_i}{\sum_{i=1}^n TP_i + FP_i} \quad \text{Equació 3.9}$$

- Sensibilitat: proporció d'objectes de cada classe que són correctament identificats mitjançant el model; és a dir, la fracció de veritables positius (Equació 3.10). Indica la capacitat del nostre estimador per donar com a casos positius els casos realment positius, o proporció de positius correctament identificats.

$$\frac{1}{n} \sum_{i=1}^n \frac{TP_i}{TP_i + FN_i} \quad \text{Equació 3.10}$$

- Especificitat: proporció d'objectes de les altres classes que són rebutjats correctament pel model; la fracció de veritables negatius (Equació 3.11). Indica la capacitat del nostre estimador per donar com a casos negatius els casos realment negatius; o proporció de negatius correctament identificats.

$$\frac{1}{n} \sum_{i=1}^n \frac{TN_i}{TN_i + FP_i} \quad \text{Equació 3.11}$$

3.4.2.3.2.2. *Models quantitativus*

D'altra banda, per dur a terme l'avaluació de la bondat d'ajust dels models quantitativus s'utilitza dos mètodes diferents. El primer consisteix en la construcció dels gràfics de comparació entre la resposta predita pel model i la concentració esperada per a cadascun dels analits, mentre que el segon es basa en el càlcul de l'error quadràtic mitjà (RMSE) o l'error quadràtic mitjà normalitzat (NRMSE).

- Recta de regressió a partir dels gràfics de comparació: un cop construït el gràfic de concentracions predites *vs.* concentracions esperades, es realitza un ajust lineal dels dos conjunts de dades i es calcula el pendent (*m*), l'ordenada a l'origen (*b*) i el coeficient de correlació (*r*); utilitzant aquests tres paràmetres com a avaluadors de la capacitat de predicció del sistema mitjançant la comparació amb els valors ideals (1, 0 i 1, respectivament).

Per tal de facilitar la tria del model que ofereix millors prestacions a partir dels gràfics de comparació, i especialment en els casos en que duem a terme la multideterminació de diferents analits, es calcula la suma de la diferència entre el valor ideal i l'obtingut per cadascun dels paràmetres d'acord amb l'Equació 3.12 i l'Equació 3.13; on "X" correspon als valors de "m" o "r", i "j" a cadascun dels analits.

$$\Delta X = \sum_j |1 - X_j| \quad \text{Equació 3.12}$$

$$\Delta b = \sum_j |b_j| \quad \text{Equació 3.13}$$

$$\Delta X = \frac{1}{j} \sum_j |1 - X_j| \quad \text{Equació 3.14}$$

$$\Delta X = \frac{1}{j} \sum_j |b_j| \quad \text{Equació 3.15}$$

Així, en el cas del coeficients de correlació, donat que el valor d'aquest es troba restringit per definició entre 0 i 1, el valor de Δr estarà restringit entre 0 i "j"; essent 0 quan l'ajust és perfecte i augmentant fins a assolir el valor de "j" a mesura que la qualitat del model decau. Similarment, tot i no ser estrictament cert per definició, el mateix passa en el cas de Δm ; els valors del qual tendeixen a estar en el rang de 0 a "j". No obstant, el rang de valors de Δb dependrà de les unitats de mesura, essent millor el model com més proper sigui aquest a 0.

Adicionalment, també podem normalitzar el valor anterior d'acord amb el nombre d'analits per tal de que aquest es trobi en el mateix rang que els respectius paràmetres inicials: Equació 3.14 i Equació 3.15.

- El RMSE és una mesura de quan propera és la recta ajustada respecte els punts dels valors originals (Equació 3.16). Per cada punt es calcula l'error entre el valor extret de l'ajust i el valor real; és a dir, la distància des del punt real (c_i) al corresponent valor ajustat per aquest (\hat{c}_i), considerat respecte l'eix d'ordenades. Es realitza la suma d'aquests valors al quadrat, per tal d'evitar que els valors negatius compensin els valors positius, és divideix pel número de punts (n) menys 1 i es fa l'arrel quadrada per tal de deixar-ho en les unitats corresponents. En cas de voler calcular el RMSE total del model, caldrà tenir en compte també el número total d'analits a determinar (k), ja que per cada un dels analits tindrem n punts.

$$RMSE = \sqrt{\frac{\sum_{ij} (c_{ij} - \hat{c}_{ij})^2}{k \cdot n - 1}} \quad \text{Equació 3.16}$$

L'avantatge d'aquest mètode radica en el fet que el seu valor representa la distància mitjana respecte l'eix d'ordenades d'un punt a la línia ajustada, en les mateixes unitats que aquest; proporcionant així, una mesura de la bondat de l'ajust directament interpretable en termes d'unitats de mesura, a diferència del coeficient de correlació. Per tant, com més petit sigui el valor del RMSE, significa que la línia ajustada és més propera a les dades, ergo millor és l'ajust i la capacitat de predicció del model.

- NRMSE: de forma similar i com a alternativa al paràmetre anterior, una altra opció és normalitzar aquest dividint pel rang de concentracions de cada analit (Equació 3.17); obtenint així, un nou indicador entre 0 i 1 que ens proporcionarà una mesura de l'error per cada analit ($j=1$) o l'error global pel conjunt dels analits ($j>1$).

$$NRMSE = \frac{1}{j} \sum_i \frac{RMSE_i}{c_{i,\max} - c_{i,\min}} \quad \text{Equació 3.17}$$

Tot i haver dit que l'avantatge del RMSE és el fet de que proporciona una mesura de la bondat de l'ajust en les mateixes unitats que les pròpies dades, això pot suposar un desavantatge en els casos que vulguem avaluar la capacitat global del model per predir diferents analits de forma simultània, ja que si el valors de concentracions són diferents per cada un dels analits, el valor del RMSE total reflectirà majoritàriament el RMSE de l'analit amb valors de concentracions més grans. D'aquesta manera, la normalització d'aquest representa una solució a aquest problema i continua donant una idea de l'error, ja que passem d'una mesura d'error absolut a una mesura d'error relatiu.

Finalment, un altre indicador que també s'ha utilitzat en alguns casos per avaluar la capacitat de predicció del model ajustat ha estat el grau de recuperació (*recovery yield*, Equació 3.18). Tot i que aquest paràmetre no proporciona en si un indicador de la bondat d'ajust del model, ja que proporciona informació de cada un dels analits i/o mostres per separat, si que és útil en el cas de les mostres reals o sintètiques per veure de forma més detallada l'error comés en cada una de les determinacions.

$$\% = \frac{\text{trobat}}{\text{esperat}} \cdot 100 \qquad \text{Equació 3.18}$$

El seu càlcul és molt senzill i consisteix en la relació entre la concentració determinada pel model i l'esperada; essent 100% quan no hi ha diferències entre ambdós, superior a aquest en els casos de sobredeterminació o inferior si els valors trobats són inferiors als esperats. A més, també tenim la possibilitat de realitzar la mitjana del valor obtingut per les diferents mostres, obtenint un valor global de la tendència del model.

3.4.3. Software

El control del potenciostat i l'adquisició de dades és va realitzar a través dels programes *GPES 4.7* i *GPES Multichannel 4.7* (Autolab, Ecochemie, Netherlands). El processament quimiomètric es va dur a terme a través de rutines específiques utilitzant diferents versions del programa Matlab (MathWorks, Natick, MA), així com els respectius paquets *Neural Network Toolbox*, *Wavelet Toolbox* i *Statistics Toolbox*.

D'altra banda, per l'anàlisi i representació dels resultats és va utilitzar el programa Sigmaplot (Systat Software Inc, California, USA).

3.5. Bibliografia

- [1] F. Céspedes, E. Martínez-Fàbregas, S. Alegret, New materials for electrochemical sensing I. Rigid conducting composites, *TrAC-Trend. Anal. Chem.* **1996**, *15*, 296.
- [2] R. Olivé-Monllau, M. Baeza, J. Bartrolí, F. Céspedes, Novel Amperometric Sensor Based on Rigid Near-Percolation Composite, *Electroanal.* **2009**, *21*, 931.
- [3] A. Mimendia, Desenvolupament i aplicació de llengües electròniques per anàlisi ambiental, Ph.D. Thesis, *Universitat Autònoma de Barcelona*, **2011**
- [4] J.M. Palacios-Santander, L.M. Cubillana-Aguilera, M. Cocchi, A. Ulrici, I. Naranjo-Rodríguez, R. Seeber, J.L. Hidalgo-Hidalgo de Cisneros, Multicomponent analysis in the wavelet domain of highly overlapped electrochemical signals: Resolution of quaternary mixtures of chlorophenols using a peg-modified Sonogel–Carbon electrode, *Chemometr. Intell. Lab.* **2008**, *91*, 110.
- [5] X. Cetó, J.M. Gutiérrez, M. Gutiérrez, F. Céspedes, J. Capdevila, S. Mínguez, C. Jiménez-Jorquera, M. del Valle, Determination of total polyphenol index in wines employing a voltammetric electronic tongue, *Anal. Chim. Acta* **2012**, *732*, 172.
- [6] S.M. Scott, D. James, Z. Ali, Data analysis for electronic nose systems, *Microchim. Acta* **2006**, *156*, 183.
- [7] Z. Boger, Selection of quasi-optimal inputs in chemometrics modeling by artificial neural network analysis, *Anal. Chim. Acta* **2003**, *490*, 31.
- [8] A. Gutiérrez, F. Céspedes, S. Alegret, M. del Valle, Determination of phenolic compounds by a polyphenol oxidase amperometric biosensor and artificial neural network analysis, *Biosens. Bioelec.* **2005**, *20*, 1668.
- [9] A. Gutiérrez, A.B. Ibáñez, M. del Valle, F. Céspedes, Automated SIA e-Tongue Employing a Voltammetric Biosensor Array for the Simultaneous Determination of Glucose and Ascorbic Acid, *Electroanal.* **2006**, *18*, 82.
- [10] H.H. Harman, *Modern Factor Analysis*, 3rd Ed. ed., University of Chicago Press, Chicago, **1976**.
- [11] J.-M. Vesin, R. Grüter, Model selection using a simplex reproduction genetic algorithm, *Signal Process.* **1999**, *78*, 321.
- [12] Z. Xiaobo, Z. Jiewen, M.J.W. Povey, M. Holmes, M. Hanpin, Variables selection methods in near-infrared spectroscopy, *Anal. Chim. Acta* **2010**, *667*, 14.
- [13] J.R.M. Smits, W.J. Melssen, L.M.C. Buydens, G. Kateman, Using artificial neural networks for solving chemical problems: Part I. Multi-layer feed-forward networks, *Chemometr. Intell. Lab.* **1994**, *22*, 165.
- [14] P.R. Bevington, D.K. Robinson, *Data reduction and error analysis for the physical sciences*, 3 ed., McGraw-Hill, Boston, MA, **2003**.
- [15] J. Simons, M. Bos, W.E. van der Linden, Data processing for amperometric signals, *Analyst* **1995**, *120*, 1009.
- [16] M. Cortina, A. Duran, S. Alegret, M. del Valle, A sequential injection electronic tongue employing the transient response from potentiometric sensors for anion multidetermination, *Anal. Bioanal. Chem.* **2006**, *385*, 1186.
- [17] R. Gutiérrez-Osuna, H.T. Nagle, A method for evaluating data-preprocessing techniques for odour classification with an array of gas sensors, *IEEE T. Syst. Man Cy. B* **1999**, *29*, 626.

- [18] V. Tchistiakov, C. Ruckebusch, L. Duponchel, J.-P. Huvenne, P. Legrand, Neural network modelling for very small spectral data sets: reduction of the spectra and hierarchical approach, *Chemometr. Intell. Lab.* **2000**, *54*, 93.
- [19] S.G. Mallat, A theory for multiresolution signal decomposition - the wavelet representation, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **1989**, *11*, 674.
- [20] Wavelets: Classification, Theory, and Applications, Nova Science Pub Inc, New York, US, **2011**.
- [21] A. Grossmann, J. Morlet, Decomposition of hardy functions into square integrable wavelets of constant shape, *SIAM J. Math. Anal.* **1984**, *15*, 723.
- [22] L. Moreno-Barón, R. Cartas, A. Merkoçi, S. Alegret, M. del Valle, L. Leija, P.R. Hernandez, R. Muñoz, Application of the wavelet transform coupled with artificial neural networks for quantification purposes in a voltammetric electronic tongue, *Sensor Actuat. B-Chem.* **2006**, *113*, 487.
- [23] M.H. Beale, M.T. Hagan, H.B. Demuth, in, MathWorks, **2012**.
- [24] R.O. Duda, P.E. Hart, D.G. Stork, Pattern classification, John Wiley & Sons, New York, USA, **2000**.
- [25] I.T. Jolliffe, Principal Component Analysis, Springer New York, New York, US, **2002**.
- [26] B.F.J. Manly, Multivariate Statistical Methods: A Primer, 3rd ed., Chapman & Hall, London, UK, **2005**.
- [27] R.A. Johnson, D.W. Wichain, Applied multivariate statistical analysis, Pearson Education, Harlow, GB, **2007**.
- [28] D.J. Hand, Discrimination and classification, John Wiley, Chichester, UK, **1981**.
- [29] E. Richards, C. Bessant, S. Saini, Multivariate Data Analysis in Electroanalytical Chemistry, *Electroanal.* **2002**, *14*, 1533.
- [30] K.H. Esbensen, Multivariate Data Analysis - in practice, CAMO Sosftware, Oslo, **2002**.
- [31] R. Bro, Multi-way analysis in the food industry. Models, algorithms and applications, Ph.D. Thesis, *University of Amsterdam*, **1998**
- [32] F. Despaigne, D. Luc Massart, Neural networks in multivariate calibration, *Analyst* **1998**, *123*, 157R.
- [33] D.R. Hush, B.G. Horne, Progress in supervised neural networks, *IEEE Signal Processing Magazine* **1993**, *10*, 8.
- [34] M.d. Valle, Chapter 30 Potentiometric electronic tongues applied in ion multidetermination, in: S. Alegret, A. Merkoçi (Eds.) Comprehensive Analytical Chemistry, Elsevier, **2007**, pp. 721.
- [35] J.W. Tukey, Bias and confidence in not-quite large samples, *Ann. Math. Stat.* **1958**, *29*, 614.
- [36] S.D. Peddada, 21 Jackknife variance estimation and bias reduction, in: C.R. Rao (Ed.) Handbook of Statistics, Elsevier North-Holland, New York, **1993**, pp. 723.
- [37] A.M. Molinaro, R. Simon, R.M. Pfeiffer, Prediction error estimation: a comparison of resampling methods, *Bioinformatics* **2005**, *21*, 3301.
- [38] J. Riu, R. Bro, Jack-knife technique for outlier detection and estimation of standard errors in PARAFAC models, *Chemometr. Intell. Lab.* **2003**, *65*, 35.

4. Resultats i discussió

4. Resultats i discussió

Amb l'objectiu de donar resposta a la creixent demanda per part del sector alimentari, i més concretament del sector vinícola, de nous mètodes analítics capaços d'oferir el màxim d'informació analítica amb el mínim "esforç" possible, i que permetin dur a terme la qualificació i/o categorització de mostres de forma anàloga a la que realitzaria un "sommelier" (tastador expert) o que permetin inclús quantificar paràmetres d'interès en aquestes, la present tesi s'ha enfocat en el desenvolupament de nous mètodes d'anàlisi basats en l'ús de llengües electròniques. Així, aquestes representen el desenvolupament de mètodes alternatius als mètodes convencionals, amb avantatges sobre aquests com pot ser el seu baix cost i temps d'anàlisi, susceptibles de ser utilitzats de forma rutinària en el control de processos de producció i per al cribratge (*screening analysis*), per tal de detectar qualsevol adulteració o contaminació del producte ja sigui durant la seva elaboració o per comprovar que compleixen els estàndards de control de qualitat.

Concretament la present tesi ha volgut donar resposta a la demanda de quatre dels sectors més importants en quant a begudes alcohòliques en el nostre país: vi, cava, cervesa i brandi; intentant satisfer les necessitats concretes de cada sector, i incrementant a més la complexitat a través de les diferents aproximacions.

D'aquesta manera, el primer cas presentat dona resposta a una problemàtica del sector vinícola, plantejada per part de l'INCAVI, a l'hora de classificar els diferents tipus de cava i detectar possibles fraus en la producció d'aquests.

De forma anàloga, sota l'assessorament de l'empresa *Miguel Torres, S.A.*, se'ns va proposar dos possibles nínxols d'aplicació en la producció dels brandis: el primer, l'avaluació de la capacitat de les ETs de predir les diferents qualitats d'aquests tal com realitzaria un panel sensorial, i el segon, la detecció de l'ús de diferents mètodes d'envelliment alternatius en el procés de producció d'aquests.

Donats els bons resultats obtinguts en el cas del cava, i amb la col·laboració de l'empresa *Art Cervecers company - Cervesa Artesanal*, es va procedir a l'aplicació del concepte de la llengua electrònica per a la discriminació dels diferents tipus de cervesa i avaluar inclús si era possible distingir aquestes en funció del mètode d'elaboració (artesanal o industrial). A més, en aquest sector i donada la importància que han adquirit els compostos fenòlics en els darrers temps, degut a les seves propietats tant

organolèptiques com saludables, es va procedir a la quantificació individual de mescles d'aquests compostos.

Finalment, i a través de col·laboracions puntuals amb l'INCAVI, es va treballar àmpliament el tema dels compostos fenòlics en vins, tant des del punt de vista individual com la quantificació d'índexs globals.

A continuació, es recullen les publicacions que formen part d'aquesta tesi doctoral, els resultats de les quals agrupats en funció del camp d'anàlisi i ordenats en complexitat creixent, es presenten i discuteixen en els següents apartats. Addicionalment, les diferents publicacions es troben íntegres en el capítol d'articles i annexos.

- i. Voltammetric Electronic Tongue in the analysis of Cava wines
Electroanalysis, **23**(1): 72-78
X. Cetó, J.M. Gutiérrez, L. Moreno-Barón, S. Alegret and M. del Valle
- ii. Cava wine authentication employing a voltammetric Electronic Tongue
manuscript
X. Cetó, F. Capdevila, S. Mínguez and M. del Valle
- iii. Voltammetric BioElectronic Tongue for the analysis of phenolic compounds in rosé cava wines
manuscript
X. Cetó, F. Capdevila, S. Mínguez and M. del Valle
- iv. Application of an Electronic Tongue towards the analysis of brandies
Analytical Methods, **5**(5): 1120-1129
X. Cetó, M. Llobet, J. Marco and M. del Valle
- v. Voltammetric Electronic Tongue for qualitative analysis of beers
manuscript
X. Cetó, J.M. Gutiérrez, A. Mimendia, F. Céspedes and M. del Valle
- vi. Assessment of individual polyphenol content in beer by means of a voltammetric BioElectronic Tongue
Electroanalysis, **25**(1): 68-76
X. Cetó, F. Céspedes and M. del Valle
- vii. Chapter 6: BioElectronic Tongues: When the sensor array incorporates biosensors
M. del Valle, X. Cetó and M. Gutiérrez
in *Multisensor systems for chemical analysis - materials and sensors* by L. Lvova, D. Kirsanov, C. Di Natale and A. Legin
Pan Stanford Publishing Pte Ltd. ISBN: 9789814411158

- viii. Determination of total polyphenol index in wines employing a voltammetric electronic tongue
Analytica Chimica Acta, **732**(0): 172-179
X. Cetó, J.M. Gutiérrez, M. Gutiérrez, F. Céspedes, J. Capdevila, S. Mínguez, C. Jiménez-Jorquera and M. del Valle
- ix. Bioelectronic Tongue for the quantification of total polyphenol content in wine
Talanta, **99**(0): 544-551
X. Cetó, F. Céspedes and M. del Valle
- x. Resolution of phenolic antioxidant mixtures employing a voltammetric bio-electronic tongue
Analyst, **137**(2): 349-356
X. Cetó, F. Céspedes, M.I. Pividori, J.M. Gutiérrez and M. del Valle
- xi. Comparison of methods for the processing of voltammetric Electronic Tongues data
Microchimica Acta, doi: 10.1007/s00604-012-0938-7
X. Cetó, F. Céspedes and M. del Valle

4.1. Anàlisi de caves

El primer camp d'aplicació escollit per avaluar la capacitat de les llengües electròniques va ser el sector del cava. Com és sabut, el cava és el vi escumós produït segons el mètode tradicional, el qual és el mateix que el mètode xampanyès (o *méthode champenoise*) que es segueix per elaborar el Xampany, adaptat a les varietats autòctones del Penedès [1]. Al llarg d'aquest procés hi ha dos etapes claus, les quals esdevenen les tres fonts de classificació més comunes del cava i que per tant en determinaran la categoria. Per tal d'entendre millor aquestes, a continuació es descriu breument aquest procés.

Així doncs, el procés de producció del cava comença amb la verema i premsat del raïm per tal d'obtenir el most flor; el qual és sotmès a la primera fermentació en la que els llevats transformen en alcohol els sucres del most, obtenint així el vi base. Aquest procés es repetit per separat per les diferents varietats de raïm, obtenint un vi base jove, afruitat i tranquil. La següent etapa, coneguda com a *cupatge*, consisteix en la barreja de les diferents varietats de vi per obtenir-ne una única melodia denominada *cupatge* o *cupada*; essent la proporció de cada una de les varietats la que permet elaborar els diferents tipus d'escumosos. A continuació, comença l'etapa de *tiratge*, en la que s'omple la botella amb el cupatge seleccionat i s'hi afegeixen els llevats que permeten

provocar la segona fermentació (aquest cop en ampolla); llavors aquestes són transportades a les caves, començant així la segona fermentació i envelliment del cava. Aquest és un punt clau, que a més esdevé una font de classificació del cava segons el temps d'envelliment d'aquest; període de criaça que pot variar d'un mínim de 9 mesos fins a 36 mesos, d'acord amb la legislació del consell regulador del cava [2].

Després de l'envelliment, segueix la *remoguda* consistent en la decantació mecànica de les ampolles per conduir els sediments cap al coll de l'ampolla, per així facilitar-ne l'eliminació en el *degollament* mitjançant congelació del coll d'aquesta. Finalment, té lloc el *dosatge*, també denominat farciment, i que consisteix a afegir-hi el licor d'expedició (sucre dissolt en vi), que incorpora diferents dosis de sucre en funció del tipus de cava que es desitja obtenir (e.g. brut, brut nature, sec...). Procedint finalment al procés de *tap i etiquetatge*, en que es col·loca el tap de suro definitiu i el morrió o grapa, i es procedeix a l'etiquetatge abans de l'embalatge final.

Per tant, el cava és classificat normalment en funció de tres paràmetres que vindran determinats al llarg del seu procés de producció (Taula 4.1):

- les varietats de raïm utilitzats en el cupatge; i.e. el seu “color”
- el temps d'envelliment (o criaça) d'aquest
- el sucre afegit després de la segona fermentació

Taula 4.1. Classificació del cava segons la normativa establerta pel consell regulador del cava i publicada en el BOE [2].

| <i>Varietats</i> (“color”) | | <i>Sucre afegit</i> (<i>sequedat</i>) | | <i>Envelliment</i> (<i>criança mínima</i>) | |
|-------------------------------|----------------|--|-------------------------|---|----------|
| <i>Blanc</i> | macabeu | <i>Brut Nature</i> | < 3 g·L ⁻¹ | <i>Jove</i> | 9 mesos |
| | parellada | | (sense sucre afegit) | <i>Reserva</i> | 15 mesos |
| | xarel·lo | <i>Extra Brut</i> | < 6 g·L ⁻¹ | <i>Gran Reserva</i> * | 30 mesos |
| | chardonnay | <i>Brut</i> | < 15 g·L ⁻¹ | | |
| | subirat parent | <i>Extra sec</i> | 12-20 g·L ⁻¹ | | |
| <i>Rosat</i> | | <i>Sec</i> | 17-35 g·L ⁻¹ | | |
| | garnatxa | <i>Semi sec</i> | 33-50 g·L ⁻¹ | | |
| | monestrell | | > 50 g·L ⁻¹ | | |
| | pinot negre | <i>Dolç</i> | | | |
| | trepal | | | | |

*A més, cal tenir en compte que la indicació “Gran Reserva” només pot ser utilitzada en el cas dels caves “Brut Nature”, “Extra Brut” i “Brut”.

En aquest context, existeix una demanda per part del sector de noves eines que permetin la ràpida identificació i control de les diferents classes de cava, evitant així possibles contaminacions durant el procés de producció o frauds en la seva comercialització. Amb aquest objectiu, el primer pas va ser abordar la classificació de diferents caves segons la quantitat de sucre afegida (la qual respon a la classificació més

utilitzada), procedint després a la quantificació d'aquest. Mentre que en una etapa posterior, incrementat el grau de complexitat, es va procedir a avaluar la capacitat de la ET per tal de dur a terme la distinció dels caves en funció del grau d'envelliment. Justament, aquesta criança provoca una sèrie de canvis de tipus fisicoquímics i/o biològics, encara molt desconeguts, però que proporcionen aromes que tenen gran acceptació pel consumidor, i que justifiquen els llargs períodes de rima (anys) al que es sotmeten els caves d'alta gamma i el seu elevat preu.

Aquesta segona classificació, respon a una demanda per part de l'INCAVI degut a la dificultat de dur a terme aquesta tasca a partir de mètodes clàssics, ja que com a tal, el grau d'envelliment no està relacionat amb cap paràmetre analític concret i respon a una suma de característiques del cava. No obstant, existeixen certs compostos químics que poden ser indicatius d'aquest envelliment tals com el vitispirà, el TDN (1,1,6-Trimetil-1,2-dihidronaftalè) o el succinat de dietil [3]; però no estrictament relacionats a nivell individual, sinó com a mers indicadors. Així doncs, la manera més habitual de dur a terme aquesta classificació actualment és mitjançant l'anàlisi sensorial per part de panels de tast; malauradament aquests tenen un alt component subjectiu.

4.1.1. Classificació dels caves

4.1.1.1. Quantitat de sucre afegida

En primer lloc, es va plantejar com a objectiu inicial aconseguir dur a terme la classificació dels diferents tipus de cava, a partir de mesures voltamperomètriques, en funció de la quantitat de sucre afegida en la producció d'aquests; la qual correspon a la classificació més senzilla des del punt de vista analític, a excepció del "color". Per aquest motiu, es van adquirir un total de 21 vins escumosos diferents entre els quals es van incloure 2 xampanyes franceses, 2 caves procedents de la regió d'Extremadura i 17 caves de la regió del Penedès; tots sota la respectiva Denominació d'Origen (DO).

Per tal d'assolir l'objectiu plantejat, demostrant a més la capacitat de les llengües electròniques com a eina biomimètica capaç d'imitar el sentit de la percepció humana, es va decidir fer ús d'una matriu de 5 sensors amb diferents modificadors electroquímics, però sense incloure entre ells cap sensor o biosensor selectiu per la detecció de sucres. Així doncs, basant-se en treballs anteriors [4], es va decidir fer ús de la llengua electrònica formada pels sensors (§3.1, Taula 3.1): GEC, Cu, Pt, PANI i PPy.

A més, per tal de simplificar el protocol de mesura i de cara a la seva possible aplicació en el sector, les mostres van ser analitzades sense dur a terme cap pretractament d'aquestes; amb la única precaució d'analitzar cada mostra després d'uns pocs minuts de l'obertura de la botella (ca. 5 min, temps equivalent a l'anàlisi de cada mostra) per tal d'evitar la formació de bombolles en la superfície de l'elèctrode. Minimitzant així la distorsió produïda per les bombolles en el procés de mesura, però sense dur a terme cap etapa específica per l'eliminació del CO₂ o l'O₂.

Per la mesura de les mostres, es va seguir el procediment descrit a §3.3, aplicant un escombrat de potencial entre -1.0V i +1.3V vs. Ag/AgCl amb un salt de potencial de 3.66mV i una velocitat d'escombrat de 100mV·s⁻¹. Com a precaució addicional, es realitzava una etapa de neteja electroquímica de la superfície dels sensors després de l'anàlisi de cada mostra per tal de minimitzar possibles "efectes història" o la contaminació entre aquestes, evitant a més el poliment dels sensors entre cada una de les mesures, fet que accelera i simplifica el protocol de mesura (§3.3). Concretament, aquesta neteja electroquímica va consistir en l'aplicació d'un potencial de +1.5V durant 40 segons en un vas que contenia 10mL d'aigua destil·lada. Finalment, com a precaució addicional (al igual que en la resta de casos al llarg de tota la present tesi), les mostres van ser analitzades de forma aleatòria per assegurar que la discriminació d'aquestes no prové de l'ordre en que han estat mesurades o efectes d'histèresi dels sensors.

Tal com es pot observar a la Figura 4.1, diferents respostes voltamperomètriques van ser obtingudes per cada tipus de sensor, amb un patró de resposta per cadascun d'ells i un comportament diferenciat per cadascuna de les mostres, generant així informació molt valuosa per a ser utilitzada com a punt de partida per la ET. Per tant, representant l'ús d'una matriu de sensors basats en compòsits grafit-epoxi modificats en *bulk*, una manera senzilla d'obtenir sensors amb una resposta diferenciada i creuada front les mostres d'interès; una situació altament desitjable en els sistemes basats en llengües electròniques.

No obstant, per tal d'extreure el màxim de informació significativa de cadascun dels voltamperogrames, es requereix d'una etapa de compressió d'aquests per tal de guanyar avantatges en l'etapa de processament, tals com la reducció en el temps d'entrenament, evitar redundància en les dades d'entrada i obtenir un model amb millor capacitat de generalització (§3.4.1) [5]. D'altra banda, també cal tenir en compte l'extrema complexitat i elevada dimensionalitat dels senyals de partida, ja que en el cas de les ETs voltamperomètriques, els tipus de senyals obtinguts impliquen el registre dels corrents

generats de les solucions sota estudi en funció dels diferents potencials aplicats i per cadascun dels sensors; podent existir zones de solapament del senyal degut a que tots els components de la solució electroquímicament actius a un determinat potencial contribuiran al corrent mesurat, essent el registre complet el que conté tota la informació de la mostra.

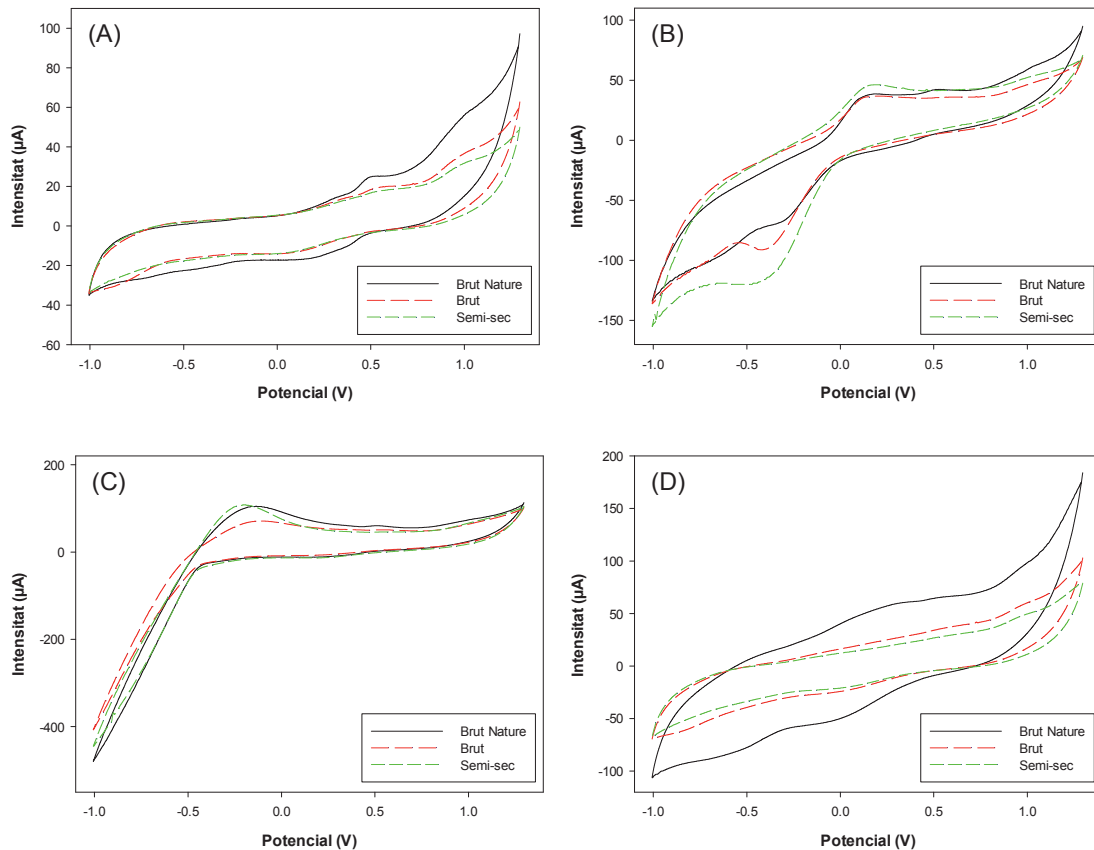


Figura 4.1. Exemple dels diferents voltamperogrames obtinguts per diferents mostres arbitràries amb quatre dels elèctrodes usats: (A) GEC, (B) Cu, (C) Pt i (D) PPy.

En aquest cas, aquesta compressió va ser assolida mitjançant l'ús de la DWT (§3.4.1.7), utilitzant la funció mare *Daubechies* i un nivell de descomposició quatre [6]; permetent la compressió de la resposta original fins a un 97.4% sense pèrdua d'informació rellevant.

A continuació, per tal de confirmar aquesta possible discriminació de les diferents mostres observada al llarg dels diferents voltamperogrames, es va procedir a l'anàlisi dels coeficients extrets en la compressió d'aquests mitjançant PCA (§3.4.2.1.1): proporcionant així el PCA una eina de visualització entre les similituds i diferències de les mostres molt útil per tal d'identificar possibles agrupacions d'aquestes.

La Figura 4.2 mostra el resultat d'aquest anàlisi, on aconseguirem representar ca. del 94.8% de la variància inicial amb només els tres primers PCs, facilitant així la interpretació de les mesures degut a que quasi tota la variància present en la informació original queda reduïda a només 3 variables. En concret, els patrons observats en el gràfic evidencien que les mostres són agrupades en base al seu tipus i regió d'origen; separant així els vins escumosos del Penedès de la resta, permetent a l'hora la categorització d'aquests també.

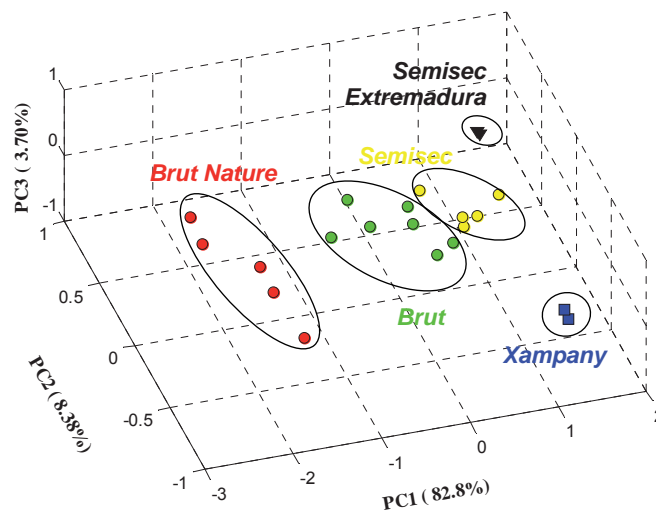


Figura 4.2. Representació dels *scores* dels tres primers components principals obtinguts després de l'anàlisi PCA. Un total de 21 mostres van ser analitzades corresponents a: (●) Brut Nature del Penedès, (●) Brut del Penedès, (●) Semisec del Penedès, (▼) Semisec d'Extremadura i (■) Xampany.

No obstant, el PCA només representa una eina de visualització que permet representar en unes poques variables la major part de variabilitat present en les mostres, fet que permet observar similituds i diferències entre aquestes, però sense poder actuar com a classificador per si sol. Així doncs, per confirmar aquestes agrupacions observades en el gràfic anterior, es va construir un model DWT-ANN (§3.4.2.2.1) que assolís la distinció i classificació de les mostres d'acord amb quatre grups diferenciats: les 3 classes considerades (*Brut*, *Brut Nature* i *Semisec*) i els *outliers* (el xampany i el cava d'Extremadura); els darrers van ser considerats com un únic grup (*outliers*) donat el seu diferent origen i la poca similitud observada respecte les altres classes en el gràfic PCA (Figura 4.2).

Com s'ha descrit (§3.4), prèviament a l'obtenció del model final, aquest ha de ser entrenat i la seva arquitectura optimitzada d'acord amb la tasca assignada. En aquest cas, després d'algunes proves preliminars, la topologia de la xarxa va quedar definida

amb 235 neurones en la capa d'entrada (47 coefs. Wavelet x 5 sensors), 20 neurones en la capa oculta i 4 neurones binàries (1/0) en la capa de sortida; utilitzant com a funció de transferència la funció *logsig* en les tres capes.

El model DWT-ANN va ser entrenat amb el 50% de les dades, utilitzant el 50% romanent per avaluar la capacitat de predicció d'aquest, obtenint així dades no esbiaixades. A partir de les classes pronosticades amb el model per aquest darrer, es va construir la matriu de confusió (Taula 4.2). Tal com es pot observar, totes les mostres van ser classificades correctament, a excepció d'una mostra pertanyen al grup *outlier* que va ser classificada com a *Semisec* pel model; no obstant, cal tenir en compte que aquesta mostra aparentment mal classificada correspon al cava semisec de la regió d'Extremadura.

Taula 4.2. Matriu de confusió construïda a partir de les classes predites mitjançant el model DWT-ANN pel subconjunt de test.

| | | <i>Predit</i> | | | |
|----------------|--------------------|---------------|-------------|----------------|-----------------|
| | | <i>BN</i> | <i>Brut</i> | <i>Semisec</i> | <i>Outliers</i> |
| <i>Esperat</i> | <i>Brut Nature</i> | 2 | 0 | 0 | 0 |
| | <i>Brut</i> | 0 | 3 | 0 | 0 |
| | <i>Semisec</i> | 0 | 0 | 3 | 0 |
| | <i>Outliers</i> | 0 | 0 | 1 | 1 |

Per tal de mesurar l'eficàcia del model es van calcular els paràmetres descrits a §3.4.2.3.2.1: grau d'encert, sensibilitat i especificitat; assolint el 90.0%, 87.5% i 85.7%, respectivament. Per tant, cal destacar l'elevat valor d'aquests tenint en compte que les mostres utilitzades per avaluar les prestacions del model no van intervenir de cap manera en l'etapa de modelatge; encara més, si considerem que la mostra incorrectament assignada, realment és un cava semisec. Així doncs, demostrant la capacitat del sistema per classificar els caves d'acord amb la quantitat de sucre afegida durant la seva producció o distingir un cava d'un xampany, però no essent capaç de distingir entre caves produïts entre diferents regions.

Addicionalment, i donada la bona discriminació observada mitjançant l'anàlisi PCA, es va procedir a la construcció d'un molde PCA-ANN ja que el PCA només presenta una millor visualització de les dades. Així, utilitzant els *scores* obtinguts de l'anàlisi PCA com a inputs del model ANN, aquest pot actuar com a classificador. Al igual que en el cas DWT-ANN, es va construir un model ANN utilitzant una divisió de les dades *train/test* del 50:50%, amb la diferència que, aquest cop, el model contemplava la

classificació de les mostres en cinc classes enlloc de quatre. Concretament, els clusters observats en el gràfic PCA: Brut Nature del Penedès, Brut del Penedès, Semisec del Penedès, Semisec d'Extremadura i Xampany.

Un cop optimitzada la topologia del model PCA-ANN, aquest va quedar format per 3 neurones en la capa d'entrada (corresponents als 3 primers PCs), 7 neurones en la capa oculta i 5 neurones binàries en la capa de sortida; utilitzant com a funció de transferència la funció *logsig* en les tres capes. Al igual que en el cas DWT-ANN, es varen calcular els paràmetres de grau d'incert, sensibilitat i especificitat; assolint el 100% pels tres estadístics. Per tant, confirmant la discriminació observada en el gràfic PCA i la capacitat de la llengua electrònica per distingir els diferents tipus de caves.

A més, és pot observar com la major reducció en el nombre d'inputs de la xarxa gracies a l'anàlisi PCA, ha conduït a l'obtenció d'un model més simple i amb major capacitat de generalització i predicció. Confirmant la importància de l'ús de mètodes de preprocessament eficaços per al tractament de les dades provinents de les matrius de sensors voltamperomètrics.

4.1.1.2. Temps de cria

Donats els bons resultats assolits en el cas anterior, el qual a priori representava el cas més senzill, és va procedir a abordar la classificació dels caves en funció del temps de cria. Tal com ja s'ha dit, aquesta representa la major demanda per part del sector degut a la manca de mètodes de referència capaços de dur a terme aquesta distinció. No obstant, l'anterior va ser escollida com a sistema model, la qual va servir per avaluar la viabilitat del sistema i establir un protocol de mesura; confirmant així la capacitat de la ET voltamperomètrica de distingir entre diferents tipus de caves.

Amb aquesta finalitat, i a través d'una col·laboració amb l'INCAVI, aquests ens van subministrar un total de 65 mostres de caves diferents; catalogades en les següents categories segons el temps de cria: 16 Gran Reserva, 24 Reserva i 25 joves. A més, al igual que en el cas anterior, també es va incorporar un xampany francès a l'estudi per tal d'avaluar el comportament del model davant de possibles *outliers*; replicant aquest quatre vegades i considerant cada un com una mostra independent.

Per a la mesura de les mostres, es va seguir el mateix protocol i condicions experimentals establertes en el cas anterior (§4.1.1.1), variant però dos paràmetres. Per una banda, és va augmentar el salt de potencial de 3.66mV a 9mV, donat que aquest

augment permetia reduir d'entrada el nombre de dades generades per cada registre ca. 2.5 vegades sense cap pèrdua d'informació (§3.4.1.2ii). D'altra banda, es va decidir incorporar un sisè sensor (CoPc) a la matriu per tal d'extreure més informació de les mostres, ja que aquest no alterava en res el protocol de mesura establert donada la capacitat de l'instrument utilitzat; quedant així la ET formada pels següents sensors (§3.1, Taula 3.1): GEC, CoPc, Pt, PANI, PPy i Cu.

El conjunt de mostres va ser doncs analitzat sota les condicions descrites de forma aleatòria, obtenint un registre voltamperomètric sencer per cadascun dels sis sensors. En aquest sentit, la Figura 4.3 mostra un extracte dels resultats obtinguts per algunes d'aquestes mostres seleccionades arbitràriament. Tal com es pot observar, diferents senyals varen ser obtinguts per cadascun dels sensors, amb un perfil de resposta similar al cas anterior (Figura 4.1), però amb major complexitat i solapament dels senyals. A més, es pot apreciar com les diferents mostres presentaren una resposta voltamperomètrica molt similar en zones específiques del voltamperograma; i com aquestes similituds es repetiren al llarg de tot el registre voltamperomètric en funció del tipus de mostra.

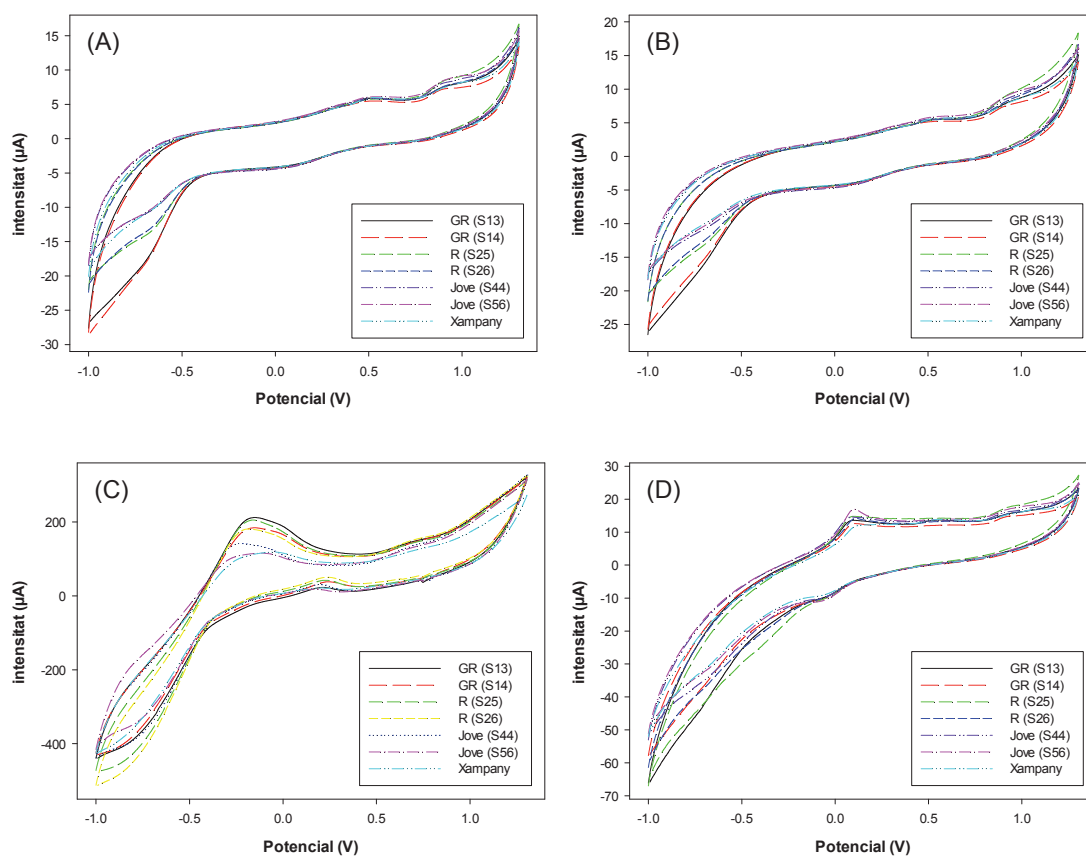


Figura 4.3. Exemple dels diferents voltamperogrames obtinguts per certes mostres arbitràries amb quatre dels elèctrodes usats: (A) GEC, (B) CoPc, (C) Pt i (D) Cu.

Per exemple, en la Figura 4.3A es pot observar com els caves *Gran Reserva*, *Reserva* i *joves* presenten una resposta catòdica clarament diferenciada en la regió de -0.5V a -1.0V; mentre que el *Xampany* presenta una resposta més similar als caves joves, tot i continuar mostrant encara una resposta lleugerament diferenciada en la regió de -0.75V a -1.0V. Aquest comportament podria ser en certa manera esperat donat que ambdós tipus de vins escumosos s'elaboren seguint el "mateix" procediment, però emprant diferents varietats de raïm; per la qual cosa s'esperen més semblances amb els caves joves, ja que a pesar de no existir aquesta classificació pel cas del *Xampany*, aquests van ser envellits pel mateix període de temps. D'altra banda, no es possible observar aquest tipus de similituds per mostres de la mateixa classe en la part anòdica, fet que fa pensar que els corrents generats en aquesta zona eren deguts a l'oxidació de compostos fenòlics (§4.4.1.1) o dels sucres; recordem però, que la qualificació *Gran Reserva* i *Reserva* només pot ser atorgada als caves en els que s'ha afegit menor quantitat d'aquest. Tendències similars són observades pels sensors CoPc (Figura 4.3B) o PPy; mentre que els sensors Pt (Figura 4.3C) i Cu (Figura 4.3D) aporten informació nova al sistema, relativa a altres paràmetres o compostos. Tanmateix encara s'observen algunes de les similituds discutides anteriorment; en aquest aspecte cal remarcar la resposta clarament diferenciada observada en el cas del sensor Pt a la regió anòdica al voltant de +1V per la mostra de xampany, induint a la possible discriminació d'aquest respecte la resta.

Al igual que en el cas anterior, i per tal d'extreure la màxima informació de cadascun dels registres voltamperomètrics, es requereix d'una etapa de preprocessament del senyal per tal de comprimir i reduir l'elevada dimensionalitat d'aquest (mostres x sensors x potencials aplicats). A diferència del cas anterior, aquest cop es va fer ús de la FFT per tal d'assolir aquesta reducció (§3.4.1.6). Per tal d'escollir el número de coeficients de Fourier que calia preservar, es va avaluar com variava el grau de reconstrucció del senyal original en funció del nombre de coeficients seleccionats; utilitzant el coeficient de determinació (R^2) i el factor f_c per a mesurar aquest últim (§3.4.1.8).

La Figura 4.4A mostra com varia la reconstrucció del senyal voltamperomètric original en funció del número de coeficients de Fourier utilitzats; tal com queda demostrat, al incrementar el número de coeficients, obtenim una millor representació del senyal original, però a expenses del factor de compressió. Per tant, fa falta escollir una situació de compromís, que permeti un elevat factor de compressió i una bona

reconstrucció del senyal; reduint així la complexitat del model, però preservant la informació rellevant per a ser utilitzada com a *input* per aquest. Per tal d'objectivar aquesta selecció, es va fer ús del grau de correlació entre el senyal reconstruït i el senyal original (Figura 4.4B) tant pel coeficient R^2 com pel factor f_c . Tal com s'observa, ambdós factors presenten la mateixa tendència, presentat però, el factor f_c una millor capacitat de discriminació. Per exemple, amb només 8 coeficients de Fourier, quan la reconstrucció encara és pobra, el valor obtingut per R^2 és de 0.9971, mentre que el valor de f_c és de 0.8994. Per tant, el número òptim de coeficients de Fourier va ser seleccionat en aquest cas com la primera vegada que el factor f_c va sobrepassar el valor de 0.95; això va ser assolit preservant els primers 16 coeficients de Fourier, el qual va permetre una compressió de la informació original del 96.9% sense cap pèrdua de informació rellevant.

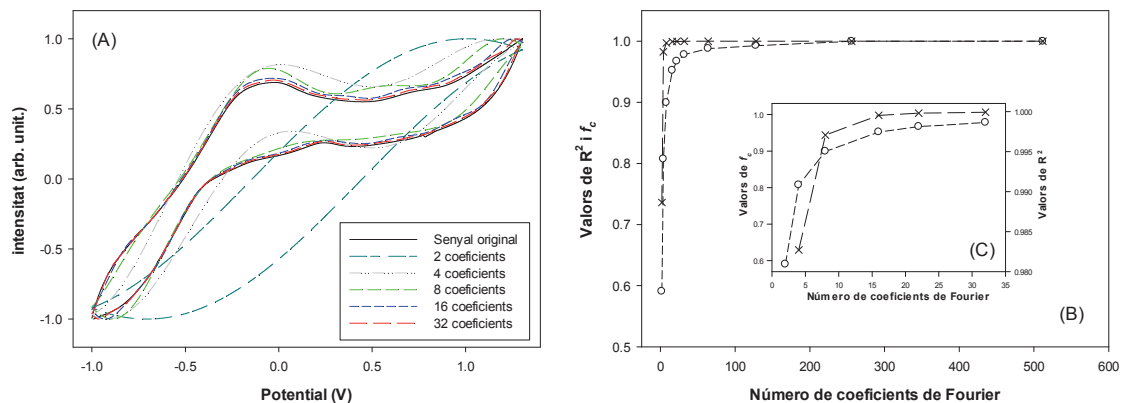


Figura 4.4. (A) Reconstrucció del senyal voltamperomètric en funció del número de coeficients de Fourier utilitzats; on el senyal del sensor Pt es mostra com a exemple. (B) Representació del coeficient de determinació (R^2 , X) i f_c (o) vs. el número de coeficients de Fourier utilitzats en comparar la resposta voltamperomètrica en cru amb el senyal reconstruït; (C) ampliació del gràfic anterior.

Per tal de confirmar aquestes tendències inicials observades en les respostes voltamperomètriques i per corroborar el seu contingut analític, els corresponents voltamogrames comprimits van ser processats mitjançant l'anàlisi PCA. Construint així un model de reconeixement preliminar que permet visualitzar les similituds entre les diferents mostres d'acord amb les mesures realitzades. El gràfic obtingut es mostra a la Figura 4.5, on amb els tres primers PCs el valor de variància acumulada és de ca. 97.8%. A més, els patrons en la figura evidencien com les mostres s'agrupen en funció del temps de cria d'aquestes, amb *clusters* que separen clarament les classes principals corresponents a: *Gran Reserva*, *Reserva*, *joves* i *xampany*.

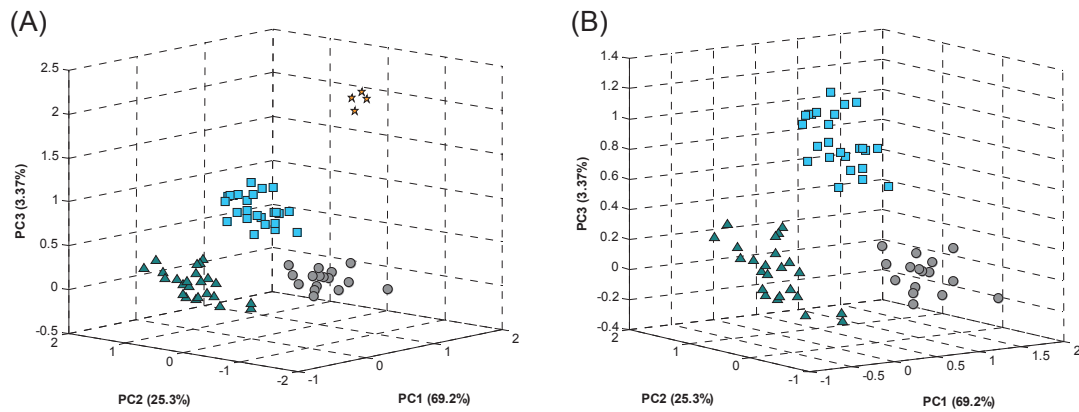


Figura 4.5. (A) Representació dels *scores* dels tres primers components principals obtinguts després de l'anàlisi PCA. Un total de 65 mostres van ser analitzades corresponents a: (●) Gran Reserva, (▲) Reserva, (■) joves i (★) Xampany. (B) Ampliació del gràfic anterior, deixant fora el xampany.

Per tal d'estimar la taxa d'èxit en la identificació de les mostres, es va construir un model LDA capaç d'actuar com a classificador, donat que el PCA només proporciona una visualització de les diferents zones d'agrupament (§3.4.2.1). Al igual que en el PCA, els coeficients de Fourier obtinguts van ser utilitzats com a *inputs* en el model, utilitzant com a *target* les diferents classes de cava. Tal com es podria esperar a partir del gràfic anterior, una bona discriminació de les classes considerades va ser obtinguda tal com es mostra a la Figura 4.6. Addicionalment, en aquest cas, dos models diferents varen ser construïts; un incloent les mostres de xampany (Figura 4.6B) i l'altre sense elles (Figura 4.6A).

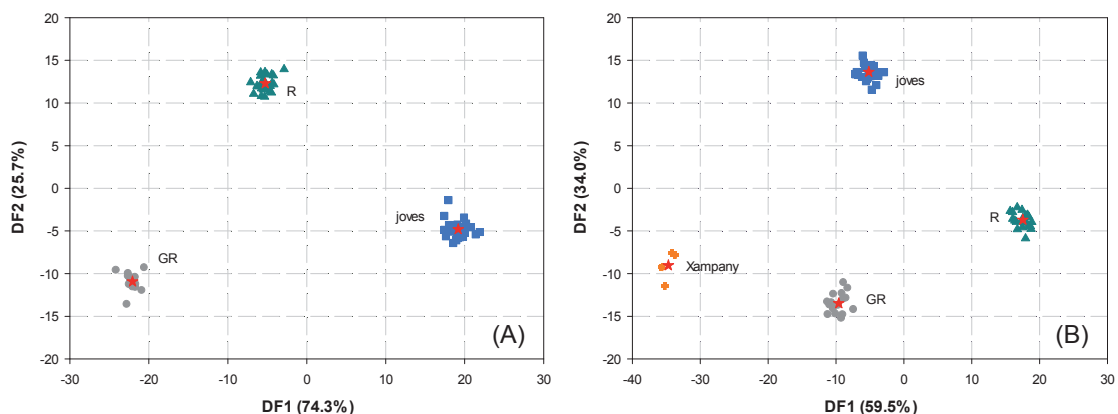


Figura 4.6. Representació dels *scores* de les dos primeres funcions discriminants obtingudes després de l'anàlisi LDA d'acord amb les classes: (●) Gran Reserva, (▲) Reserva, (■) joves i (★) xampany. Addicionalment, també s'ha representat el centroe de cada classe (★). Models construïts (A) sense incloure les mostres de xampany i (B) incloent aquestes.

El model LDA va ser entrenat i validat fent ús del mètode de validació creuada *k-fold* amb $k=6$ (§3.4.2.3.1). A continuació, la matriu de confusió va ser construïda a

partir dels resultats obtinguts pel subconjunt de validació (Taula 4.3), permetent calcular l'eficàcia del model a través de tres indicadors diferents (§3.4.2.3.2.1): grau d'encert, sensibilitat i especificitat; assolint el 100% pels tres indicadors estadístics. Demostrant per tant, la capacitat de la ET de discriminar també els caves en funció del temps de criança; i per tant, proporcionant una eina molt útil al sector capaç de resoldre la manca de mètodes analítics capaços de resoldre aquesta problemàtica de manera objectiva.

Taula 4.3. Matriu de confusió construïda d'acord el temps de criança pronosticat mitjançant el model FFT-LDA, utilitzant el mètode de validació creuada *k-fold* amb *k*=6.

| | | <i>Predit</i> | | |
|----------------|---------------------|---------------|----------|--------------|
| | | <i>GR</i> | <i>R</i> | <i>Joves</i> |
| <i>Esperat</i> | <i>Gran Reserva</i> | 16 | 0 | 0 |
| | <i>Reserva</i> | 0 | 24 | 0 |
| | <i>Joves</i> | 0 | 0 | 25 |

4.1.2. Quantificació del sucre afegit

Donats els bons resultats obtinguts en l'anàlisi qualitativa dels caves (§4.1.1.1) i per tal de confirmar els patrons de resposta observats, es va procedir a l'anàlisi quantitativa de la quantitat de sucre afegida, la qual determinarà la classificació del cava (Taula 4.1). Així, en un primer intent, es va fer ús del mètode de l'addició estàndard de segon ordre (SOSAM) per tal de confirmar aquesta tendència; mentre que posteriorment, a partir de les mesures dutes a terme en §4.1.1.2, es va procedir a la quantificació directa d'aquest.

4.1.2.1. Ús del mètode de l'addició estàndard de segon ordre

A partir dels resultats obtinguts en §4.1.1.1, es pot assumir el fet que els sensors són capaços de distingir la quantitat de sucre afegida en cadascun dels caves durant el procés de producció, ja que és aquest paràmetre el que determina el tipus de cava obtingut, i se n'ha assolit la correcta classificació. No obstant, també ha quedat palès, que aquest no és l'únic paràmetre que és capaç de distingir la ET, ja que aquesta també ha estat capaç de distingir al mateix temps el xampany francès o els caves de la regió d'Extremadura.

Així doncs, per tal de confirmar aquesta possible resposta, es va procedir a l'anàlisi quantitativa de la quantitat de sucre afegida en diferents caves. Amb aquest fi, es van analitzar quatre caves diferents fent ús del mètode de l'addició estàndard de segon ordre (SOSAM), el qual va ser usat per primera vegada per Booksh et al. a partir de dades espectrofotomètriques [7].

El mètode SOSAM apareix com a alternativa a l'addició estàndard tradicional, a causa de l'elevada dimensionalitat de les dades de partida que fan aquesta inviable. D'aquesta manera, el primer pas consisteix en la descomposició de la matriu de dades de tercer ordre i la seva reducció, per a continuació, identificar quins components (*loadings*) estan relacionats amb l'analit d'interès, i finalment graficar els *scores* d'aquests en front de la quantitat d'analit afegida per tal d'extrapolar linealment la concentració d'aquest, de la mateixa manera que fariem en el cas l'addició estàndard univariable.

En aquest cas, es va preparar una solució estàndard de sacarosa en aigua destil·lada de $750 \text{ g}\cdot\text{L}^{-1}$, i al igual que fariem en el cas convencional, diferents quantitats d'aquesta van ser afegides a cadascun dels caves sota estudi. La concentració de sucre (sacarosa) afegit va ser escollida d'acord amb el rang de les diferents classificacions definides pel consell regulador del cava (és a dir, entre 0 i $80 \text{ g}\cdot\text{L}^{-1}$; Taula 4.1). A continuació, les mostres fortificades van ser tractades com a mostres independents i mesurades de forma aleatòria tal com s'ha descrit a §4.1.1.1. Un cop analitzades totes les mostres, per tal de dur a terme la quantificació del sucre, les dades van ser primer analitzades mitjançant l'ús del PCA, representant després els *scores* del PC1 front la concentració de sucre afegida, obtenint així el típic gràfic d'addició estàndard.

Tal com s'observa a la Figura 4.7A, es confirma la tendència observada anteriorment, on la matriu de sensors semblava ser capaç de discriminar entre la diferent quantitat de sucre afegida en el procés de producció dels caves; demostrant a més que és possible establir una relació lineal entre la variació de la resposta i la quantitat de sucre afegida (Figura 4.7B).

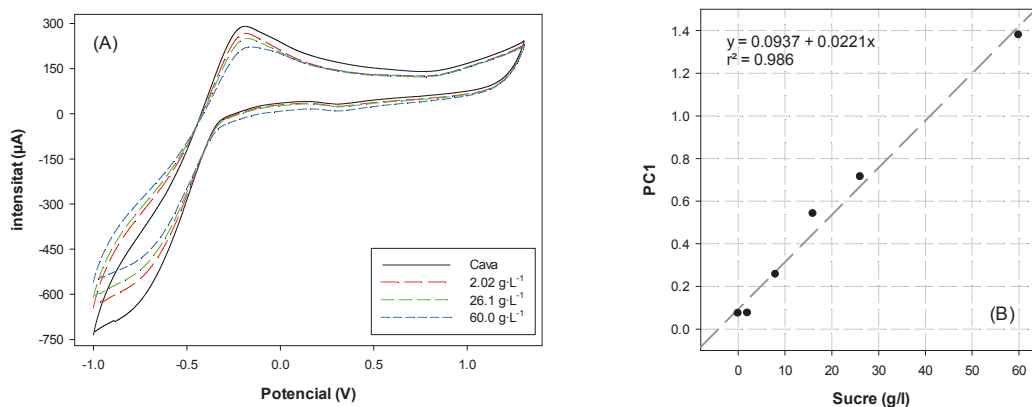


Figura 4.7. Addició estàndard de segon ordre (SOSAM); el cava *Codorniu Extra Brut Nature* es mostra com a exemple. (A) Resposta voltamperomètrica del sensor Pt front les diferents mostres fortificades; els valors de la llegenda corresponen a la concentració de sucre afegida en el cava. (B) Gràfic del primer component del PCA vs. la quantitat de sucre afegida.

Aquest procés va ser repetit per 4 mostres diferents per tal de confirmar que no es tractava d'un cas aïllat, ajustant la recta de regressió per cada un dels gràfics d'addició estàndard i extrapolant el valor de sucre afegit en la mostra original. La Taula 4.4 mostra les equacions ajustades i els valors extrapolats per les diferents mostres avaluades; tal com es pot deduir dels coeficients de correlació, podem confirmar que existeix una relació lineal entre el PC1 i la quantitat de sucre afegida. A més, s'observa com els valors trobats són similars als valors indicats pel fabricant, fet que també corrobora la relació entre la resposta de la matriu de sensors i la quantitat de sucre en la mostra.

Taula 4.4. Quantitat de sucre present obtingut amb el mètode de l'addició estàndard de segon ordre (SOSAM) comparada amb els valors declarats pels productors.

| <i>Mostra</i> | <i>Recta de regressió</i> | <i>r</i> | <i>Sucre trobat (g·L⁻¹)</i> | <i>Sucre declarat a l'etiqueta (g·L⁻¹)</i> |
|-----------------------------------|---------------------------|----------|--|---|
| <i>Codorniu Extra Semi-sec</i> | $y=0.0182x+0.744$ | 0.946 | 40.9 | 35 |
| <i>Anna de Codorniu Brut</i> | $y=0.0228x+0.0869$ | 0.908 | 3.81 | 6-8 |
| <i>Codorniu Extra Brut Nature</i> | $y=0.0221x+0.0937$ | 0.993 | 4.24 | <3 |
| <i>Mont Marçal Brut Nature</i> | $y=0.0210x+0.0398$ | 0.980 | 1.67 | <2 |

4.1.2.2. Ajust d'un model multivariable

Finalment, fruit de la col·laboració amb l'INCAVI, es va procedir a la quantificació de la quantitat de sucres totals i l'extracte sec del conjunt de 65 caves mitjançant els mètodes de referència [8], amb l'objectiu d'avaluar la capacitat de la llengua electrònica per predir la concentració d'aquest de forma quantitativa.

Es tracta doncs, d'un doble objectiu: per una banda, una vegada confirmada la resposta de la ET davant la quantitat de sucre mitjançant el mètode de l'addició estàndard, seria molt interessant poder assolir la seva quantificació de forma directa; de l'altra banda, cal recordar que un altre requisit per part del Consell Regulador del cava per atorgar la denominació de *Gran Reserva*, és que la quantitat de sucre afegida a de ser inferior a 15g·L⁻¹ [2].

Així doncs, a partir del conjunt de respostes voltamperomètriques obtingut pel segon conjunt de caves (§4.1.1.2), i utilitzant una xarxa neuronal (§3.4.2.2.1) com a eina de modelatge enlloc del LDA, es va procedir a la construcció d'un model quantitatiu que permetés predir la quantitat total de sucre present en els caves; utilitzant el 76% de les dades (48 mostres) per l'entrenament d'aquest, i el 24% romanent (15 mostres) per la seva avaluació.

El primer pas en la construcció d'un model ANN és la selecció de la topologia o arquitectura d'aquesta (§3.4.2.2.1). Tal com ja s'ha explicat, i donades les dificultats de predir la seva configuració per endavant, es tracta d'un procés d'assaig-error on els diferents paràmetres d'aquesta són optimitzats. Aquest procés, consisteix en la construcció i entrenament d'un model neuronal per cada una de les possibles configuracions, i l'avaluació posterior de les prestacions d'aquest amb les mostres del subconjunt d'entrenament. A continuació, s'avalua la bondat de l'ajust del model a partir dels paràmetres descrits anteriorment (§3.4.2.3.2.2): el RMSE, i els valors de pendent, ordenada a l'origen i coeficient de correlació calculats a partir de l'anàlisi de regressió dels gràfics de concentració predita vs. esperada.

A tall d'exemple, la Figura 4.8 mostra els resultats obtinguts com a cas il·lustratiu del procés d'optimització i selecció de la topologia del model ANN; a pesar de que aquest procés va ser repetit de forma anàloga per tots els casos posteriors, no s'en presentaran els resultats, describint només el model seleccionat després del procés.

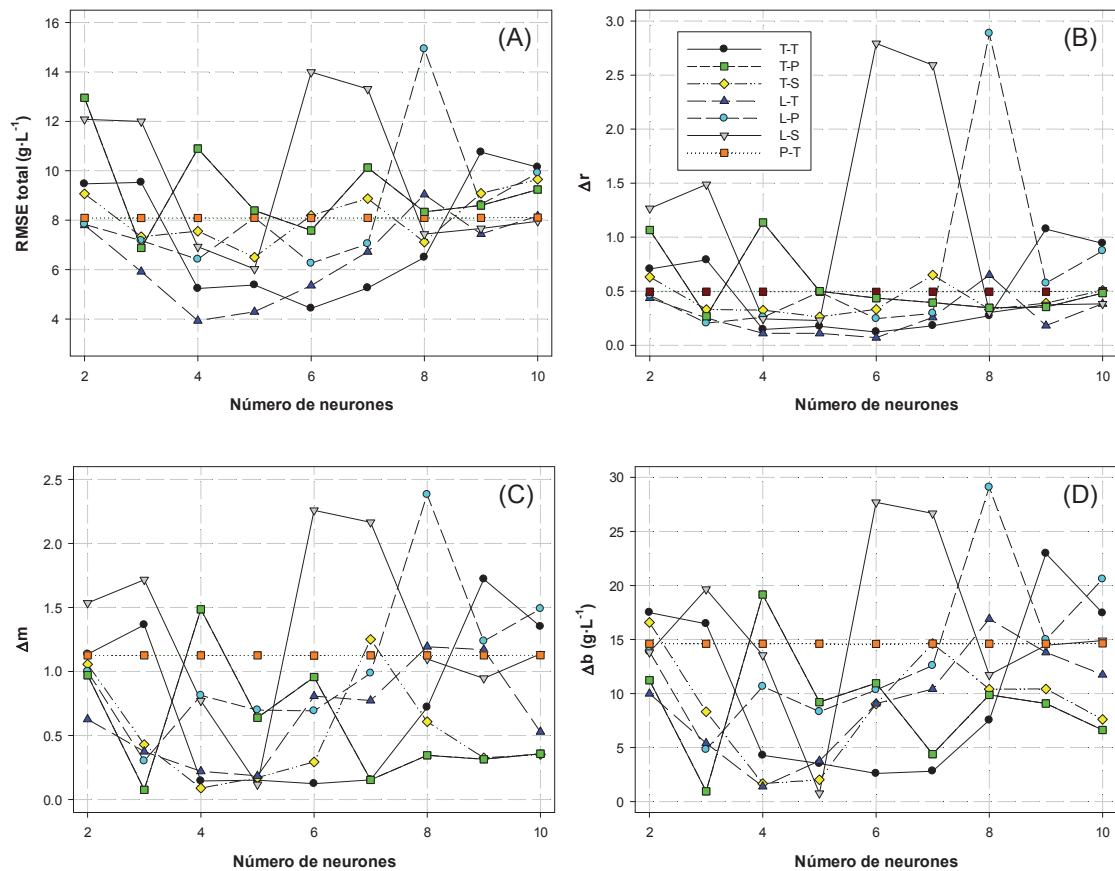


Figura 4.8. Resultats detallats de l'optimització de la FFT-ANN. Els valors obtinguts de (A) RMSE, (B) coeficient de correlació, (C) pendent i (D) ordenada a l'origen dels gràfics de comparació d'obtingut vs. esperat són graficats front el número de neurones en la capa oculta, utilitzant a més diferents combinacions de les funcions de transferència en la capa oculta i la capa de sortida: (L) *logsig*, (T) *tansig* i (P) *purelin*; presentant només les configuracions que presentaren millors resultats.

Així doncs, l'elecció del model final vindrà determinada pel model que presenti un menor RMSE juntament amb valors de pendent i coeficients de correlació propers a 1, i ordenada a l'origen propera a 0. En el nostre cas, per tal de facilitar aquesta interpretació, es va calcular la suma de la diferència entre el valor ideal i l'obtingut per cadascun dels paràmetres. Per tant, el model òptim serà el que presenti valors més baixos (propers a 0) pels quatre paràmetres calculats.

Després de l'optimització de l'arquitectura de la ANN mitjançant l'estudi sistemàtic de les diferents configuracions d'aquesta (Figura 4.8), la topologia final del model neuronal va ser la formada per 96 neurones (6 sensors x 16 coefs. provinents de l'anàlisi FFT) en la capa d'entrada, 4 neurones i la funció de transferència *tansig* en la capa oculta, i dos neurones i la funció de transferència *tansig* a la capa de sortida, viz. la quantitat total de sucre i l'extracte sec.

L'eficàcia del model generat va ser llavors avaluada front el subconjunt de mostres denominades com a subconjunt de test, les quals no varen participar en tot el procés de construcció del model, essent la seva concentració predita a partir de la resposta dels sensors. A continuació, per tal de comprovar i avaluar fàcilment la capacitat de predicció del model, es van construir els gràfics comparació de concentració predita vs. esperada, tant pel conjunt d'entrenament com pel de test (Figura 4.9).

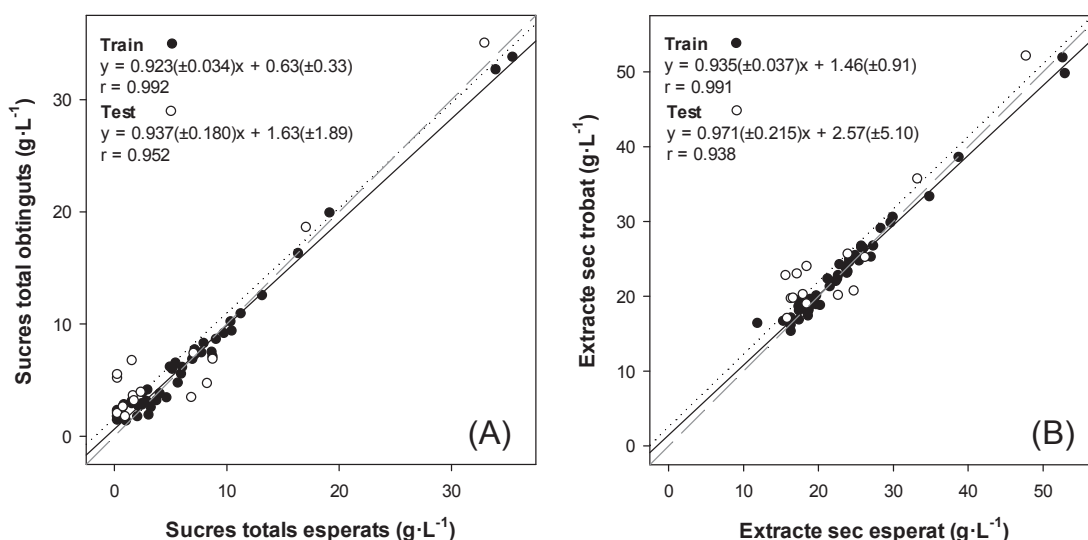


Figura 4.9. Capacitat de modelatge del model FFT-ANN. Gràfics de comparació de les concentracions obtingudes vs. les esperades pels (A) sucres totals i (B) l'extracte sec, tant pel subconjunt d'entrenament (●, línia sòlida) com de test (○, línia de punts). La línia discontinua correspon a la diagonal teòrica.

Tal com es pot observar, s'assoleix una tendència satisfactòria pels dos índexs, amb línies de regressió molt properes a les teòriques. A més, tal com és habitual en els

models ANN, s'observa una menor dispersió i error pel subconjunt d'entrenament, tal com caldria esperar ja que les dades del subconjunt de test no participen en absolut durant el procés de construcció del model, representant així la seva bondat d'ajust una mesura objectiva de l'eficàcia del model.

Adicionalment, també es van calcular els paràmetres de la regressió, essent també aquests propers als valors teòrics; a saber, ordenada a l'origen propera a 0 i valors de pendent i coeficients de correlació propers a 1, denotant que no hi ha diferències significatives entre els valors predits per l'ANN i els valors esperats (els valors obtinguts mitjançant el mètode de referència). Per tant, demostrant la capacitat de la ET per predir de forma quantitativa la quantitat de sucre present en les mostres de cava.

Així doncs, s'ha demostrat que la combinació de les mesures voltamperomètriques amb les eines quimiomètriques constitueix una eina analíticament poderosa per la caracterització de les mostres de cava, tant des del punt de vista qualitatiu com quantitatiu. D'aquesta manera, el sistema presentat podria ser considerat com una alternativa atractiva a altres mètodes clàssics o a l'ús de panels sensorials, realment adequats per a anàlisis de cribratge (*screening analysis*).

4.1.3. Quantificació de compostos fenòlics en caves rosats

Per últim, i donada la importància dels compostos fenòlics, també es va avaluar la seva quantificació en els caves. Els compostos fenòlics inclouen un ampli grup de compostos químics que afecten al sabor, color i paladar del vi i derivats [9]. Entre aquests compostos s'inclouen els àcids fenòlics, estilbens, flavonols, dihidroflavonols, antocians i flavanols tant en forma de monòmers (catequines) com de polímers (proantocianidines); els quals poden ser separats en dos categories principals: flavonoides i no-flavonoides. Els flavonoides inclouen els antocians i els tanins, els quals contribueixen al color i paladar del vi [2], mentre que els no-flavonoides inclouen els estilbens tals com el resveratrol, i àcids fenòlics com els àcids benzoic, cinàmic i cafèic.

A banda del contingut total de polifenols, els tanins i els antocians són dues classes de flavonoides amb gran importància, sobretot en el cas dels vins rosats i negres, ja que tenen una clara influència en el seu color i paladar, sobretot en la seva astringència [10]. Per una banda, els tanins poden afectar el color, la capacitat d'envelliment i la textura del vi. Tot i que aquests no aporten cap olor o sabor, sí que poden ser percebuts durant

la degustació dels vins per la sensació de sequedat i d'amargor que poden deixar en la boca. D'altra banda, els antocians són compostos inodors i gairebé insípid, contribuint al gust del vi com una sensació moderadament astringent; a més, la seva concentració dicta el color del vi.

Més enllà de la seva contribució a les característiques sensorials del vi, la majoria d'aquests compostos són potents antioxidants amb beneficis per a la salut derivats de la seva acció com a captadors de radicals lliures i inhibidors de l'oxidació de les lipoproteïnes [11]. Com a conseqüència de tot plegat, donada la seva importància, diversos mètodes per quantificar els diferents compostos fenòlics (ja sigui el contingut total o la identificació individual) són reportats a la literatura [12]; entre els que destaquen l'espectrofotometria i la cromatografia com els mètodes més comuns (§4.4.1 i §4.4.2). No obstant, aquests procediments requereixen etapes de preparació addicionals, l'ús d'instruments de laboratori pesats, i no són adequats per a l'anàlisi *in-situ*.

En aquest sentit, un dels objectius de la present tesi serà la quantificació d'aquests compostos en els diferents tipus de begudes analitzades. Centrant-nos en aquest cas en la quantificació del contingut total de diferents classes (antocians i tanins), però també en el contingut total de compostos fenòlics.

Amb aquesta finalitat, i a través d'una col·laboració amb l'INCAVI, aquests ens van subministrar un total de 20 ampolles de diferents caves rosats; el contingut en compostos fenòlics dels quals va ser analitzat segons els mètodes de referència per tal d'avaluar la capacitat de la llengua electrònica per modelar aquests. Per una banda, es van utilitzar tres índexs diferents relacionats amb el contingut total de polifenols: Folin-Ciocalteu, I_{280} i I_{320} (§4.4.1) [13]; mentre que per l'altra banda, el contingut total de tanins i antocians van ser també quantificats utilitzant mètodes estàndard [8].

Per a la mesura de les mostres, es va seguir el mateix protocol que en els casos anteriors, sols que aquest cop es va decidir fer ús de la BioET formada pels (bio)sensors (§3.1, Taula 3.1): GEC, Tyr, Lac i Cu. Per tant, per la mesura de les mostres, es va seguir el procediment descrit a §3.3 aplicant un escombrat de potencial entre $-0.4V$ i $+0.8V$ vs. Ag/AgCl amb un salt de potencial de $9mV$ i una velocitat d'escombrat de $100 mV \cdot s^{-1}$; fent ús d'una etapa de neteja electroquímica entre mostra i mostra, consistent en l'aplicació d'un potencial de $+1.0V$ durant 40 segons en un vas que contenia 25 mL d'aigua destil·lada.

El conjunt de mostres va ser doncs analitzat sota les condicions descrites de forma

aleatòria, obtenint un registre voltamperomètric sencer per cadascun dels quatre sensors. En aquest sentit, la Figura 4.10 mostra un extracte dels resultats obtinguts per algunes d'aquestes mostres seleccionades arbitràriament. Tal com es mostra, diferents senyals són obtinguts per cadascun dels sensors, obtenint majors intensitat de corrent per les mostres amb major contingut fenòlic; observant, com a tendència general, l'oxidació dels compostos fenòlics sobre la superfície de l'elèctrode en tots els casos, mentre que en els cas dels dos biosensors i el sensor Cu també s'obtenen corrents reductors a la regió pròxima a 0 V a causa del seu efecte catalític.

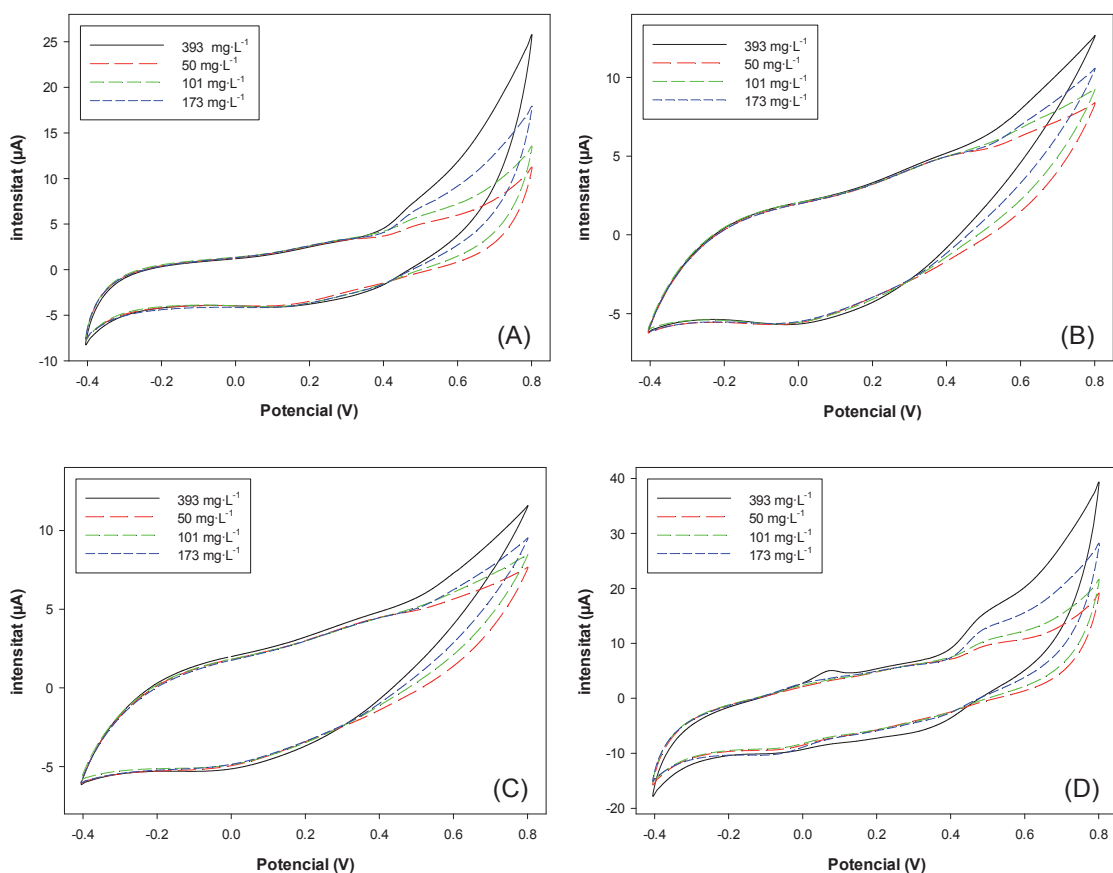


Figura 4.10. Exemple dels diferents voltamperogrames obtinguts per diferents mostres arbitràries amb els (bio)sensors: (A) GEC, (B) Tyr, (C) Lac i (D) Cu. Els valors expressats corresponen a la concentració de tanins mesurada mitjançant el mètode de referència.

Al igual que en el casos anteriors, i per tal d'extreure la màxima informació de cadascun dels registres voltamperomètrics, es requereix d'una etapa de preprocessament del senyal per tal de comprimir i reduir l'elevada dimensionalitat d'aquest (mostres \times sensors \times potencials aplicats). A diferència del cas anterior, aquest cop es va fer ús de la *integral de finestra seccionada* per tal d'assolir aquesta reducció (§3.4.1.4); escollint el número de seccions en base als resultats d'altres estudis (§4.3.2).

A partir de la resposta voltamperomètrica, es podria afirmar que la matriu de (bio)sensors de la BioET està responent al diferent contingut fenòlic de les mostres. No obstant, per tal de detectar altres similituds entre aquestes o capacitats de la BioET per discernir altres factors relatius a les mostres, el conjunt de voltamperogrames comprimits va ser analitzat mitjançant PCA (Figura 4.11). Tot i no poder ser considerat com un mètode de reconeixement de patrons, ja que només proporciona una representació visual de les relacions entre les mostres i les variables, és una eina molt útil ja que proporciona una idea de com a partir de les variables mesurades, algunes mostres presenten un comportament similar, o per contra difereixen l'una de l'altra.

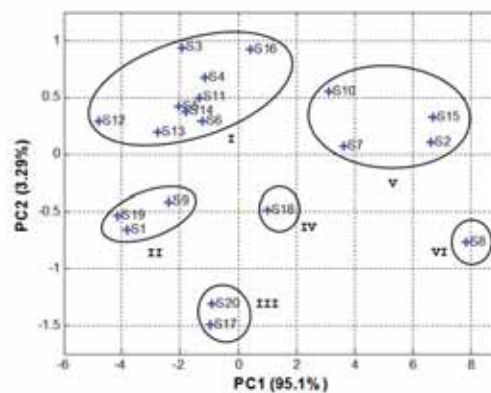


Figura 4.11. Representació dels scores dels dos primers components principals obtinguts després de l'anàlisi PCA; un total de 20 caves rosats van ser analitzats.

En primer lloc, cal denotar que amb només els dos primers PCs, la variància acumulada és ca. el 98.4%, un valor elevat que significa que gairebé tota la variància continguda en la informació original està representada ara per només aquestes dos noves coordenades. En segon lloc, com es pot veure pel simple anàlisi visual del gràfic, s'observa l'aparició d'algunes zones d'agrupament després d'aquesta transformació, la qual cosa indica algunes similituds entre les mostres.

En aquest sentit, es pot observar com les mostres sembla que s'agrupin en funció del seu contingut fenòlic, per exemple, el *cluster* II sembla agrupar les mostres amb un contingut fenòlic més baix, mentre que el *cluster* V sembla implicar les mostres amb el major contingut. D'altra banda, la mostra S18 correspon a un cava de la classe *Brut Nature*, mentre que gairebé tots els altres pertanyen a la classe *Brut*, per tant sembla que la BioET també sigui capaç de distingir aquest fet parcialment (§4.1.1.1).

A més a més, a part del contingut fenòlic, la BioET sembla ser capaç de distingir altres característiques com ara l'envelliment o la regió. Per exemple, la mostra S20

pertany a una cava molt poc envellit, el que podria indicar que la mostra S17 és també un cava jove. Una altra particularitat de la mostra S8 és la seva regió, l'Empordà, mentre que la resta de mostres són principalment de la regió del Penedès. Per tant, mostrant així una àmplia resposta i sensibilitat per a detectar petites discrepàncies entre les mostres per part de la BioET.

Després de la construcció d'aquest model preliminar que va permetre reconèixer alguns patrons inicials i similituds entre les mostres, confirmant al mateix temps que la BioET també estava responen al contingut fenòlic de les mostres, el següent pas va ser la construcció d'un model neuronal que permetés la predicció quantitativa del contingut fenòlic de les mostres de cava rosat.

Després d'una avaluació sistemàtica de les diferents topologies, l'arquitectura del model ANN final estava formada per 44 neurones en la capa d'entrada (4 sensors x 11 coefs. obtinguts en l'etapa de compressió del senyal), 10 neurones i la funció de transferència *purelin* en la capa oculta, i 5 neurones (una per cadascun dels índexs considerats) i la funció de transferència *purelin* en la capa de sortida.

Per avaluar l'eficàcia del model generat i la capacitat de la BioET, es va fer ús del mètode de validació creuada *leave-one-out* degut al baix nombre de mostres disponibles (§3.4.2.3.1). Construint a continuació els gràfics de comparació de concentració obtinguda vs. esperada per cadascun dels cinc índexs, tant pel subconjunt d'entrenament com pel de test. Com a exemple d'aquests, la Figura 4.12 mostra el gràfic obtingut en el cas del contingut de tanins de les mostres de cava rosats.

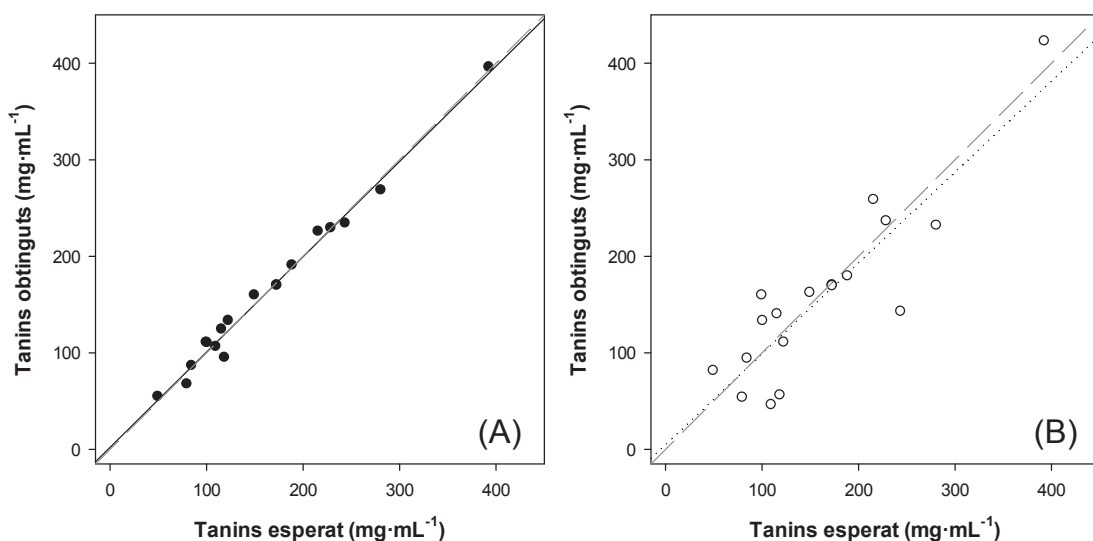


Figura 4.12. Capacitat de modelatge del model Int-ANN. Gràfics de comparació de les concentracions obtingudes vs. les esperades pel contingut de tanins total, tant pel (A) subconjunt d'entrenament (●, línia sòlida) com (B) de test (○, línia de punts). La línia discontinua correspon a la diagonal teòrica.

Tal com es pot observar, s'obté una tendència satisfactòria, amb línies de regressió pràcticament indistingibles de les teòriques; a més, tal com és habitual, obtenint una millor tendència pel subconjunt d'entrenament. Anàlogament, els mateixos gràfics varen ser construïts pels altres quatre índexs i s'ajustaren les respectives línies de regressió, els paràmetres de les quals es troben resumits en la Taula 4.5. Tal com es podria esperar del gràfic anterior, s'obté una tendència satisfactòria en general, malgrat els baixos coeficients de correlació obtinguts en determinats casos. Per tant, indicant la capacitat de la BioET per assolir aquesta tasca, però requerint d'un major nombre de mostres per poder dur a terme una avaluació i validació més extensiva.

Taula 4.5. Paràmetres de les rectes de regressió ajustades dels gràfics de comparació entre valors obtinguts vs. esperats pel model Int-ANN proposat, pels cinc índexs avaluats, utilitzant el mètode de validació creuada *leave-one-out*. Interval de confiança calculats amb un nivell de confiança del 95%.

| | <i>Subconjunt d'entrenament</i> | | | | |
|--|---------------------------------|--------------------|-------------------|-------------|--------------------|
| | <i>Pendent</i> | <i>Ord. Origen</i> | <i>Correlació</i> | <i>RMSE</i> | <i>Total NRMSE</i> |
| <i>Índex FC (mg·L⁻¹)</i> | 1.018±0.147 | -1.4±29.9 | 0.965 | 15.1 | |
| <i>Índex I₂₈₀ (arb. unit)</i> | 0.955±0.109 | 0.45±1.14 | 0.977 | 0.55 | |
| <i>Índex I₃₂₀ (arb. unit)</i> | 0.896±0.236 | 0.61±1.27 | 0.895 | 0.49 | 0.077 |
| <i>Tanins (mg·L⁻¹)</i> | 0.987±0.060 | 2.1±11.0 | 0.993 | 9.50 | |
| <i>Antocians (mg·L⁻¹)</i> | 0.869±0.167 | 1.2±2.2 | 0.940 | 2.36 | |
| | <i>Subconjunt de test</i> | | | | |
| | <i>Pendent</i> | <i>Ord. Origen</i> | <i>Correlació</i> | <i>RMSE</i> | <i>Total NRMSE</i> |
| <i>Índex FC (mg·L⁻¹)</i> | 0.977±0.257 | 8.9±52.4 | 0.896 | 26.6 | |
| <i>I₂₈₀ index (arb. unit)</i> | 0.973±0.308 | 0.51±3.22 | 0.858 | 1.52 | |
| <i>I₃₂₀ index (arb. unit)</i> | 0.872±0.358 | 0.82±1.92 | 0.791 | 0.74 | 0.151 |
| <i>Tanins (mg·L⁻¹)</i> | 0.942±0.261 | 4.8±47.7 | 0.886 | 41.4 | |
| <i>Antocians (mg·L⁻¹)</i> | 0.739±0.340 | 1.9±4.6 | 0.755 | 4.89 | |

4.2. Anàlisi de brandis

A arrel dels resultats assolits en el cas dels caves, en que havíem estat capaços de classificar aquests de forma anàloga a la que faria un “sommelier”; des de l'empresa *Miguel Torres, S.A.*, se'ns va proposar dos problemàtiques a abordar amb interès en el sector dels brandis: la predicció de la qualitat d'aquests d'acord amb la classificació realitzada per un panel de tast i la detecció de l'ús de diferents mètodes d'envelliment en el procés de producció d'aquests. Per tal d'entendre la base d'aquesta problemàtica i la procedència del conjunt de mostres utilitzat, a continuació es descriu breument el procés d'elaboració del brandi.

El brandi és un aiguardent produït a partir de destil·lats de vi amb un abv ca. 36-45%; la producció del qual és un procés complex que consta de diversos passos

importants, incloent la producció del vi base, selecció d'aquests, primera i segona destil·lació, envelliment i cupatge.

Entre els diferents brandis produïts, destaca el *Cognac*, el qual és el més conegut de tots. Tot i que aquest només es refereix a un tipus de brandi produït en aquesta regió francesa, molts productors de diferents països intenten emular els seus mètodes de producció. En particular, el *Cognac* ha de ser elaborat a partir de varietats de raïm específiques, obtingut a partir d'una doble destil·lació del vi base en alambins de coure i envellit durant almenys dos anys en bótes de roure francès de certes regions. En aquest sentit, el brandi madura de la mateixa manera com ho fa el whisky i el vi quan són envellits en bótes; per tant, és bastant habitual entre els productors el fet de deixar envellir aquests per períodes inclús més llargs que el requerit legalment per tal de millorar les seves característiques organolèptiques. En general, tant abans de començar el procés d'envelliment com al final d'aquest, els destil·lats són tastats per tal de classificar-los i avaluar-ne la qualitat.

Aquesta classificació és generalment realitzada per un panel sensorial expert, donades les dificultats per dur a terme aquesta avaluació mitjançant les tècniques d'anàlisi clàssiques. Això és degut a que els principals compostos responsables de l'estil del brandi no han estat encara identificats, el que a la vegada dificulta l'ús dels mètodes d'anàlisi clàssics ja que implicaria la detecció dels compostos responsables d'aquestes sensacions. Per contra, tal com s'ha descrit a §1.4, les ETs combinen tota la informació extreta de la matriu de sensors per generar un model de forma anàloga a la que ho fa el cervell, evitant centrar-se en quins són els compostos específics responsables d'aquestes sensacions; però, tot i així, reconeixent-ne la seva empremta, la qual en permet la seva identificació o quantificació. Per tant, convertint-les en unes candidates idònies per la solució d'aquesta problemàtica.

D'altra banda, cal destacar l'aparició recent de mètodes alternatius d'envelliment, els quals tenen com a objectiu assolir la reducció en el temps de cria en bóta, intentant imitar i preservar el seu efecte en el brandi. Una d'aquestes tècniques consisteix en l'addició de petites peces de fusta, les quals proporcionen una major superfície de contacte, accelerant d'aquesta manera el procés [14]. Conseqüentment, generant una nova problemàtica en el sector, que és la detecció de l'ús d'aquesta pràctica. Malauradament, al igual que en el cas anterior, aquesta situació és detectada a través de panels de tast ja que existeix una manca de mètodes capaços de dur a terme aquesta tasca.

4.2.1. Destil·lats de vi

La resolució dels dos objectius plantejats en aquest sector, es va abordar per separat; a saber, a partir de la mesura de dos subconjunts de mostres diferenciats. Aquesta divisió de la problemàtica va ser necessària, ja que tal com s'ha dit, en el moment de dissenyar una aplicació d'una ET, les mostres utilitzades han de ser representatives de la situació que volem abordar. Per tant, en una primera etapa es va estudiar la correlació de la resposta de la ET amb el panel de tast (A), mentre que posteriorment es va procedir a la detecció de diferents mètodes d'envelliment (B); utilitzant dos conjunts de mostres especialment seleccionades per cada un dels casos: 41 mostres pel primer cas (conjunt A) i 33 pel segon (conjunt B). En tots els casos, les mostres analitzades corresponien a destil·lats de vi, és a dir, mostres preses durant el procés d'elaboració del brandi.

Les mostres del conjunt A van ser prèviament seleccionades i tastades per un panel de tast, el qual les va classificar en tres categories diferents d'acord amb els seus atributs organolèptics. D'aquesta manera, les mostres més apreciades van ser assignades a la classe "tast 1", mentre que les més dolentes van ser assignades a la classe "tast 3". Assignant a la classe 1 les mostres que tenen un millor aroma i sabor en boca, sense cap defecte apreciable; la classe 2, les mostres que presenten algun lleuger defecte, requerint l'ús de mètodes estàndards pel tal de garantir que certs compostos indicadors de defectes es troben per sota el llindar desitjat, però que un cop confirmat poden encara ser utilitzades per l'elaboració de brandis, controlant-ne el cupatge; i finalment, la classe 3, contenint les mostres que no assoleixen els estàndards exigits. Addicionalment, aquestes mostres van ser caracteritzades mitjançant mètodes clàssics per tal d'obtenir major informació relativa a la concentració de certs compostos indicadors de defectes que facilités la comprensió de les respostes obtingudes.

El conjunt B de mostres es trobava format per destil·lats de vi de la mateixa collita i envellits utilitzant diferents mètodes, viz. l'ús de peces de fusta (xips, barretes, blocs i llistons) en el procés d'envelliment dels brandis per tal d'imitar l'envelliment en bóta i accelerar el procés. Per tant, també es van incloure un conjunt de brandis envellits en bóta, però d'una collita anterior donat que l'objectiu de l'ús de peces de fusta és accelerar el procés. Addicionalment, analitzant també el seu contingut en polifenols per investigar si existeix alguna relació clara entre el contingut total d'aquests i el mètode d'envelliment.

Al igual que en el cas dels caves, el primer pas va ser establir un protocol de mesura

que garantís la qualitat dels resultats obtinguts. En aquest sentit, es va prendre certes precaucions per tal d'evitar que paràmetres fora de control o el medi de mesura poguessin alterar el model o la integritat dels sensors; a saber, dilució de les mostres, control del abv i reducció de les fonts de variabilitat en les mostres.

La primera precaució va consistir en la dilució de les mostres amb KCl 0.1M, i tenia com a fi augmentar la conductivitat del medi, evitant a la vegada que l'alt contingut alcohòlic pugui malmetre els sensors. A més, aquest procés de dilució també va ser utilitzat per a igualar el diferent abv de les mostres, el qual va ser mesurat amb anterioritat mitjançant el mètode de referència [15], per tal d'evitar que aquest esdevingui una font de discriminació; fenomen que han reportat altres autors [16]. Finalment, en el cas de la detecció de l'ús de mètodes alternatius d'envelliment, les mostres considerades van ser seleccionades de la mateixa collita i es va mesurar l'índex de polifenols donat que alguns autors suggereixen que aquesta pot ser una de les fonts de variabilitat [14].

Donat els bons resultats assolits en el cas dels caves, i basant-nos en el coneixement adquirit en aquest, es va decidir fer ús de la llengua electrònica formada aquest cop pels sensors (§3.1, Taula 3.1): GEC, CoPc, PPy, Cu, Pt i Ag. Substituint així el sensor PANI pel sensor Ag, i mantenint el sensor CoPc. Al igual que en el cas anterior, les mostres van ser analitzades sense dur cap més pretractament que la seva dilució en KCl 0.1M.

Per la mesura de les mostres, es va seguir el procediment descrit a §3.3, aplicant un escombrat de potencial entre -1.0V i +1.3V vs. Ag/AgCl amb un salt de potencial de 9mV i una velocitat d'escombrat de $100\text{mV}\cdot\text{s}^{-1}$; fent ús d'una etapa de neteja electroquímica entre mostra i mostra, consistent en l'aplicació d'un potencial de +1.5V durant 40 segons en un vas que contenia 25mL d'aigua destil·lada.

Exemples de la resposta obtinguda per les diferents mostres dels dos conjunts mitjançant la ET es mostren a la Figura 4.13; on els voltamperogrames de l'esquerra corresponent al conjunt A (tast) i els de la dreta al conjunt B (envelliment).

En el primer cas, es pot veure com algunes mostres presenten respostes voltamperomètriques molt similars en regions específiques; i com a més, semblances anàlogues es van repetint al llarg de tot el voltamperograma, però entre diferents mostres. Per exemple, a la Figura 4.13A (sensor GEC) es pot observar com les mostres A16, A23 i A29 mostren una resposta anòdica equivalent a la regió al voltant de +1V; mentre que en la resposta catòdica (ca. 0V), la mostra A23 mostra una resposta diferenciada, però mantenint encara una resposta similar les mostres A16 i A29, més la

mostra A27, la qual presenta la mateixa resposta en aquesta regió, però no en l'altra. A més a més, obtenint diferents tendències per als diferents sensors que formen la ET; per exemple, en el cas del sensor Pt (Figura 4.13C) es pot observar con només les mostres A1 i A4 mostren una resposta similar tant a la regió anòdica com catòdica entre -0.5V i -1V, comportament que estaria d'acord amb la classificació atorgada pel panel de tast. En definitiva, s'observa una gran font de variabilitat i resposta creuada per les diferents mostres i els diferents sensors, una situació molt desitjable en els cas de les ETs.

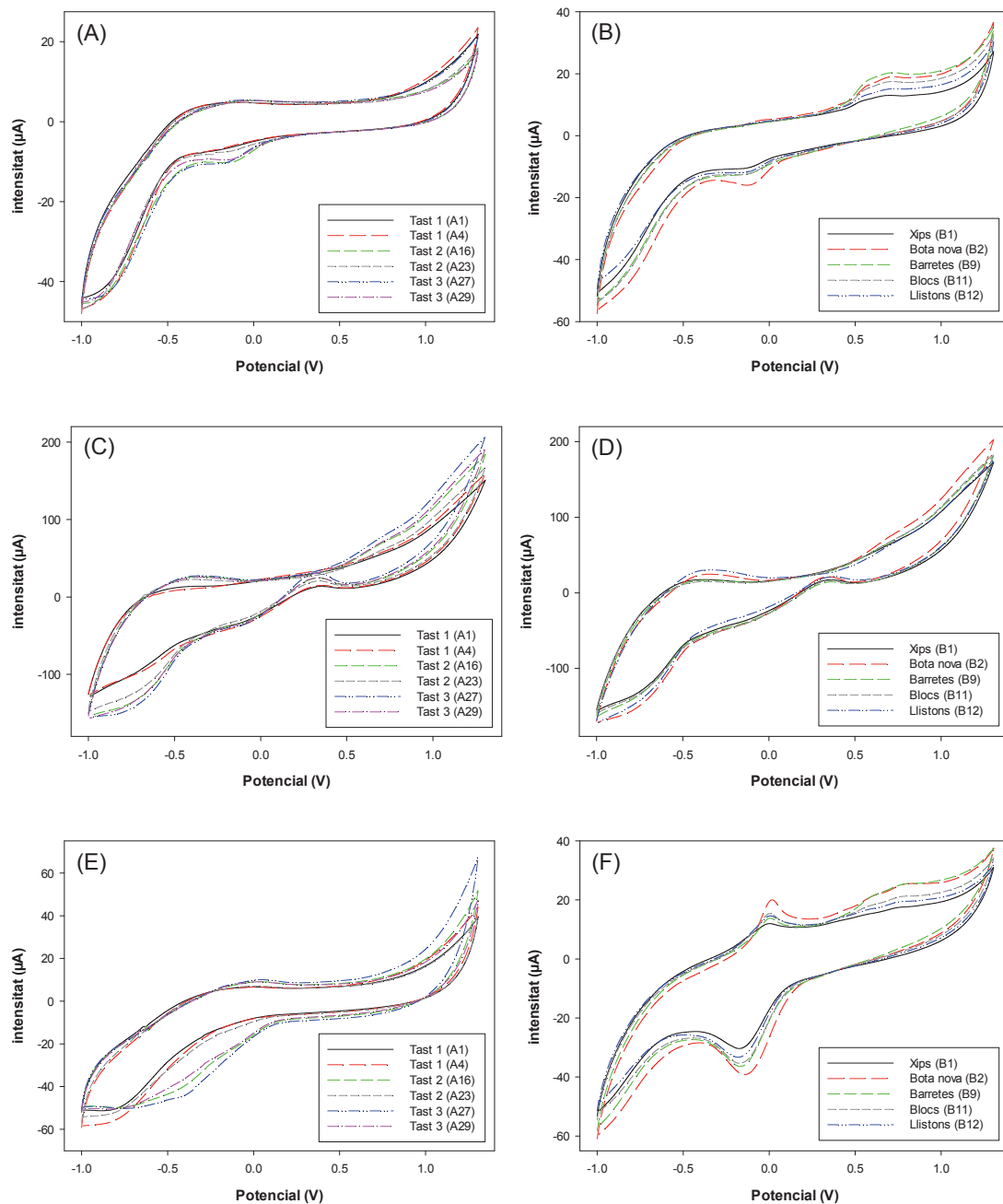


Figura 4.13. Exemple dels diferents voltamperogrames obtinguts per certes mostres arbitràries amb tres dels sensors usats: (A,B) GEC, (C,D) Pt i (E,F) Cu. Els gràfics de l'esquerra corresponen a l'estudi de la qualitat dels brandis (conjunt A), mentre que els de la dreta corresponen als brandis elaborats fent ús de diferents mètodes d'envelliment (conjunt B).

De la mateixa manera, per les mostres del segon conjunt, es pot observar un comportament similar i consistent; mostrant com a principal diferència l'aparició d'alguns senyals oxidatius que poden ser deguts al major contingut de compostos fenòlics de certes mostres (§4.4.1.1); tot i que com es veurà més endavant, aquesta no esdevé la base de la discriminació entre les diferents categories (§4.2.4, Figura 4.17).

Per tal d'explotar al màxim tota la informació present en els registres voltamperomètrics, i abans de construir el model, es requereix d'una etapa de compressió dels senyals enregistrats. En aquest cas, es va fer ús de la FFT, la qual va permetre reduir cada registre voltamperomètric a només 16 coeficients sense cap pèrdua de informació significativa, assolint un ratio de compressió del 96.9%.

4.2.2. Correlació amb el panel de tast

Per tal de comprovar si la ET era capaç de catalogar les diferents mostres de destil·lats (conjunt A) d'acord amb les classes assignades pel panel de tast, aquestes van ser primer "analitzades" pel panel, els qual les van classificar en 3 classes: de "1" a "3", és a dir, de més bo a més dolent. Procedint posteriorment a la construcció d'un model LDA; el qual utilitza els coeficients extrets del conjunt de voltamperogrames comprimits mitjançant FFT com a *inputs*, i les classes assignades pel panel com a *sortida*.

La Figura 4.14 mostra el resultat del model construït, on tal com és pot observar, s'assoleix una clara distinció per les tres classes de destil·lats; a més, mostrant una ordenació de les mostres al llarg de la DF1 segons la qualitat d'aquestes, d'acord amb l'ordre assignat pel panel. La qualitat del model va ser avaluada fent ús del mètode de validació creuada *leave-one-out* (Taula 4.6); assolint la correcta classificació de totes les mostres, tal com es podia esperar a partir del gràfic. Addicionalment, l'eficàcia del model es va quantificar a partir de 3 indicadors diferents (§3.4.2.3.2.1): grau d'incert, sensibilitat i especificitat; assolint el 100% pels tres indicadors estadístics. Demostrant per tant, la capacitat de la ET de classificar els destil·lats de forma anàloga al panel de tast.

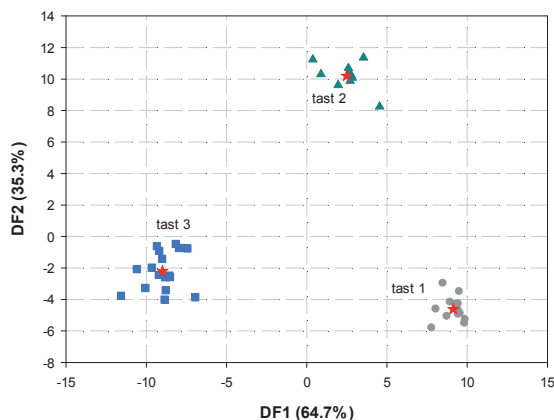


Figura 4.14. Representació dels *scores* de les dos primeres funcions discriminants obtingudes després de l'anàlisi LDA d'acord amb les classes: (●) tast 1, (▲) tast 2 i (■) tast 3. Addicionalment, també s'ha representat el centroide de cada classe (★).

Taula 4.6. Matriu de confusió construïda d'acord amb la nota de tast pronosticada mitjançant el model FFT-LDA, utilitzant el mètode de validació creuada *leave-one-out*.

| | <i>Predit</i> | | | |
|----------------|---------------|----------|----------|----|
| | <i>1</i> | <i>2</i> | <i>3</i> | |
| <i>Esperat</i> | <i>1</i> | 14 | 0 | 0 |
| | 2 | 0 | 10 | 0 |
| | 3 | 0 | 0 | 17 |

4.2.3. Quantificació compostos indicadors de defectes en els brandis

Per tal de complimentar la informació qualitativa referent a la categoria de les mostres, la identificació de compostos indesitjats va ser també avaluada de forma semi-quantitativa. Posant així més atenció a la detecció de compostos específics, identificats com a responsables d'induir alguns defectes en els destil·lats de vi i els brandis; intentant la seva quantificació per mitjà de la ET.

Tot i que l'objectiu inicial era únicament la categorització de les mostres segons la seva classe; la identificació d'aquests compostos és important com a complement d'aquesta, ja que com s'ha dit, i especialment en les mostres assignades a la categoria "tast 2", es requereix de la identificació d'aquests mitjançant mètodes clàssics per tal d'assegurar que no sobrepassen un valor llindar. Així doncs, seria molt interessant que a partir d'una única mesura, fos possible no tan sols d'assignar la classe, sinó també proporcionar el nivell d'aquests ja que permetria prendre la decisió de forma automàtica sense haver d'esperar a posteriors anàlisis.

Generalment, els alcohols amb més de dos àtoms de carboni, coneguts com alcohols superiors i produïts principalment durant la fermentació, així com també els seus èsters, tenen olors intenses que juguen un paper important en els aromes del vi. A baixes concentracions, contribueixen a la seva complexitat aromàtica; mentre que a nivells més elevats, les seves olors penetrants emmascaren la seva finor aromàtica [10, 17].

Particularment, el butan-2-ol, l'acetat d'etil, l'acetaldehid i el butan-1-ol van ser els compostos específics estudiats. Entre ells, el butan-2-ol com a tal no té cap incidència organolèptica però és un marcador d'alteracions bacterianes sovint provinents de les mares (lies) en vins mal conservats; l'acetaldehid té un caràcter eteri, és un compost intermedi de la fermentació alcohòlica, podent augmentar la seva concentració a causa de diversos factors, com ara l'addició de SO₂; l'acetat d'etil indueix una olor estranya (similar a la cola *imedio*) i es genera en vins picats (avinagrats) a partir de l'àcid acètic i l'etanol.

Per estimar la presència d'aquests marcadors, els nivells de concentracions de cada un dels compostos es van dividir en dominis específics en una escala de tres nivells (a saber: -1 per a nivells normals, 0 per un lleuger defecte i +1 en cas d'un clar defecte) d'acord amb els requisits de la indústria alimentària (Taula 4.7) [18], valors que van ser llavors modelats emprant lògica difusa (*fuzzy logic*).

Taula 4.7. Rangs de concentració, expressats en mg·L⁻¹, utilitzats per la divisió de les mostres en una escala de tres nivells per dur-ne a terme la seva determinació de forma semi-quantitativa.

| | <i>butan-2-ol</i> | <i>acetat d'etil</i> | <i>acetaldehid</i> | <i>butan-1-ol</i> |
|----|-------------------|----------------------|--------------------|-------------------|
| -1 | < 1 | < 350 | < 30 | < 3 |
| 0 | 1 ≤ X < 2.5 | 350 ≤ X < 450 | 30 ≤ X < 50 | 3 ≤ X < 6.6 |
| +1 | ≥ 2.5 | ≥ 450 | ≥ 50 | ≥ 6.6 |

Tot i que els nivells d'alguns compostos específics poden ser utilitzats com a marcadors d'alguns defectes del brandi, s'ha de tenir en compte que són només indicadors i que aquests no determinen la categoria assignada organolèpticament. És a dir, a pesar que algunes mostres puguin tenir nivells relativament elevats d'un d'aquests compostos, això no implica que la seva classificació serà degradada, donat que la seva qualitat és també afectada a causa de la presència d'altres compostos no identificats o quantificats. En aquest sentit, alguns exemples poden trobar-se en la Taula S1 de §7.1, per exemple, les mostres A36 i A37 ambdós tenen nivells baixos dels quatre compostos considerats, i tot i així són assignades pel panel de tast a la classe de pitjor qualitat en quant a gust, mentre que la situació oposada es dona per exemple per la mostra A8.

D'aquesta manera, es va procedir a la construcció d'un model ANN semi-quantitatiu per a la quantificació ordinal dels quatre compostos diferents emprant el 75% de les dades per a l'entrenament d'aquest, i utilitzant el 25% de les dades restant (subconjunt de test) per a l'avaluació de la capacitat del model; repetint aquesta subdivisió 4 vegades, emprant el mètode de validació creuada *k-fold* (§3.4.2.3.1). És a dir, les dades

es van distribuir inicialment en quatre grups a l'atzar, i a continuació, es va construir un model utilitzant tres dels subconjunts pel seu entrenament, reservant el quart per avaluar la seva eficàcia; repetint aquest procés quatre vegades, deixant fóra un subconjunt diferent cada vegada per tal d'avaluar les seves capacitats de predicció. Després, una vegada que tots els models són construïts, les dades són agrupades depenen de si han intervingut en el procés d'entrenament o en el de validació.

Després de l'optimització del model, aquest estava format per 96 neurones en la capa d'entrada (corresponents als 16 coefs. extrets de cada voltamperograma x 6 sensors), 6 neurones i la funció de transferència *tansig* en la capa oculta, i quatre neurones i la funció de transferència *purelin* en la capa de sortida, proporcionant els nivells de cadascun dels compostos estudiats en les diferents mostres.

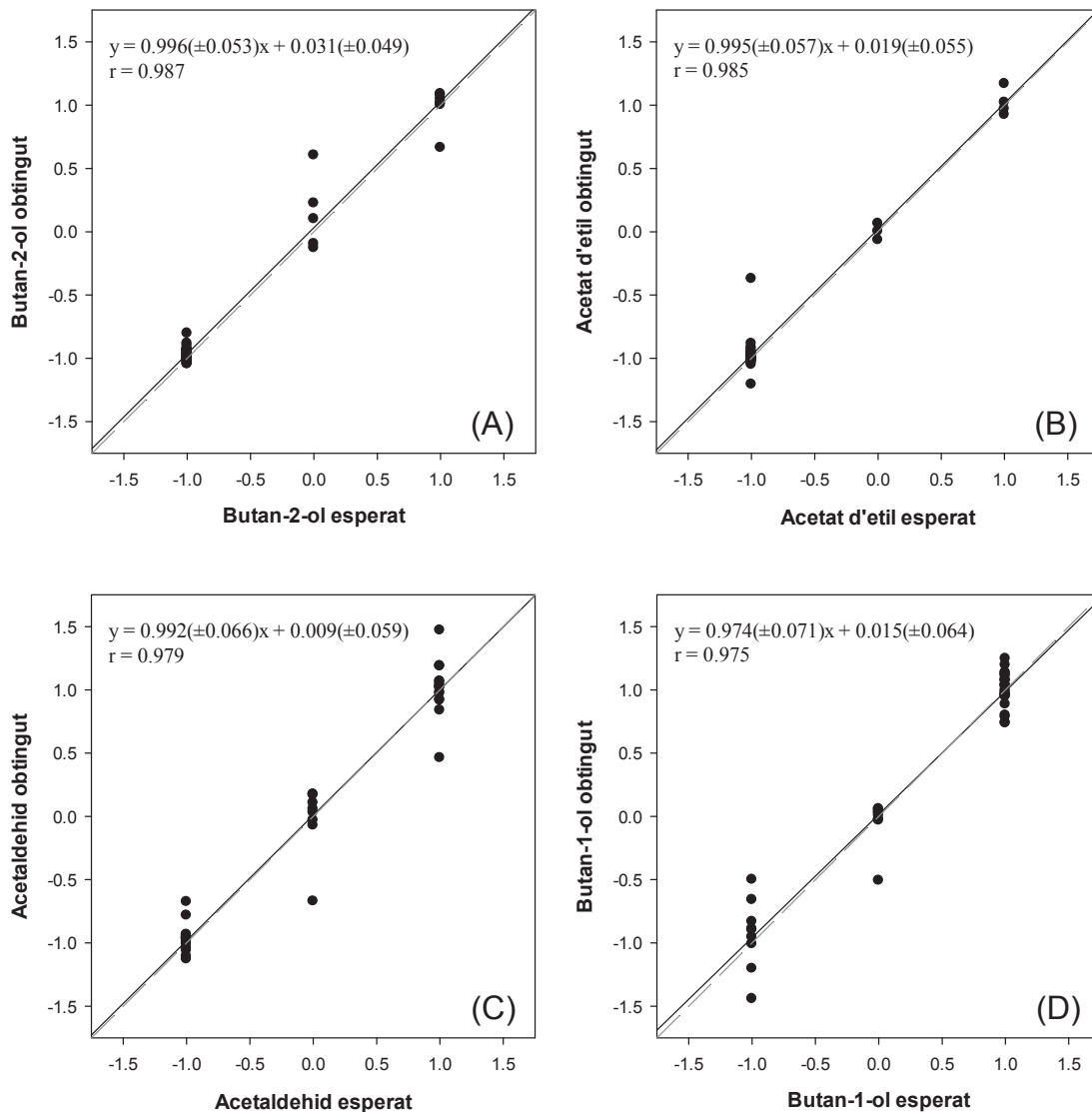


Figura 4.15. Capacitat de modelatge del model FFT-ANN per la predicció semi-quantitativa de quatre compostos utilitzats com a marcadors de defectes en els brandis. Gràfics de comparació dels valors obtinguts vs. els esperats pel subconjunt de test, on la línia discontinua correspon a la diagonal teòrica.

Com es pot veure en la Figura 4.15, on es mostren els gràfics de comparació dels nivells predits vs. esperats per a cada un dels compostos, va ser possible assolir una bona capacitat de predicció per part model. Tot i la major dispersió obtinguda per algunes mostres, s'observa una bona tendència en general, on les línies de regressió ajustades són gairebé indistingibles de les teòriques.

4.2.4. Detecció de l'ús de mètodes d'envelliment alternatius

Assolit el primer objectiu, es va procedir mitjançant les mesures realitzades amb el segon conjunt de mostres (conjunt B), a comprovar si el sistema era capaç de distingir l'ús de peces de fusta durant el procés d'envelliment amb la finalitat d'accelerar aquest, imitant al mateix temps el procés que té lloc en les bótes i les característiques que aquesta aporta. Al igual que en el cas anterior, es va procedir a la construcció d'un model LDA utilitzant com a *inputs* els coeficients extrets en l'etapa de compressió del registre voltamperomètric, i els diferents mètodes seguits com a sortida d'aquest (les classes a distingir).

Donat que alguns autors han descrit que un paràmetre clau en aquest procés és la mida i forma de les peces de fusta utilitzades [14], és van avaluar quatre formes diferents d'aquestes i comparar amb el mètode convencional d'envelliment en bóta nova. Per tant, les mostres es van distribuir en els següents cinc grups: bóta nova, xips, barretes, blocs i llistons.

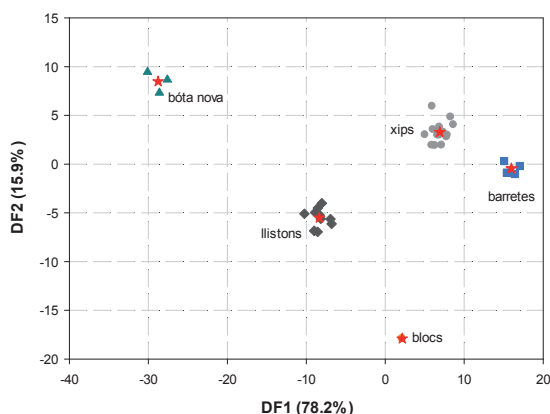


Figura 4.16. Representació dels *scores* de les dos primeres funcions discriminants obtingudes després de l'anàlisi LDA d'acord amb les classes: (●) xips, (▲) bóta nova, (■) barretes, (●) blocs i (◆) llistons. Addicionalment, també s'ha representat el centre de cada classe (★).

Taula 4.8. Matriu de confusió construïda d'acord el mètode d'envelliment pronosticat mitjançant el model FFT-LDA, utilitzant el mètode de validació creuada *leave-one-out*.

| | | <i>Predit</i> | | | | |
|----------------|------------------|---------------|-----------|------------|--------------|--------------|
| | | <i>Xip</i> | <i>BN</i> | <i>Bar</i> | <i>Blocs</i> | <i>Llist</i> |
| <i>Esperat</i> | <i>Xip</i> | 15 | 0 | 0 | 0 | 0 |
| | <i>Bóta Nova</i> | 0 | 3 | 0 | 0 | 0 |
| | <i>Barretes</i> | 0 | 0 | 4 | 0 | 0 |
| | <i>Blocs*</i> | 0 | 0 | 0 | 0 | 1 |
| | <i>Llistons</i> | 0 | 0 | 0 | 0 | 10 |

*Només es disposava d'1 mostra

Tal com es mostra a la Figura 4.16, la baixa dispersió dels *clusters* evidència la capacitat de la ET per discriminar entre els 5 mètodes d'envelliment diferents utilitzats. Analitzant el gràfic més detalladament, s'observa com les mostres envellides en bóta són les més distants respecte els altres *clusters*; mentre que la resta de *clusters* són més propers els uns als altres. En aquest sentit, la distància de cada grup al grup corresponent a les mostres envellides en bóta nova, pot proporcionar una idea de quant semblant és el mètode proposat al mètode original.

Per tal d'avaluar l'eficàcia del model, es va fer ús del mètode de validació creuada *leave-one-out* i es va construir la matriu de confusió a partir dels resultats de les mostres del subconjunt de test (Taula 4.8). Tal com es podria esperar del gràfic LDA, gairebé totes les mostres van ser ben classificades, assolint un grau d'encert del 97.0%, mentre que els valors de sensibilitat i d'especificitat assolits varen ser del 88.0% i el 99.1%, respectivament. Cal remarcar que únicament una sola mostra va resultar mal classificada, la qual correspon a la classe "blocs"; aquest error és degut a que només disposàvem d'una mostra per fer l'estudi. Per tant, en el moment de validar el model mitjançant el mètode *leave-one-out*, l'ús d'aquesta mostra en el subconjunt de validació i l'absència d'una altra mostra similar en el conjunt d'entrenament, n'impedeix la seva correcta classificació; resultant en l'assignació d'aquesta al *cluster* més proper per part de l'algorisme de reconeixement de patrons, en aquest cas la classe "llistons".

4.2.5. Quantificació de compostos fenòlics

Per tal d'assegurar que la discriminació dels diferents mètodes d'envelliment no era deguda al diferent contingut en compostos fenòlics, el contingut total d'aquests va ser mesurat utilitzant l'índex I_{280} com a mètode de referència (§4.4.1). Aquest índex proporciona un indicador de manera ràpida i simple del contingut total de polifenols en el vi o derivats. Fruit d'aquesta quantificació, es va poder confirmar que aquesta no era la font de discriminació entre els diferents mètodes d'envelliment, ja que no existeix cap relació entre l'índex total i el mètode utilitzat; i encara més, amb quasi el mateix contingut promig per totes les classes (Figura 4.17).

No obstant, la capacitat de la ET per predir aquest índex va ser també avaluada; una proposta ja provada exitosament en el cas dels vins (§4.4.1). Amb aquesta finalitat, es va construir i optimitzar una xarxa neuronal utilitzant com a *inputs* els coeficients de Fourier extrets de les respostes voltamperomètriques i una única sortida, l'índex I_{280} .

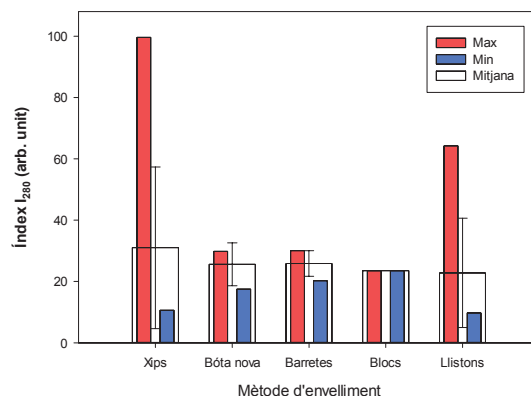


Figura 4.17. Contingut en compostos fenòlics de cada classe mesurat segons el mètode I_{280} . El contingut promig i la seva desviació estàndard es mostren per cada classe, així com el seu màxim i mínim.

Després de l'optimització de la topologia d'aquesta, la ANN estava formada per 96 neurones en la capa d'entrada (corresponents als 16 coefs. extrets de cada voltamperograma x 6 sensors), 6 neurones i la funció de transferència *tansig* en la capa oculta, i una única neurona i la funció de transferència *purelin* en la capa de sortida.

L'eficàcia del model generat va ser avaluada fent ús del mètode *repeated random sub-sampling validation* (§3.4.2.3.1) per tal d'evitar la dependència dels valors predits de la subdivisió específica de les dades; utilitzant-ne el 75% (25 mostres seleccionades aleatòriament) per a l'ajust del model i validant aquest amb el 25% romanent (8 mostres). D'aquesta manera, la divisió de dades entre els dos subconjunts (*train/test*) va ser repetida 25 vegades de forma aleatòria, permetent així obtenir dades no esbiaixades, independents de la subdivisió de les dades i calcular a més les incerteses per cada punt.

A continuació, i agrupant les mostres en funció de si havien intervingut en el procés d'entrenament o de validació, es van construir els gràfics de comparació de resposta obtinguda *vs.* esperada per tal de comprovar l'eficàcia del model ANN (Figura 4.18). Tal com es pot observar, la línia de regressió obtinguda és quasi indistingible de la teòrica; amb els paràmetres de la regressió també propers als valors teòrics de pendent, ordenada a l'origen i coeficient de correlació (1, 0 i 1, respectivament). Per tant, concloent que no hi ha diferències significatives entre els valors predits mitjançant la ANN i els obtinguts amb el mètode de referència, demostrant així la capacitat de la ET per assolir-ne la seva quantificació; permetent al seu torn, la identificació de l'ús de mètodes d'envelliment i la quantificació de l'índex de polifenols a partir d'una única mesura de la mostra.

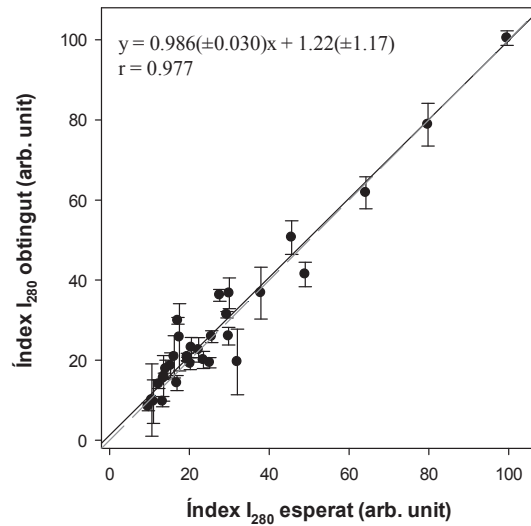


Figura 4.18. Capacitat de modelatge del model FFT-ANN per la predicció de l'índex I₂₈₀. Gràfics de comparació dels valors obtinguts vs. els esperats pel subconjunt de test, on la línia discontinua correspon a la diagonal teòrica. Els resultats proporcionats corresponen a la mitjana dels valors obtinguts per cadascuna de les mostres després de la repetició del càlcul 25 vegades, utilitzant una divisió aleatòria de les mostres pels subconjunts d'entrenament i test cada vegada. Interval de confiança calculats amb un nivell de confiança del 95%.

4.2.6. Identificació de l'anyada

Finalment, donada la importància i influència de l'anyada en la qualitat del producte obtingut en el cas del vi, i tenint en compte que els brandis són obtinguts a partir de la seva destil·lació, l'efecte de l'anyada també va ser estudiat. De fet, es va examinar l'existència de qualsevol empremta electroquímica relacionada amb l'anyada detectable mitjançant la ET. Per tant, fent ús de les mostres del conjunt A, un nou model LDA va ser construït tenint en compte aquest paràmetre; distribuït, en aquest cas, les mostres en cinc grups corresponents a l'any de collita, a saber, des del 2008 al 2012.

Tal com mostra la Figura 4.19, on les mostres apareixen clarament agrupades en funció de l'anyada, és possible la discriminació d'aquest factor. Cal remarcar la distància de les mostres corresponents als brandis del 2012, fet que pot ser explicat donat que aquests són els més joves. Una altra vegada, l'eficàcia del model es va avaluar emprant el mètode de validació creuada *leave-one-out*, els resultats del qual es resumeixen a la Taula 4.9. Com era d'esperar a partir del gràfic obtingut, gairebé totes les mostres van ser classificades correctament, amb un grau d'incert del 95.1%, mentre que els valors de sensibilitat i d'especificitat assolits varen ser del 78.8% i el 98.7%, respectivament. Una vegada més, cal tenir en compte que només hi va haver dues mostres mal classificades, una d'elles corresponent a una classe on només hi havia una mostra disponible per a l'estudi, fet que en va impedir la seva correcta identificació.

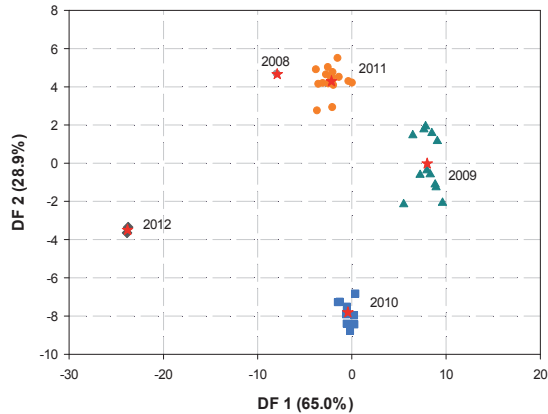


Figura 4.19. Representació dels *scores* de les dos primeres funcions discriminants obtingudes després de l'anàlisi LDA d'acord amb l'anyada: (●) 2008, (▲) 2009, (■) 2010, (○) 2011 i (◆) 2012. Addicionalment, també s'ha representat el centre de cada classe (★).

Taula 4.9. Matriu de confusió construïda d'acord amb l'anyada pronosticada mitjançant el model FFT-LDA, utilitzant el mètode de validació creuada *leave-one-out*.

| | | <i>Predit</i> | | | | |
|----------------|-------|---------------|------|------|------|------|
| | | 2008 | 2009 | 2010 | 2011 | 2012 |
| <i>Esperat</i> | 2008* | 0 | 0 | 0 | 1 | 0 |
| | 2009 | 0 | 12 | 0 | 0 | 0 |
| | 2010 | 0 | 0 | 9 | 0 | 0 |
| | 2011 | 1 | 0 | 0 | 16 | 0 |
| | 2012 | 0 | 0 | 0 | 0 | 2 |

*Només es disposava d'1 mostra

Per tant, la ET ha demostrat ser una eina analíticament útil capaç de proporcionar informació tant de l'ús de mètodes alternatius d'envelliment, la detecció de l'anyada o per categoritzar els destil·lats de vi d'acord amb la puntuació atorgada per un panel de tast; sobrepasant així els problemes trobats amb les tècniques d'anàlisi clàssiques, l'aplicació de les quals es troba dificultada per l'absència de coneixement sobre quins són els compostos responsables d'aquestes sensacions. A més, ha estat capaç de proporcionar la detecció semi-quantitativa de certs compostos utilitzats com a marcadors de defectes en la producció de brandis.

4.3. Anàlisi de cerveses

El tercer sector d'estudi va ser el món cerveser. La cervesa és la beguda alcohòlica més consumida i probablement antiga del món; essent la tercera beguda més popular després de l'aigua i el té [19]. Aquesta és produïda a partir de la fermentació de midons, principalment derivats de grans de cereals; habitualment malta, tot i que el blat, el blat de moro o l'arròs entre altres també poden ser utilitzats. La majoria de les cerveses són aromatitzades amb llúpul, el qual aporta amargor i actua com a conservant natural, encara que altres additius com fruites o herbes poden ser inclosos com a complement aromàtic. Un dels paràmetres clau en la seva producció és l'aigua, la qual és l'ingredient principal, i tot i ser un compost insípid, la composició química d'aquesta té una gran

influència en el tast final; fins al punt, que és definida per molts cervesers com “l'ingredient més important de la cervesa” [20].

En aquest sentit, existeix un esforç per diferenciar i classificar les cerveses en funció de diversos factors tals com el color, el sabor, l'aroma, el cos, els ingredients, el mètode de producció, el tipus de fermentació, l'origen, etc. Juntament amb l'aigua, un altre dels paràmetres que major influència tindrà en les propietats finals de la cervesa és el llevat; fins al punt que cada marca té la seva pròpia soca d'aquest, i esdevé el seu bé més apreciat [21]. Bàsicament, existeixen dos tipus de llevats diferents, a partir dels quals la cervesa és classificada en un dels dos grans estils: Ales (llevats d'alta fermentació, que actuen a temperatures d'entre 12-24°C i puguen a la superfície durant la fermentació) i Lagers (llevats de baixa fermentació, que actuen a temperatures d'entre 7-13°C i tendeixen a depositar-se al fons del fermentador).

Per tant, una tasca important durant l'elaboració de la cervesa és caracteritzar-ne el seu gust, així com la seva amargor, per tal d'identificar les diferents cerveses. Amb aquesta finalitat, diferents mètodes han estat proposats per dur a terme la caracterització d'aquestes de forma objectiva tals com són l'ús de: l'escala de “graus Lovibond”, la qual mesura el color de la cervesa [15]; l'escala de Plato, la qual mesura el cos de la cervesa [15]; o l'escala EBU o IBU (*European/International Bitternes Units*), que proporciona una mesura de l'amargor de la cervesa, relacionada majoritàriament amb el contingut d'iso- α -àcids [15]. A pesar de tots els esforços realitzats, hi continua havent una manca de mètodes capaços de realitzar aquesta classificació, requerint en general el seu tast per part d'un panel de tast per tal d'assolir-ne la seva correcta classificació.

Per tant, al igual que en els casos anterior, l'aplicació de les ETs en aquest sector representa un nínxol d'aplicació, on aquests dispositius presenten grans atractius capaços de satisfer les demandes del sector proporcionant una eina que en permeti la seva classificació de forma objectiva.

4.3.1. Classificació de cerveses

Juntament amb l'empresa *Art Cervecers company – Cervesa Artesanal*, es va preparar un conjunt de mostres amb l'objectiu d'avaluar la capacitat de la ET per discriminar-ne la seva classe.

Tal com s'ha dit les cerveses són classificades en dos grans tipus, però dins d'aquestes existeixen altres estils que poden ser distingits. D'aquesta manera, les

cerveses tals com les *IPA* (*Indian Pale Ale*) i les *Stout* són varietats de *Ales*, mentre que les *Pilsner* o les cerveses d'estil americà són classificades com a *Lagers*. D'altra banda, i davant el ressorgiment d'un nombre elevat de productors que advoquen a l'obtenció d'aquesta mitjançant mètodes tradicionals, el mètode de producció ha esdevingut també un factor de distinció important, degut al valor afegit que aquest aporta tant des del punt de vista sensorial com econòmic.

Amb aquesta finalitat, es van adquirir un total de 25 cerveses de diferents varietats i productors; seleccionades d'acord amb dos paràmetres diferents: el procés d'elaboració (industrial o artesanal) i la classe (*Lager*, *Stout* o *IPA*). Així, el conjunt estava format per 15 cerveses industrials i 10 d'artesans; o tenint en compte el tipus de cervesa, format per 17 *Lagers*, 5 *Stouts* i 3 *IPAs*.

En base als bons resultats assolits anteriorment, i basant-nos en el coneixement adquirit en aquests casos, es va decidir fer ús de la llengua electrònica clàssica formada aquest cop pels sensors (§3.1, Taula 3.1): GEC, CoPc, PPy, Cu, Pt i GOX; incorporant així el biosensor de glucosa, motivat per la detecció de les cerveses artesans. Per tant, podent considerar aquesta també com una llengua (bio)electrònica.

Tot i que a priori pugui semblar una mica estrany, doncs no associem la cervesa amb el sabor dolç o quelcom que pugui contenir sucre; en el cas de les cerveses més artesans, aquestes són envasades amb addicions de sucre i de llevat fresc (o simplement aprofitant el llevat romanent després del filtrat), per tal de provocar una segona fermentació en la botella, generant així alcohol i CO₂, responsable de l'efervescència de les cerveses. Per tant, la detecció de sucre que no hagi fermentat pot ser un indicador d'una producció més artesanal de la cervesa.

Per tal de simplificar el procediment de mesura, no es va realitzar cap pretractament de les mostres, llevat de la seva mera dilució en aigua destil·lada. D'aquesta manera, cada mostra de cervesa es va obrir just abans d'iniciar la mesura, per així evitar la pèrdua de qualsevol compost volàtil o l'oxidació de components de la cervesa. Procedint, a la seva dilució mitjançant la barreja amb aigua destil·lada en la proporció 15:35, respectivament [22]; amb l'objectiu de minimitzar l'aparició de bombolles a la superfície de l'elèctrode durant la mesura electroquímica, i al mateix temps, per reduir qualsevol possible efecte matriu. Per finalment realitzar la mesura voltamperomètrica immediatament després de la dilució de la mostra.

La mesura de les mostres, es va dur a terme seguint el procediment descrit a §3.3, aplicant un escombrat de potencial entre -1.0V i +1.2V vs. Ag/AgCl amb un salt de

potencial de 9mV i una velocitat d'escombrat de $100\text{mV}\cdot\text{s}^{-1}$; fent ús d'una etapa de neteja electroquímica entre mostra i mostra, consistent en l'aplicació d'un potencial de +1.4V durant 45 segons en un vas que contenia 50 mL d'aigua destil·lada. Degut a la incorporació del biosensor GOX, i per tal de preservar millor l'activitat enzimàtica d'aquest, es va decidir reduir lleugerament la finestra de potencials escombrats.

La Figura 4.20 mostra la resposta mitjana obtinguda per les diferents classes de cerveses mitjançant la ET. Tal com es pot observar, les diferències obtingudes són menors en aquest cas, fet que pot ser en part degut a la dilució de la mostra. No obstant, si calculem la diferència entre la resposta obtinguda per les diferents classes, és possible observar amb major facilitat les diferències entre aquestes al llarg de tot l'escombrat de potencial. Per exemple, en el cas del sensor Pt podem veure com aquest mostra una resposta diferenciada en la zona anòdica propera a -0.25V entre les mostres industrials i artesanals; o com el sensor PPy presenta una resposta clarament diferenciada per la classe "lager artesanals" a valors de potencial propers a +1.0V.

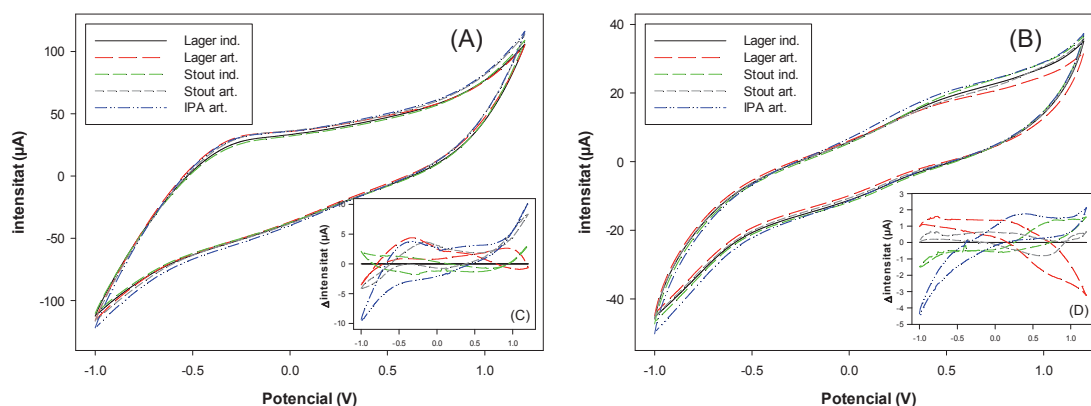


Figura 4.20. Exemple dels diferents voltamperogrames obtinguts per les diferents classes de mostres amb dos dels sensors usats: (A) Pt i (B) PPy; les respostes mitjanes de les diferents classes de cervesa es mostren en funció del tipus de cervesa (Lager, Stout o IPA) i el procés de fabricació (industrial o artesanal). (C, D) Diferència de les respostes mitjanes de cada classe respecte la resposta de la classe "Lager industrial".

A pesar de les petites diferències observades, cal tenir en compte l'elevada capacitat de les diferents eines quimiomètriques a l'hora de buscar i modelar aquestes. És a dir, fins i tot en casos on l'ull entrenat no és capaç de detectar un patró de resposta entre les diferents mostres i sensors, els models són capaços de fer-ho. A més, el pretractament i compressió del senyal, ajuda a eliminar les possibles fonts de soroll i detectar amb major facilitat aquestes diferències. Aquests són doncs, mètodes poderosos capaços d'extreure informació dels segments de senyal en que no podria ser assolit sense l'ús

d'algorismes avançats tals com el LDA o el PLS-DA.

Abans de procedir al modelatge, degut a l'elevada dimensionalitat de les dades, la qual en dificulta el seu tractament, i per tal d'extreure tota la informació obtinguda de cada voltamperograma, és va dur a terme una etapa de preprocessament per a la reducció del volum d'aquestes. En aquest cas, és va fer ús de la FFT, preservant els primers 32 coeficients, el qual va permetre una compressió de la informació original del 93.5% sense cap pèrdua d'informació rellevant.

Per tal de confirmar les tendències inicials observades en les respostes voltamperomètriques i per corroborar el seu contingut analític, els corresponents voltamperogrames comprimits van ser processats mitjançant l'anàlisi PCA; construint així un model de reconeixement preliminar que permeti visualitzar les similituds entre les diferents mostres d'acord amb les mesures realitzades. El gràfic obtingut es mostra a la Figura 4.21, on amb els tres primeres PCs el valor de variància acumulada és de ca. 94.1%. A pesar d'aquest gran valor, no és va poder assolir una discriminació suficientment clara entre les diferents cerveses, tot i que si que és possible observar l'aparició d'alguns patrons inicials. Per exemple, les cerveses de producció industrial semblen agrupar-se a la part superior dreta del gràfic, mentre que les de producció artesanal apareixen agrupades al mig; o el cas de les cerveses IPA que apareixen en una regió propera.

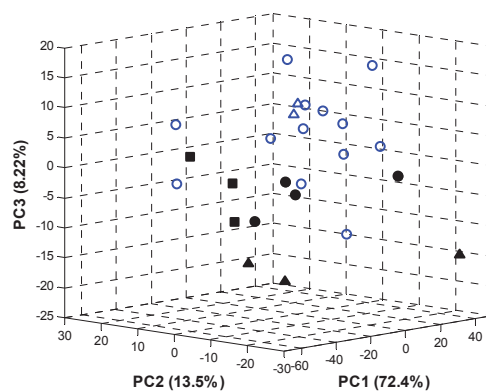


Figura 4.21. Representació dels *scores* dels tres primers components principals obtinguts després de l'anàlisi PCA. Un total de 25 mostres van ser analitzades corresponents a les diferents classes: (●) Lager, (▲) Stout i (■) IPA, i procés de fabricació: (blau, buit) industrial o (negre, ple) artesanal.

Per tant, per tal de confirmar aquestes tendències, el següent pas és l'ús d'algorismes de reconeixement de patrons supervisats tals com el PLS-DA o el LDA. La diferència respecte el PCA, és que aquets construeixen un model que maximitzi la discriminació de les mostres d'acord amb les classes que volem separar; mentre el PCA

només respon a una visualització de les semblances/diferències de les mostres en les direccions de màxima variància, sense tenir en compte la classificació a priori d'aquestes. A efectes pràctics i de forma breu, en el cas dels algorismes supervisats haurem de construir un model per cada un dels criteris de distinció que vulguem avaluar.

En aquest cas, i per tal d'assegurar la validesa de la resposta del model utilitzat, la classificació de les mostres es va dur a terme utilitzant dos models diferents: PLS-DA i LDA; i dos mètodes de validació diferents: validació creuada *leave-one-out* (LOO) i *split sample resampling* (SSR; utilitzant 6 mostres diferents per la validació del model cada vegada i repetint aquest procés 15 vegades), permetent el darrer a la vegada, el càlcul de les incerteses associades als estimadors de la bondat del model.

4.3.1.1. Tipus de cervesa

El primer objectiu va ser doncs la classificació de les cerveses d'acord amb el seu estil, a saber: Lager, Stout i IPA. Utilitzant els coeficients de Fourier extrets dels diferents voltamperogrames com a entrades del model i les tres classes com a valor de sortida, es va procedir, primer de tot, a la construcció del model LDA fent ús del mètode de validació creuada *leave-one-out*. La Figura 4.22 mostra el resultat d'aquest, on els patrons en la figura evidencien que les mostres s'agrupen en funció del tipus de cervesa, podent observar una clara discriminació entre les tres varietats de cervesa.

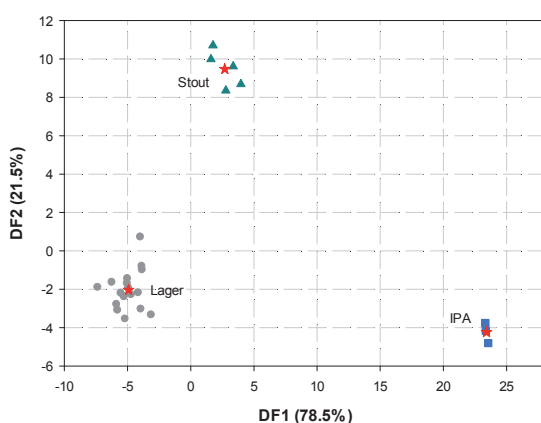


Figura 4.22. Representació dels *scores* de les dos primeres funcions discriminants obtingudes després de l'anàlisi LDA d'acord amb les classes: (●) Lager, (▲) Stout i (■) IPA. Addicionalment, també s'ha representat el centroid de cada classe (★).

Taula 4.10. Resum de les prestacions dels models construïts per a la classificació de les diferents cerveses d'acord amb la seva classe.

| | <i>G. Enc.</i> | <i>Sensib.</i> | <i>Espec.</i> |
|-----------------------|----------------|----------------|---------------|
| <i>PLS-DA&LOO</i> | 100% | 100% | 100% |
| <i>LDA&LOO</i> | 100% | 100% | 100% |
| <i>LDA&RRS*</i> | 100±0% | 100±0% | 100±0% |

*Les incerteses expressades corresponen a la desviació estàndard calculada a partir de la repetició del modelatge amb diferent mostreig

Adicionalment, el modelatge de les respostes mitjançant el mètode LDA va ser repetit amb un mètode de validació més exigent (*repeated random sub-sampling validation, RRS*), utilitzant 6 mostres per la validació del model cada vegada i repetint el procés de construcció del model 15 vegades seleccionant de forma aleatòria les mostres a utilitzar en el conjunt de validació. Encara més, es va utilitzar el mètode PLS-DA com a contrast per tal de verificar que el conjunt de dades extretes amb la ET eren suficientment riques per assolir la tasca desitjada. Així, la Taula 4.10 resumeix els valors dels paràmetres utilitzats per avaluar la capacitat de cada model després de construir les respectives matrius de confusió; viz. grau d'encert, sensibilitat i especificitat. Tal com es podia esperar a partir del gràfic LDA, en tots tres casos es va aconseguir una correcta classificació de les mostres, demostrant un altre cop la capacitat de la ET com a eina biomimètica.

4.3.1.2. Procés d'elaboració

Com s'ha comentat, més enllà de la discriminació dels tres tipus de cervesa, l'altre objectiu era avaluar la capacitat de la ET per discriminar les cerveses segons el procés d'elaboració: industrial o artesanal. Igual que el cas anterior, un nou model LDA va ser construït i avaluat utilitzant el mètode de validació creuada *leave-one-out*. Donat que en aquests cas només dues classes van ser considerades, el model LDA resultant estarà format només per una funció discriminant (§3.4.2.1.2). Per tant, per tal de visualitzar millor els resultats, els *scores* obtinguts van ser representats front el número de mostra.

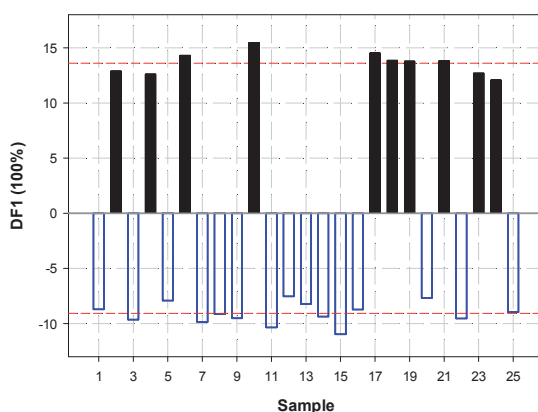


Figura 4.23. Representació dels *scores* de la funció discriminant obtinguda després de l'anàlisi LDA d'acord amb les classes: (□) industrial i (■) artesanal, vs. el número de mostra. Addicionalment, també s'ha representat el centre de cada classe (- - -).

Taula 4.11. Resum de les prestacions dels models construïts per a la classificació de les diferents cerveses d'acord amb el procés d'elaboració

| | <i>G. Enc.</i> | <i>Sensib.</i> | <i>Espec.</i> |
|-----------------------|----------------|----------------|---------------|
| <i>PLS-DA&LOO</i> | 100% | 100% | 100% |
| <i>LDA&LOO</i> | 100% | 100% | 100% |
| <i>LDA&RRS*</i> | 100±0% | 100±0% | 100±0% |

*Les incerteses expressades corresponen a la desviació estàndard calculada a partir de la repetició del modelatge amb diferent mostreig

Tal com es pot observar a la Figura 4.23, una clara discriminació és obtinguda entre les dos classes considerades. A més a més, tal com es va fer en el cas anterior, les respectives matrius de confusió es va construir tant pel model LDA com PLS-DA fent ús del mètode de validació creuada *leave-one-out* i *repeated random sub-sampling validation*, calculant a continuació els paràmetres indicadors de la bondat de l'ajust del model, els resultats del qual es mostren a la Taula 4.11. Per tant, demostrant la capacitat de la ET per tal de discriminar també les mostres de cervesa segons el procés d'elaboració.

4.3.1.3. Discriminació simultània d'ambdós efectes

Finalment, donada la bona discriminació assolida amb els models anteriors, es va procedir a la construcció d'un nou model tenint en compte ambdós paràmetres (tipus de cervesa i mètode d'elaboració). Així, en aquest cas les mostres es van distribuir en els següents cinc grups: Lager industrial, Lager artesanal, Stout industrial, Stout artesanal i IPA artesanal.

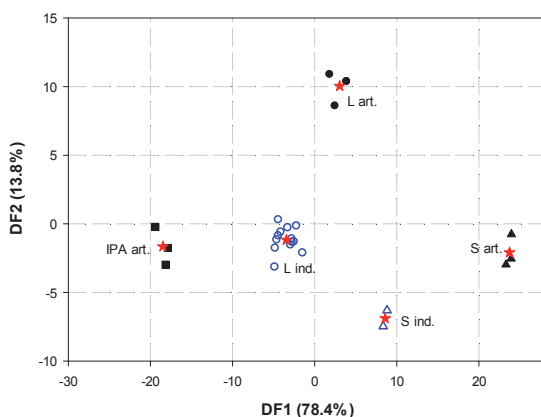


Figura 4.24. Representació dels *scores* de les dos primeres funcions discriminants obtingudes després de l'anàlisi LDA d'acord amb les classes: (○) Lager industrial, (●) Lager artesanal, (△) Stout industrial, (▲) Stout artesanal i (■) IPA artesanal. Addicionalment, també s'ha representat el centroide de cada classe (★).

Taula 4.12. Resum de les prestacions dels models construïts per a la classificació de les diferents cerveses d'acord amb la seva classe i el procés d'elaboració.

| | <i>G. Enc.</i> | <i>Sensib.</i> | <i>Espec.</i> |
|----------------------|----------------|----------------|---------------|
| <i>PLS-DA&LO</i> | 88.0% | 81.7% | 95.8% |
| <i>O</i> | | | |
| <i>LDA&LOO</i> | 100% | 100% | 100% |
| <i>LDA&RRS*</i> | 93.3±8.5% | 95.0±7.3% | 98.5±1.9% |

*Les incerteses expressades corresponen a la desviació estàndard calculada a partir de la repetició del modelatge amb diferent mostreig

Tal com era d'esperar a partir dels resultats anteriors, les mostres van ser agrupades correctament com es pot veure a la Figura 4.24. A més, aquests resultats són consistents amb els anteriors donats que els grups proposats I i II, que corresponen a les cerveses Lager industrials i artesanals respectivament, estan molt a prop un de l'altre, i separats

principalment per la segona funció discriminant; obtenint un comportament similar en el cas de les cerveses Stout (grups III i IV). Per tant, a partir de la unió d'aquests *clusters* (I amb II i III amb IV) i considerant cada un d'aquests com una sola classe enlloc de dos, el model LDA quedaria simplificat a la discriminació entre els diferents tipus de cervesa (Figura 4.22).

Un cop més, l'eficàcia de la ET va ser avaluada a partir de la construcció dels tres models diferents proposats, els resultats dels quals es resumeixen a la Taula 4.12. Com era d'esperar a partir del gràfic LDA, totes les mostres van ser classificades correctament, en el cas del mètode de validació *leave-one-out*; no obstant, en aquest cas, sí que es va observar una lleugera degradació en els resultats en els altres dos models. Aquest fet podria ser explicat, donada la major complexitat del cas i la menor proporció de mostres disponibles per a cada classe usades pel model per extreure l'empremta característica d'aquestes. A més, aquest fet confirmaria el fet que a vegades l'ús del mètode de validació creuada *leave-one-out* pot conduir a resultats lleugerament sobreoptimistes [23].

No obstant això, aquest fet no representa un gran inconvenient, ja que la correcta classificació de les mostres es podria aconseguir per separat a partir dels models individuals; simplement introduint els coeficients obtinguts com a entrada en dos models diferents sense necessitat de dur a terme cap mesura extra de la mostra. Per tant, podent confirmar la capacitat de la ET per distingir les diferents cerveses en funció del seu tipus o el mètode d'elaboració, proporcionant una eina objectiva d'anàlisi, la qual proposa una solució analítica a una problemàtica existent dins d'aquest sector. A més, i de forma anàloga, la ET podria ser entrenada no sols per distingir entre aquestes classes, sinó per fer-ho també entre altres tipus a partir d'un conjunt de mostres adequat.

4.3.2. Quantificació de compostos fenòlics en cervesa

La cervesa conté una àmplia varietat de compostos fenòlics de diferents classes, els quals actuen com a antioxidants naturals (§4.4.1 i §4.4.2). No obstant, el contingut individual per a cada un dels compostos fenòlics és més aviat baix; tot i això, aquests tenen un paper important en el sabor de la cervesa (l'amargor, l'astringència i la duresa), el color i l'estabilitat. Per tant, el seu control és també un paràmetre de clar interès en el sector; la determinació dels quals mitjançant mètodes clàssics es troba descrita a §4.4.1 i §4.4.2.

Donat que la quantificació de compostos fenòlics ja s'havia estudiat en el cas del vi i que un dels objectius de la present tesi era el desenvolupament d'una llengua (bio)electrònica que permetés assolir-ne la seva determinació, es va decidir avaluar l'aplicabilitat del mateix sistema desenvolupat en el cas dels vins (§4.4.2.3), però adaptat als compostos fenòlics majoritaris presents en la cervesa. Així, es va caracteritzar la resposta de la BioET front dels àcids ferúlic, gál·lic i sinàpic (Figura 4.25), tant de forma individual com per la seva multideterminació.

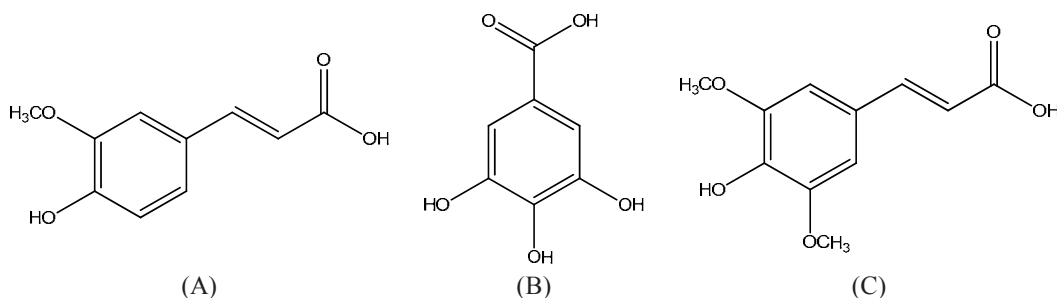


Figura 4.25. Estructura química dels compostos fenòlics estudiats: (A) àcid ferúlic, (B) àcid gál·lic i (C) àcid sinàpic

Per tant, es va decidir fer ús de la BioET formada pels (bio)sensors: (§3.1, Taula 3.1): GEC, Tyr, Lac i Cu. Per la mesura de les mostres, es va seguir el procediment descrit a §3.3, aplicant un escombrat de potencial entre $-0.4V$ i $+0.8V$ vs. Ag/AgCl amb un salt de potencial de $9mV$ i una velocitat d'escombrat de $100 mV \cdot s^{-1}$; fent ús d'una etapa de neteja electroquímica entre mostra i mostra, consistent en l'aplicació d'un potencial de $+1.0V$ durant 40 segons en un vas que contenia 10 mL d'aigua destil·lada.

En primer lloc, abans de la seva aplicació a la resolució de mescles de compostos fenòlics, es va procedir a la caracterització de la resposta de cada un dels (bio)sensors que formen la BioET front a solucions estàndard de tres dels principals compostos fenòlics trobats a la cervesa (àcids ferúlic, gál·lic i sinàpic) per tal de confirmar que els diferents sensors proporcionaven respostes diferenciades entre ells i per cadascun dels analits. Per tant, extraient la suficient informació de la mostra per a ser utilitzada com a punt de partida en la construcció del model multivariable.

Com a exemple, les respostes voltamperomètriques del sensor Cu es mostren en la Figura 4.26A. En general, s'observa com dos processos tenen lloc per tots els sensors corresponents a l'oxidació del fenol corresponent a la forma quinona, i la reducció de la quinona a la forma fenòlica. A més, és possible observar l'aparició d'alguns pics que poden atribuir-se a l'oxidació dels grups $-OCH_3$ dels compostos fenòlics. Així, tal com

es podria esperar, s'obtenen dos pics en el cas de l'àcid sinàpic, un per l'àcid ferúlic i no s'observa cap pic en el cas de l'àcid gál·lic (el qual no té cap grup metoxi) (Figura 4.25).

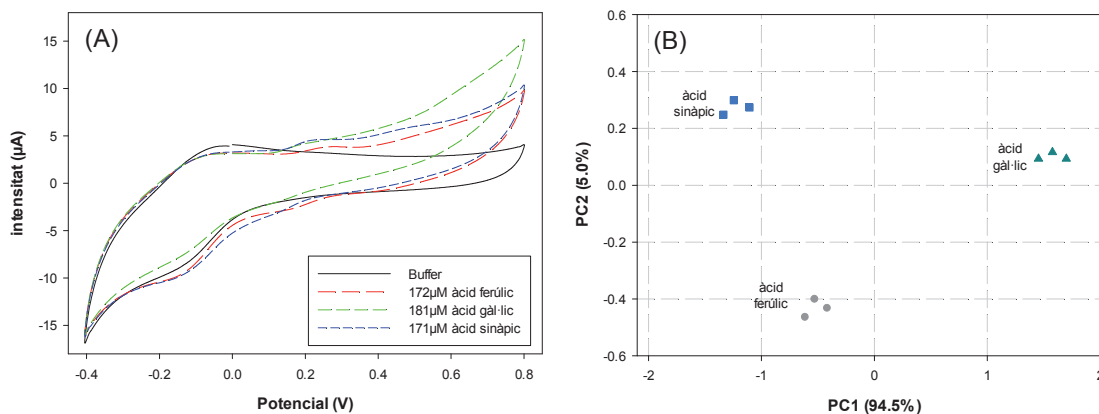


Figura 4.26. (A) Exemple de la resposta voltamperomètrica obtinguda pel sensor Cu front els diferents compostos sota estudi. (B) Representació dels *scores* dels dos primers components principals obtinguts després de l'anàlisi PCA dels diferents estàndards.

Per tal de confirmar aquesta resposta diferenciada, i com a etapa prèvia a la construcció del model quantitatiu, les respostes voltamperomètriques en cru d'aquests estàndards van ser analitzades mitjançant l'anàlisi PCA (Figura 4.26B); assolint una clara discriminació pels tres compostos. Inclús és possible observar com les mostres s'ordenen al llarg del PC1 en funció del número de grups metoxi que presenten; amb l'àcid gál·lic a la dreta distant de la resta, i els àcids ferúlic i sinàpic propers l'un a l'altre (contenint 1 i 2 grups metoxi respectivament) i separats majoritàriament pel PC2.

Un cop confirmada la capacitat de la BioET per discriminar els tres analits, el següent pas va ser l'avaluació més profunda de la resposta de cada un dels sensors en front de cada un dels analits. Així, es va procedir a la mesura de solucions de concentració creixent de cadascun dels tres compostos fenòlics per tal de construir les respectives rectes de calibrat.

La Figura 4.27 mostra com a exemple la resposta del sensor Cu front a diferents solucions estàndard d'àcid sinàpic de concentració creixent. Tal com era d'esperar, la intensitat de corrent associada als fenols incrementa amb la concentració d'aquests. Per tant, per tal de caracteritzar la sensibilitat de cada (bio)sensor per cadascun dels compostos, els corrents dels pics associats als fenols es van representar davant de la concentració de cada espècie. Tot i que la resposta màxima difereix lleugerament per a cada compost, es va escollir el mateix potencial per a cada sensor a l'hora de construir

la corba de calibratge, tal com es faria de forma habitual en amperometria (intensitat mesurada a un únic potencial).

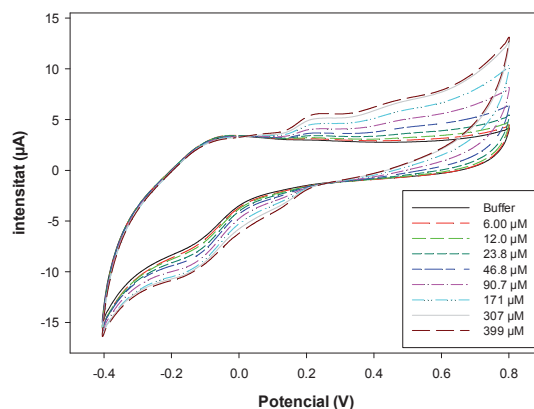


Figura 4.27. Resposta voltamperomètrica del sensor Cu front a diferents solucions estàndard d'àcid sinàpic de concentració creixent.

En aquest sentit, els valors de sensibilitat calculats a partir de les corbes de calibratge per a cadascun dels (bio)sensors de la BioET es resumeixen a la Taula 4.13; confirmant així la resposta creuada pels diferents analits estudiats i mostrant una resposta diferenciada per cadascun dels (bio)sensors de la BioET, una situació molt desitjable com a punt de partida en les llengües electròniques.

Taula 4.13. Caracterització de la resposta de la BioET front els diferents compostos fenòlics estudiats.

| <i>Sensor</i> | | <u>àcid ferúlic</u> | | <u>àcid gàl·lic</u> | | <u>àcid sinàpic</u> | |
|---------------|---|---------------------|------------|---------------------|------------|---------------------|------------|
| | | <i>an</i> | <i>cat</i> | <i>an</i> | <i>cat</i> | <i>an</i> | <i>cat</i> |
| <i>GEC</i> | <i>E (V)</i> | 0.28 | 0.08 | 0.28 | 0.08 | 0.28 | 0.08 |
| | <i>Sensibilitat (µA·M⁻¹)</i> | 1756 | -2104 | 4918 | - | 2428 | -2108 |
| <i>Tyr</i> | <i>E (V)</i> | 0.40 | -0.01 | 0.40 | -0.01 | 0.40 | -0.01 |
| | <i>Sensibilitat (µA·M⁻¹)</i> | 1276 | 1227 | 4633 | - | 1780 | -954 |
| <i>Lac</i> | <i>E (V)</i> | 0.40 | -0.01 | 0.40 | -0.01 | 0.40 | -0.01 |
| | <i>Sensibilitat (µA·M⁻¹)</i> | 1114 | 1104 | 3630 | - | 1692 | -790 |
| <i>Cu</i> | <i>E (V)</i> | 0.31 | 0.08 | 0.31 | 0.08 | 0.31 | 0.08 |
| | <i>Sensibilitat (µA·M⁻¹)</i> | 3381 | -3891 | 8555 | - | 5228 | -4234 |

Per tal d'avaluar i demostrar les capacitats de la BioET de dur a terme la quantificació simultània de mesclades dels diferents compostos fenòlics, es van preparar un conjunt de 42 mostres amb un rang de concentració per a les tres espècies de 0 a 200µM per a cada àcid. El conjunt de mostres va ser dividit en dos subconjunts: el conjunt d'entrenament format per 27 mostres (64% del total) distribuïdes en un disseny cúbic (Figura 3.2,) i utilitzades per construir el model de resposta; més 15 mostres addicionals (36%) distribuïdes aleatòriament al llarg del domini experimental i usades per avaluar la capacitat de predicció del model generat.

Seguint el procediment descrit, el conjunt de mostres va ser analitzat de forma aleatòria, obtenint un registre complet per cadascun dels sensors (Figura 4.28). Tot i el clar solapament observat entre els senyals deguts als diferents compostos de les mescles, encara és possible observar l'aparició dels pics deguts als grups metoxi caracteritzats anteriorment.

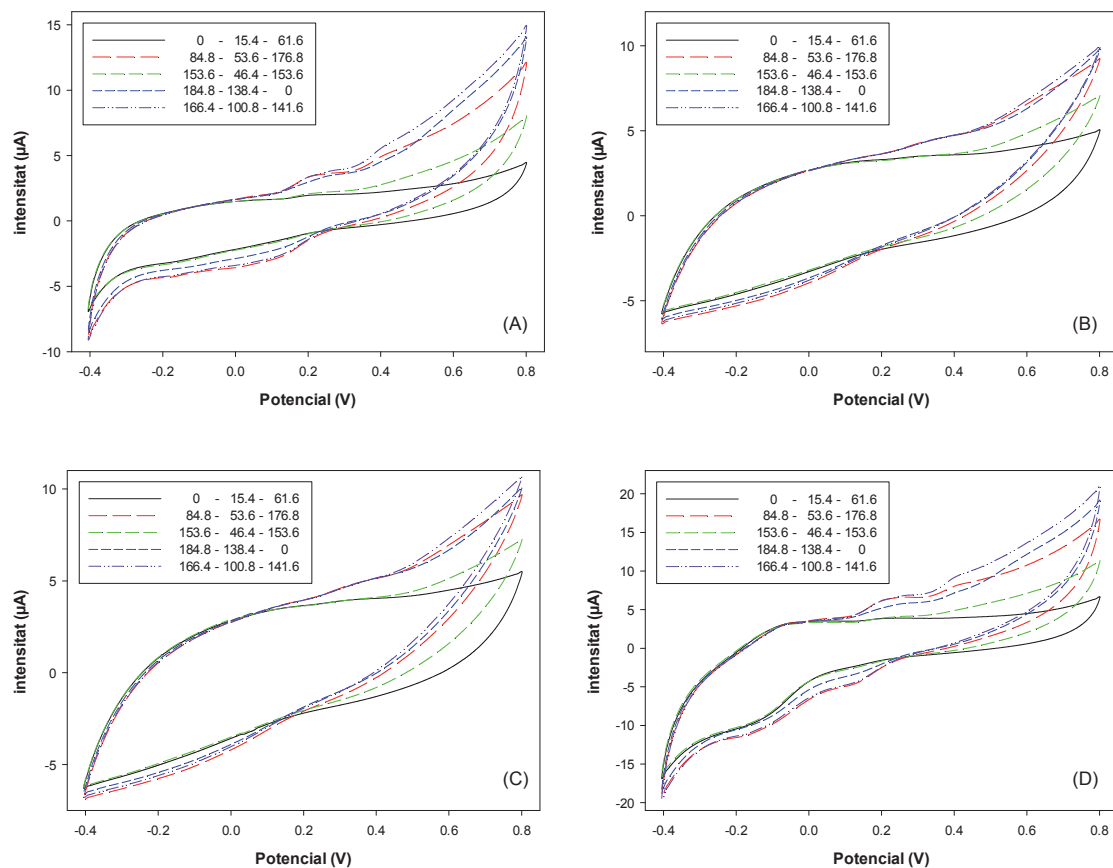


Figura 4.28. Exemple dels diferents voltamperogrames obtinguts per certes mescleres ternàries arbitràries amb els (bio)sensors: (A) GEC, (B) Tyr, (C) Lac i (D) Cu. Els valors de la llegenda corresponen a les concentracions en μM de cada un dels compostos fenòlics: àcid ferúlic, gal·lic i sinàpic, respectivament.

Com ja s'ha discutit bastament, abans de procedir a la construcció del model cal realitzar una etapa de compressió del senyal per tal de reduir-ne l'elevada dimensionalitat i guanyar avantatges en l'etapa de modelatge. En aquest cas, es va proposar l'ús d'una nova metodologia basada en el càlcul de l'àrea sota la corba voltamperomètrica: la *integral de finestra seccionada* (§3.4.1.4). Breument, aquest mètode es basa en dividir el voltamperograma en k seccions i calcular l'àrea de cada una d'aquestes, valor el qual serà després utilitzat com a *input* en el model generat; al incrementar el nombre de seccions, augmenta la representació del senyal, però a expenses del grau de compressió. Per tant, requerint d'una situació de compromís.

Així doncs, el principal inconvenient d'aquest mètode és la dificultat de predir a priori el nombre de seccions k a utilitzar, i per tant aquest paràmetre va ser optimitzat juntament amb la topologia de la xarxa. D'aquesta manera, un estudi exhaustiu i sistemàtic de l'arquitectura de la xarxa va ser dut a terme variant: el nombre de seccions (k) entre 1 i 20; el nombre de neurones en la capa oculta entre 1 i 12; i provant diferents combinacions de les funcions de transferència *tansig*, *logsig*, *purelin* i *satlins* en la capa oculta i la capa de sortida. Avaluant així un total de 2880 models diferents; no obstant, val a dir que aquest elevat nombre de combinacions provades va ser degut a que era la primera vegada que s'utilitzava aquest mètode i per tant, es pretenia analitzar detalladament l'efecte obtingut al variar el nombre de seccions. Extraient així el suficient coneixement del mètode, que permeti relegar a un menor nombre el número de seccions a avaluar en un futur.

Per la tria del millor model ANN, es va calcular el valor de RMSE total i els paràmetres de la regressió extrets dels gràfics de comparació dels valors obtinguts *vs.* els esperats (§3.4.2.3.2.2). Així, el millor model tindrà un baix valor de RMSE, i valors de Δm , Δr i Δb propers a 0; és a dir, que els paràmetres calculats al fer la regressió són propers als valors ideals. La Figura 4.29, il·lustra una part d'aquests resultats; no és d'estranyar que al augmentar el número de seccions, la resposta del model millori doncs li estem aportant més informació gràcies a la millor representació de les dades inicials; observant una millora substancial al principi i menors diferències amb valors de k superiors. D'altra banda, si que és destacable l'elevat grau de compressió que permet aquest mètode donat que amb tan sols 5 coeficients per cada voltamperograma ja és possible obtenir resultats força satisfactoris; tot i que els millors resultats van ser obtinguts amb un valor de k igual a 11.

Cal remarcar doncs, el fet que l'ús d'aquest mètode va permetre la compressió de la informació original en un 95.9% sense pèrdua d'informació significativa (11 coefs. *vs.* 268 intensitats de corrent en el registre voltamperomètric original). Aquest és un valor significatiu tenint en compte la senzillesa del mètode, tant des del punt de vista teòric com a nivell de càlcul.

Després de l'estudi exhaustiu entre les diferents configuracions, el model Int-ANN final tenia 44 neurones (corresponents a 4 sensors x 11 coefs.) a la capa d'entrada, 5 neurones i la funció de transferència *tansig* en la capa oculta, i 3 neurones i la funció de transferència *purelin* a la capa de sortida (una per a la concentració de cada compost fenòlic).

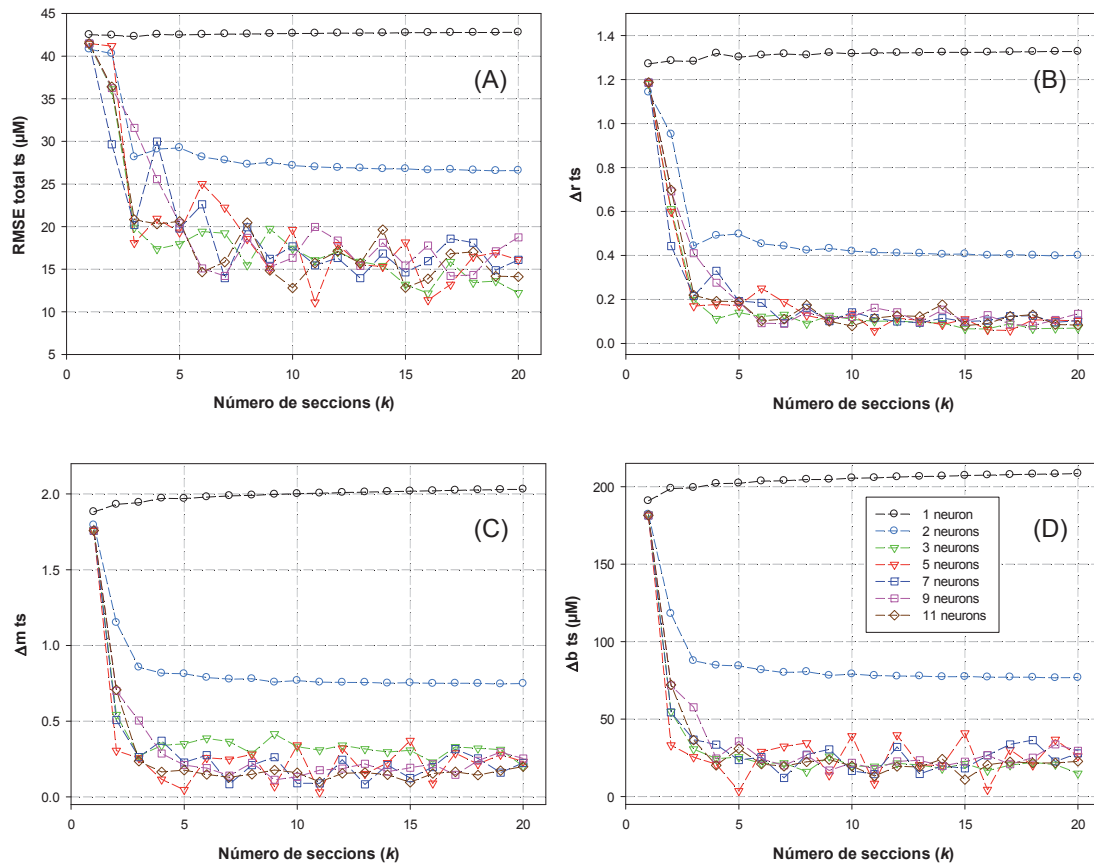


Figura 4.29. Resultats detallats de l'optimització de la Int-ANN. Els valors obtinguts de (A) RMSE, (B) coeficient de correlació, (C) pendent i (D) ordenada a l'origen dels gràfics de comparació de concentració obtinguda vs. esperada són graficats front el número de seccions usades amb el mètode de la *integral de finestra seccionada*. Els resultats mostrats corresponen a diferents arquitectures de la ANN, variant tant el número de neurones en la capa d'entrada (número de coeficients extrets) com en la capa oculta; però en tots els casos utilitzant les funcions de transferència *tansig* i *purelin*, respectivament, per connectar les diferents capes.

La Figura 4.30 mostra els resultats obtinguts amb el model descrit per la concentració de cadascun dels tres analits considerats, tant pel conjunt d'entrenament com pel de test, on la capacitat d'aquest per dur a terme la resolució de les mesclades queda demostrada; obtenint línies de regressió gairebé indistingibles de les teòriques.

Cal recordar que les mostres pertanyents al subconjunt de test no són utilitzades en absolut per a la construcció del model, per tant la seva bondat d'ajust és una mesura objectiva de la capacitat d'aquest. Així, la Taula 4.14 resumeix els paràmetres obtinguts de la regressió, mostrant una bona tendència lineal per tots els casos, però amb millors ajustos en el cas del subconjunt d'entrenament degut a la menor dispersió dels punts; una situació habitual en aquests casos. A pesar d'això, els resultats obtinguts pels dos subconjunts són propers als valors ideals: amb ordenades a l'origen properes a 0, i pendents i coeficients de correlació propers a 1.

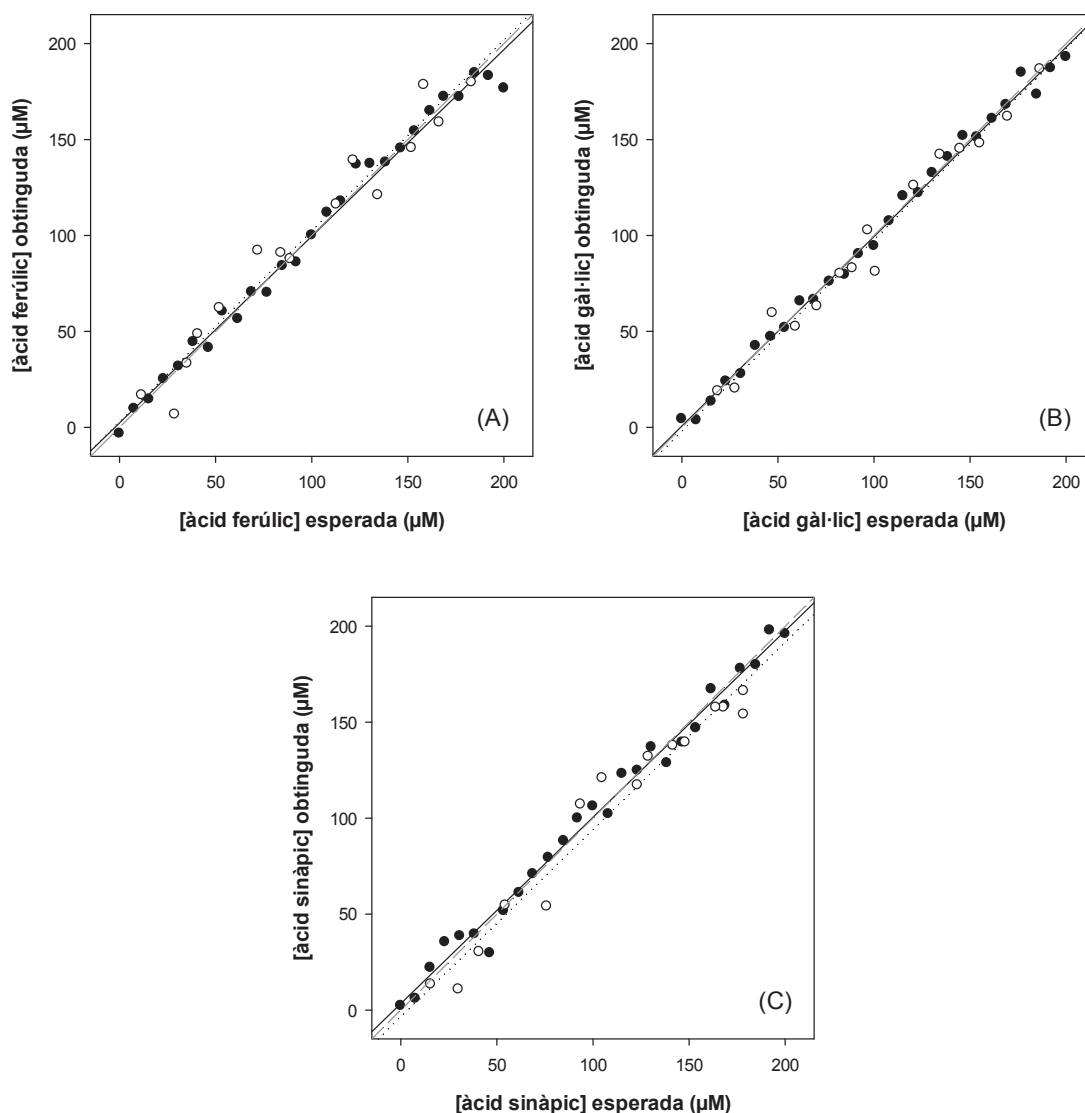


Figura 4.30. Capacitat de modelatge del model Int-ANN. Gràfics de comparació de les concentracions obtingudes vs. les esperades pels àcids (A) ferúlic, (B) gàl·lic i (C) sinàpic, tant pel subconjunt d'entrenament (●, línia sòlida) com pel subconjunt de test (○, línia de punts). La línia discontinua correspon a la diagonal teòrica.

Taula 4.14. Paràmetres de les rectes de regressió ajustades dels gràfics de comparació entre valors obtinguts vs. esperats pel model Int-ANN proposat, tant pel subconjunt d'entrenament com de test i pels tres compostos considerats. Intervals de confiança calculats amb un nivell de confiança del 95%.

| <i>Subconjunt d'entrenament</i> | | | | | |
|---------------------------------|-------------------|----------------|-------------------------|-----------------|------------------------|
| | <i>Correlació</i> | <i>Pendent</i> | <i>Ord. Origen (μM)</i> | <i>RMSE(μM)</i> | <i>RMSE total (μM)</i> |
| <i>àcid ferúlic</i> | 0.994 | 0.973±0.044 | 2.3±5.1 | 8.82 | |
| <i>àcid gàl·lic</i> | 0.998 | 0.988±0.029 | 0.7±3.3 | 5.75 | 8.00 |
| <i>àcid sinàpic</i> | 0.994 | 0.972±0.045 | 3.3±5.2 | 9.02 | |
| <i>Subconjunt de test</i> | | | | | |
| | <i>Correlació</i> | <i>Pendent</i> | <i>Ord. Origen (μM)</i> | <i>RMSE(μM)</i> | <i>RMSE total (μM)</i> |
| <i>àcid ferúlic</i> | 0.977 | 0.995±0.131 | 2.9±14.4 | 11.97 | |
| <i>àcid gàl·lic</i> | 0.988 | 0.998±0.094 | -1.9±10.5 | 8.03 | 11.10 |
| <i>àcid sinàpic</i> | 0.978 | 0.976±0.123 | -3.5±15.1 | 12.71 | |
| <i>Mostres fortificades</i> | | | | | |
| | <i>Correlació</i> | <i>Pendent</i> | <i>Ord. Origen (μM)</i> | <i>RMSE(μM)</i> | <i>RMSE total (μM)</i> |
| <i>àcid ferúlic</i> | 0.986 | 0.986±0.098 | 3.2±11.6 | 9.21 | |
| <i>àcid gàl·lic</i> | 0.981 | 1.001±0.119 | 2.4±12.9 | 10.07 | 10.46 |
| <i>àcid sinàpic</i> | 0.976 | 0.994±0.134 | 4.6±15.3 | 11.92 | |

Un cop confirmada la capacitat de modelatge de la BioET per assolir la correcta quantificació dels tres compostos fenòlics a partir de solucions estàndard, i per tal d'avaluar-ne la seva aplicabilitat en mostres reals, es va procedir a la preparació d'algunes mostres de cervesa fortificades, amb concentracions d'aquests dins del rang experimental, addicionant quantitats variables dels tres compostos.

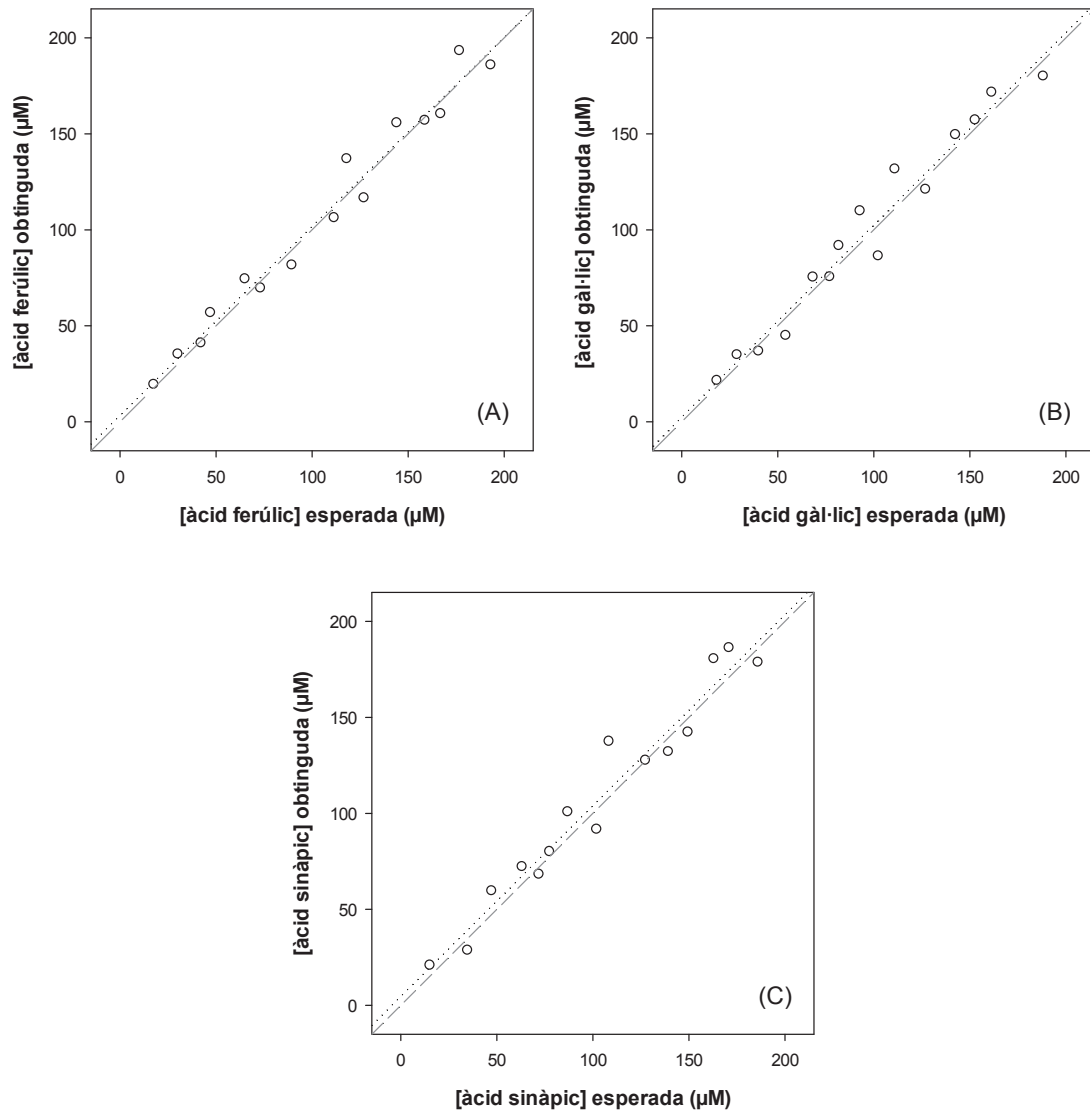


Figura 4.31. Capacitat de modelatge del model Int-ANN. Gràfics de comparació de les concentracions obtingudes vs. les esperades pels àcids (A) ferúlic, (B) gàl·lic i (C) sinàpic, pel subconjunt de test (\circ , línia de punts). La línia discontinua correspon a la diagonal teòrica.

Utilitzant la mateixa arquitectura de la ANN i el mètode de validació creuada *leave-one-out*, degut al reduït nombre de mostres i per assegurar que totes les mostres intervenen en el procés de validació, el sistema va ser reentrenat; avaluant així la resposta d'aquest en mostres de cervesa. Igual que en el cas anterior, els gràfics de

comparació de concentracions obtingudes *vs.* esperades van ser construïts per cadascun dels compostos fenòlics (Figura 4.31). Com es pot observar, també en aquest cas s'obté una bona tendència pels tres compostos fenòlics, amb línies de regressió quasi indistingibles de les teòriques. De nou, els paràmetres de la regressió són resumits a la Taula 4.14; quedant continguts els valors teòrics de la regressió dins dels respectius intervals de confiança en tots els casos, demostrant la bondat de l'ajust.

4.4. Anàlisi de vins

Per últim, donada la llarga tradició vinícola del nostre país, és va decidir també aplicar les llengües electròniques en aquest sector; focalitzant-se en el vi com a tal, i no en els seus derivats com són els casos del cava o brandi, els quals ja s'han presentat anteriorment.

Concretament en aquest cas es va voler abordar el tema dels compostos fenòlics o polifenols [24], la majoria dels quals són poderosos antioxidants amb grans beneficis per la salut; actuant com a segrestadors de radicals lliures (*free radical scavengers*) i inhibidors de l'oxidació de les lipoproteïnes [11], proporcionant així un efecte protector contra les patologies degudes a l'edat tals com les malalties cardiovasculars o mutacions cancerígenes. Atribuïnt així a aquests compostos la majoria de les propietats beneficioses per la salut derivades del vi, i essent el seu control de clar interès.

D'aquesta manera, es va treballar tant en la discriminació i identificació de diferents compostos fenòlics tant a nivell individual (§4.4.2) com en la quantificació de l'índex global d'aquests (§4.4.1); cal però, tenir en compte la diferència en el rang de concentracions en ambdós casos, així com la discrepància entre ambdós.

A saber, en el cas del contingut total de polifenols, expressats com a equivalents d'àcid gàl·lic, el seu valor depenent del tipus de vi serà al voltant de $320 \text{ mg}\cdot\text{L}^{-1}$ en el cas dels vins blancs, $820 \text{ mg}\cdot\text{L}^{-1}$ pels rosats i $2160 \text{ mg}\cdot\text{L}^{-1}$ pels negres [25]. Mentre que els valors individuals de cada un dels compostos fenòlics serà molt més baix que els valors anteriors; e.g. l'àcid cafèic en el rang $5\text{-}13 \text{ mg}\cdot\text{L}^{-1}$ pels vins negres i $1\text{-}4$ pels vins blancs, o la catequina amb un contingut de $120\text{-}390 \text{ mg}\cdot\text{L}^{-1}$ pels vins negres i $16\text{-}46 \text{ mg}\cdot\text{L}^{-1}$ pels vins blancs.

D'altra banda, si realitzem la suma dels diferents compostos fenòlics determinats a partir de mètodes cromatogràfics, i comparem aquest valor amb el contingut total determinat a partir del mètode de Folin-Ciocalteu, el primer és inferior. Aquesta

discrepància pot ser en part explicada per la falta de coneixement dels diferents compostos i per la dificultat de mesurar aquests, per una banda. Així com a una sobreestimació del valor final per part del mètode de Folin per l'altra; ja que aquest tot i ser el mètode de referència, presenta com a inconvenient una manca de selectivitat que pot induir a la resposta d'altres compostos reductors no fenòlics.

Al mateix temps, es va decidir aprofitar aquest estudi per tal de realitzar un altre estudi paral·lel en el que es pretenia comparar i avaluar els beneficis derivats de l'ús de biosensors en la matriu d'elèctrodes que formen la llengua (bio)electrònica. Com ja s'ha explicat (§1.4.3.1), aquesta incorporació respon a la última tendència en el disseny de llengües electròniques; en que els sistemes derivats només es distingeixen dels convencionals en la incorporació d'un o més biosensors en la matriu de sensors que formarà part de la llengua electrònica, normalment compartint el mateix principi de detecció per tal de facilitar la compatibilitat entre ells. Així, es postulat que mitjançant aquesta nova modalitat, és possible millorar la resposta del sistema degut a la combinació de la millor resposta dels biosensors respecte els sensors convencionals (major selectivitat i especificitat) i els beneficis propis de les eines quimiomètriques (§1.3; capacitat de modelar i extreure informació útil a partir de dades complexes, i construcció de models multivariable). Obtenint així una eina que combina les ANNs per tal de solucionar els problemes de interferència o efecte matriu dels biosensors, i els biosensors com l'eina per solucionar la falta de selectivitat pròpia de les llengües electròniques; combinant per tant els avantatges de cada un d'ells per separat, per solucionar els desavantatges l'un de l'altre.

4.4.1. Índexs de polifenols totals

El primer intent que es va dur a terme va consistir en la quantificació de l'índex total de polifenols a partir de la comparació amb dos mètodes de referència diferents: Folin-Ciocalteu (FC) i absorbància a 280nm (I_{280}). Per tant, intentant predir ambdós a partir de les mesures voltamperomètriques tant de la ET com de la BioET (§3.1, Taula 3.1).

El primer mètode consisteix en un assaig colorimètric que mesura la quantitat de fenols necessària per inhibir l'oxidació del reactiu de Folin-Ciocalteu (una barreja de fosfomolibdat i fosfotungstè, que són reduïts als respectius òxids) [26]. Aquest mètode presenta com a inconvenient, el fet que aquest reactiu no mesura únicament la quantitat

de fenols totals, reaccionant amb qualsevol substància reductora present en la mostra. Per tant, mesura la capacitat reductora total d'una mostra, no només el nivell de compostos fenòlics d'aquesta.

El segon índex és una mesura directa de l'absorbància a 280nm [27]. La relació entre l'índex I_{280} i la concentració fenòlica és deguda al fet que tots els compostos fenòlics absorbeixen la llum UV, i encara més, tots ells presenten certa absorbància a la zona de 280nm. Un problema d'aquest mètode és que cada classe de substàncies fenòliques té una absortivitat diferent (coeficient d'extinció, ϵ) a 280nm [28], i per tant els resultats no es poden relacionar amb cap estàndard o compost concret, sinó que són presentats directament en unitats d'absorbància o unitats arbitràries (arb. units). Malgrat que aquest mètode és menys sensible i més inespecífic, el seu ús ha crescut en els últims anys a causa de la seva simplicitat i baix cost; especialment en el sector industrial.

4.4.1.1. *Quantificació mitjançant la Llengua Electrònica*

En aquest cas, a través d'un projecte conjunt entre la UAB, el IMB-CNM i l'INCAVI, es va procedir a l'anàlisi d'un total de 20 mostres de vins diferents per tal d'abordar la quantificació de l'índex global de polifenols mitjançant l'ús d'una llengua electrònica i mesures voltamperomètriques.

Per tal d'enriquir la informació de partida i amb motiu de la col·laboració entre la UAB i el IMB-CNM, es va decidir fer ús d'un microsensor d'or desenvolupat pel CNM (μAu) [29]. Així en aquest cas la ET utilitzada estava formada per 5 sensors de la UAB (§3.1, Taula 3.1) més el microsensor d'Au del CNM: GEC, CoPc, PPy, Cu, Pt i μAu . Les mesures es van dur a terme de la forma descrita a §3.3, amb un escombrat de potencial entre -1.0V i +1.3V vs. Ag/AgCl (-0.5V i +1.6V en el cas del sensor μAu), una velocitat d'escombrat de $100\text{mV}\cdot\text{s}^{-1}$ i un salt de potencial de 9mV; aplicant un potencial de condicionament de +1.5V durant 40s en un vas que contenia 25mL d'aigua destil·lada en l'etapa de neteja electroquímica. A més, les mostres van ser analitzades de forma aleatòria, i sense realitzar cap regeneració física dels elèctrodes.

Tal com s'observa a la Figura 4.32, diferents respostes són obtingudes per cadascun dels diferents sensors, els quals presumiblement contribuiran de diferent manera en el model generat. En general, s'observa com els corrents obtinguts per les diferents mostres augmenten a mesura que augmenta l'índex FC d'aquestes, demostrant així una clara relació entre aquest i la resposta de cada sensor. Al mateix temps, es pot observar

com cada sensor presenta el seu propi perfil de resposta, generant informació molt útil per ser utilitzada com a punt de partida. Per exemple, els sensors modificats amb nanopartícules semblen presentar senyals d'oxidació segurament deguts a l'oxidació catalítica de sacàrids i/o compostos fenòlics. De forma similar, el sensor modificat amb el polímer conductor aporta nova informació amb una major resposta al llarg de tot el registre voltamperomètric. Mentre que el GEC i el μAu proporcionen una resposta redox més genèrica de la mostra.

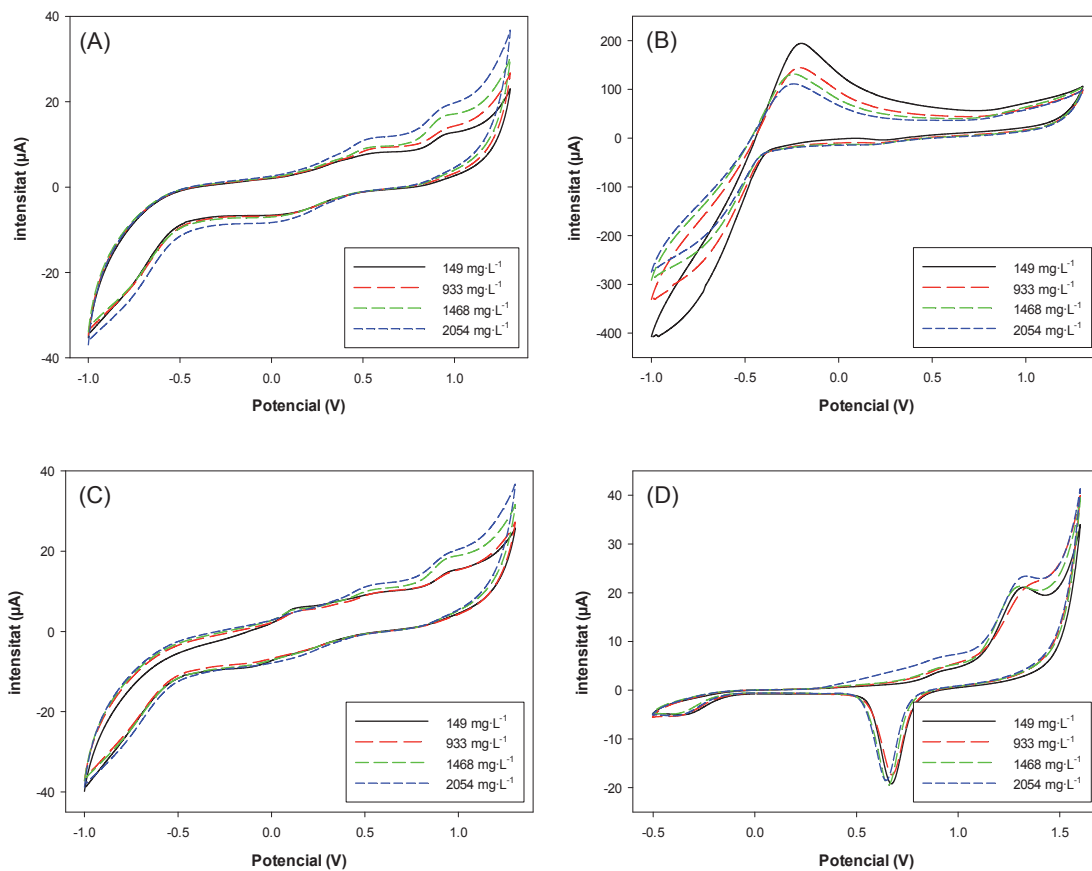


Figura 4.32. Exemple dels diferents voltamperogrames obtinguts per certes mostres arbitràries amb els (bio)sensors: (A) GEC, (B) Pt, (C) Cu i (D) μAu . Els valors expressats corresponen a l'índex de Folin-Ciocalteu expressat com a equivalents d'àcid gal·lic mesurat mitjançant el mètode de referència.

No obstant, i tal com ja s'ha comentat reiteradament, previ a la construcció del model de resposta i per tal de millorar el comportament d'aquest, cal realitzar una etapa de compressió del senyal, la qual va ser assolida en aquest cas mitjançant l'ús de la DWT utilitzant la funció mare *Daubechies* i un nivell de descomposició sis. Assolint en aquest cas una compressió del senyal original del 97.3%, sense pèrdua de informació

significativa. Llavors, els coeficients obtinguts van ser utilitzats com a *inputs* en el model ANN per tal d'assolir la predicció dels dos índexs.

Un cop definits els inputs del model neuronal, el següent pas va ser la optimització de la topologia de la xarxa. En aquest cas, el model escollit estava format per 84 neurones en la capa d'entrada (corresponents als 14 coeficients Wavelet obtinguts per cadascun dels 6 sensors), 10 neurones en la capa oculta i 2 neurones en la capa de sortida (corresponents als índexs FC i I_{280}), emprant respectivament les funcions *logsig* i *purelin* per connectar les diferents capes.

Per a l'avaluació de l'eficàcia del model es va emprar el mètode *repeated random sub-sampling validation* (descriu a §3.4.2.3.1) per tal d'evitar la dependència dels resultats obtinguts de la subdivisió específica de les dades. Així doncs, el procés de divisió de les dades pels dos subconjunts (*train/test*) va ser repetit aleatòriament 25 vegades, emprant cada vegada el 75% de les dades (15 mostres) per la construcció del model (subconjunt d'entrenament) i el 25% romanent (5 mostres) per l'avaluació de la resposta d'aquest (subconjunt de *test*). Finalment, els resultats de les diferents repeticions van ser agrupats depenen de quin dels dos subconjunts havien intervingut en cada cas i es van construir les rectes de comparació de valors obtinguts *vs.* esperats.

En aquest cas, i per tal de donar el mateix pes a tots els punts, independentment de quin sigui el valor de l'índex, es va fer ús de la regressió ponderada. El principal avantatge d'aquest mètode és la seva robustesa, ja que elimina la dependència de la regressió dels valors dels índexs, proporcionant així paràmetres que mitiguin els efectes de possibles *outliers*, igualant la contribució de tots els punts.

Així doncs, la Figura 4.33 mostra els resultats obtinguts mitjançant el model DWT-ANN per la predicció dels índexs de Folin-Ciocalteu i I_{280} , respectivament. En ambdós casos, es representa la mitjana dels valors obtinguts amb el mètode *repeated random sub-sampling validation* i l'estimació dels errors de predicció calculats a partir de la dispersió de les rèpliques (IC calculats amb un nivell de confiança del 95%).

Tal com és habitual en el cas dels mètodes de calibrat multivariable, millors resultats varen ser obtinguts pel subconjunt d'entrenament que pel de test; i.e. millor concordança dels valors predits amb els valors esperats i menor dispersió (IC de cada un dels punts). No obstant, la tendència general és bona en ambdós casos, malgrat l'increment observat en els errors de predicció individuals pel subconjunt de test, especialment per certes mostres més dependents de la subdivisió específica de dades que s'utilitza en el procés d'entrenament. La Taula 4.15 resumeix els resultats de les

regressions ponderades, on destaquen els bons coeficients de correlació obtinguts, així com els baixos errors associats a aquests (IC calculats en el nivell de confiança del 95%). A més, els valors dels diferents paràmetres de la regressió són propers als valors ideals, amb ordenades a l'origen properes a 0 i pendents i coeficients de correlació propers a 1, mostrant una bona tendència i concordança en general entre els valors predits i esperats. Demostrant així, la capacitat de la ET proposada per assolir la determinació quantitativa de l'índex total de polifenols d'acord amb els dos mètodes de referència proposats.

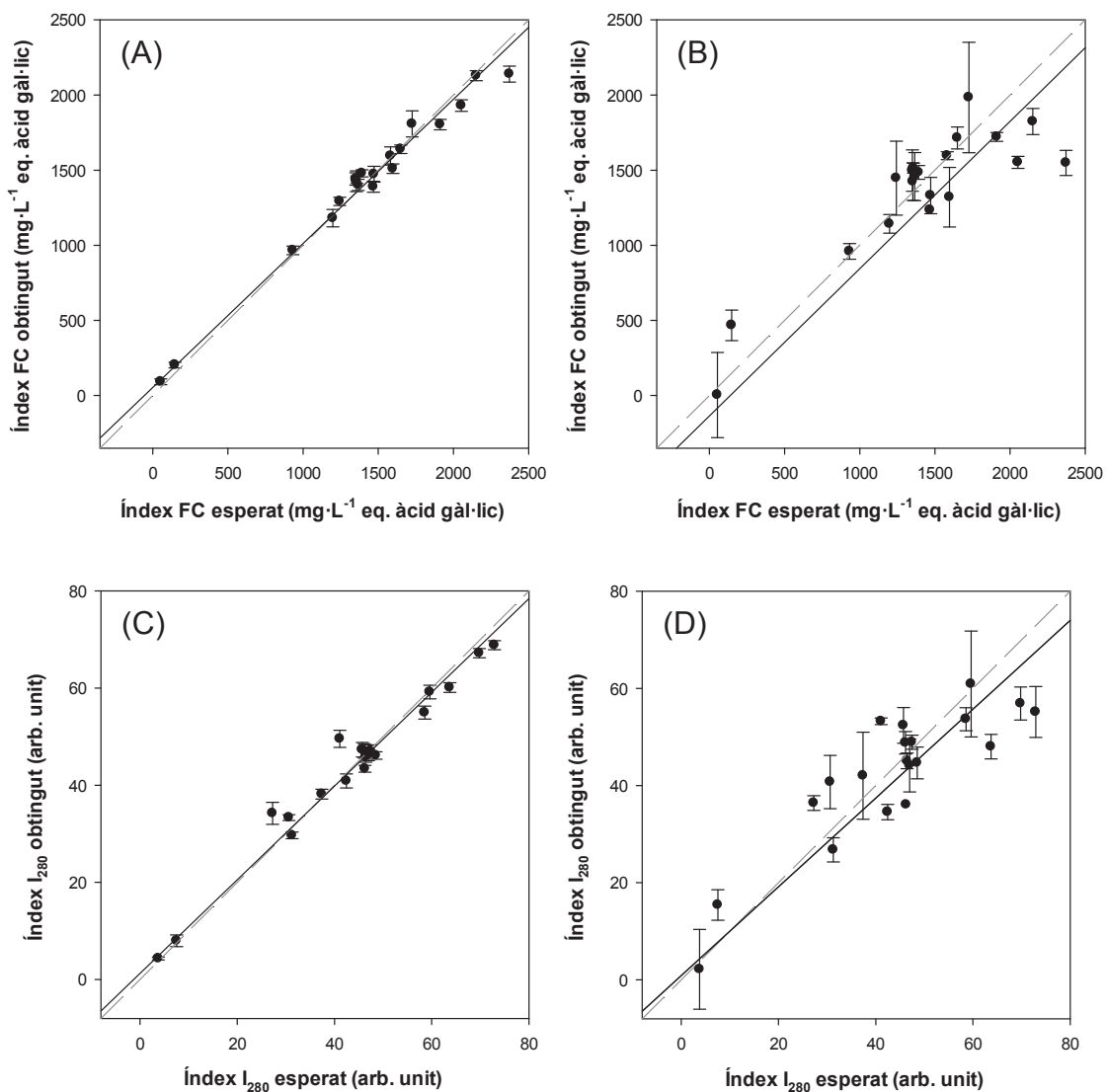


Figura 4.33. Capacitat de modelatge del model DWT-ANN utilitzant com a mètode de validació *el repeated random sub-sampling validation*. Gràfics de comparació de les concentracions obtingudes vs. les esperades pels índexs (A,B) Folin-Ciocalteu i (C,D) I₂₈₀, tant pel subconjunt d'entrenament (A,C - esquerra) com de test (B,D - dreta). La línia discontinua correspon a la diagonal teòrica. Els resultats proporcionats corresponen a la mitjana dels valors obtinguts per cadascuna de les mostres després de la repetició del càlcul 25 vegades, utilitzant una divisió aleatòria de les mostres pels subconjunts d'entrenament i test cada vegada. Interval de confiança calculats amb un nivell de confiança del 95%.

Taula 4.15. Paràmetres de les rectes de regressió ajustades dels gràfics de comparació entre els valors obtinguts vs. els esperats pel model DWT-ANN proposat, pels índexs Folin-Ciocalteu i I_{280} , promitjant els valors de cada mostra dels 25 replicats dependent de si pertanyien al conjunt d'entrenament o de test. Interval de confiança calculats amb un nivell de confiança del 95%.

| Folin-Ciocalteu index(54-2372 mg·L⁻¹) | | | | |
|---|-------------------|----------------|--|---------------------------------|
| | <i>Correlació</i> | <i>Pendent</i> | <i>Ord. Origen (mg·L⁻¹)</i> | <i>RMSE (mg·L⁻¹)</i> |
| <i>Entrenament</i> | 0.998 | 0.959±0.029 | 53.0±30.6 | 89 |
| <i>Test</i> | 0.979 | 0.952±0.093 | -43.6±28.7 | 366 |
| I_{280} index(3.78-72.91 arb. unit) | | | | |
| | <i>Correlació</i> | <i>Pendent</i> | <i>Ord. Origen (arb. unit)</i> | <i>RMSE (arb. unit)</i> |
| <i>Entrenament</i> | 0.994 | 0.965±0.049 | 1.24±1.68 | 3.5 |
| <i>Test</i> | 0.963 | 0.914±0.119 | 0.87±3.77 | 10.1 |

4.4.1.2. Quantificació mitjançant la Llengua (Bio)Electrònica

De forma anàloga al cas anterior, però aquest cas amb un total de 29 vins i utilitzant la BioET, es va procedir a la determinació de l'índex de polifenols d'aquests mitjançant mesures voltamperomètriques i l'ús de (bio)sensors desenvolupats per la determinació d'aquests compostos per tal d'avaluar si s'observava cap millora en el model de resposta. És a dir, avaluar la resposta de la ET front de la BioET.

En aquest cas la BioET estava formada pels següents (bio)sensors (§3.1, Taula 3.1): GEC, Tyr, Lac i Cu. L'escombrat de potencial va comprendre el rang de -0.4V a +0.8V vs. Ag/AgCl amb un salt de potencial de 9mV i una velocitat d'escombrat de 100mV·s⁻¹; aplicant un potencial de condicionament de +1.0V durant 40s en un vas que contenia 10mL d'aigua destil·lada en l'etapa de neteja electroquímica entre les mesures de les diferents mostres.

D'una banda, la tirossinassa i la lacassa van ser escollides ja que aquests són els dos enzims més utilitzats de forma habitual per al desenvolupament de biosensors enfocats a la detecció de compostos fenòlics. De l'altra, la tria de les nanopartícules de coure va ser deguda a que tant la tirossinassa com la lacassa són enzims redoxs cuprífers. És a dir, es tracta d'enzims que tenen àtoms de coure en el seu centre actiu (dos en el cas de la tirossinassa i tres en el cas de la lacassa), el qual interacciona amb el dioxigen per formar un producte intermedi molt reactiu químicament, que a continuació oxida el substrat; així doncs, es va pensar que les nanopartícules de coure podien tenir un efecte catalític similar.

Aquests enzims pertanyen a la classe de les cupro-oxidases, les quals catalitzen la reducció de l'oxigen molecular mitjançant diferents donadors d'electrons. En aquestes reaccions, l'oxigen es reduït directament a aigua sense la formació intermèdia de peròxid d'hidrogen. Concretament, la tirossinassa catalitza la hidroxilació dels

monofenols a catecols, els quals a la vegada són oxidats de nou a *o*-quinones, utilitzant oxigen en ambdós casos; llavors les *o*-quinones generades en la reacció enzimàtica són reduïdes electroquímicament a *o*-difenois, resultant en el reciclatge dels difenois en aplicar un potencial lleugerament negatiu. De forma similar, la lacassa catalitza l'oxidació dels fenols generant espècies fenoxiradicals, les quals són convertides a quinones en el segon pas de l'oxidació; llavors, al igual que en el cas de la tirosinassa, aquestes quinones poder ser electroquímicament reduïdes de nou a fenols [30].

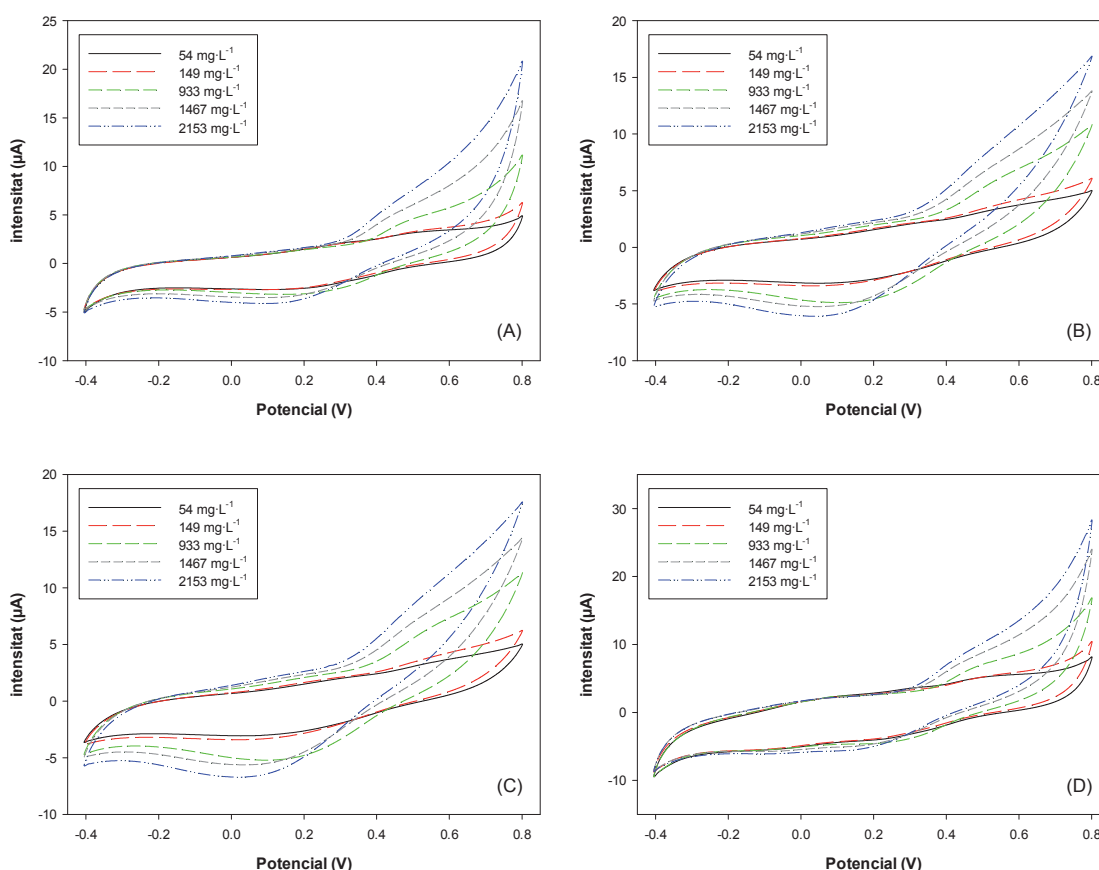


Figura 4.34. Exemple dels diferents voltamperogrames obtinguts per certes mostres arbitràries amb els (bio)sensors: (A) GEC, (B) Tyr, (C) Lac i (D) Cu. Els valors expressats corresponen a l'índex de Folin-Ciocalteu expressat com a equivalents d'àcid gàl·lic mesurat mitjançant el mètode de referència.

A tall d'exemple, la Figura 4.34 mostra la resposta de la BioET obtinguda per diferents de les mostres de vins analitzades. Al igual que en el cas anterior, es pot observar com hi ha una clara resposta per part dels sensors a l'increment en l'índex de polifenols, i.e. obtenint majors corrents per als vins amb un major índex. En general, dos processos són observats per tots els sensors, corresponents a l'oxidació del corresponent fenol a la respectiva quinona i la reducció de la quinona a la forma fenòlica. No obstant, tal com es pot observar, majors intensitat de corrent són obtingudes per als biosensors que contenen els enzims tirosinassa (Figura 4.34B) i

lacassa (Figura 4.34C), especialment en la zona de reducció propera a 0V on la resposta obtinguda és ca. 3 vegades major. En aquest mateix sentit, cal destacar també l'efecte catalític observat pel sensor modificat amb nanopartícules de Cu; confirmant així la hipòtesi exposada anteriorment.

Abans de procedir a la construcció del model de resposta i per tal de millorar el comportament d'aquest, cal realitzar una etapa de compressió del senyal, la qual va ser assolida en aquest cas mitjançant l'ús de la FFT, conservant només els primers 32 coeficients obtinguts per cada sensor; el qual va permetre assolir un nivell de compressió del senyal original del 88.1% sense pèrdua d'informació significativa.

Després de dur a terme la compressió del senyal, el següent pas va ser l'optimització del model ANN per tal de dur a terme la quantificació dels dos índexs, utilitzant com a *inputs* els coeficients obtinguts. Mitjançant un procés iteratiu, es va procedir a la selecció de la topologia òptima per la ANN per aquest cas, la qual va ser la formada per 128 neurones en la capa d'entrada (corresponents als 32 coeficients obtinguts de l'anàlisi FFT per cadascun dels 4 sensors), 6 neurones en la capa oculta i 2 neurones en la capa de sortida (corresponents als índexs FC i I₂₈₀), emprant respectivament les funcions de transferència *logsig* i *tansig* per connectar les diferents capes.

Al igual que en el cas anterior, es va aplicar el mètode de validació *repeated random sub-sampling* (descriu a §3.4.2.3.1) per a l'avaluació de la capacitat predictiva del model, ja que aquest presenta com a principals avantatges el fet que elimina la dependència dels resultats obtinguts de la subdivisió específica de les dades, permetent a la vegada extreure avaluadors estadístics de la qualitat d'aquests. Així doncs, es va repetir el procés de divisió de les dades pels dos subconjunts (*train/test*) de forma aleatòria 29 vegades (tantes com mostres), emprant cada vegada el 80% de les dades (23 mostres) per la construcció del model (subconjunt d'entrenament) i el 20% romanent (6 mostres) per l'avaluació de la resposta d'aquest (subconjunt de test).

Llavors, una vegada obtingudes les respostes de tots els models construïts, els valors predits per cada model es van agrupar segons si la mostra havia estat utilitzada en el procés d'entrenament o de validació. Finalment, es calculen els valors mitjans per a cada mostra, fet que ens permet calcular les incerteses del model i obtenir dades imparcials. A partir d'aquests valors, es van construir els gràfics de comparació dels valors obtinguts vs. els esperats pels índexs de Folin-Ciocalteu (Figura 4.35A&B) i I₂₈₀ (Figura 4.35C&D), tant pel conjunt d'entrenament com pel de test, per tal de verificar la capacitat de predicció del model ANN generat.

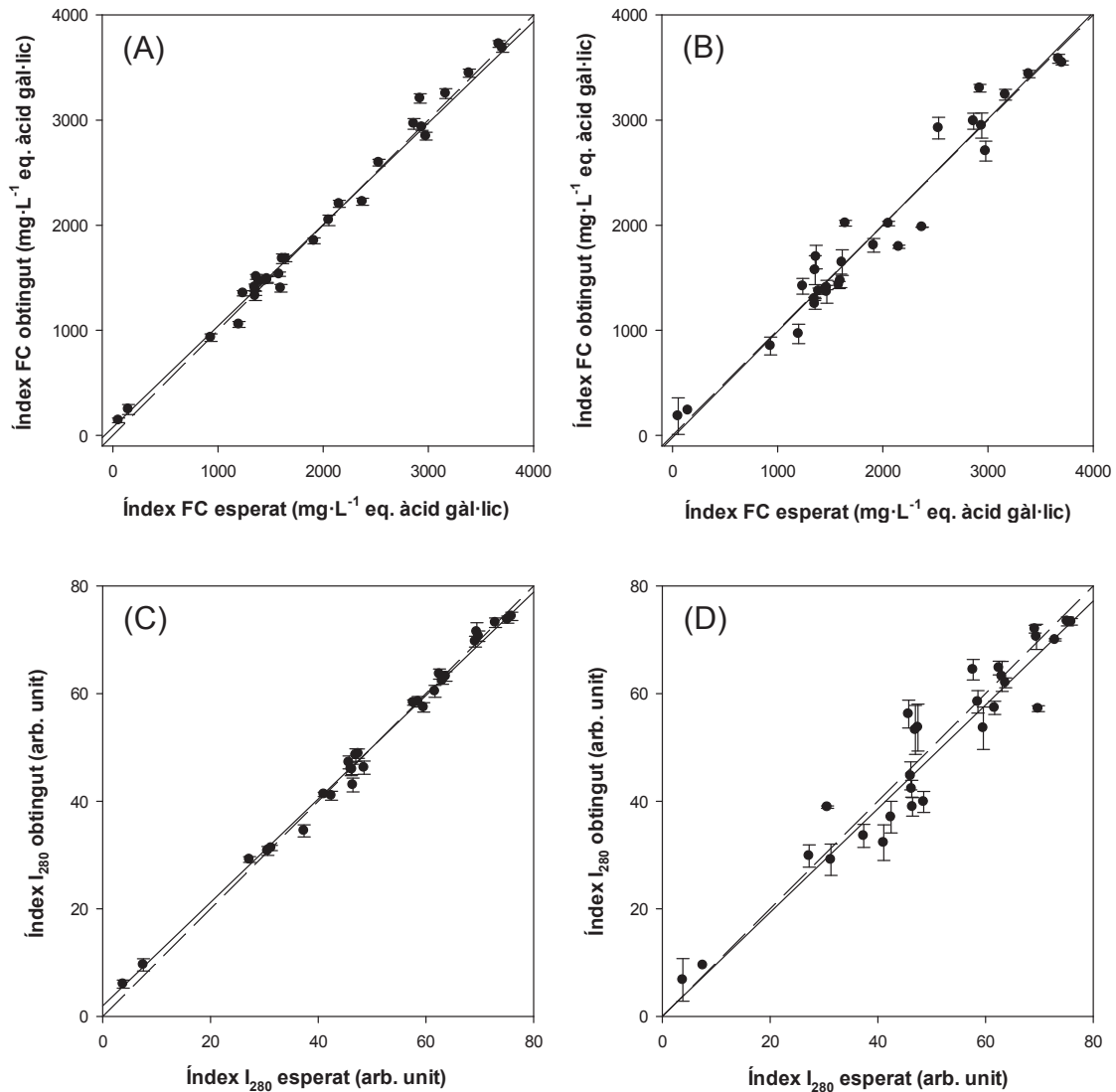


Figura 4.35. Capacitat de modelatge del model FFT-ANN utilitzant com a mètode de validació *repeated random sub-sampling*. Gràfics de comparació de les concentracions obtingudes vs. les esperades pels índexs (A,B) Folin-Ciocalteu i (C,D) I₂₈₀, tant pel subconjunt d'entrenament (A,C - esquerra) com de test (B,D - dreta). La línia discontinua correspon a la diagonal teòrica. Els resultats proporcionats corresponen a la mitjana dels valors obtinguts per cadascuna de les mostres després de la repetició del càlcul 29 vegades, utilitzant una divisió aleatòria de les mostres pels subconjunts d'entrenament i test cada vegada. Interval de confiança calculats amb un nivell de confiança del 95%.

Tal com es pot observar, per ambdós índexs s'obté una tendència satisfactòria, amb línies de regressió gairebé indistingibles de les teòriques; obtenint com és habitual en els models ANN, una menor dispersió i incertesa dels resultats pels subconjunts d'entrenament. La Taula 4.16 mostra els paràmetres obtinguts mitjançant la regressió ponderada, per tal de facilitar la comparació amb els resultats anteriors (§4.4.1.1, Taula 4.15). Tal com es pot esperar dels gràfics de comparació de resposta obtinguda vs. esperada, una bona tendència va ser assolida en tots els casos, però amb una millor correlació pels subconjunts d'entrenament degut a la menor dispersió dels resultats. No obstant, els resultats obtinguts pels dos subconjunts són propers als valors ideals, amb

valors d'ordenada a l'origen propers a 0 i valors de pendents i coeficients de correlació propers a 1.

Taula 4.16. Paràmetres de les rectes de regressió ajustades dels gràfics de comparació entre els valors obtinguts *vs.* els esperats pel model FFT-ANN proposat, pels índexs Folin-Ciocalteu i I_{280} , promitjant els valors de cada mostra dels 29 replicats depenent de si pertanyien al conjunt d'entrenament o de test. Interval de confiança calculats amb un nivell de confiança del 95%.

| Folin-Ciocalteu index (54-3705 mg·L⁻¹) | | | | |
|--|-------------------|----------------|--|---------------------------------|
| | <i>Correlació</i> | <i>Pendent</i> | <i>Ord. Origen (mg·L⁻¹)</i> | <i>RMSE (mg·L⁻¹)</i> |
| <i>Entrenament</i> | 0.997 | 0.966±0.002 | 75.3±3.4 | 102 |
| <i>Test</i> | 0.989 | 1.011±0.005 | -21.4±7.0 | 205 |
| I_{280} index (3.78-75.93 arb. unit) | | | | |
| | <i>Correlació</i> | <i>Pendent</i> | <i>Ord. Origen (arb. unit)</i> | <i>RMSE (arb. unit)</i> |
| <i>Entrenament</i> | 0.998 | 0.961±0.018 | 1.99±0.74 | 1.85 |
| <i>Test</i> | 0.974 | 0.964±0.044 | 0.07±2.09 | 5.17 |

Per tant, quedant palesos en aquest cas (i.e. determinació del'índex global de polifenols) els beneficis derivats de la incorporació de biosensors en la matriu de sensors, és a dir, l'ús de la BioET *vs.* la ET; presentant avantatges més enllà de l'obtenció d'un millor model. Per exemple, reducció en el temps de mesura degut a l'escombrat d'una menor finestra de potencials, fet que comporta una reducció en la dimensionalitat de les dades; generació de dades més riques per a ser utilitzades com a punt de partida; reducció dels potencials en els que s'observa la resposta dels sensors, reduint així possibles interferències i proporcionant senyals amb una major relació S/N; etc. No obstant, sense menystenir la capacitat de la ET, la qual ha demostrat ser capaç d'assolir també la correcta quantificació d'ambdós índexs.

4.4.2. Discriminació individual de compostos fenòlics

Actualment, existeix una creixent demanda de mètodes analítics altament sensibles i selectius que a banda de ser capaços de proporcionar un índex global del contingut de polifenols totals, siguin capaços també de dur a terme la determinació individual d'aquests [31, 32]. En aquest aspecte, tot i la gran quantitat de recerca que s'està duent a terme en aquest camp, la separació i quantificació dels diferents compostos fenòlics segueixen sent difícil, especialment la determinació simultània d'aquells de subgrups químics diferents.

Per aquest motiu i donats els bons resultats obtinguts en la predicció del contingut total de polifenols, es va voler anar un pas més enllà i intentar dur a terme la discriminació individual d'aquests compostos en el vi. Així, es va procedir primer a

l'anàlisi qualitativa de mostres fortificades amb diferents compostos fenòlics per tal d'avaluar la capacitat d'ambdues llengües electròniques en la discriminació d'aquests. Mentre que en un segon intent, es va procedir a la resolució i quantificació de mescleres ternàries d'aquests per tal d'avaluar la capacitat de la BioET en aquesta tasca.

En ambdós casos, per a la preparació de les mostres fortificades, es va utilitzar un vi amb un baix contingut en polifenols com a matriu de partida. Els compostos considerats van ser seleccionats d'acord al seu contingut promig en el vi [9, 24]. Tanmateix, cal tenir en compte que el contingut total de polifenols mitjà mesurat pel mètode de Folin és major que el contingut individual de cada compost individual [25, 33]. Per exemple, el contingut total de polifenols en vi negre és de l'ordre de $2160\text{mg}\cdot\text{L}^{-1}$, $820\text{mg}\cdot\text{L}^{-1}$ en el vi rosat i $320\text{mg}\cdot\text{L}^{-1}$ per als vins blancs; mentre que la concentració individual de compostos fenòlics és molt més baixa, amb valors per sota de $300\text{mg}\cdot\text{L}^{-1}$ per al vi negre i $50\text{mg}\cdot\text{L}^{-1}$ pels vins blancs [25].

4.4.2.1. Anàlisi qualitativa mitjançant la Llengua Electrònica

Emprant els mateixos sensors i condicions de mesura que s'havien utilitzat en el cas de la quantificació dels índex globals de polifenols (§4.4.1.1), es van mesurar un total de 35 mostres sintètiques, les quals contenien 6 compostos fenòlics diferents (Figura 4.36): àcid gàl·lic, catequina, àcid *p*-coumàric, àcid cafèic, catecol i *m*-cresol. Aquestes mostres estaven distribuïdes en 7 grups diferents (un grup per cadascun dels 6 compostos estudiats més un altre pel vi original).

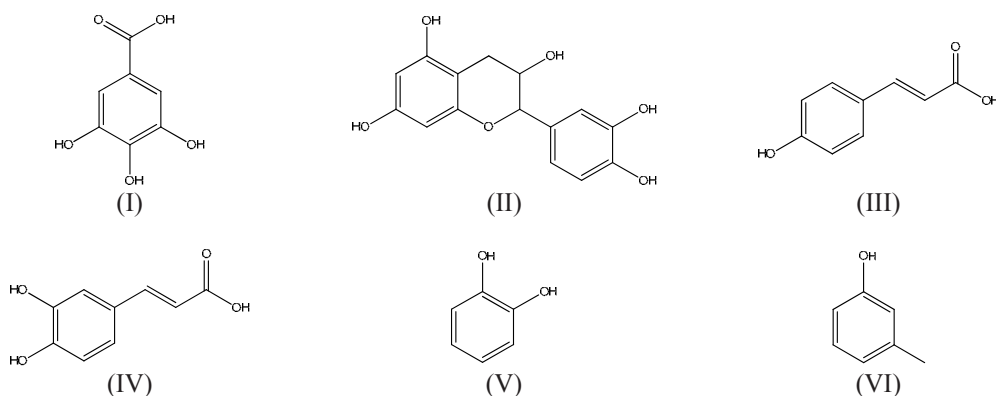


Figura 4.36. Estructura química dels compostos fenòlics sota estudi: (I) àcid gàl·lic, (II) catequina, (III) àcid *p*-coumàric, (IV) àcid cafèic, (V) catecol, i (VI) *m*-cresol.

Per tal de confirmar que la discriminació realment provenia dels diferents compostos afegits, la mateixa concentració va ser utilitzada per cadascun d'ells, preparant a més 5 mostres diferents per cada una de les classes, les quals van ser tractades com a mostres independents i analitzades de forma aleatòria. Per tant, per cadascuna de les classes, 5 μmol s del compost estudiat van ser afegits a 25mL de vi, el qual representa un increment de la concentració de ca. $36\text{mg}\cdot\text{L}^{-1}$ ($200\mu\text{M}$).

Tal com es pot observar a la Figura 4.37A, on es mostra com a exemple la resposta del sensor PPy obtinguda front les diferents classes de mostres analitzades, existeix un evident patró de resposta diferenciat per cadascun dels compostos (especialment en la zona catòdica). Per tal de confirmar aquesta discriminació, la resposta voltamperomètrica obtinguda del conjunt sensors va ser preprocessada emprant la DWT i avaluada mitjançant l'ús del PCA. Finalment, degut a que el PCA només és una eina de visualització i no proporciona informació sobre la classificació de les mostres per si mateix, les mostres van ser agrupades utilitzant eines d'agrupament de dades (*cluster analysis tools*).

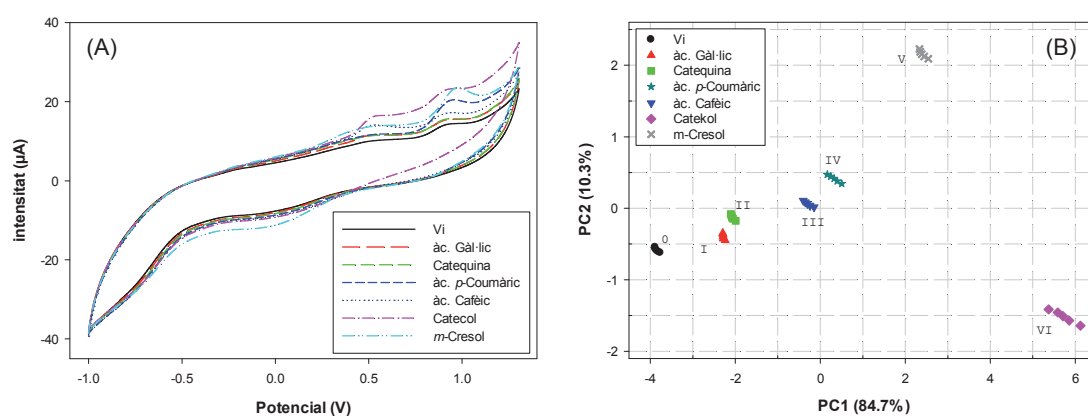


Figura 4.37. (A) Exemple de la resposta voltamperomètrica obtinguda pel sensor PPy per les diferents classes de mostres analitzades. (B) Representació dels *scores* dels dos primers components principals obtinguts després de l'anàlisi PCA del conjunt de respostes de tots els sensors. Un total de 35 mostres van ser analitzades corresponents a: (0) vi, (I) àcid gàl·lic, (II) catequina, (III) àcid *p*-coumàric, (IV) àcid cafèic, (V) catecol i (VI) *m*-cresol.

Aquestes eines permeten la construcció de *clusters* en un conjunt de dades a partir la distància entre diferents observacions (normalment la distància euclidiana), com a unió per tal de construir l'arbre d'agrupament jeràrquic (*hierarchical cluster tree*) [34]. D'aquesta manera, el conjunt d'objectes són assignats en grups de tal manera que els objectes en un mateix grup són més similars (en un sentit o un altre) entre si que amb els dels altres grups.

La Figura 4.37B mostra el resultat de l'anàlisi PCA, on es pot observar com els patrons de la figura evidencien el fet que les mostres es troben agrupades en funció del compost fenòlic afegit; separant així de forma clara les diferents classes de mostres. Cal destacar el fet que amb només els dos primers PCs, la variància acumulada és de ca. 95.0%; un elevat valor que significa que quasi tota la variància continguda en les dades originals es troba representada utilitzant només aquestes dos noves coordenades.

El fet que els *clusters* corresponents a l'àcid gàl·lic i la catequina es trobin situats tant propers entre ells, i també al *cluster* del vi, pot ser degut al fet que aquests dos compostos són els que es troben en major concentració generalment en el vi. Tanmateix, els *clusters* corresponents als àcids *p*-coumàric i cafèic es troben més separats del *cluster* del vi, però propers entre ells, tal com es podria esperar ja que ambdós són àcids hidroxicinàmics que només difereixen per un grup alcohol (Figura 4.36). Finalment, els *clusters* corresponents al *m*-cresol i el catecol es troben clarament separats de la resta de classes i representats pel PC2. La posició dels quals podria explicar-se pel fet que aquests components es troben en menor proporció en el vi i la seva estructura química és molt diferent a la dels altres (Figura 4.36).

Així doncs, de forma resumida, podem concloure que s'assoleix una correcta discriminació entre els diferents compostos fenòlics estudiats, obtenint grups individuals per cadascun d'ells; on a més a més, s'observa que compostos amb certa similitud estructural es troben propers entre ells, significat que hi ha alguna relació de fons entre ells.

No obstant, tot i els bons resultats obtinguts amb les mostres fortificades, cal tenir en compte que la quantificació individual d'aquests compostos en una mostra complexa com el vi és molt difícil; no obstant, els resultats obtinguts suggereixen aquesta possibilitat. En qualsevol cas, l'estudi demostra el potencial de la ET proposada, tant en la quantificació del contingut de polifenols totals com en la discriminació dels diferents compostos fenòlics presents en el vi.

4.4.2.2. Anàlisi qualitativa mitjançant la Llengua (Bio)Electrònica

Anàlogament als experiments duts a terme en l'apartat anterior, es va procedir a la preparació i mesura de mostres fortificades amb diferents compostos fenòlics, però aquest cop utilitzant la matriu de (bio)sensors de la BioET (§4.4.1.2). En aquest cas, i degut a la millor resposta de la BioET, es van considerar la discriminació d'un total de

10 compostos diferents (Figura 4.38): àcid gàl·lic, catequina, àcid *p*-coumàric, àcid cafèic, catecol, fenol, *m*-cresol, àcid ferúlic, àcid clorogènic i quercetina. Al igual que en el cas anterior, també es van preparar 5 mostres per cadascun dels compostos fenòlics considerats (5µmols de cada un dels compostos van ser afegits a 25mL de vi, el qual representa un increment en la concentració de ca. 36mg·L⁻¹ (200µM)) més 5 mostres corresponents al vi de partida, obtenint d'aquesta manera un total de 55 mostres distribuïdes en 11 classes. A més, totes les mostres van ser analitzades de forma aleatòria per tal d'assegurar que l'agrupació no és conseqüència de l'ordre en que aquestes són mesurades.

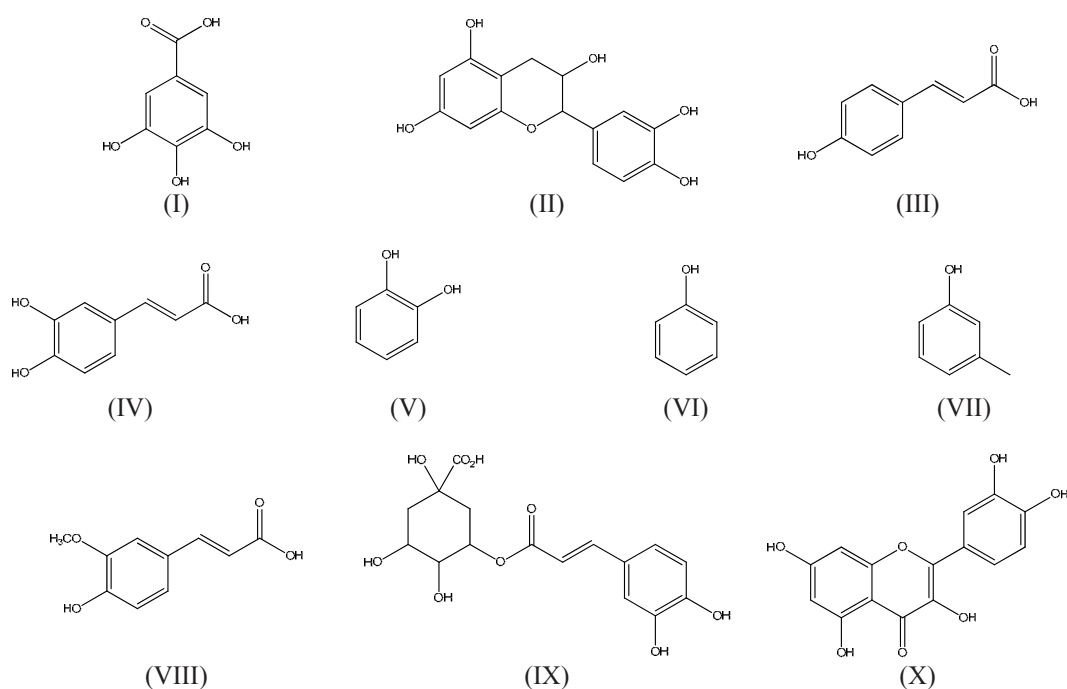


Figura 4.38. Estructura química dels compostos fenòlics sota estudi: (I) àcid gàl·lic, (II) catequina, (III) àcid *p*-coumàric, (IV) àcid cafèic, (V) catecol, (VI) fenol, (VII) *m*-cresol, (VIII) àcid ferúlic, (IX) àcid clorogènic i (X) quercetina.

Un cop mesurades totes les mostres, el conjunt de voltamperogrames obtinguts va ser preprocessat utilitzant la FFT i analitzat utilitzant el PCA per tal d'obtenir una visualització dels patrons de resposta. Degut a que el PCA no pot actuar per si mateix com a classificador, els resultats d'aquest van ser explorats amb eines d'agrupament de dades (*cluster analysis tools*) per tal d'establir les agrupacions en les mostres.

La Figura 4.39 mostra el resultat de l'anàlisi PCA, on es pot observar com els patrons en la figura evidencien el fet que les mostres es troben agrupades en funció del compost afegit; separant així de forma clara les diferents classes de mostres. Cal

destacar el fet que amb només els dos primers PCs, la variància acumulada és de ca. 98.3%; un valor molt elevat que garanteix la pràctica representació de quasi tota la informació present en les dades originals.

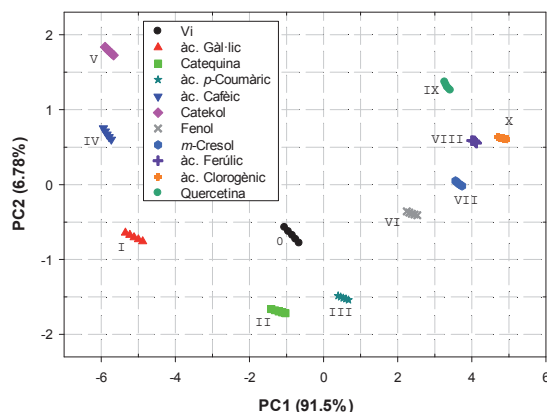


Figura 4.39. Representació dels *scores* dels dos primers components principals obtinguts després de l'anàlisi PCA. Un total de 55 mostres van ser analitzades corresponents a: (0) vi, (I) àcid gàl·lic, (II) catequina, (III) àcid *p*-coumàric, (IV) àcid cafèic, (V) catecol, (VI) fenol, (VII) *m*-cresol, (VIII) àcid ferúlic, (IX) àcid clorogènic i (X) quercetina

Analitzant més detalladament els resultats obtinguts, és pot observar com a diferència del cas anterior (Figura 4.37B) existeix una major separació entre els diferents *clusters*, trobant-se gairebé en la posició central el *cluster* corresponent al vi sense cap compost afegit, i essent el punt de partida a partir del qual s'observen els canvis induïts pels diferents compostos. Per una banda, al costat esquerra (valors de PC1 negatius) es troben els compostos fenòlics presents en major concentració de forma habitual en el vi (*clusters* I a V); mentre que a l'altra banda (valors de PC1 positius), i més propers entre ells, es troben la resta de compostos estudiats i menys abundants de forma habitual. Finalment, cal també tenir en compte que la posició dels *clusters* en aquest cas està també altament influenciada per la diferent selectivitat dels enzims per cadascun dels compostos.

Al igual que en el cas anterior, comparant els resultats obtinguts en el cas de la ET i de la BioET, es pot observar una millora en les prestacions del model degudes a la incorporació de biosensors en la matriu d'elèctrodes que formen la llengua. I encara més, degut a la bona discriminació assolida entre els diferents compostos en les mostres fortificades en el cas de la BioET, intuïnt la possibilitat de dur a terme la resolució i quantificació de mesclades de compostos fenòlics.

4.4.2.3. Resolució de mescles ternàries mitjançant la Llengua BioElectrònica

Donats els bons resultats obtinguts en la discriminació qualitativa de diferents compostos fenòlics, i per tal de confirmar la capacitat de la BioET per assolir aquesta tasca, assolint-ne inclús la seva quantificació individual en presència d'altres compostos fenòlics, és va procedir a la resolució de mescles ternàries d'àcid cafèic, catecol i catequina (Figura 4.40); tres dels compostos fenòlics majoritaris en vins.

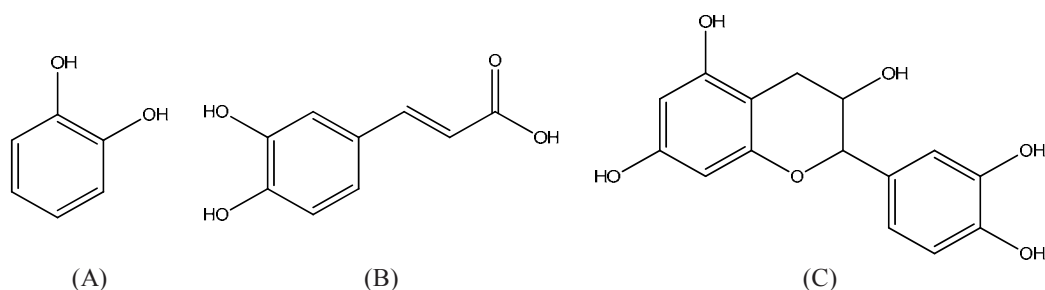


Figura 4.40. Estructura química dels compostos fenòlics sota estudi: (A) àcid cafèic, (B) catecol i (C) catequina.

Així, primer és va avaluar la resolució de les mescles d'aquests tres compostos en mostres sintètiques preparades en buffer per tal d'avaluar la viabilitat de la proposta; procedint posteriorment a l'anàlisi de noves mostres preparades utilitzant un vi genèric com a matriu per tal de comprovar que aquesta no representa cap problema.

Per a la preparació de les mostres es va optar per la utilització d'un disseny cúbic (Figura 3.2, amb un rang de concentracions de 0 a 200 μM per cadascun dels compostos avaluats), el qual requereix de 27 mostres per a la construcció del model de resposta (conjunt d'entrenament; 73%), més 10 mostres addicionals distribuïdes al llarg del domini experimental definit per les anteriors, les quals van ser utilitzades per tal d'avaluar la capacitat de predicció d'aquest (conjunt de test; 27%). Finalment, es van preparar 10 mostres addicionals suplementàries, però utilitzant vi com a matriu enlloc de buffer per tal d'avaluar-ne la seva aplicabilitat en el sector.

Per la mesura de les mostres, el conjunt de 4 sensors que formen part de la BioET va ser utilitzat (§3.1, Taula 3.1): GEC, Tyr, Lac i Cu; realitzant un escombrat de potencial en el rang de -0.4V a $+0.8\text{V}$ vs. Ag/AgCl amb un salt de potencial de 9mV i una velocitat d'escombrat de $100\text{mV}\cdot\text{s}^{-1}$; aplicant una etapa de neteja electroquímica entre la mesura de les diferents mostres consistent en l'aplicació d'un potencial de $+1.0\text{V}$ durant 40s en un vas que contenia 10mL d'aigua destil·lada.

A tall d'exemple, la Figura 4.41 mostra la resposta de la BioET per algunes de les diferents mostres analitzades seleccionades de forma arbitrària. Tal com es pot observar, i degut al solapament que existeix entre els senyals dels diferents compostos, no és possible dur a terme l'assignació específica del potencial d'oxidació o reducció per cada compost fenòlic.

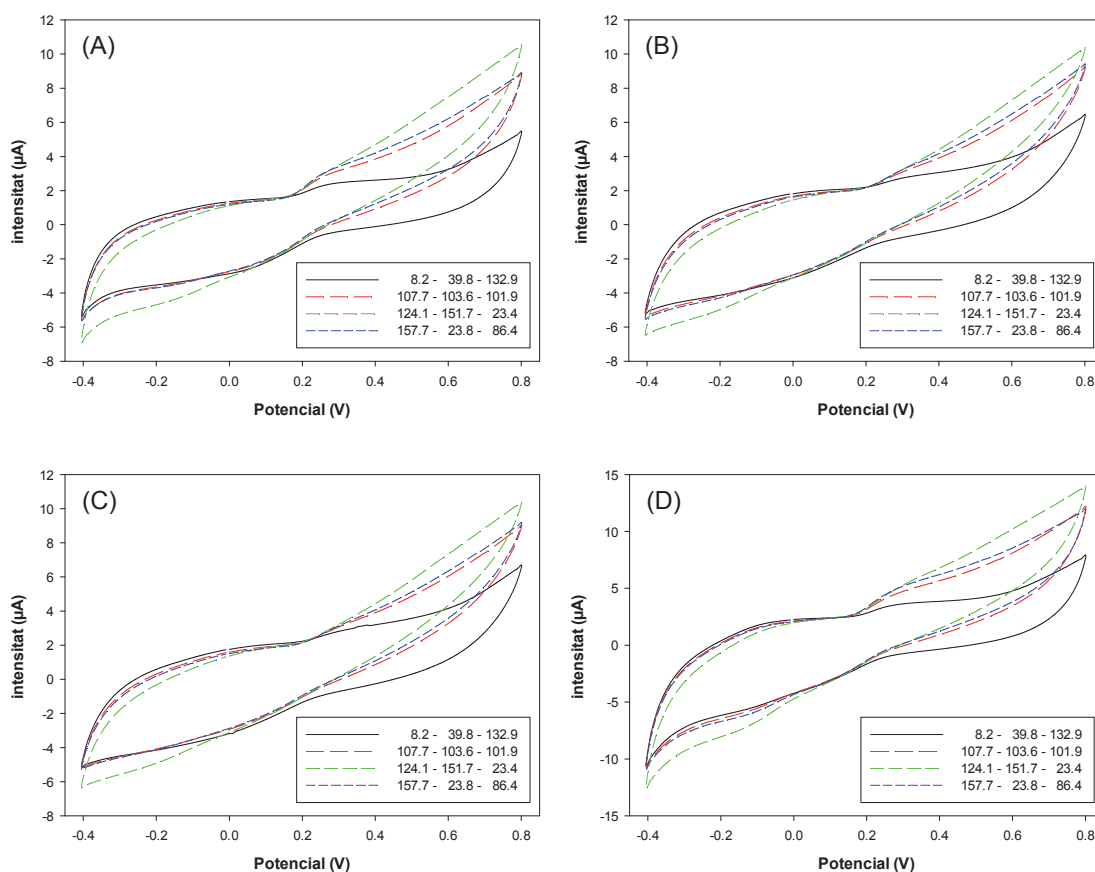


Figura 4.41. Exemple dels diferents voltamperogrames obtinguts per quatre mescles ternàries amb els (bio)sensors: (A) GEC, (B) Tyr, (C) Lac i (D) Cu. Els valors de la llegenda corresponen a les concentracions en μM de cada un dels compostos fenòlics: catecol, àcid cafèic i catequina, respectivament.

No obstant, tot i no ser possible assolir aquesta discriminació en els senyals, i a pesar de la semblança visual entre la resposta dels diferents (bio)sensors (la diferència en la resposta dels quals s'accentua en el cas de les mostres reals, Figura 4.34), cal remarcar que la diferència de sensibilitat mostrada d'aquests en front de cada compost és lleugerament diferent (e.g. Taula 4.13); essent aquesta diferent resposta deguda a la diferent afinitat de cada enzim pels diferents compostos fenòlics [30], o la pròpia activitat electroquímica en el cas del sensor Cu o del GEC.

Aquest cas, on hi ha una clara resposta creuada per part dels diferents (bio)sensors, generant respostes voltamperomètriques complexes i superposades, però que contenen informació global sobre els compostos analitzats al llarg de tot el voltamperograma, el converteix en una situació molt desitjable i avantatjosa per ser estudiada mitjançant l'aproximació i els conceptes de les llengües electròniques.

Així, en aquest cas, es va optar per fer ús del mètode CI (§3.4.1.2xi) per dur a terme la selecció dels diferents punts de cada voltamperograma que seran utilitzats com a inputs en el model neuronal; permetent al mateix temps demostrar la diferent sensibilitat dels diferents (bio)sensors vers els diferents compostos estudiats. Cal destacar que el conjunt d'*inputs* seleccionat no és únic i que aquest pot variar lleugerament en funció del grau de reducció desitjat, la pròpia repetició o com a conseqüència d'usar diferents arquitectures en la xarxa (variació de les funcions de transferència o el número de neurones en la capa oculta).

D'aquesta manera, del total de intensitats de corrent del conjunt de voltamperogrames (268 punts x 4 sensors) es van seleccionar només un total de 23 corrents (97.9% del total), el qual significa una elevada reducció en el nombre d'*inputs* de la ANN sense pèrdua en les capacitats de modelatge d'aquesta. Concretament, els potencials als que corresponen les intensitats de corrent seleccionades varen ser les següents (on ^{“ab”} denota la intensitat de corrent corresponent a l'ona anòdica i ^{“c”} a l'ona catòdica):

- 4 pel sensor GEC: +0.252^c, +0.576^c, -0.018^a i -0.315^a V
- 5 pel biosensor Tyr: +0.693^c, +0.756^c, +0.513^a, -0.135^a i -0.369^a V
- 5 pel biosensor Lac: +0.333^c, +0.729^c, +0.531^a, -0.144^a i -0.369^a V
- 9 pel sensor Cu: +0.549^c, +0.576^a, +0.306^a, +0.225^a, -0.054^a, -0.216^a, -0.252^a, -0.108^c i -0.054^c V

A continuació, un cop seleccionats els *inputs*, el següent pas va ser l'optimització de la topologia de la xarxa i la construcció del model; procés després del qual va quedar definida de la següent manera: 23 neurones en la capa d'entrada (corresponents a les 23 intensitats de corrent associades als potencials seleccionats), 5 neurones en la capa oculta i 3 neurones en la capa de sortida (corresponents a la concentració de cadascun dels compostos estudiats), emprant respectivament les funcions *logsig* i *purelin* per connectar les diferents capes.

Cal recordar que aquest cop, la divisió de les mostres entre el conjunt d'entrenament i de test va quedar definida pel disseny experimental utilitzat, així doncs no requerint de

cap precaució extra en aquest procés. No obstant, per tal de proporcionar informació relativa a la reproductibilitat del model, tots els models van ser reentrenats 5 vegades amb una reinicialització aleatòria dels pesos cada vegada.

La Figura 4.42 mostra els resultats obtinguts amb el model seleccionat per cadascun dels tres analits, tant pel subconjunt d'entrenament com de test, on les barres d'error varen ser calculades a partir dels resultats dels diferents reentrenaments. Tal com es pot observar, existeix una bona correlació entre els valors esperats i els calculats a partir del model, obtenint rectes de regressió molt properes als valors ideals; demostrant la bona capacitat de predicció del model obtingut.

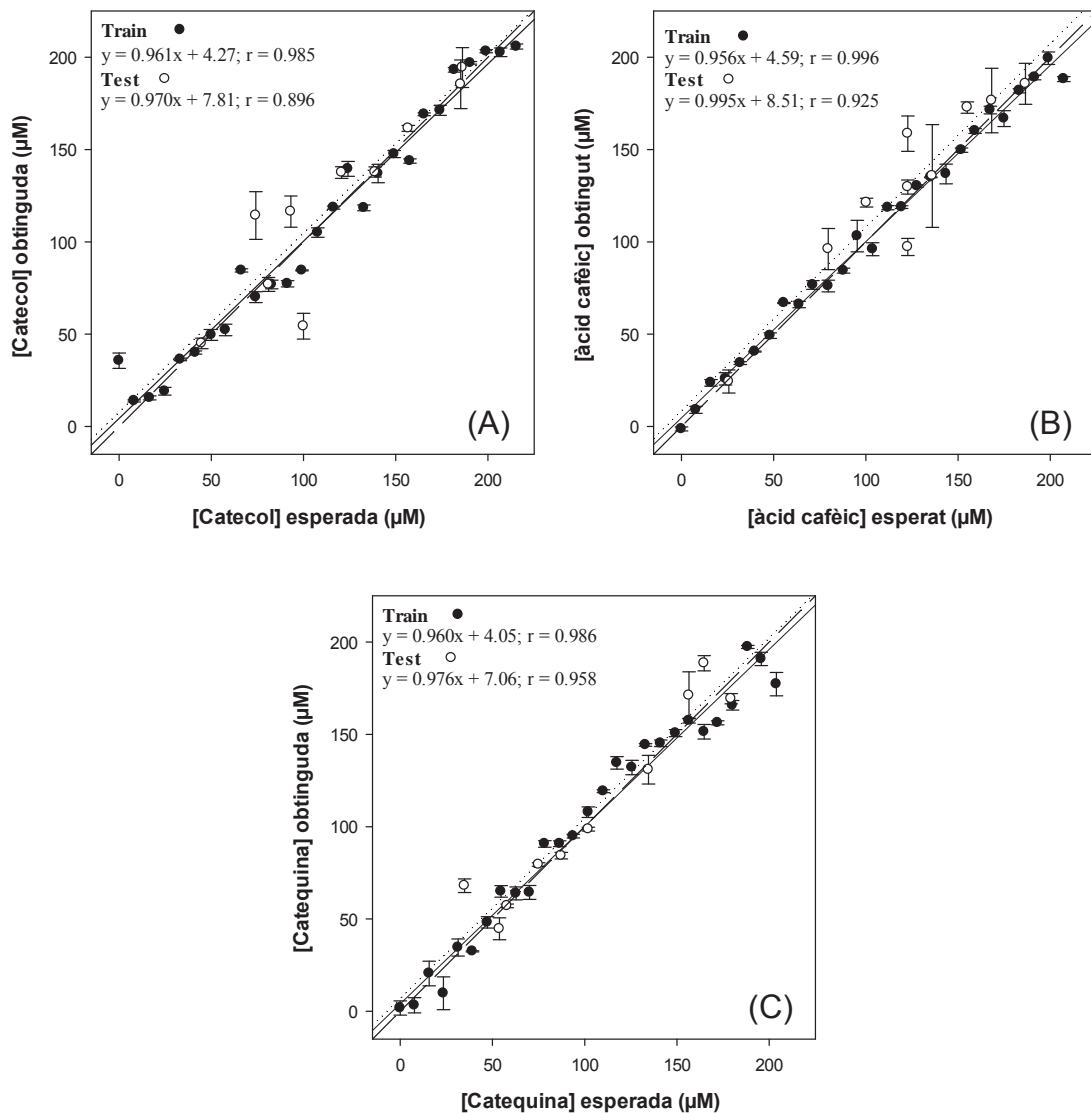


Figura 4.42. Capacitat de modelatge del model CI-ANN. Gràfics de comparació de les concentracions obtingudes vs. les esperades per (A) catecol, (B) l'àcid cafèic i (C) catequina, tant pel subconjunt d'entrenament (\bullet , línia sòlida) com de test (\circ , línia de punts). La línia discontinua correspon a la diagonal teòrica. Intervalls de confiança calculats amb un nivell de confiança del 95% a partir del reentrenament del model cinc vegades amb reinicialització aleatòria dels pesos.

A continuació, un cop seleccionats els *inputs* i definida l'arquitectura del model, es va procedir a l'anàlisi de les mostres de vi fortificades amb diferents quantitats dels analits d'interès. El procediment de mesura i de tractament del senyal va ser el mateix, amb la única precaució de reentrenar el model per tal de compensar les diferències entre les solucions tamponades i la matriu del vi. Així, utilitzant com a mètode de validació l'estratègia *leave-one-out* degut al menor número de mostres disponibles i per tal d'assegurar que totes les mostres intervenien en el procés de validació, es va procedir a l'avaluació de la resposta de la BioET en mostres de vi.

La Taula 4.17 mostra els valors del grau de recuperació obtinguts per cadascuna de les mostres analitzades (§3.4.2.3.2.2); obtenint resultats satisfactoris en general pels 3 compostos, amb uns valors mitjans de recuperació de 104.0%, 116.7% i 121.9% pel catecol, l'àcid cafèic i la catequina, respectivament.

Taula 4.17. Resultats detallats dels valors proporcionats per la BioET per les mostres fortificades i els valors reals d'aquestes pels tres compostos considerats; addicionalment, el grau de recuperació també es va calcular per cada mostra.

| | <i>Catecol</i> | | | <i>Àcid cafèic</i> | | | <i>Catequina</i> | | |
|-----|-----------------------------------|------------------------------------|-------------------------|-----------------------------------|------------------------------------|-------------------------|-----------------------------------|------------------------------------|-------------------------|
| | <i>BioET</i> (μM) | <i>Afegit</i> (μM) | <i>Recuperat</i> (%) | <i>BioET</i> (μM) | <i>Afegit</i> (μM) | <i>Recuperat</i> (%) | <i>BioET</i> (μM) | <i>Afegit</i> (μM) | <i>Recuperat</i> (%) |
| #1 | 38.49 | 49.98 | 77.0 | 115.20 | 108.60 | 106.1 | 170.18 | 177.78 | 95.7 |
| #2 | 178.17 | 178.37 | 99.9 | 82.07 | 92.85 | 88.4 | 37.60 | 12.64 | 297.5 |
| #3 | 17.26 | 17.23 | 100.1 | 66.36 | 64.67 | 102.6 | 132.18 | 200.61 | 65.9 |
| #4 | 115.06 | 75.83 | 151.7 | 159.35 | 197.31 | 80.8 | 168.71 | 161.47 | 104.5 |
| #5 | 148.71 | 143.04 | 104.0 | 131.62 | 131.82 | 99.8 | 119.19 | 116.61 | 102.2 |
| #6 | 151.34 | 157.69 | 96.0 | 58.25 | 30.26 | 192.5 | 26.62 | 26.10 | 102.0 |
| #7 | 92.37 | 93.92 | 98.3 | 40.92 | 19.48 | 210.0 | 69.25 | 58.72 | 117.9 |
| #8 | 49.66 | 36.84 | 134.8 | 44.16 | 41.45 | 106.5 | 108.70 | 98.67 | 110.2 |
| #9 | 117.84 | 118.05 | 99.8 | 123.46 | 156.69 | 78.8 | 123.70 | 75.84 | 163.1 |
| #10 | 158.04 | 202.50 | 78.0 | 177.40 | 174.10 | 101.9 | 83.49 | 140.26 | 59.5 |

Per tant demostrant entre aquest cas i el de les cerveses, la capacitat de la BioET per dur a terme de manera satisfactòria la resolució i quantificació de mescleres ternàries de diferents compostos fenòlics, tant en mostres tamponades com en mostres fortificades; provant així que la matriu tant del vi com de la cervesa no representen cap problema. A més, essent aplicable a la determinació de mescleres més complexes o de compostos diferents simplement a partir del reentrenament del sistema utilitzant un conjunt de mostres apropiat.

4.5. Comparativa de les diferents estratègies de tractament del senyal

Finalment, per tal de donar resposta a un dels objectius de la present tesi que consistia en el desenvolupament de nous mètodes de processament del senyal i la cerca de la millor metodologia en cada cas, es va creure oportú realitzar una macrocomparativa que englobes les diferents estratègies utilitzades al llarg de la present tesi, comparant a més amb els diferents mètodes de modelatge clàssics. D'aquesta manera, l'estudi presenta els resultats obtinguts de la comparació de cinc mètodes diferents per al pretractament del senyal (CI, Int, PCA, FFT i DWT) i de quatre mètodes de modelatge (PCR, PLS, nPLS i ANN), resultant en un total de 17 combinacions diferents avaluades.

En aquest sentit, aquest estudi representa per un costat, un resum de les diferents estratègies seguides a la present tesi, proporcionant una breu introducció teòrica, explicant el fonament i les diferències de cada mètode (§1.3 i §3.4); mentre que per l'altra, representa l'aplicació de les diferents estratègies en un mateix cas, proporcionant així la comparació numèrica de tots ells, facilitant l'extracció de conclusions que permetin l'elecció del mètode més adequat en cada cas. No obstant, cal tenir en compte, que aquest estudi representa només l'aplicació en un únic cas, i que per poder extreure conclusions més definitives caldria estendre el mateix estudi a altres situacions: diferents analits, sensors, etc.

i. Cas d'estudi

Així doncs, per tal d'il·lustrar les diferències entre els diferents mètodes de preprocessament i de modelatge, la seva eficàcia va ser avaluada en la discriminació i quantificació de mesclades de compostos fenòlics amb les dades recollides a partir de la BioET (§4.3.2). Aquest cas fou escollit com a cas d'estudi degut principalment a dos factors: el tipus de disseny escollit en la preparació de les mostres i els bons resultats aconseguits.

És a dir, en primer lloc, el disseny de les mostres es feu de manera que quedava ben definit el domini experimental del model, i la divisió *train/test* en el procés de modelatge; a més, aquest proporcionava una situació suficientment complexa (solapament dels senyals, resposta creuada dels sensors, etc.) per a requerir l'ús del concepte de llengua electrònica. En segon lloc, calia garantir que el cas d'estudi era

susceptible de ser modelat, és a dir, que la informació de partida era suficientment rica i contenia la suficient relació amb les mostres per permetre la correcta identificació mitjançant les eines adequades; avaluant així la capacitat d'extracció de cada model.

Per tant, el cas d'estudi va ser la resolució de mesclades de tres fenols formades pels àcids ferúlic, gàl·lic i sinàpic. El conjunt de mostres estava format per un total de 42 mostres, amb un interval de concentració per a les tres espècies de 0 a 200 μ M per a cada un. Aquest va ser dividit en dos subconjunts de dades: el subconjunt d'entrenament format per 27 mostres (64%), distribuïdes en un disseny cúbic i utilitzat per establir el model de resposta (§3.2, Figura 3.2); més 15 mostres addicionals (36%) pel subconjunt de test, distribuïdes aleatòriament al llarg del domini experimental, i usades per avaluar la capacitat de predicció dels models. D'aquesta manera, el conjunt de mostres va ser mesurat emprant la matriu de 4 sensors de la BioET, registrant un voltamperograma cíclic per a cada un dels sensors.

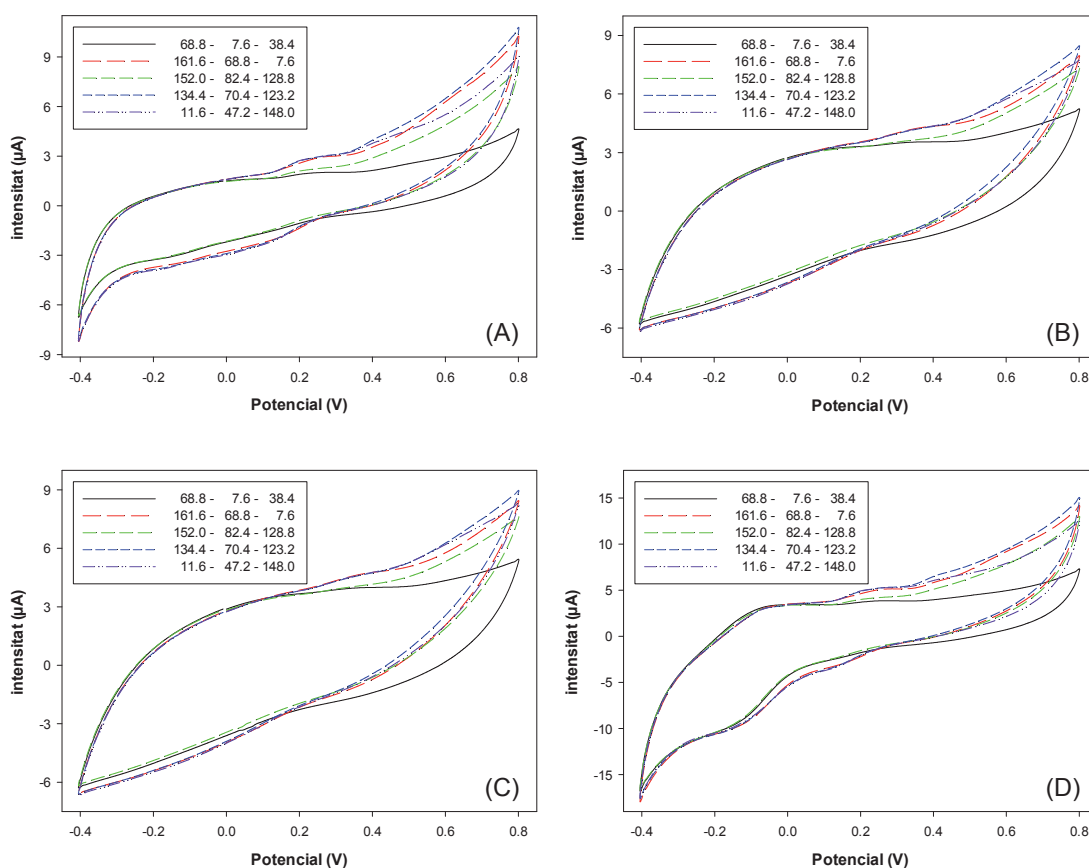


Figura 4.43. Exemple dels diferents voltamperogrames obtinguts per certes mesclades ternàries arbitràries amb els (bio)sensors: (A) GEC, (B) Tyr, (C) Lac i (D) Cu. Els valors de la llegenda corresponen a les concentracions en μ M de cada un dels compostos fenòlics: àcids ferúlic, gàl·lic i sinàpic, respectivament.

Exemples dels diferents senyals obtinguts per a algunes de les barreges de compostos fenòlics es mostren a la Figura 4.28 i Figura 4.43. Com es pot observar, diferents senyals són obtinguts pels diferents tipus de (bio)sensors, amb respostes voltamperomètriques complexes i superposades que contenen informació global sobre els compostos analitzats al llarg de tot el voltamperograma. No obstant això, la quantitat d'informació i la dimensionalitat dels senyals (4 sensors x 268 potencials aplicats x 42 mostres) dificulten l'obtenció d'un model de resposta. Per tant, per tal d'aprofitar plenament tota la informació obtinguda de cada voltamograma, una etapa de compressió del senyal és requerit.

ii. *Preprocessament del senyal - Compressió*

El primer paràmetre considerat va ser el grau de compressió assolit mitjançant cada mètode, així com la representativitat dels coeficients extrets de les dades inicials en els casos que fos possible avaluar aquest factor. En tots els casos, els paràmetres dels diferents algorismes van ser ajustats per tal d'assegurar que pràcticament tota la informació present en les dades originals era preservada, tal com es demostra pels casos del PCA, la FFT i la DWT, on aquest factor va poder ser objectivat. Així doncs, la Taula 4.18 resumeix els resultats obtinguts després d'aplicar les diferents estratègies de compressió del senyal.

Taula 4.18. Nombre de característiques (*features*) utilitzades com a inputs en l'etapa de modelatge i grau de compressió assolit mitjançant els diferents mètodes de preprocessament del senyal.

| <i>Mètode</i> | <i>Número de features</i> | <i>Número d'inputs</i> | <i>Grau de compressió</i> | <i>Variància acumulada</i> | R^2 | f_c |
|-----------------|---------------------------|------------------------|---------------------------|----------------------------|-------|-------|
| <i>CI-ANN</i> | - | 25 | 97.7% | - | - | - |
| <i>Integral</i> | 11 | 44 | 95.9% | - | - | - |
| <i>PCA</i> | 4 | 16 | 98.5% | 99.9% | - | - |
| <i>FFT</i> | 16 | 64 | 94.0% | - | 0.999 | 0.975 |
| <i>DWT</i> | 23 | 92 | 91.4% | - | 0.997 | 0.969 |

El primer aspecte a destacar, és l'elevat grau de compressió assolit en tots els casos, essent superior al 90%; valor que permetrà reduir tant el temps d'entrenament com la complexitat del model, així com permetre l'obtenció de models amb millor capacitat de generalització. Més enllà del grau de compressió proporcionat per cada mètode, també cal tenir en compte els avantatges i desavantatges de cada mètode i conseqüentment, depenen del cas, seleccionar el mètode més apropiat.

Per exemple, l'ús del CI (§3.4.1.2iv) és un procediment relativament lent, però a la vegada, és una forma molt útil d'identificar els trets més significatius en el perfil de

resposta; és a dir, identificar els sensors que proporcionen informació més significativa, i encara més, quins són els potencials que estan associats amb els processos electroquímics que tenen lloc. Malauradament, el seus principals desavantatges són el fet que no proporciona una solució única, ja que el conjunt de variables seleccionades vindrà determinada per la topologia de la xarxa utilitzada, essent aquesta la que determinarà l'assoliment d'un punt òptim.

Com a exemple, la Figura 4.44 il·lustra el procés de reducció del número d'*inputs* del model i l'error assolit per aquest en cada iteració. Tal com es pot observar, en les primeres iteracions hi ha una major reducció del número de variables a utilitzar, i una major variació en l'error assolit al model degut a la dificultat de construir un model robust amb un nombre tan elevat d'entrades; assolint una situació més estable en l'etapa final del procés. No obstant, cal recordar, que un cop finalitzada l'etapa de selecció de variables, al igual que en la resta de casos cal dur a terme una optimització de l'arquitectura del model utilitzant el conjunt d'*inputs* reduït.

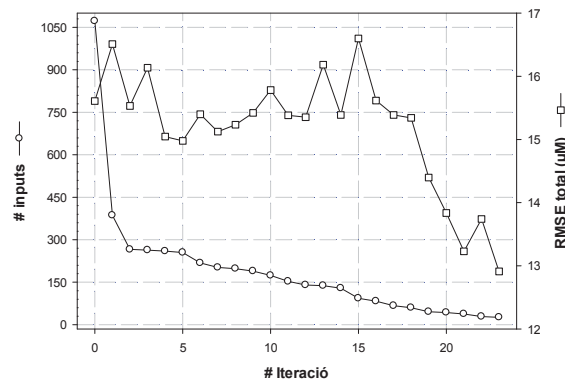


Figura 4.44. Procés de selecció i reducció del número d'inputs mitjançant el mètode CI-ANN. En cada iteració s'ajusta un model, es calcula el seu RMSE pel subconjunt de test, i es decarten els *inputs* en els que el model ha assignat un menor pes, repetint aquest procés fins a assolir el grau de reducció desitjat o un increment en el RMSE del model durant diverses iteracions successives.

En el cas de la *integral de finestra seccionada* (§3.4.1.4), el nombre de seccions és arbitrari, per tant, en el nostre cas, es va escollir un valor de $k=11$, tot i que fins i tot amb menys coeficients va ser possible assolir bons resultats (Figura 4.29). No obstant, aquest valor va ser preferit ja que era més comparable al dels altres mètodes en què es va utilitzar un nombre similar de coeficients, i a més, aquest proporcionava millors resultats. En aquest sentit, i segons el cas, aquest paràmetre pot ser variat amb la finalitat de millorar el grau de detall en la representació del senyal original o per reduir el nombre d'*inputs* del model.

L'ús de PCA és una altra manera simple de reduir l'alta dimensionalitat del senyal d'entrada, tot i que aquest mètode no es recomana en els casos en que hi hagi indicis d'alguna deriva en les mesures ja que aquest és molt sensible a aquesta, proporcionant així coeficients més relacionats amb la deriva que amb el contingut analític de les mostres [35]. A més, el seu ús no es recomanat en tots els casos, ja que aquest es centra en la màxima variació entre les mostres, per tant, donant menys importància a les petites diferències que en alguns casos poden ser significatives. És a dir, si en el nostre conjunt tenim un cas que és molt diferent dels altres, la variància d'aquest no ens permetrà veure les diferències més petites entre la resta de les mostres (en termes més col·loquials, el bosc no ens deixa veure els arbres).

Finalment, tot i no ser tan palès en aquest cas, cal tenir en compte que en respostes complexes i altament superposades, mètodes tals com la FFT o la DWT són més eficients [36]. La raó es deu al fet que els coeficients obtinguts són una representació de tot el senyal voltamperomètric, el qual a la vegada és filtrat (eliminant el soroll) durant el procés d'extracció de característiques. Entre ambdós, la DWT presenta com a avantatge la seva major polivalència; mentre que la FFT és especialment eficaç en la compressió de senyals de tipus sinusoidal.

Per últim, també cal tenir en compte la riquesa del conjunt de dades utilitzat en aquest estudi; a saber, l'ús d'una matriu de (bio)sensors amb alta selectivitat pels compostos analitzats i l'ús d'un cas model; proporcionant així una resposta altament relacionada amb els analits i sense presència d'interferents. Aquest fets contribueixen a explicar els excel·lents resultats assolits en el cas del PCA (variància acumulada ca. 100% amb tan sols 4 PCs).

Precisament en aquests casos, on la informació relacionada amb la mostra pot ser més fàcilment ubicada en regions específiques del voltamperograma, alguns grups de recerca afirmen que l'ús de només una part de la corba és suficient, donat que en el cas d'un procés electroquímic reversible, la mateixa informació es pot obtenir tant de l'ona catòdica com de l'anòdica [37]. En aquest sentit, els mètodes de preprocessament descrits podrien ser aplicats centrant-se només en una part d'aquesta resposta, tant en la part catòdica com l'anòdica, o fins i tot en una regió més petita.

iii. Processament de la informació - Modelatge

Un cop obtinguts els diferents coeficients resultat d'aplicar les diferents estratègies de compressió del senyal a les dades en cru, el següent pas és la utilització d'aquests

com a *inputs* dels diferents mètodes de modelatge. Addicionalment, en el cas dels mètodes lineals (PCR, PLS i nPLS), aquests també varen ser ajustats utilitzant la resposta voltamperomètrica en cru; això no va ser possible en el cas de la ANN ja que aquest mètode és més sensible al número d'*inputs* utilitzat, obtenint models menys robustos [38]. No obstant, en el seu lloc, és va fer ús del mètode CI, el qual es basa en l'ús de la ANN com una eina de selecció de variables enlloc de com a eina de quantificació (§3.4.1.2iv). D'altra banda, donat que l'ús del PCA com a eina de preprocessament és una etapa implícita en el cas del PCR, aquest no va ser considerat; de forma anàloga, i donades les similituds entre el PCR i el PLS si aquest fos utilitzat, el seu ús també va ser descartat.

D'aquesta manera es va procedir a l'ajust dels diferents models utilitzant els diferents coeficients obtinguts com a *inputs* i entrenant aquests amb les dades del subconjunt d'entrenament, utilitzant les dades del subconjunt de test per evaluar-ne la seva eficàcia i seleccionar el model òptim en cada cas (i.e. arquitectura de la ANN, o número de PCs o variables latents requerides).

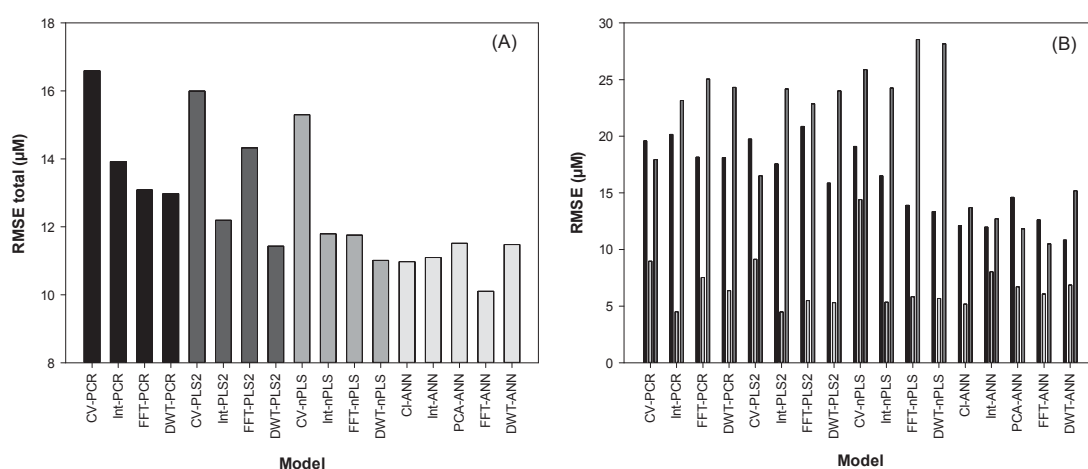


Figura 4.45. Valors de RMSE obtinguts pel subconjunt de test i cadascun dels models avaluats. (A) RMSE total i (B) RMSE individual per cadascun dels compostos analitzats: àcids (■) ferúlic, (■) gàl·lic i (■) sinàpic.

Un cop finalitzada l'etapa de modelatge, els RMSEs per cadascun dels models construïts van ser calculats i comparats entre ells (Figura 4.45). A partir de la figura es pot observar com els majors errors van ser assolits en els casos en que es va utilitzar la resposta voltamperomètrica en cru, amb resultats similars per la resta de casos; confirmant i demostrant així els avantatges descrits anteriorment (§3.4.1). Tanmateix, es va poder observar com en general, els millors resultats s'obtingueren en utilitzar la DWT com a mètode de preprocessament per les diferents eines de modelatge. D'altra

banda, cal destacar els bons resultats que es van obtenir emprant la *integral de finestra seccionada*, el qual representa l'aplicació d'un mètode de preprocessament nou i simple, amb elevades perspectives.

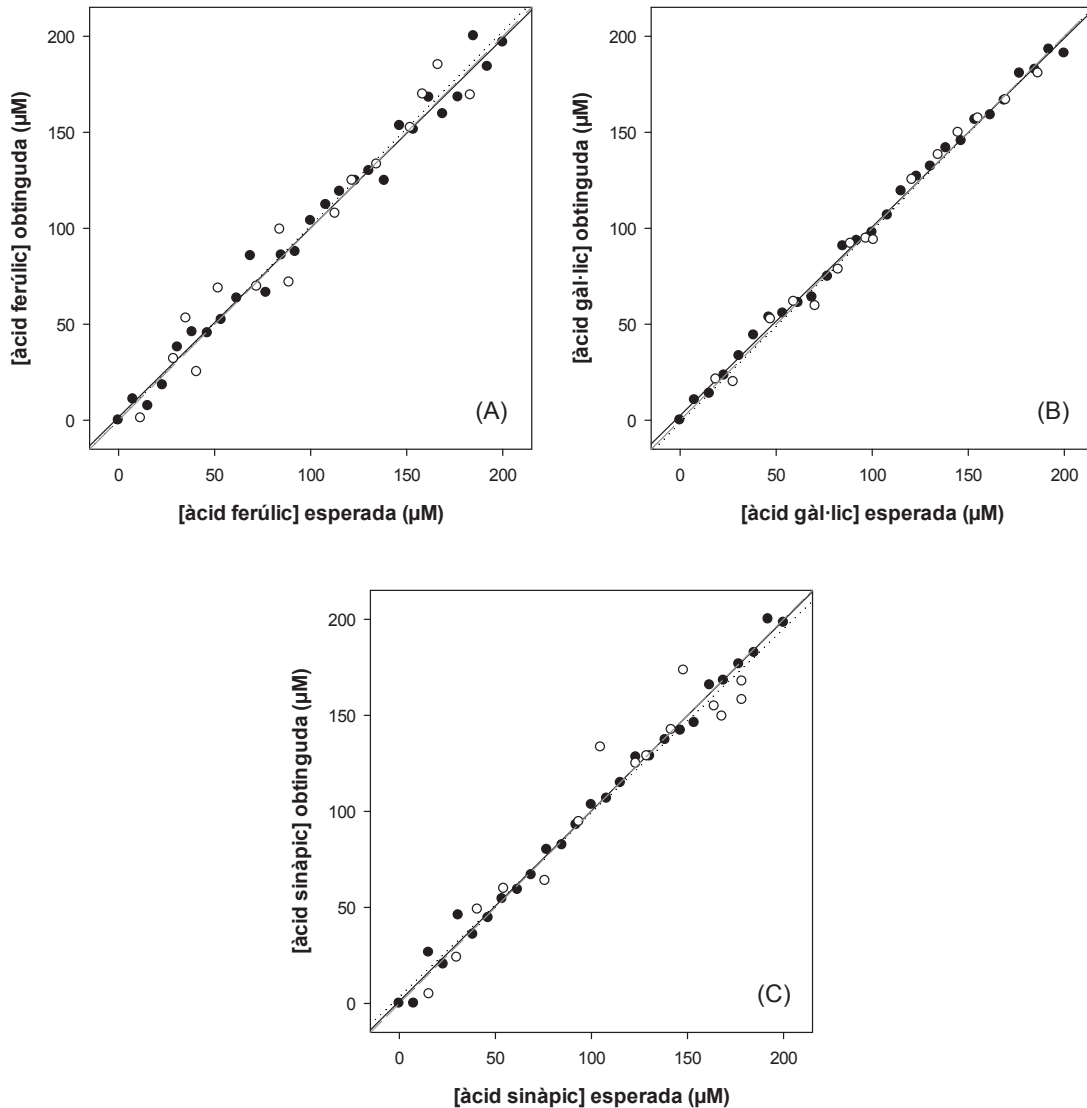


Figura 4.46. Capacitat de modelatge del model FFT-ANN. Gràfics de comparació de les concentracions obtingudes vs. les esperades pels àcids (A) ferúlic, (B) gàl·lic i (C) inàpic, tant pel subconjunt d'entrenament (●, línia sòlida) com pel subconjunt de test (○, línia de punts). La línia discontinua correspon a la diagonal teòrica.

Adicionalment, els gràfics de comparació de concentració predita vs. esperada van ser construïts pels tres compostos fenòlics i per cada un dels models, tant pel conjunt d'entrenament com pel de test. Calculant a continuació els valors de pendent, ordenada a l'origen i els coeficients de correlació de les comparacions, i utilitzant aquests per tal d'ajudar en la tria de la millor configuració. Un exemple d'aquests gràfics es mostra a la Figura 4.30 i la Figura 4.46. Es pot veure com s'obté una tendència satisfactòria en tots

els casos, amb línies de regressió gairebé indistingibles de les teòriques. També, com és habitual en els models de calibratge multivariable, assolint menor dispersió i incerteses per als subconjunts d'entrenament.

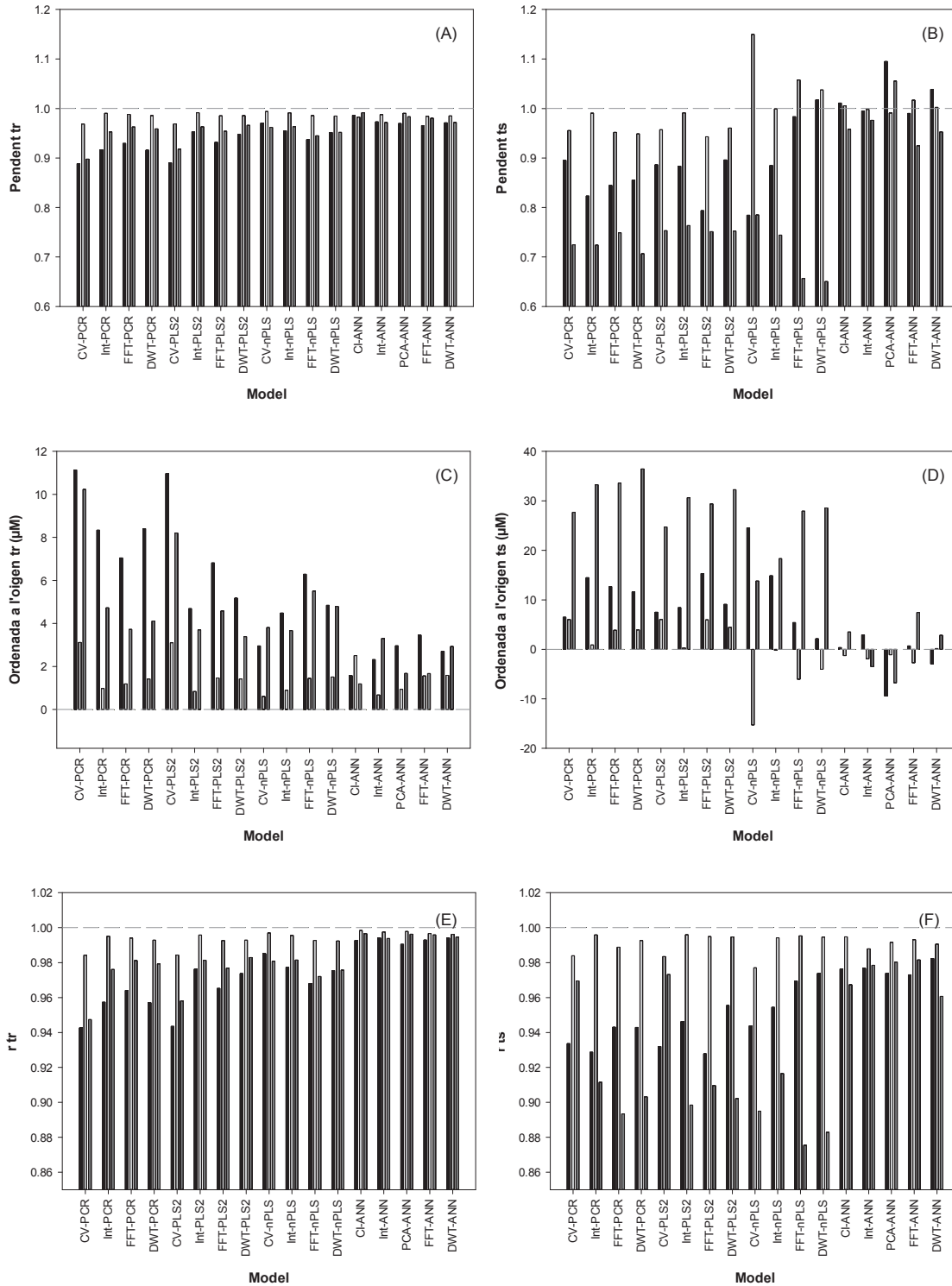


Figura 4.47. Paràmetres de les rectes de regressió ajustades dels gràfics de comparació entre concentracions obtingudes vs. esperades pels diferents models considerats i cadascun dels tres compostos estudiats (àcids (■) ferúlic, (■) gál·lic i (■) sinàpic) tant pel subconjunt d'entrenament com de test, respectivament. (A,B) pendents, (C,D) ordenades a l'origen i (E,F) coeficients de correlació.

D'aquesta manera, els gràfics de comparació van ser construïts per tots els casos; calculant els diferents paràmetres de les regressions, els quals es troben resumits a la Figura 4.47. Com ja s'ha comentat, es pot observar com en tots els casos s'obté una millor tendència pel subconjunt d'entrenament, on els valors de pendents, dels coeficients de correlació i d'ordenada a l'origen són més propers als valors teòrics (1, 1 i 0 respectivament), presentant menors diferències entre els diferents models avaluats. No obstant, com és sabut, la capacitat del model no pot ser avaluada a partir dels resultats del conjunt d'entrenament ja que aquests no presenten una valoració objectiva. Per tant, cal fer ús del subconjunt de test.

A pesar dels baixos i similars valors de RMSE obtinguts a partir dels diferents models construïts pel subconjunt de test, majors diferències van ser assolides en els gràfics de comparació. En primer lloc, es va poder veure que l'àcid sinàpic és el que presentava els pitjors resultats en tots els casos, implicant que el seu senyal es troba emmascarat pels altres compostos. A pesar d'això, es va poder veure que en tots els casos la correlació era encara bona ($r > 0.875$ en el pitjor dels casos), però trobant la principal diferència en el cas dels pendents i interseccions (valors tan distants com ca. 0.65 i 36.5 μM , respectivament), on s'obtingué un clar descens particularment amb els models lineals. En aquest sentit, les ANNs són les que presenten els millors resultats amb pendents superiors a 0.92 i ordenades a l'origen inferiors a 9.5 μM en el pitjor dels casos (PCA-ANN).

Per tant, per la selecció del model òptim, tant el valor de RMSE com els paràmetres calculats a partir dels gràfics de comparació han de ser considerats; presentant així el model òptim un baix RMSE, el que significa que no hi ha diferències entre els valors pronosticats pel model i els reals; i valors de correlació i del pendent propers a 1, amb un valor de l'ordenada a l'origen proper a 0, descartant així qualsevol error del model.

En aquest cas, els millors resultats globals varen ser obtinguts utilitzant la FFT-ANN (Figura 4.46); mentre que en el cas dels mètodes lineals, els millors resultats varen ser obtinguts utilitzant la DWT com a eina de preprocessament del senyal. Tal com s'ha indicat, aquesta millora dels resultats deriven del fet que ambdós mètodes proporcionen una bona compressió del senyal, preservant la informació rellevant i utilitzant el registre complet, el qual a la vegada és filtrat. Tanmateix, l'obtenció de millors resultats amb els models ANN que amb els mètodes lineals, pot ser explicat donat la capacitat de les ANNs d'ajustar casos tant lineals com no lineals i a la seva major flexibilitat en l'arquitectura, facilitant la seva adaptació a casos més complexos.

Finalment, remarcar també els bons resultats obtinguts tant en el cas de la *integral de finestra seccionada*, el qual a pesar de la seva senzillesa ha demostrat ser un mètode eficaç, com del CI. El primer, presentat com a principal avantatge la seva simplicitat, i com a principal desavantatge l'arbitrarietat en la selecció del número de coeficients a emprar, la qual requereix de coneixement o de la realització d'alguns assajos previs. Mentre que el darrer, tot i presentar com a principals inconvenients que el conjunt d'*inputs* no és únic i que dependrà de la topologia de la ANN, a més del major temps requerit per completar el procés d'extracció; presenta com a gran avantatge el fet de permetre la identificació dels sensors i punts del voltamperograma que més contribueixen a la discriminació i quantificació dels compostos sota estudi. I conseqüentment, permetent la selecció de quins sensors cal preservar en estudis futurs o la modificació dels paràmetres de mesura.

Per últim, recordar que aquesta comparativa no busca establir la cerca del millor mètode en tots els casos (o la gran majoria), la qual cosa requeriria un estudi molt més extensiu, sinó il·lustrar la capacitat de cada un d'ells i proporcionar elements que permetin ajudar en la decisió de l'eina quimiomètrica més adequada en cada cas.

4.6. Bibliografia

- [1] S. Buxaderas, E. López-Tamames, WINES | Production of Sparkling Wines, in: C. Editor-in-Chief: Benjamin (Ed.) Encyclopedia of Food Sciences and Nutrition (Second Edition), Academic Press, Oxford, **2003**, pp. 6203.
- [2] Ministerio de Agricultura, Pesca y Alimentación, in: BOE-A-1991-28079, Secretaria General Técnica, Madrid, 20 November **1991**, pp. 37587.
- [3] J. Bosch Fusté, in: Departament de Nutrició i Bromatologia, Universitat de Barcelona, **2007**.
- [4] J.M. Gutiérrez, L. Moreno-Barón, M.I. Pividori, S. Alegret, M. del Valle, A voltammetric electronic tongue made of modified epoxy-graphite electrodes for the qualitative analysis of wine, *Microchim. Acta* **2010**, *169*, 261.
- [5] Z. Boger, Selection of quasi-optimal inputs in chemometrics modeling by artificial neural network analysis, *Anal. Chim. Acta* **2003**, *490*, 31.
- [6] L. Moreno-Barón, R. Cartas, A. Merkoçi, S. Alegret, M. del Valle, L. Leija, P.R. Hernandez, R. Muñoz, Application of the wavelet transform coupled with artificial neural networks for quantification purposes in a voltammetric electronic tongue, *Sensor Actuat. B-Chem.* **2006**, *113*, 487.
- [7] K. Booksh, J.M. Henshaw, L.W. Burgess, B.R. Kowalski, A 2nd-order standard addition method with application to calibration of a kinetics-spectroscopic sensor for quantitation of trichloroethylene, *J. Chemometr.* **1995**, *9*, 263.
- [8] Office International de la Vigne et du Vin (OIV), Recueil des méthodes internationales d'analyse des vins et des moûts, OIV, Paris, **2005**.

- [9] P. Ribéreau-Gayon, Y. Glories, A. Maujean, D. Dubourdieu, Phenolic Compounds, in: Handbook of Enology, John Wiley & Sons, Ltd, Chichester (UK), **2006**, pp. 141.
- [10] R.S. Jackson, Wine tasting: a professional handbook, Academic Press, London, **2002**.
- [11] C. Sánchez-Moreno, J.A. Larrauri, F. Saura-Calixto, Free radical scavenging capacity and inhibition of lipid oxidation of wines, grape juices and related polyphenolic constituents, *Food Res. Int.* **1999**, *32*, 407.
- [12] D. De Beer, J.F. Harbertson, P.A. Kilmartin, V. Roginsky, T. Barsukova, D.O. Adams, A.L. Waterhouse, Phenolics: A comparison of diverse analytical methods, *Am. J. Enol. Viticult.* **2004**, *55*, 389.
- [13] E. Elejalde Caravaca, Application of spectroscopic methods to the study of chromatic characteristics of polyphenic components in wines., *Formula* **1999**, *5*, 39.
- [14] M. Del Alamo Sanza, I. Nevares Domínguez, Wine aging in bottle from artificial systems (staves and chips) and oak woods: Anthocyanin composition, *Anal. Chim. Acta* **2006**, *563*, 255.
- [15] A.J. Buglass, D.J. Caven-Quantrill, N.J. Fuller, J.A. Hudson, C.G. Lee, S.H. Lee, M. McKay, Handbook of Alcoholic Beverages: Technical, Analytical and Nutritional Aspects, John Wiley & Sons, Ltd, **2011**.
- [16] W. Novakowski, M. Bertotti, T.R.L.C. Paixão, Use of copper and gold electrodes as sensitive elements for fabrication of an electronic tongue: Discrimination of wines and whiskies, *Microchem. J.* **2011**, *99*, 145.
- [17] P. Ribéreau-Gayon, Y. Glories, A. Maujean, D. Dubourdieu, Handbook of Enology, Volume 2, The Chemistry of Wine Stabilization and Treatments, 2nd Edition ed., Wiley-Blackwell, London, **2000**.
- [18] J. Sourisseau, Étude chimique de la distillation du cognac, *Bull. Union Physiciens* **2002**, *96*, 881.
- [19] M. Nelson, The Barbarian's Beverage: A History of Beer in Ancient Europe, Taylor and Francis, Abingdon, Oxon, **2008**.
- [20] S. Snyder, The brewmaster's bible: the gold standard for home brewers, William Morrow Cookbooks, New York, NY, **1997**.
- [21] E.J. Lodolo, J.L.F. Kock, B.C. Axcell, M. Brooks, The yeast *Saccharomyces cerevisiae*—the main character in beer brewing, *FEMS Yeast Research* **2008**, *8*, 1018.
- [22] A. Rudnitskaya, E. Polshin, D. Kirsanov, J. Lammertyn, B. Nicolai, D. Saison, F.R. Delvaux, F. Delvaux, A. Legin, Instrumental measurement of beer taste attributes using an electronic tongue, *Anal. Chim. Acta* **2009**, *646*, 111.
- [23] D. Kirsanov, O. Mednova, V. Vietoris, P.A. Kilmartin, A. Legin, Towards reliable estimation of an “electronic tongue” predictive ability from PLS regression models in wine analysis, *Talanta* **2012**, *90*, 109.
- [24] J. Garrido, F. Borges, Wine and grape polyphenols—A chemical perspective, *Food Res. Int.* **2011**, *44*, 3134.
- [25] E.N. Frankel, A.L. Waterhouse, P.L. Teissedre, Principal Phenolic Phytochemicals in Selected California Wines and their Antioxidant Activity in Inhibiting Oxidation of Human Low-Density Lipoproteins, *J. Agr. Food Chem.* **1995**, *43*, 890.
- [26] V.L. Singleton, J.A. Rossi, Colorimetry of total phenolics in grapes and wine with phosphomolybdic -phosphotungstic acid reagents, *Am. J. Enol. Viticult.* **1965**, *16*, 144.

- [27] E. Mataix, M.D.L. de Castro, Simultaneous (or sequential) determination of the total polyphenol index (or I-280) and density in wines by flow injection, *Analyst* **2001**, *126*, 251.
- [28] A.L. Waterhouse, Determination of Total Phenolics, in: R.E. Wrolstad (Ed.) *Current Protocols in Food Analytical Chemistry*, John Wiley & Sons, Inc., New York, US, **2001**.
- [29] J. Orozco, G. Suarez, C. Fernandez-Sanchez, C. McNeil, C. Jimenez-Jorquera, Characterization of ultramicroelectrode arrays combining electrochemical techniques and optical microscopy imaging, *Electrochim. Acta* **2007**, *53*, 729.
- [30] R. Solná, P. Skládal, Amperometric flow-injection determination of phenolic compounds using a biosensor with immobilized laccase, peroxidase and tyrosinase, *Electroanal.* **2005**, *17*, 2137.
- [31] I. Ignat, I. Volf, V.I. Popa, A critical review of methods for characterisation of polyphenolic compounds in fruits and vegetables, *Food Chem.* **2011**, *126*, 1821.
- [32] A.S. Arribas, M. Martínez-Fernández, M. Chicharro, The role of electroanalytical techniques in analysis of polyphenols in wine, *TrAC-Trend. Anal. Chem.* **2012**, *34*, 78.
- [33] V. Rastija, G. Srećnik, S. Marica Medic, Polyphenolic composition of Croatian wines with different geographical origins, *Food Chem.* **2009**, *115*, 54.
- [34] M. Marina, Comparing clusterings—an information based distance, *J. Multivariate Anal.* **2007**, *98*, 873.
- [35] E. Tønning, S. Sapelnikova, J. Christensen, C. Carlsson, M. Winther-Nielsen, E. Dock, R. Solna, P. Skladal, L. Nørgaard, T. Ruzgas, J. Emnéus, Chemometric exploration of an amperometric biosensor array for fast determination of wastewater quality, *Biosens. Bioelec.* **2005**, *21*, 608.
- [36] L. Moreno-Barón, R. Cartas, A. Merkoçi, S. Alegret, J.M. Gutiérrez, L. Leija, P.R. Hernández, R. Muñoz, M. del Valle, Data compression for a voltammetric electronic tongue modelled with artificial neural networks, *Anal. Lett.* **2005**, *38*, 2189.
- [37] C. Apetrei, I.M. Apetrei, J.A.D. Saja, M.L. Rodriguez-Mendez, Carbon Paste Electrodes Made from Different Carbonaceous Materials: Application in the Study of Antioxidants, *Sensors* **2011**, *11*, 1328.
- [38] D.R. Hush, B.G. Horne, Progress in supervised neural networks, *IEEE Signal Processing Magazine* **1993**, *10*, 8.

5. Conclusions

5. Conclusions

Durant la realització del present treball s'ha comprovat l'àmplia aplicabilitat de les llengües electròniques en el món de les begudes alcohòliques, tant en la classificació d'aquestes en funció de diferents criteris com la quantificació de diferents paràmetres d'interès analític; presentant de forma gradual en complexitat creixent la versatilitat d'aquests sistemes. Demostrant així, la viabilitat d'utilitzar aquestes eines com a alternativa als mètodes convencionals, sobretot en anàlisis de cribratge (*screening analysis*), amb grans avantatges sobre els darrers com poden ser el seu baix cost, la no-necessitat de pretractament de la mostra, la seva ràpida resposta o la capacitat de dur a terme tasques biomimètiques reproduint el sentit humà del gust.

Tot seguit, s'especifiquen les conclusions particulars que es poden extreure d'aquest treball, agrupades en funció de cada una de les aplicacions concretes i seguint el mateix esquema establert en els capítols precedents. Finalment, i com a resum global i comparació dels diferents mètodes utilitzats, s'exposen de forma breu les conclusions més generals i perspectives futures.

Aplicació en el món del cava

1. S'ha assolit el desenvolupament d'una llengua electrònica capaç de dur a terme la classificació del cava a partir de dos dels factors utilitzats dins el sector; a saber, a partir de la quantitat de sucre afegida durant la seva producció i segons el grau d'envelliment o criaça.
2. A partir de l'ús del mètode de l'addició estàndard de segon ordre (SOSAM), ha estat possible demostrar la resposta de la llengua electrònica en front al sucre, i predir la quantitat afegida d'aquest en els caves durant la seva producció; demostrant la capacitat de les llengües electròniques, ja que aquesta no incorporava cap sensor selectiu a aquest. A més, representant l'aplicació del mètode SOSAM en llengües electròniques per primer cop, fins on arriba el nostre coneixement.
3. Aquesta resposta ha estat confirmada a partir de la construcció d'un model neuronal capaç de dur a terme la quantificació del sucre total present en les mostres de cava mitjançant comparació amb el mètode de referència i sense necessitat de fer ús del mètode de l'addició estàndard.
4. El sistema proposat representa l'obtenció d'una nova eina d'anàlisi, capaç de satisfer els requeriments del sector en quan a rapidesa, senzillesa, fiabilitat i baix cost, que

dóna resposta a una demanda d'aquest davant la manca de mètodes capaços de determinar el temps de criaça dels caves. És a dir, la distinció entre caves *joves*, *Reserva* i *Gran Reserva*.

Aplicació en el món dels brandis

5. S'ha descrit per primera vegada l'aplicació d'una llengua electrònica voltamperomètrica per l'anàlisi de destil·lats de vi, els quals són el producte base en la producció dels brandis.
6. En aquest camp s'ha demostrat la capacitat del sistema proposat per dur a terme la classificació d'aquests en funció de la seva qualitat de tast, tal com realitzaria un "sommelier".
7. A més, s'ha aconseguit dur a terme la detecció i distinció de les mostres en les que s'han utilitzat mètodes per tal d'accelerar i simular el procés d'envelliment dels destil·lats en bóta.
8. Tal com passa en el vi, s'ha observat que tot i el procés de destil·lació dut a terme, és possible observar una variació en les mostres segons l'anyada d'aquestes.
9. Finalment, a nivell semi-quantitatiu, ha estat possible dur a terme la detecció de compostos indicadors de diferents defectes en els destil·lats segons els criteris utilitzats en el sector: bo, lleuger defecte i defecte.

Aplicació en el món de la cervesa

10. El sistema proposat ha estat capaç de distingir tant entre diferents varietats de cervesa (i.e. lager, stout i IPA) com en el procés d'elaboració d'aquestes (i.e. industrial o artesanal), així com la distinció dels dos efectes simultàniament.
11. S'ha assolit la quantificació individual i simultània de tres dels compostos fenòlics majoritaris presents en les cerveses; a saber, els àcids ferúlic, gál·lic i sinàpic.

Aplicació en el món del vi

12. S'han avaluat els beneficis derivats de la incorporació de biosensors en la matriu de sensors de la llengua electrònica mitjançant comparació dels resultats d'una llengua electrònica amb els obtinguts amb una llengua bioelectrònica.
13. Ambdues matrius de sensors han estat aplicades a la discriminació qualitativa de diferents compostos fenòlics presents en el vi, demostrant la capacitat d'aquestes per dur a terme la seva discriminació entre fins a deu compostos diferents.

14. Amb motiu dels bons resultats assolits en la discriminació dels diferents tipus de compostos fenòlics, s'ha procedit a la quantificació de mescles ternàries d'aquests mitjançant l'ús de la llengua bioelectrònica; utilitzant com a situació model, mescles de catecol, catequina i àcid cafeic.
15. Paral·lelament, s'ha avaluat la capacitat d'ambdós sistemes per dur a terme la quantificació del nivell de compostos fenòlics totals presents en el vi d'acord amb l'índex de Folin-Ciocalteu i l'índex I_{280} , els dos índexs més utilitzats.

Utilització de les diferents estratègies de tractament del senyal

16. S'ha realitzat una extensa comparativa a nivell quantitatiu dels resultats obtinguts mitjançant la combinació de diferents estratègies de preprocessament del senyal i modelatge d'aquest.
17. S'ha proposat la utilització d'un nou mètode de preprocessament basat en el càlcul de l'àrea sota la corba voltamperomètrica, presentant com a principals avantatges la seva senzillesa i eficàcia.
18. S'ha proposat l'ús de la transformada ràpida de Fourier com a eina alternativa i eficaç per a la compressió de la resposta voltamperomètrica, una eina rarament utilitzada. A més, s'ha demostrat la seva capacitat per dur a terme aquesta tasca a partir de la compressió i reconstrucció del senyal original.
19. S'ha demostrat que l'ús d'una etapa de preprocessament del senyal és una estratègia eficaç per tal de millorar la capacitat de predicció i generalització del sistema, tant en mètodes lineals com no lineals.
20. S'ha demostrat l'elevada adaptabilitat dels models basats en xarxes neuronals per a la resolució de problemes complexos en comparació amb mètodes tals com el PLS.
21. S'han aplicat noves eines quimiomètriques, tals com l'anàlisi discriminant lineal (LDA) o algorismes de selecció de variables tals com el *Causal Index* (CI), per primera vegada dins el nostre grup de recerca.

Conclusions generals

22. S'ha demostrat la viabilitat de les llengües electròniques com a eina de tast artificial, proporcionant una resposta equivalent i objectiva a la d'un "sommelier"; convertint-les en una eina molt interessant com a mètode de cribratge.
23. S'ha demostrat la incorporació de diferents modificadors en sensors compòsits com una manera fàcil i senzilla d'obtenir sensors amb una resposta diferenciada i

- creuada; característiques molt interessants per al desenvolupament de sistemes basats en llengües electròniques.
24. S'ha demostrat els avantatges derivats de la incorporació de biosensors en la matriu de la llengua electrònica; és a dir, l'ús de llengües bioelectròniques.
 25. S'ha demostrat l'ús de partícules de coure com a modificador útil per al desenvolupament de sensors per a la determinació de compostos fenòlics; en el que podria ser un mecanisme catalític similar al de les fenol oxidases, les quals tenen àtoms de coure en els seus centres catalítics.
 26. La interacció amb altres grups d'investigació, ha permès la transferència de coneixements i l'aprenentatge de noves tècniques i metodologies de treball, complimentant així la formació com a investigador. A més, permetent treballar amb altres tècniques i temes no relacionats amb el món de les begudes, però que han permès demostrar l'àmplia aplicabilitat de les llengües electròniques.

Perspectives futures

1. Miniaturització de la matriu de sensors desenvolupada, mitjançant substitució dels elèctrodes basats en compòsits per elèctrodes serigrafiats. Augmentant el grau de portabilitat del sistema, reduint el volum de mostra requerit i el cost de fabricació.
2. Implantació d'algun dels sistemes proposats en dispositius portàtils tipus FPGA (*Field-Programmable Gate Array*) que permetin dur a terme mesures *on-field* de forma més senzilla i còmoda.
3. Avaluar i desenvolupar protocols que permetin assolir la correcció dels models construïts en funció del temps o del *batch* de sensors utilitzat; evitant així el reentrenament complet del sistema, facilitant per tant la seva implantació en el sector industrial. Un procés en estudi i conegut com a *model transfer*.
4. Desenvolupament de nous sistemes basats en llengües bioelectròniques, aplicades a la determinació d'altres analits; e.g. desenvolupament d'una llengua bioelectrònica basada en enzims deshidrogenassa.
5. Continuar treballant en la resolució de noves demandes per part del sector; com pot ser la determinació dels iso- α -àcids en el cas de les cerveses.
6. Aplicació dels principis de les llengües electròniques en altres camps tals com la seguretat ciutadana, e.g. detecció explosius.
7. Desenvolupament de noves estratègies i alternatives de preprocessament dels senyals voltamperomètrics.

6. Publicacions que formen part del compendi de la tesi doctoral

6. Publicacions que formen part del compendi de la tesi doctoral

- 6.1. *Voltammetric Electronic Tongue in the analysis of cava wines*
X. Cetó, J.M. Gutiérrez, L. Moreno-Barón, S. Alegret and M. del Valle
Electroanalysis, **23**(1): 72-78
- 6.2. *Resolution of phenolic antioxidant mixtures employing a voltammetric bio-electronic tongue*
X. Cetó, F. Céspedes, M.I. Pividori, J.M. Gutiérrez and M. del Valle
Analyst, **137**(2): 349-356
- 6.3. *Determination of total polyphenol index in wines employing a voltammetric electronic tongue*
X. Cetó, J.M. Gutiérrez, M. Gutiérrez, F. Céspedes, J. Capdevila, S. Mínguez, C. Jiménez-Jorquera and M. del Valle
Analytica Chimica Acta, **732**(0): 172-179
- 6.4. *Bioelectronic Tongue for the quantification of total polyphenol content in wine*
X. Cetó, F. Céspedes and M. del Valle
Talanta, **99**(0): 554-551
- 6.5. *Assessment of individual polyphenol content in beer by means of a voltammetric BioElectronic Tongue*
X. Cetó, F. Céspedes and M. del Valle
Electroanalysis, **25**(1): 68-76
- 6.6. *Chapter 6: BioElectronic Tongues: when the sensor array incorporates biosensors*
M. del Valle, X. Cetó and M. Gutiérrez
in *Multisensor systems for chemical analysis - materials and sensors* by L. Lvova, D. Kirsanov, C. Di Natale and A. Legin
Pan Stanford Publishing Pte Ltd. ISBN: 9789814411158

Article 1

Voltammetric Electronic Tongue in the Analysis of Cava Wines

X. Cetó, J. M. Gutiérrez, L. Moreno-Barón, S. Alegret and M. del Valle

Electroanalysis, 23(1): 72-78

Voltammetric Electronic Tongue in the Analysis of Cava Wines

Xavier Cetó, Juan Manuel Gutiérrez, Laura Moreno-Barón, Salvador Alegret, Manel del Valle*

Sensors & Biosensors Group, Department of Chemistry, Universitat Autònoma de Barcelona, Edifici Cn, 08193 Bellaterra, Spain
tel: +34935811017; fax: +34935812379

*e-mail: manel.delvalle@uab.es; salvador.alegret@uab.es

Received: July 15, 2010

Accepted: September 21, 2010

Abstract

This work demonstrates the use of a voltammetric electronic tongue formed by five modified graphite-epoxy electrodes in the qualitative and quantitative analysis of cava wines. The different samples were analyzed using cyclic voltammetry without any sample pretreatment. Recorded data were evaluated by Principal Component Analysis and Discrete Wavelet Transform in order to compress and extract significant features from the voltammetric signals. The preprocessed information was evaluated by an Artificial Neural Network that accomplishes the qualitative classification. Moreover, a preliminary study related to the quantification of sugar amount present was assessed by Second-Order Standard Addition Method.

Keywords: Voltammetric sensors, Electronic tongue, Wine analysis, Cava wine

DOI: 10.1002/elan.201000439

Presented at the 13th International Conference on Electroanalysis, ESEAC 2010, Gijón, Spain

1 Introduction

Cava wine is a sparkling wine which is produced in north-east Spain (Catalonia) distinguished by its pale colour, intense aroma and floral bouquet. Its flavour is slightly acidic with a fruit taste different from French Champagne wine which is sweeter and floral in flavour. The name “cava” comes from the place where the bottles are fermented; a dark and damp cave. Different kinds of Cava wines are made following the “*méthode champenoise*” which is the same method used to make Champagne, but for the Spanish production it is called Traditional Method [1]. Cava is produced in different levels of dryness; these ratings refer to the amount of sugar added after second fermentation, riddling and disgorging and before the stoppering and labelling of the bottles to reach the market. In this way, this addition of sugar is the last step in the cava wine production. Hence, produced wine is classified as content of sugar (sucrose) added: Brut nature $< 3 \text{ gL}^{-1}$ (only residual sugars), extra brut $0\text{--}6 \text{ gL}^{-1}$, brut $< 15 \text{ gL}^{-1}$, extra dry $12\text{--}20 \text{ gL}^{-1}$, dry $17\text{--}35 \text{ gL}^{-1}$, medium dry $33\text{--}50 \text{ gL}^{-1}$ and sweet $> 50 \text{ gL}^{-1}$ [1].

In order to provide the corresponding denomination of origin (DO) is needed a control over the entire process of production and development, guaranteeing the origin and quality of Cava wine through the processes that ensure its analytical and organoleptic characteristics [1]. In this way, the current demand of analytical methods, reliable and economical has grown in the food industry. The present work proposes an objective new technical method to

reveal some of the parameters used to characterize cava wines. The proposed method is known as Electronic Tongue (ET); this approach is formed by an array of sensors capable of giving a wide and complete response of the analyzed species, plus a chemometric processing tool, capable of extracting meaningful data from the complex readings [2–4]. Although the use of electronic tongues in the analysis of liquids has been widely described over the past decade, there are only some papers directly related to the world of wine [5–9]. The few studies which involve the study of sparkling wines as Cava often use other analysis techniques instead of ET, such as: gas chromatography (GC) [10], mass spectrometry (MS) [11] and some other atomic spectrometry techniques [12]. These studies were focused in analyzing either the influence of vineyard yield on the phenolic, volatile and nitrogen compounds, on the foam [13] or the evolution of volatile compounds during long ageing times [14]. Other works identify the influence of bentonite to facilitate the riddling process and its effects on the foam properties [15]. There are also works that develop strategies to differentiate between samples of French Champagne and Spanish Cava according to their mineral content [12].

The system chosen in this application is a voltammetric electronic tongue, with modified sensors capable of presenting differentiated response. While the voltammetric sensors provide a set of multivariate signals, the use of advanced chemometric models facilitates the data interpretation. In this way, Artificial Neural Networks (ANNs) are one of the most widely processing tools in

chemical data interpretation; their use has been popularized in electronic noses and tongues applications, due to their ability in the interpretation, modelling and calibration of complex analytical signals [16–20]. However, processing a set of voltammetric signals with ANN requires, given its high dimensionality, a preprocessing stage of data reduction, such as Principal Component Analysis (PCA) or Discrete Wavelet Transform (DWT). This data preprocessing is needed in order to gain advantages in training time, to avoid redundancy in input data, and to obtain a model with better generalization ability. This fact is mainly due to the extreme complexity of these signals. In this sense, processing with DWT is particularly interesting because of its ability to compress and denoise data [21].

The aim of this work is therefore to establish the foundations for the future use of voltammetric electronic tongue based on bulk-modified composites, in order to identify qualitatively different classes of cava wines. Additionally, different standard additions of sugar were performed, in order to establish the original amount of sugar present in the samples and to compare with the values given by the cava producer. Given the multi-way nature of the data generated in this stage, a Second-Order Standard Addition Method (SOSAM) was employed to correlate the sought analytical information. This strategy was chosen due to the difficulty in generating standards close enough to a cava wine.

2 Experimental

2.1 Sparkling Wines Under Study

A total of 21 samples of sparkling wine of different brands were acquired at the local supermarket, 19 with Spanish Cava DO and 2 Champagne (French) samples. These samples were chosen according with the principal kinds of cava. In this way, the formed set includes 7 Brut, 5 Brut Nature and 7 Medium-dry Cava wines. Most of the samples come from the Penedès area (Catalonia region), except of two medium-dry samples which are made in the Extremadura region. In addition, two Champagne (French) samples were included as outliers in the set. Table 1 summarizes detailed information about the Spanish sparkling wines used.

2.2 Electrodes and Measurement Cell

An array of 5 graphite-epoxy voltammetric sensors made with different modifiers added to the bulk mixture were selected according with previous study in our laboratory with wines (Figure 1) [5]. Four of them were modified by adding components as nanoparticles of copper, platinum or conducting polymer in powder like polyaniline and polypyrrole (Sigma-Aldrich, St. Louis, USA) – one component per sensor [22–25]. Standard graphite-epoxy composites were prepared using 50- μm particle size graphite powder (BDH laboratory Supplies, UK) and Epotek H77

Table 1. Spanish Cava wine samples under study.

| Variety | Spanish DO | Producer |
|-------------|-------------|-------------------|
| Brut | Penedès | Freixenet |
| | | Mascaró Classic |
| | | Castellblanch |
| | | Rondel Blue |
| | | Anna de Codorniu |
| Brut Nature | | Juvé & Camps |
| | | Blancher |
| | | Jaume Serra |
| | | Castell del Llac |
| | | Codorniu Extra |
| Medium-dry | Penedès | Lavit |
| | | Mont Marçal |
| | | Canals i Nubiola |
| | | Extra de Codorniu |
| | | Delapierre |
| | Extremadura | Dubois |
| | | Jaume Giró |
| | | Vía de la Plata |
| | | Ex de Extremadura |
| | | |

resin and hardener (both from Epoxy Technology, USA) [26,27]. The final electrode area was 28 mm². The measurement cell was formed by the 5-sensor voltammetric array and a reference double junction Ag/AgCl electrode (Thermo Orion 900200) plus a commercial platinum counter electrode (Model 52-67, Crison Instruments). A sample volume of 10 mL of sparkling wine was immediately used for each measurement, directly without any sample pretreatment.

2.3 Electrochemical Technique and Procedure

The electrodes were cycled for 3–5 times in distilled water in order to get stable voltammetric responses

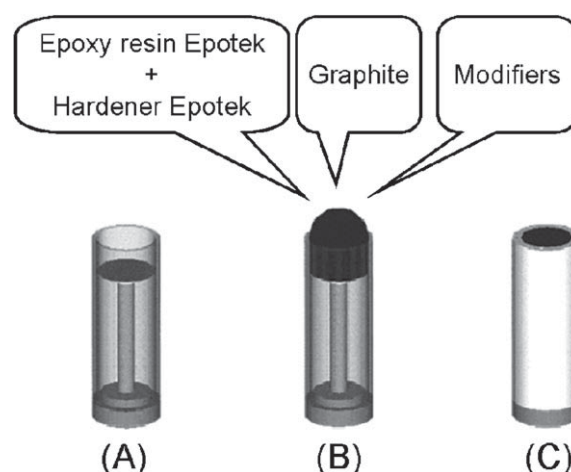


Fig. 1. Graphite-epoxy composite electrodes construction scheme. (A) Copper disk soldered to the connector and assembled into the PVC tube. (B) Preparation of the graphite-epoxy mixture and incorporation of modifier. (C) Final look after hardening and polishing.

before performing the measurements with cava samples. Cyclic voltammetry measurements were taken using a 6-channel AUTOLAB PGSTAT20 (Ecochemie, Netherlands). Electroanalytical experiments were carried out at room temperature (25 °C) and cyclic voltammograms were made under quiescent condition. For the sample measurements scan rate was fixed at 0.1 V s⁻¹ with a step potential of 3.66 mV. The potential sweep was between -1.0 and 1.3 V vs. Ag/AgCl. No pretreatment of sample was performed; once cava bottles were opened, measures were carried out after a few minutes (ca. 5 min at room temperature) in order to avoid the formation of bubbles into the electrode surface. That is, no specific removal of oxygen and CO₂ was done, but minimization of distortions by air bubbles was attempted.

Apart, all experiments were carried out without performing any physical surface regeneration of the working electrodes. In order to prevent the accumulative effect of impurities on the working electrode surfaces, an electrochemical cleaning stage was done between each measurement applying a conditioning potential of +1.5 V for 40 s after each experiment, in a cell containing 10 ml of distilled water.

2.4 Standard Addition

A solution of sucrose standard (750 g/L) in distilled water was prepared and different amounts of this solution were added to the selected cava wines chosen from the complete set, to perform the standard addition determination of sugar content. The concentration of added sugar (sucrose) was in the range from 0 to 80 g/L according to the different ratings by the DO commission regulation. Then, the spiked samples were treated as independent samples and the measurements were performed as previously described.

2.5 Data Processing

Chemometric processing was done by specific routines in MATLAB 7.0 (MathWorks, Natick, MA) programmed by the authors, and using either Neural Network Toolbox (v.4.0) or Wavelet Toolbox (3.0). From the five scans performed in each determination, only the last one is used as input data. The whole cyclic voltammograms from the five sensor array were included in the different data processing stages.

3 Results and Discussion

3.1 Voltammetric Responses

Different obtained signals coming from the voltammetric sensors used are shown in Figure 1. It is possible to observe differentiated signals obtained for each kind of sensor. Catalytic oxidative signals seem to originate from the sensors using metal nanoparticles, which may be due to a catalytic oxidation of saccharides (e.g. glucose) pres-

ent. Similarly, sensors modified using conducting polymers bring new information depending on the type of cava considered [5].

3.2 Electronic Tongue Preliminary Recognition

As deduced from the voltammograms in Figure 2, the incorporation of these modifiers in the graphite-epoxy voltammetric electrodes is responsible for the generation of a clear response variability, a good departure point for an ET approach [3,25]. In voltammetric electronic tongues, the nature of the signals involves the recording of currents generated in the solutions under study related to an applied potential. Voltammetric signals contain hundreds of measures and usually overlapping regions with non-stationary characteristics [28], since all components in the solution which are electrochemical active below a specific potential, all contribute to the current measure. In this sense, the voltammograms should be firstly analyzed in order to corroborate its analytical content. In order to reduce the huge data amount generated in each measurement, a preprocessing stage employing DWT was used [29]. In this way, the compression of the original sensor information was achieved up to 97.34% without any loss of relevant information using Daubechies wavelet and a fourth decomposition level. In this way, and using the proposed sensor array, the corresponding compressed voltammograms were processed applying PCA analysis. This method allows to summarize almost all variance contained in the departure information on a fewer number of directions (the PCs) with new coordinates called scores, obtained after data transformation [5]. Thus, a preliminary recognition model was performed in order to display the different kinds of cava, according to measurements done. The PCA plot with the different clusters is shown in Figure 3 with the three first PCs, the explained variance accumulated ca. 94%. The patterns in the figure evidence that cava samples are grouped based on type and production region. These well established clusters clearly separate the main classes of samples corresponding to: (1) Brut Nature, (2) Brut, (3) Medium-dry from Penedès region (4) Medium-dry from Extremadura region and (5) French Champagne wine (the outliers).

3.3 Classification of Cava Wines

The data compressed with DWT and next processed using PCA, only provides a visualization of the different grouping regions. In this sense, an actual classifier was built employing an ANN. This model was able to perform properly the qualitative classification of the three main classes of cava as well as the considered outliers. The samples of Champagne and Extremadura Cava were considered as outliers due to their reduced relationship with the rest of samples.

In order to find the appropriate ANN model, significant effort is needed to optimize the configuration details that determine its operation. Normally, this is a trial-and-

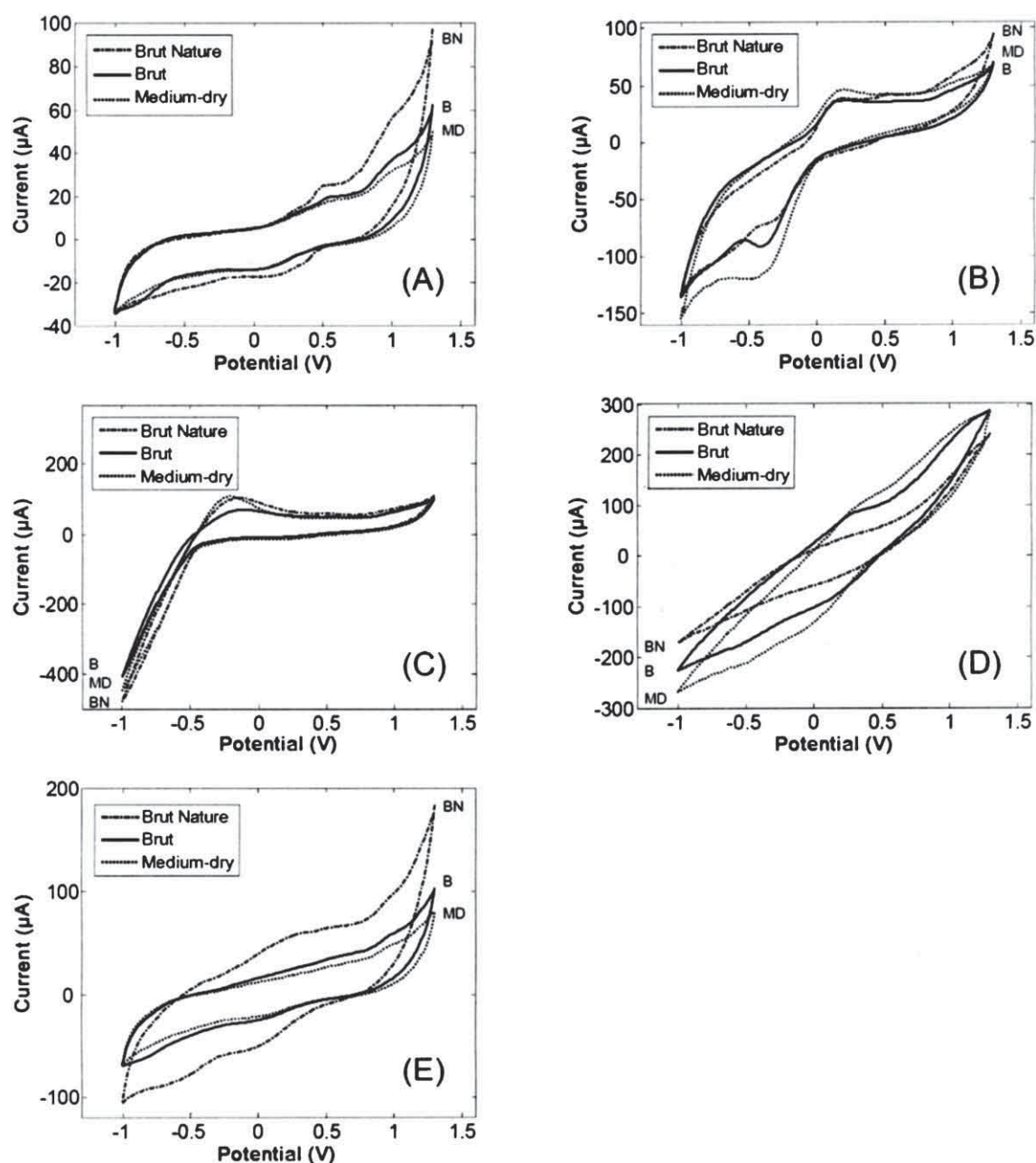


Fig. 2. Sensor array and some cyclic voltammograms obtained with five of the electrodes used: (A) graphite-epoxy sensor, (B) nanoparticle copper sensor, (C) nanoparticle platinum sensor, (D) polyaniline sensor and (E) polypyrrole sensor. Different Cava wines responses are shown: Brut-Nature (BN), Brut (B) and Medium-dry (MD).

error process, where several parameters (training algorithms, number of hidden layers, transfer functions, etc.) are fine-tuned in order to find the best configuration to optimize the performance of the model. After some preliminary tests, the final ANN architecture model had 235 input neurons (corresponding to the 47 wavelet approximation coefficients obtained from wavelet analysis of each of the 5 sensor signals), 20 neurons in the hidden layer and four binary (1/0) output neurons (one for each cava class and the outliers). After a systematic evaluation of topologies, the best training results were obtained employing the “*logsig*” function in the three layers. The

backpropagation algorithm was used to train the network and the expected output error was programmed to reach a value of 0.025. Figure 4 shows a simplified scheme of this procedure.

After training with 80% of the data, a confusion matrix was built using the information of the testing subset (remaining 20% of the data) in order to characterize the accuracy of the identification model and obtain unbiased data. Table 2 shows the results of the classification, expressed as the corresponding confusion matrix; its content displays expected vs. found response as calculated from the DWT-ANN model. As it can be observed, cor-

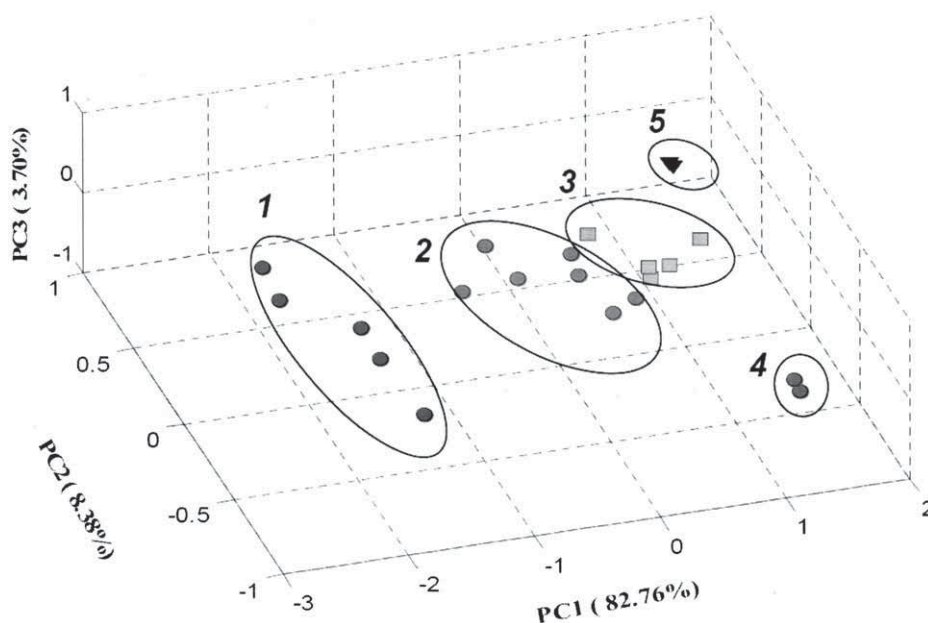


Fig. 3. Principal component analysis plot of the first three components. A total of 21 samples were analyzed. As can be observed, clear discrimination is obtained for the different types of cava wines: (1) Brut Nature, (2) Brut, (3) Medium-dry from Penedès region (4) Medium-dry from Extremadura region and (5) French Champagne wine (the outliers).

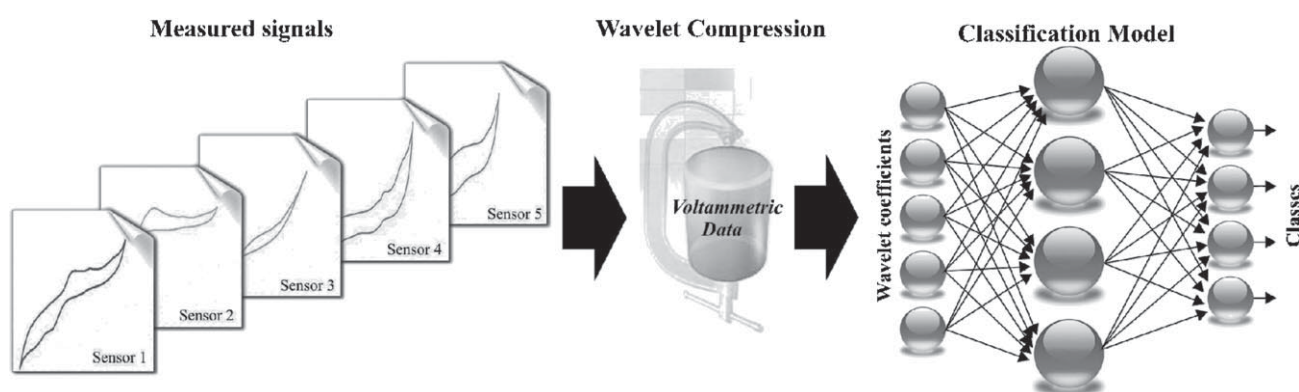


Fig. 4. Processing scheme for building the classification model.

rect classification for Brut nature, Brut and Medium-dry was obtained, while half of outliers were misclassified as medium-dry variety. It is important to note that this outlier apparently misclassified is a sample of medium-dry cava elaborated in the Extremadura region. The percentage of correct classifications was estimated from individual sample calculation in the test subset as 90.0%, a highly significant value taking into account that these samples did not participate at all in the training process. The efficiency of the classification obtained was also evaluated according to its sensitivity, i.e. the percentage of objects of each class identified by the classifier model, and to its specificity, the percentage of objects from different classes correctly rejected by the classifier model. The value of sensitivity, averaged for the four classes considered, was 87.5% and that of specificity was 85.7%.

3.4 Quantification of Sugar by Standard Addition

Due to the good results obtained with the qualitative application shown above, it can be assumed that the sensors were able to distinguish between the different amounts of sugar (sucrose) added in the cava production and which determines the variety of the cava wine.

For the quantification purposes a Second Order Standard Addition Method (SOSAM), which was first introduced by Booksh et al. for spectrophotometric data, was used [30]. This is caused by the high dimensionality of departure data, that makes the traditional approach unfeasible. With this method the first step is to decompose the third order data array and to reduce it, afterwards to identify which loadings are related with the analyte of interest and finally to plot the scores versus the amount of analyte added in order to linearly extrapolate the analyte

Table 2. Confusion matrix obtained using as classifier an ANN with 20 neurons in the hidden layer. The model was able to discriminate correctly between principal classes of Cava wines in the test. The classification rate decreased slightly because the test set included samples of Champagne and Extremadura Cava, outliers in the model, one of them misclassified as Medium-dry.

| Expected | Found | | | | Wrongly classified (%) |
|--------------------------------------|--------------|------|------------|---------|------------------------|
| | Brut Natural | Brut | Medium-dry | Outlier | |
| Brut Natural | 100 | 0 | 0 | 0 | 0 |
| Brut | 0 | 100 | 0 | 0 | 0 |
| Medium-dry | 0 | 0 | 100 | 0 | 0 |
| Outlier | 0 | 0 | 50 | 50 | 50 |
| Percentage classification (by class) | | | | | 87.5% |

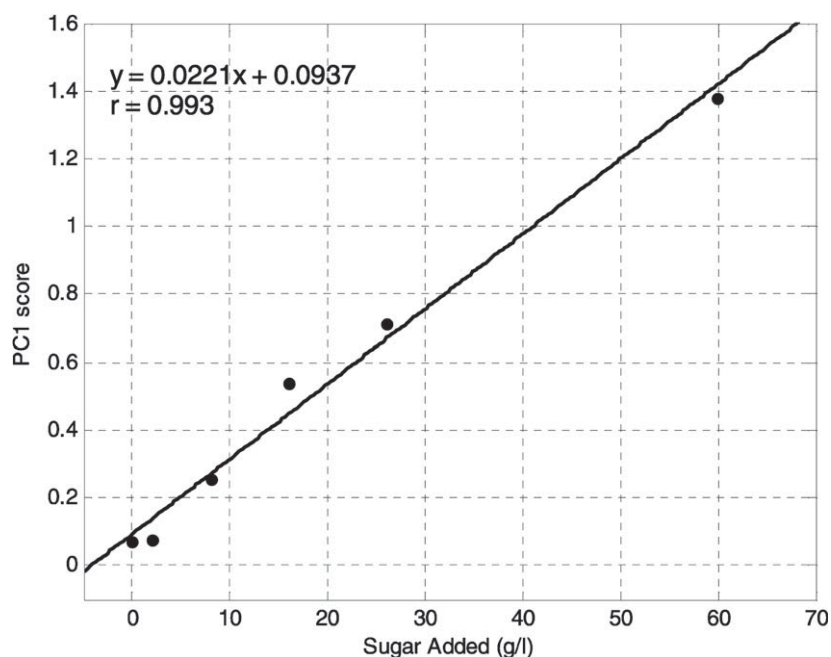


Fig. 5. Standard Addition plot of the first component of the PCA against amount of sugar added for Codorniu Extra Brut Nature. SOSAM methodology followed to obtain the amount of residual sugar in the cava wine.

Table 3. Residual sugar obtained with the standard addition determination, compared with the values given by the cava producer.

| Sample | Regression Plot | r | Total error | Found sugar (g/L) | Producer declared (g/L) |
|----------------------------|------------------------|-------|-------------|-------------------|-------------------------|
| Codorniu Extra Medium-dry | $y = 0.0182x + 0.744$ | 0.946 | 0.484 | 40.9 | 35 |
| Anna de Codorniu Brut | $y = 0.0228x + 0.0869$ | 0.908 | 0.442 | 3.81 | 6–8 |
| Codorniu Extra Brut Nature | $y = 0.0221x + 0.0937$ | 0.993 | 0.235 | 4.24 | <3 |
| Mont Marçal Brut Nature | $y = 0.0210x + 0.0398$ | 0.980 | 0.330 | 1.67 | <2 |

content, in the same way as in the univariate standard addition.

In this way, the quantification of sugar (sucrose) was performed first carrying a PCA and plotting the scores of the PCA first component against the amount of sugar added, obtaining a standard addition plot (Figure 5). The equations and extrapolated values for the samples tested can be seen in Table 3; as deduced from the correlation coefficients, a linear behaviour was obtained in all the cases, demonstrating that a linear relation between the first component of the PCA and the amount of sugar

added exists. Also, found values are not significantly different from the values given by the manufacturer which also corroborates the relation between the sensors array response and the amount of sugar (sucrose) in the sample.

4 Conclusions

An electronic tongue system based on voltammetric sensors with different modifiers (metallic nano-sized particles

and conducting polymers), was developed in order to create a tool capable of distinguishing between different kinds of cava wines. Thanks to the different redox catalytic properties of the modifiers, different oxidation and reduction peaks were obtained giving information of the different compounds present on the cava wine. Preliminary analysis using PCA were useful to identify the different varieties; however, when the data clusters were not directly identified, the use of an ANN classifier allowed us to improve the recognition ability. For a multiple electrode voltammetric electronic tongue, the complexity of generated data is very high, difficulting the subsequent applications of chemometric tools. In this sense, the use of discrete wavelet transform provides reasonable data compression preserving relevant information of the original signals, allowing to obtain a robust ANN model with good generalization ability.

In a second relevant application, we have been able to correlate the sensors signal with the amount of sugar content in the cava by using SOSAM, intended for quantification of the amount of sugar present in the original sample. To our knowledge, this is the first time SOSAM has been employed with an electronic tongue. The linearities obtained with the different samples confirm the goodness of the followed approach.

In summary, we have demonstrated the power of the electronic tongue formed by an array of non-specific sensors with the use of chemometric tools in tasks such as classification of different cava samples and quantification of the concentration of sugar (a specific compound).

Acknowledgements

Financial support for this work was provided by the *Spanish Ministry of Science and Innovation* through the project TEC2007-68012-C03-02 and by the *Mexican National Council of Science and Technology CONACYT* (México) through Postdoctoral scholarship for J. M. Gutiérrez and X. Cetó thanks the support of *Dept. d'Innovació, Universitats i Empresa de la Generalitat de Catalunya* for the predoctoral grant.

References

- [1] S. Buxaderas, E. Lopez-Tamames, in *Encyclopedia of Food Sciences and Nutrition* (Eds: B Caballero, L Trugo, PM Finnglas), Elsevier Science, St Louis **2003**, pp. 6203–6209.
- [2] M. Holmberg, M. Eriksson, C. Krantz-Rülcker, T. Artursson, F. Winqvist, A. Lloyd-Spetz, I. Lundström, *Sens. Actuators B*, **2004**, *101*, 213.
- [3] M. del Valle, *Electroanalysis* **2010**, *22*, 1539.
- [4] J. Zeravik, A. Hlavacek, K. Lacina, P. Skladal, *Electroanalysis* **2009**, *21*, 2509.
- [5] J. M. Gutiérrez, L. Moreno-Barón, M. I. Pividori, S. Alegret, M. del Valle, *Microchim. Acta* **2010**, *169*, 261.
- [6] A. Legin, A. Rudnitskaya, Y. Vlasov, C. D. Natale, E. Mazzone, A. D'Amico, *Electroanalysis* **1999**, *11*, 814.
- [7] A. Legin, A. Rudnitskaya, L. Lvova, Y. Vlasov, C. Di Natale, A. D'Amico, *Anal. Chim. Acta* **2003**, *484*, 33.
- [8] S. Buratti, S. Benedetti, M. Scampicchio, E. C. Pangerod, *Anal. Chim. Acta* **2004**, *525*, 133.
- [9] V. Parra, Á. A. Arrieta, J.-A. Fernández-Escudero, M. L. Rodríguez-Méndez, J. A. De Saja, *Sens. Actuators B* **2006**, *118*, 448.
- [10] E. Campo, J. Cacho, V. Ferreira, *J. Agric. Food Chem.* **2008**, *56*, 2477.
- [11] J. Bosch-Fuste, M. Riu-Aumatell, J. M. Guadayol, J. Caixach, E. Lopez-Tamames, S. Buxaderas, *Food Chem.* **2007**, *105*, 428.
- [12] A. Jos, I. Moreno, A. G. González, G. Repetto, A. M. Cameán, *Talanta* **2004**, *63*, 377.
- [13] M. A. Pozo-Bayón, M. C. Polo, P. J. Martín-Álvarez, E. Pueyo, *Food Chem.* **2004**, *86*, 413.
- [14] M. Riu-Aumatell, J. Bosch-Fusté, E. López-Tamames, S. Buxaderas, *Food Chem.* **2006**, *95*, 237.
- [15] G. Vanrell, R. Canals, M. Esteruelas, F. Fort, J. M. Canals, F. Zamora, *Food Chem.* **2007**, *104*, 148.
- [16] F. Despagne, D. L. Massart, *Analyst* **1998**, *123*, 157R.
- [17] A. Legin, A. Rudnitskaya, B. Seleznev, Y. Vlasov, *Int. J. Food Sci. Technol.* **2002**, *37*, 375.
- [18] R. N. Bleibaum, H. Stone, T. Tan, S. Labreche, E. Saint-Martin, S. Isz, *Food Qual. Prefer.* **2002**, *13*, 409.
- [19] R. Martínez-Mañez, J. Soto, E. Garcia-Breijo, L. Gil, J. Ibáñez, E. Llobet, *Sens. Actuators B* **2005**, *104*, 302.
- [20] I. Campos, L. Gil, R. Martínez-Mañez, J. Soto, J.-L. Vivanos, *Electroanalysis* **2010**, *22*, 1643.
- [21] J. M. Gutiérrez, L. Moreno-Barón, M. del Valle, L. Leija, R. Munoz, *Neural Netw. World* **2009**, *19*, 53.
- [22] J. Wu, J. Liu, M. Fu, G. Li, Z. G. Lou, *Sensors* **2005**, *5*, 529.
- [23] V. Martina, K. Ionescu, L. Pigani, F. Terzi, A. Ulrici, C. Zanardi, R. Seeber, *Anal. Bioanal. Chem.* **2007**, *387*, 2101.
- [24] L. Pigani, G. Foca, K. Ionescu, V. Martina, A. Ulrici, F. Terzi, M. Vignali, C. Zanardi, R. Seeber, *Anal. Chim. Acta* **2008**, *614*, 213.
- [25] M. L. Rodríguez-Méndez, V. Parra, C. Apetrei, S. Villanueva, M. Gay, N. Prieto, J. Martinez, J. A. de Saja, *Microchim. Acta* **2008**, *163*, 23.
- [26] F. Céspedes, E. Martínez-Fàbregas, S. Alegret, *TrAC – Trends Anal. Chem.* **1996**, *15*, 296.
- [27] R. Olivé-Monllau, M. Baeza, J. Bartrolí, F. Céspedes, *Electroanalysis* **2009**, *21*, 931.
- [28] J. M. Gutiérrez, L. Moreno-Barón, F. Céspedes, R. Muñoz, M. del Valle, *Electroanalysis* **2009**, *21*, 445.
- [29] L. Moreno-Baron, R. Cartas, A. Merkoci, S. Alegret, M. del Valle, L. Leija, P. R. Hernandez, R. Munoz, *Sens. Actuators B* **2006**, *113*, 487.
- [30] K. Booksh, J. M. Henshaw, L. W. Burgess, B. R. Kowalski, *J. Chemometr.* **1995**, *9*, 263.

Article 2

Resolution of phenolic antioxidant mixtures employing a voltammetric bio-electronic tongue

X. Cetó, F. Céspedes, M. I. Pividori, J. M. Gutiérrez and M. del Valle

Analyst, **137**(2): 349-356

Cite this: *Analyst*, 2012, **137**, 349

www.rsc.org/analyst

PAPER

Resolution of phenolic antioxidant mixtures employing a voltammetric bio-electronic tongue†‡

Xavier Cetó,^a Francisco Céspedes,^a Maria Isabel Pividori,^a Juan Manuel Gutiérrez^b and Manel del Valle^{*a}

Received 3rd June 2011, Accepted 18th October 2011

DOI: 10.1039/c1an15456g

This work reports the application of a Bio-Electronic Tongue (BioET) system made from an array of enzymatic biosensors in the analysis of polyphenols, focusing on major polyphenols found in wine. For this, the biosensor array was formed by a set of epoxy-graphite biosensors, bulk-modified with different redox enzymes (tyrosinase and laccase) and copper nanoparticles, aimed at the simultaneous determination of the different polyphenols. Departure information was the set of voltammograms generated with the biosensor array, selecting some characteristic features in order to reduce the data for the Artificial Neural Network (ANN). Finally, after the ANN model optimization, it was used for the resolution and quantification of each compound. Catechol, caffeic acid and catechin formed the three-analyte case study resolved in this work. Good prediction ability was attained, therefore allowing the separate quantification of the three phenols with predicted *vs.* expected slope better than 0.970 for the external test set ($n = 10$). Finally, BioET has been also tested with spiked wine samples with good recovery yields (values of 104%, 117% and 122% for catechol, caffeic acid and catechin, respectively).

Introduction

Polyphenols are a group of naturally occurring compounds that can be defined as molecules with more than one hydroxyl group on one or more phenol units per molecule. Most of these compounds are powerful antioxidants needed for the functioning of plant cells, which can be found in fruits and vegetables such as apples and onions, or beverages such as tea or wine.¹

There is evidence that wine has antioxidant properties, that these are due to polyphenolic components, and that wine without these components loses these properties.² The total content of polyphenols in wines is directly correlated with their antioxidant capacity;^{3,4} its exact composition is conditioned by the grape variety, geographical location and winemaking technology. The total concentration of polyphenols in wine varies between 1.80 and 4.06 g l⁻¹ equivalents in gallic acid for red wines and

between 0.16 and 0.33 g l⁻¹ for white wines; being gallic acid, catechin, epicatechin, *p*-coumaric acid, caffeic acid and catechol the major polyphenols found in wine.^{5–7} However, individual polyphenols concentration values are much lower than the total ones stated above, *e.g.* caffeic acid with a content of 5–13 mg l⁻¹ (27–72 μM) for red wines and 1–4 mg l⁻¹ (5–22 μM) for white wines or catechin with a content of 120–390 mg l⁻¹ (413–1343 μM) for red wines and 16–46 mg l⁻¹ (55–158 μM) for white wines are some of the typical contents.

On the one hand, several methods to quantify total phenols and polyphenols have been described in the literature.⁸ Common determinations of polyphenols include procedures such as the reaction with 4-aminoantipyrine,^{9,10} Folin–Ciocalteu reagent¹¹ or cyclic voltammetry analysis;¹² all these methods yield a total phenol content value, and therefore do not allow for the discrimination between individual phenols. On the other hand, there are techniques such as HPLC^{13–16} or GC^{17,18} which allow individual identification of polyphenols, but these require specific equipment, laboratory conditions and/or trained personnel. Finally, we cannot forget the use of amperometric biosensors which represent an attractive alternative to classical analytical methods for the detection of polyphenolic compounds; for this aim, biosensors incorporating enzymes such as tyrosinase,^{19,20} laccase^{21,22} or peroxidase^{23,24} have been employed. Even, the immobilization of two enzymes in the same biosensor has been reported.^{25,26}

In the recent years, hybrid intelligent analytical systems have started to play an important role in the food industry; one example of this is what is called an Electronic Tongue (ET).^{27–29}

^aSensors and Biosensors Group, Department of Chemistry, Universitat Autònoma de Barcelona, Edifici Cn, 08193 Bellaterra, Barcelona, Spain. E-mail: manel.delvalle@uab.cat; Fax: +34 93 5812379; Tel: +34 93 5811017

^bBioelectronics Section, Department of Electrical Engineering, CINVESTAV, 07360 Mexico D.F., Mexico

† This article is part of a web theme in *Analyst* and *Analytical Methods* on Future Electroanalytical Developments, highlighting important developments and novel applications. Also in this theme is work presented at the Eirelec 2011 meeting, dedicated to Professor Malcolm Smyth on the occasion of his 60th birthday.

‡ Electronic supplementary information (ESI) available: Table containing detailed composition of each electrode, BioET response towards individual components of the mixture and detailed results of the ANN optimization. See DOI: 10.1039/c1an15456g

These biomimetic systems have provided alternatives to many classical methods in the food analysis field.^{30–32} ETs have been already used for the detection of phenols in water from a number of impedimetric sensors interrogated at specific frequencies,³³ or even to carry out their quantification from the overlapped voltammetric signal obtained with a carbon electrode³⁴ or with a tyrosinase amperometric biosensor.³⁵ Even, the phenolic content of extra virgin olive oils was predicted from an array of chemically modified voltammetric electrodes;³⁶ and the generic pollution indicator, phenol index, has been deduced from the signal obtained with a laccase amperometric biosensor.³⁷

The ET is a new trend in the chemical analysis field: this approach uses a system formed by an array of sensors capable of giving a wide and complete response of the analyzed species, plus a chemometric processing tool, capable of extracting meaningful data from the complex readings provided by the sensor array.^{38–40} In this way, Artificial Neural Networks (ANNs), broadly used in chemical data-mining, are usually employed in electronic noses and electronic tongues.^{41,42} This is because ANNs are powerful modelling tools for deriving the meaning from complex or imprecise data; their superior performance makes them irreplaceable in chemometrics tasks such as multivariate calibration or pattern recognition.^{43,44}

Thus, the main contribution presented in this work is the use of an array of voltammetric enzyme biosensors. This is why the approach can be named Bio-Electronic Tongue (BioET). This approach has been rarely used in the literature, probably because the processing of a set of voltammetric signals is a case of high dimensionality and complexity, especially if ANNs are to be used. Similar cases required a multi-way treatment⁴⁵ or a pre-processing step for data reduction, such as Feature Extraction, Principal Component Analysis (PCA) or Discrete Wavelet Transform (DWT).⁴⁶ This data pre-processing, needed to make this case compatible with ANNs, permits one also to gain advantages in training time, to avoid redundancy in input data and to obtain a model with better generalization ability.

The BioET employed here is formed by an array of amperometric enzymatic biosensors based on bulk-modified biocomposites for the simultaneous determination of catechol, caffeic acid and (±)-catechin, in the concentration range found in wines. Using the combined electrochemical responses obtained from the set of biosensors' voltammograms to complement the departure information, performing specific feature extraction and employing an ANN model, it was possible to carry out the resolution of the mixture both in synthetic samples and in spiked wine samples, achieving their correct quantification.

Experimental

Apparatus

Electroanalytical experiments were carried out at room temperature (25 °C) using a 6-channel AUTOLAB PGSTAT20 (Ecochemie, Netherlands) in a multichannel electrode configuration under quiescent conditions. A double junction electrode Ag/AgCl Orion 900200 (Thermo Electron Corporation, Beverly, MA, USA) and a platinum-based electrode (Crison 52-67, Barcelona, Spain) were used as reference and auxiliary electrodes respectively.

Reagents and solutions

All reagents used were analytical grade and all solutions were prepared using deionised water from a Milli-Q system (Millipore, Billerica, MA, USA). Tyrosinase from mushroom (EC 1.14.18.1, 4276 U mg⁻¹), laccase from *Trametes versicolor* (EC 1.10.3.2, 21 U mg⁻¹), catechol, caffeic acid, (±)-catechin hydrate, potassium dihydrogenphosphate and potassium monohydrogenphosphate were purchased from Sigma-Aldrich (St Louis, MO, USA). KCl was purchased from Merck KGaA (Darmstadt, Germany). Copper nanoparticles (50 nm) were provided by Nanotek SA (Argentina).

Determinations were carried out in a 0.1 M phosphate buffer at pH 6.50, prepared in KCl 0.1 M, for ensuring a high electrical conductivity. 1 mM phenolic stock solutions were prepared by dissolving the right amount of pure substance (catechol, caffeic acid and (±)-catechin) into the buffer solution. After preparation of the stock solutions, training and testing samples were prepared by taking different microvolumes of phenol stock solutions and adjusting to a final volume of 25 ml. Concentrations ranged from 0 to 200 μM for each polyphenol (Fig. 1); 27 (73% of the total) samples distributed in a cubic design⁴⁷ were used for the training set and 10 additional samples, distributed randomly along the experimental domain, were used for the external test set (27% of the total).

Wine spiked samples

In order to perform the application with real samples and to verify if the wine matrix represents a problem for the resolution and quantification of individual polyphenols, 10 extra samples were prepared adding different quantities of a stock solution of each polyphenol to a standard table wine. As the ones prepared for the external test set, its concentration was ranged and distributed randomly along the experimental domain (Fig. 1).

Voltammetric sensor array

An array of 4 different graphite–epoxy voltammetric biosensors was prepared following the conventional methodology

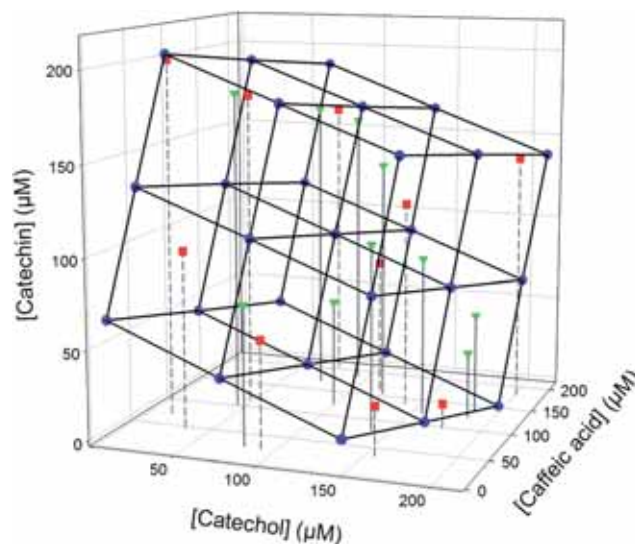


Fig. 1 Experimental design for the training subset (●), the external test subset (▼) and the spiked wine samples (■).

previously established in our research group.⁴⁸ First, resin Epo-Tek H77 (Epoxy Technology, Billerica, MA, USA) and its corresponding hardener compound were mixed in the ratio 20 : 3 (w/w). Then, each electrode was prepared adding 15% of graphite (w/w) and 2% of either the enzyme (tyrosinase or laccase) or the modifier (w/w) (copper nanoparticles) into the epoxy resin before hardening⁴⁹—one component per electrode, and the fourth electrode without any modifier (detailed composition of each sensor is summarized in Table S1 included in the ESI†). Afterwards, the biocomposite was manually homogenised for 60 min. Finally, the biocomposite paste electrode was allowed to harden for 7 days at 40 °C. The electrode surface was then polished with different sandpapers of decreasing grain size, with a final electrode area of 28 mm².

Tyrosinase catalyzes the hydroxylation of monophenols to catechols, which in turn are further oxidized to *o*-quinones, both using molecular oxygen; then *o*-quinones produced in the enzymatic reaction are electrochemically reduced to *o*-diphenols, resulting in recycling of the diphenols at the applied negative potential. Laccase catalyzes the oxidation of phenols giving phenoxyradical species which are converted to quinones in the second stage of the oxidation; then as in tyrosinase, these quinones can be electrochemically reduced again to phenols. Finally, copper-nanoparticles were chosen given that both

tyrosinase and laccase are copper-containing enzymes and it was thought that they could also have some catalytic effect as demonstrated with the results obtained.

Procedures

The measurement cell was formed by the 4-sensor voltammetric array plus the reference and auxiliary electrodes. Working electrodes were cycled 3–5 times in buffer solution in order to get stable voltammetric responses before performing the measurements with prepared samples. The potential was swept between +0.8 V and –0.4 V vs. Ag/AgCl, with a scan rate fixed at 100 mV s⁻¹ and a step potential of 9 mV. No pre-treatment of the sample was performed; and all experiments were carried out without performing any physical surface regeneration of the working electrodes. In order to prevent the accumulative effect of impurities on the working electrode surfaces, an electrochemical cleaning stage was done between each measurement applying a conditioning potential of +1.0 V for 40 s after each experiment in a cell containing 10 ml of distilled water.⁵⁰

Data processing

Chemometric processing was done by specific routines in MATLAB 7.0 (MathWorks, Natick, MA) written by the

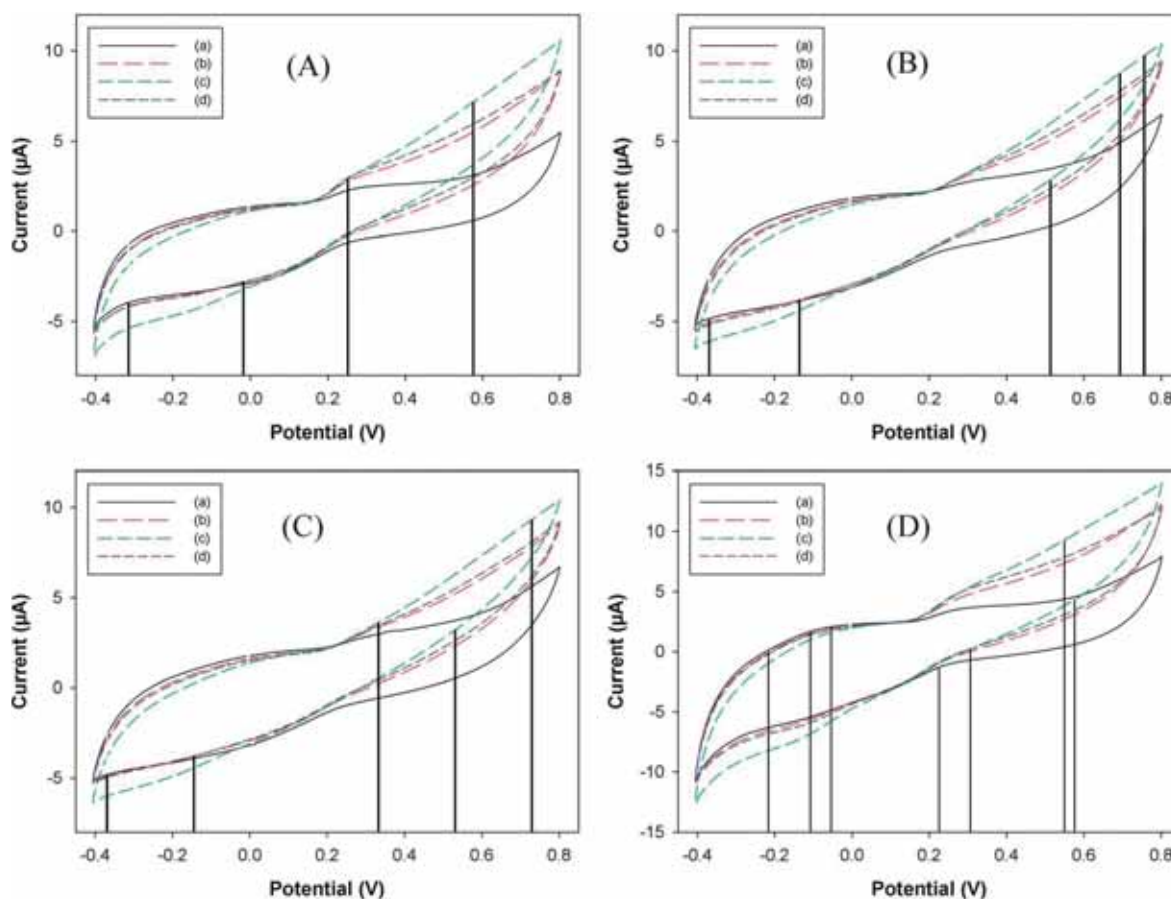


Fig. 2 Examples of the different obtained voltammograms for four mixtures of the three phenols. Concentrations for each phenol are: (a) 8.2, 39.8, 132.9 μM ; (b) 107.7, 103.6, 101.9 μM ; (c) 124.1, 151.7, 23.4 μM and (d) 157.7, 23.8, 86.4 μM for catechol, caffeic acid and (\pm)-catechin, respectively. The signals with different sensors are also shown: (A) graphite-epoxy sensor, (B) tyrosinase biosensor, (C) laccase biosensor and (D) copper nanoparticle modified sensor.

authors, and using Neural Network Toolbox (v.4.0). Sigmaplot 2000 (Systat Software Inc, California, USA) was used for graphic representations of data and results.

A total set of 37 samples, which were previously divided into two datasets, were measured: the 27 samples of the cubic design for the training set and 10 samples randomly distributed for the testing set were used. The set of voltammograms were recorded for each sample, from which some features were extracted in order to reduce the extreme complexity of these signals. The used training algorithm was the Bayesian regularization algorithm. This algorithm has the particularity that it avoids overfitting without the need to monitor the fitness degree of an internal validation subset,⁵¹ thus this precaution is not performed.

For feature extraction selection the Back-Propagation Artificial Neural Network (BP-ANN) method was used. This method usage for variable selection is the finding of an optimal set of inputs that can successfully classify or predict the desired outputs. It is a feed-forward network and combines a back-propagation algorithm which is used to train the network according to a learning rule.⁵²

An exhaustive study in the ANN architecture and configuration was done in order to optimize the separate quantification of the three phenols considered. Assessment of learning accomplished was estimated from the residual values calculated between expected and found concentration values, for each sample (i) and for each of the three analytes (j) considered:

$$\text{RMSE} = \sqrt{\frac{\sum_{ij} (c_{ij} - \hat{c}_{ij})^2}{3n - 1}} \quad (1)$$

Results and discussion

Voltammetric responses

Examples of the different obtained signals coming from the voltammetric biosensors for different phenolic compounds are shown in Fig. 2. Differentiated signals were observed for different kinds of biosensors, but due to signal overlapping it was not possible to assign a specific reduction or oxidation peak potential to each phenolic compound. Despite it being not possible to differentiate potentials, the fact is that the sensitivity towards each compound is slightly different; current intensities increase in different ways, which was confirmed when performing calibration curves of the different biosensors (as shown in the

ESI, Fig. S1†), obtaining different slopes for each compound.²³ The different relative responses of each phenolic compound arise in the different enzyme affinity towards the compound or in the electrochemical activity in the case of the copper modified bio-composite. This situation, with marked mix-response for the different involved compounds and with differentiated response from the different biosensors used, is highly desirable for studies with BioET systems.

Feature extraction

As already mentioned, it was very difficult to discriminate between individual peak signals of the three phenols in the voltammograms when the mixture was measured; when the stock solutions of each compound were treated independently it was possible to observe at which potentials there were appearing some maxima and the sensitivity differences for each one. Therefore, to carry out the dimensionality reduction and feature selection, BP-ANN was used.^{53,54}

First of all, the number of neurons in the hidden layer is selected in advance to represent the intrinsic dimensionality of the dataset. This represents a limit on the number of hidden neurons in the initial ANN. Based on experience and literature, the suggested number of neurons in the hidden layer should be already very close to the optimal size before the network reduction procedures are applied, usually five neurons is a good number. Thus, most of the effort is directed at reducing the number of network inputs since the number of outputs is fixed by the requirements of the model.

Afterwards, the next step is to determine the importance of network inputs. The relative contributions of the inputs to the variance of the output layer in the trained network are calculated. Once the ANN model is trained, the analysis of its connection weights can easily identify the important inputs. Inputs that make relatively small contributions to the variance can be discarded. These small contribution values indicate that either the input does not change significantly or that the training of the ANN has allocated small connection weights to scale the input. Repeating the process of training the ANN model with the reduced input set and the selection of the most relevant inputs can proceed until a near-optimal, small, set of inputs is identified.⁵⁵ Some of the resulting minimal sets are not unique, depending on the ANN architecture used in the training.

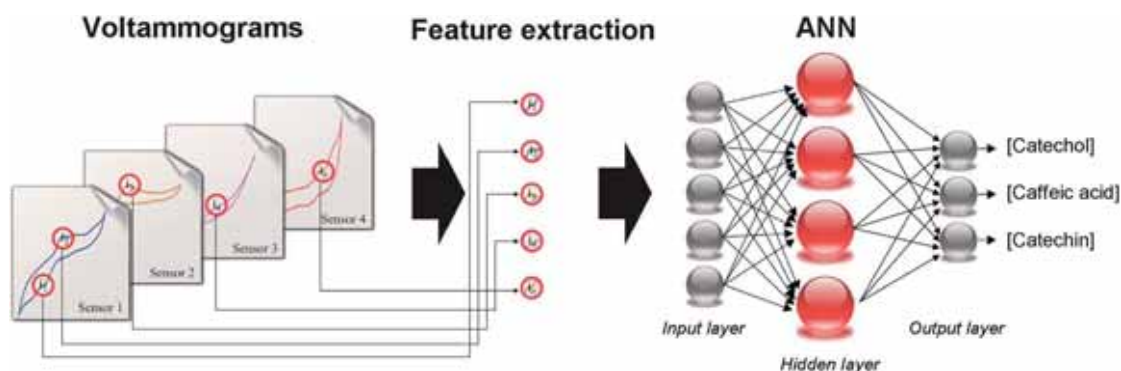


Fig. 3 Voltammetric BioET approach. Selected current intensities of the voltammograms from each biosensor are taken as the input in the ANN. Appropriate weights and biases are applied by the learning algorithm until the targets are reached within the established error.

There are advantages in identifying and removing the less-relevant inputs even if a large ANN is well trained: better modelling accuracy and higher robustness than the initial model; knowledge gain in identifying the important inputs; larger ratio of examples to ANN connection weights (thus a better generalization capacity); easier model understanding.⁵⁶ Thus, better ANN models were trained with the reduced input set, and this process was repeated until the model accuracy was degraded.

Out of the whole set of voltammograms only a number of 23 currents from the biosensor data matrix were selected, achieving an important reduction of data for the ANN input and without any loss of ANN modelling abilities. The selected potentials to which the current intensities (where “a” corresponds to anodic ones and “c” for cathodic) correspond for each electrode were: four currents for the graphite sensor (0.252^c, 0.576^c, -0.018^a and -0.315^a V), five for the tyrosinase biosensor (0.693^c, 0.756^c, 0.513^a, -0.135^a and -0.369^a V), five for the laccase biosensor (0.333^c, 0.729^c, 0.531^a, -0.144^a and -0.369^a V) and nine for copper nanoparticles sensor (0.549^c, 0.576^a, 0.306^a, 0.225^a, -0.054^a, -0.216^a, -0.252^a, -0.108^c and -0.054^c V).

Quantification of the three phenolic mixtures

According to the above, 23 current intensities at specific potentials for the array of sensors were selected as input vector in the ANN, the corresponding concentrations being the targets that the modelling should reach. The architecture of the ANN used was defined by these data; the input layer consisted of 23 neurons with a linear transfer function and the output layer consisted of three neurons, namely the three concentrations sought. We used a single intermediate layer, known as the hidden layer, whenever its number of neurons was specifically studied. Final data modelling was carried out using a feedforward architecture and Bayesian Regularization, as this training algorithm showed good capacity for training and prediction. Fig. 3 shows the scheme of this ANN based approach.

Once the algorithm was selected, an exhaustive search of the ANN architecture was performed. For this proposal a systematic study of the number of neurons in the hidden layer and combinations of functions in both hidden and output layers were tested. The number of neurons was changed from 1 to 10 and combinations of *tansig* (hyperbolic tangent sigmoid transfer function), *logsig* (log sigmoid transfer function), *purelin* (linear transfer function), *satlin* (saturating linear transfer function) and *satlins* (symmetric saturating linear transfer function) transfer functions were tested.

Once the modelling was completed, its RMSE value was calculated and an overall comparison was performed, as shown in Fig. S2, ESI†. In addition, the predicted ANN concentration values were plotted against the real ones for each analyte, linear least-squares regression was fit and then slope values of the comparison and correlation coefficients were taken into account in order to help in deciding the best configuration. Therefore, the optimum model would have a small RMSE, meaning there are no differences between the values predicted by the ANN and the real ones. Also the comparison graphs should display a correlation value and slope close to 1 and intercept close to 0, giving the comparison trend between the results obtained by the BioET and the real ones and the model capacity. Each architecture was

retrained five times, from a different random reinitialization of its weights, this being the source of the interval errors included for each point in the figures.

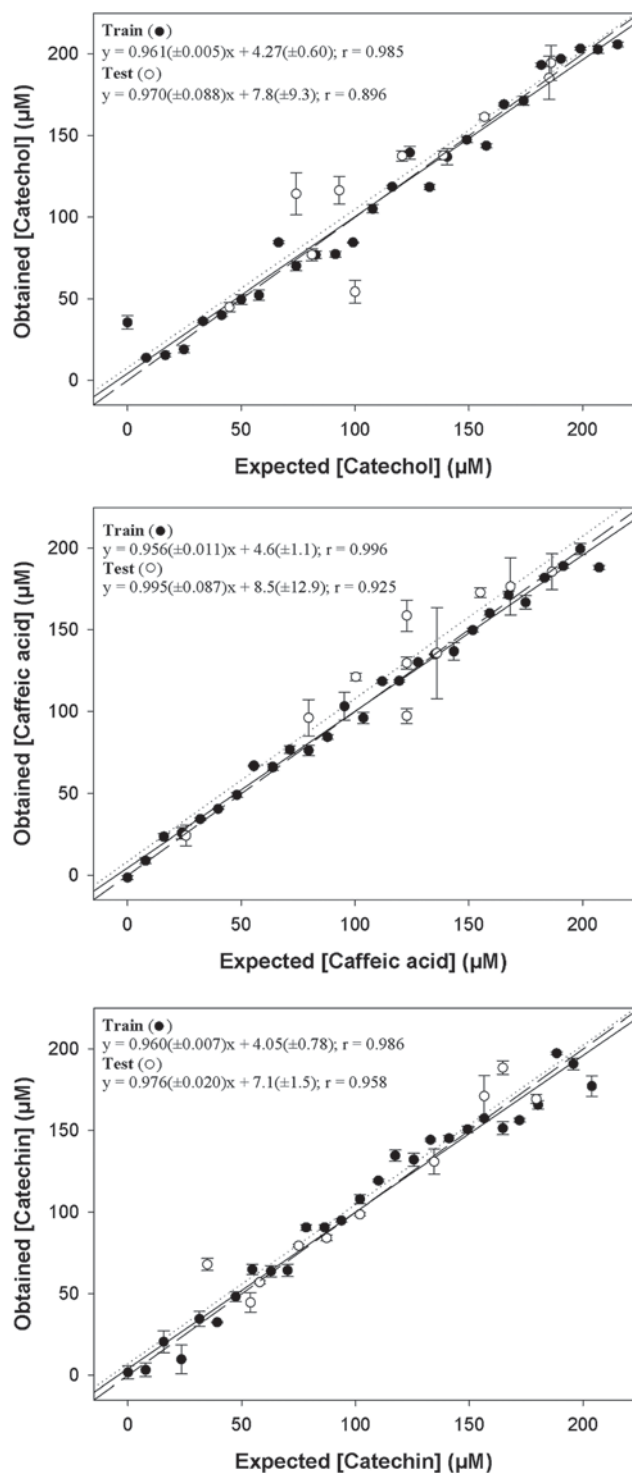


Fig. 4 Modelling ability of the optimized ANN. Training (●, solid line) and external test (○, dotted line) set adjustments of the expected concentration vs. obtained concentrations for (top) catechol, (middle) caffeic acid and (bottom) (±)-catechin. Dashed line corresponds to theoretical diagonal line. Error bars correspond to 5 different retrains with random reinitialization of weights for the final architecture.

Table 1 Detailed results obtained for some synthetic samples against the real concentration. Recovery yield is also expressed for each synthetic sample

| | Catechol | | | Caffeic acid | | | (±)-Catechin | | |
|----------|--------------|--------------------------------|-----------------------|--------------|--------------------------------|-----------------------|--------------|--------------------------------|-----------------------|
| | BioET/ μM | Spiked concentration/ μM | Recovery yield (%) | BioET/ μM | Spiked concentration/ μM | Recovery yield (%) | BioET/ μM | Spiked concentration/ μM | Recovery yield (%) |
| Sample 1 | 45.02 | 44.81 | 100.5 | 121.27 | 100.31 | 120.9 | 169.31 | 179.41 | 94.4 |
| Sample 2 | 76.92 | 81.00 | 95.0 | 24.34 | 25.70 | 94.7 | 79.47 | 75.02 | 105.9 |
| Sample 3 | 194.42 | 186.12 | 104.5 | 135.73 | 135.96 | 99.8 | 44.65 | 53.82 | 83.0 |
| Sample 4 | 54.33 | 99.96 | 54.4 | 97.25 | 122.70 | 79.3 | 57.08 | 57.90 | 98.6 |
| Sample 5 | 137.59 | 120.64 | 114.1 | 172.78 | 155.03 | 111.4 | 130.89 | 134.56 | 97.3 |

Checking the RMSE values obtained, it was concluded that the models using the *logsig* function in the output layer yielded higher RMSE values and also they had comparison slopes lower than 0.3, therefore these configurations were discarded. When *satlin* function was used in both hidden and output layer, unsatisfactory models with also bad regression behaviour were observed, which were also discarded. The same behaviour was obtained for *satlins* function, except when it was used in the output layer combined with *tansig*, *logsig* or *purelin* functions in the hidden layer. It can also be seen that the best RMSE values were obtained with three neurons in the hidden layer with similar RMSE values for the remaining configurations, except for *purelin-purelin*, *purelin-satlins* and *purelin-tansig* which were also discarded.

For the final selection of the best ANN we also considered the comparison of modelling performance for each analyte. In this way, despite the best RMSE values being obtained with three neurons in the hidden layer, there was a high improvement in both slope and correlation coefficient when the number of neurons was increased to 5. For all these reasons, the optimum configuration chosen was the one with 23 neurons in the input layer, 5 neurons and *logsig* transfer function in the hidden layer and 3 neurons and *purelin* transfer function in the output layer.

Fig. 4 shows the training and testing performances for the three analytes with the optimal ANN configuration, where its modelling and prediction capability is demonstrated. It must be remarked that the external test subset data are not employed at all for the modelling, so its goodness of fit is a measure of the

accomplished modelling performance. In both cases very good correlation was obtained, with regression lines almost indistinguishable from the theoretical ones.

Error bars plotted and stated uncertainties of fitted lines correspond to 5 different retrainings with random reinitialization of weights for the final architecture, giving information about modelling reproducibility. When statistical analysis is done on each regression separately, information more related to the data dispersion is obtained. Calculating the complete regression statistics for one of the above retraining cases, slopes were also very close to 1 and intercept close to 0, containing both values in the confidence interval, and with no significant variation in the correlation coefficient. Confidence intervals for training are one order of magnitude higher (e.g. catechin: $y = 0.956(\pm 0.073)x + 4.5(\pm 8.7)$; $r = 0.983$), and the ones for testing are even slightly greater (e.g. catechin: $y = 0.996(\pm 0.233)x + 5.8(\pm 26.9)$; $r = 0.961$). These larger uncertainty values are explained by the observed dispersion in the graphs and the lower number of points in this subset. Prediction capability of the model could be considered very good, as is summarized in Table 1, in which the recovery yield for the three polyphenols and random selected test samples is specified.

Furthermore, in order to demonstrate the abilities of the proposed BioET, some wine samples were spiked with variable amounts of the three phenolic compounds randomly distributed in the range of the experimental design. These were prepared and analyzed employing the same ET and data treatment (selected features extraction plus optimized ANN architecture). In this

Table 2 Detailed results obtained for the spiked wine samples against added concentrations of the three phenolics considered. Recovery yield is also expressed for every wine sample

| | Catechol | | | Caffeic acid | | | (±)-Catechin | | |
|---------|--------------|-------------------------------|-----------------------|--------------|-------------------------------|-----------------------|--------------|-------------------------------|-----------------------|
| | BioET/ μM | Added concentration/ μM | Recovery yield (%) | BioET/ μM | Added concentration/ μM | Recovery yield (%) | BioET/ μM | Added concentration/ μM | Recovery yield (%) |
| Wine 1 | 38.49 | 49.98 | 77.0 | 115.20 | 108.60 | 106.1 | 170.18 | 177.78 | 95.7 |
| Wine 2 | 178.17 | 178.37 | 99.9 | 82.07 | 92.85 | 88.4 | 37.60 | 12.64 | 297.5 |
| Wine 3 | 17.26 | 17.23 | 100.1 | 66.36 | 64.67 | 102.6 | 132.18 | 200.61 | 65.9 |
| Wine 4 | 115.06 | 75.83 | 151.7 | 159.35 | 197.31 | 80.8 | 168.71 | 161.47 | 104.5 |
| Wine 5 | 148.71 | 143.04 | 104.0 | 131.62 | 131.82 | 99.8 | 119.19 | 116.61 | 102.2 |
| Wine 6 | 151.34 | 157.69 | 96.0 | 58.25 | 30.26 | 192.5 | 26.62 | 26.10 | 102.0 |
| Wine 7 | 92.37 | 93.92 | 98.3 | 40.92 | 19.48 | 210.0 | 69.25 | 58.72 | 117.9 |
| Wine 8 | 49.66 | 36.84 | 134.8 | 44.16 | 41.45 | 106.5 | 108.70 | 98.67 | 110.2 |
| Wine 9 | 117.84 | 118.05 | 99.8 | 123.46 | 156.69 | 78.8 | 123.70 | 75.84 | 163.1 |
| Wine 10 | 158.04 | 202.50 | 78.0 | 177.40 | 174.10 | 101.9 | 83.49 | 140.26 | 59.5 |

case, given that the wine matrix is quite different from the standard buffered solution which was used to build the first response model, the system was retrained to compensate for any matrix effect. Then, employing a leave-one-out validation method forced by the reduced number of samples and to ensure each sample in the set is used once for validation, the prediction capability of the new response model for wine samples was evaluated. Also in this case, the recovery yield for the three polyphenols was calculated, which is summarized in Table 2. As can be seen, the obtained results are satisfactory with a good recovery yield in general (average values of 104.0%, 116.7% and 121.9% for catechol, caffeic acid and catechin, respectively).

Compared with previous works in which departure information was obtained from a single sensor, in this work, an array of voltammetric biosensors has been used which represents a novelty in the field of ETs. To confirm that the approach was necessary, and that no redundant information was used, the ANN modelling was also attempted using the information from a single voltammogram. In this case, and with little differences depending on the sensor used, reasonable models were still obtained, only they showed a problem in the quantification of caffeic acid. In all cases, comparison slopes were lower than 0.9 (even as bad as 0.3), correlation coefficients were as low as 0.90 in general, or even 0.80 for caffeic acid; finally RMSE values were also greater (between 9 and 13 μM for the best cases, instead of 5–6 μM which are the values obtained with the biosensor array). These final details, especially when compared with the BioET approach, can serve as a contrast of its utility, and validate the presented approach. Moreover, when dealing with wine samples the gain provided by the BioET which helps avoiding all matrix effects and providing a useful tool to carry out resolution of phenolic compound mixtures is clearer.

Conclusions

In summary, a very good quantification of the three phenolic antioxidant compounds has been achieved, by using a bio-Electronic Tongue which employs the combined response from different enzyme biosensors to complement the analytical departure information and ANNs as the tool for building the response model. Resolution of signal overlapping by this chemometric tool permitted the quantification of three important phenols in the beverage field such as catechol, caffeic acid and catechin, in an application comparable to more complex analytical techniques such as HPLC. The developed BioET has been evaluated with aqueous and wine samples, with good results in both cases, demonstrating its potential as a tool for phenolic mixtures resolution in this field. With a proper retraining samples set, the system may be alternatively applied for the quantification of analogous or different phenolic mixtures in real wine samples or other beverages.

Acknowledgements

Financial support for this work was provided by *Spanish Ministry of Science and Innovation*, MCINN (Madrid) through projects TEC2007-68012-C03-02 and CTQ2010-17099, and by program ICREA Academia. X. Cetó thanks the support of *Dept.*

d'Innovació, Universitat i Empresa de la Generalitat de Catalunya for the predoctoral grant.

References

- 1 J. E. Kinsella, E. Frankel, B. German and J. Kanner, *Food Technol.*, 1993, **47**, 85–89.
- 2 R. AbuAmsha, K. D. Croft, I. B. Puddey, J. M. Proudfoot and L. J. Beilin, *Clin. Sci.*, 1996, **91**, 449–458.
- 3 C. A. RiceEvans, J. Miller and G. Paganga, *Trends Plant Sci.*, 1997, **2**, 152–159.
- 4 M. Sato, N. Ramarathnam, Y. Suzuki, T. Ohkubo, M. Takeuchi and H. Ochi, *J. Agric. Food Chem.*, 1996, **44**, 37–41.
- 5 V. Rastija, G. Srećnik and S. Marica Medic, *Food Chem.*, 2009, **115**, 54–60.
- 6 N. Paixao, R. Perestrelo, J. C. Marques and J. S. Camara, *Food Chem.*, 2007, **105**, 204–214.
- 7 R. Ramos, P. B. Andrade, R. M. Seabra, C. Pereira, M. A. Ferreira and M. A. Faia, *Food Chem.*, 1999, **67**, 39–44.
- 8 D. De Beer, J. F. Harbertson, P. A. Kilmartin, V. Roginsky, T. Barsukova, D. O. Adams and A. L. Waterhouse, *Am. J. Enol. Vitic.*, 2004, **55**, 389–400.
- 9 M. Fayyad, M. Tutunji and L. Abdel-Nour, *Dirasat: Nat. Sci.*, 1983, **10**, 107–114.
- 10 G. Bieniek, T. Bochenska, G. Pieron and R. Skorek, *Chem. Anal.*, 1987, **32**, 1019–1023.
- 11 J. D. Box, *Water Res.*, 1983, **17**, 511–525.
- 12 J. Piljac, S. Martinez, T. Stipcevic, Z. Petrovic and M. Metikos-Hukovic, *Am. J. Enol. Vitic.*, 2004, **55**, 417–422.
- 13 U. Vrhovsek, A. Rigo, D. Tonon and F. Mattivi, *J. Agric. Food Chem.*, 2004, **52**, 6532–6538.
- 14 H. L. Zou, P. A. Kilmartin, M. J. Inglis and A. Frost, *Aust. J. Grape Wine Res.*, 2002, **8**, 163–174.
- 15 T. Heberer and H. J. Stan, *Anal. Chim. Acta*, 1997, **341**, 21–34.
- 16 A. Escarpa, M. D. Morales and M. C. Gonzalez, *Anal. Chim. Acta*, 2002, **460**, 61–72.
- 17 Y. Shao, P. Marriott and H. Hugel, *Chromatographia*, 2003, **57**, S349–S353.
- 18 B. Levaj, V. Dragovic-Uzelac, K. Delonga, K. K. Ganic, M. Banovic and D. B. Kovacevic, *Food Technol. Biotechnol.*, 2010, **48**, 538–547.
- 19 V. C. Sanz, M. L. Mena, A. Gonzalez-Cortes, P. Yanez-Sedeno and J. M. Pingarron, *Anal. Chim. Acta*, 2005, **528**, 1–8.
- 20 V. C. Dall'Orto, J. M. Vago, R. R. Carballo and I. N. Rezzano, *Anal. Lett.*, 2005, **38**, 19–33.
- 21 S. Gomes, J. M. F. Nogueira and M. J. F. Rebelo, *Biosens. Bioelectron.*, 2004, **20**, 1211–1216.
- 22 M. I. Pividori, S. Gomez-Montes and S. Alegret, *New Biotechnol.*, 2009, **25**, S352.
- 23 R. Solna and P. Skladal, *Electroanalysis*, 2005, **17**, 2137–2146.
- 24 S. Imabayashi, Y. T. Kong and M. Watanabe, *Electroanalysis*, 2001, **13**, 408–412.
- 25 M. R. Montekali, L. Della Seta, W. Vastarella and R. Pilloton, *J. Mol. Catal. B: Enzym.*, 2010, **64**, 189–194.
- 26 M. Elkouit, I. Naranjo-Rodriguez, K. R. Temsamani, M. D. La Vega and J. L. H. De Cisneros, *J. Agric. Food Chem.*, 2007, **55**, 8011–8018.
- 27 J. Zeravik, A. Hlavacek, K. Lacina and P. Skladal, *Electroanalysis*, 2009, **21**, 2509–2520.
- 28 F. Inquist, *Microchim. Acta*, 2008, **163**, 3–10.
- 29 M. Ghasemi-Varnamkhasi, S. S. Mohtasebi and M. Siadat, *J. Food Eng.*, 2010, **100**, 377–387.
- 30 L. Escuder-Gilabert and M. Peris, *Anal. Chim. Acta*, 2010, **665**, 15–25.
- 31 L. Pigani, A. Culet, A. Ulrici, G. Foca, M. Vignali and R. Seeber, *Food Chem.*, 2011, **129**, 226–233.
- 32 R. H. Labrador, R. Masot, M. Alcañiz, D. Baigts, J. Soto, R. Martínez-Mañez, E. García-Breijo, L. Gil and J. M. Barat, *Food Chem.*, 2010, **122**, 864–870.
- 33 C. A. Olivati, A. Riul, D. T. Balogh, O. N. Oliveira and M. Ferreira, *Bioprocess Biosyst. Eng.*, 2009, **32**, 41–46.
- 34 A. Gutiérrez, A. B. Ibáñez, F. Céspedes, S. Alegret and M. del Valle, *Anal. Bioanal. Chem.*, 2005, **382**, 471–476.
- 35 A. Gutiérrez, F. Céspedes, S. Alegret and M. del Valle, *Biosens. Bioelectron.*, 2005, **20**, 1668–1673.

- 36 M. L. Rodríguez-Mendez, C. Apetrei and J. A. de Saja, *Electrochim. Acta*, 2008, **53**, 5867–5872.
- 37 J. S. Torrecilla, M. L. Mena, P. Yanez-Sedeno and J. Garcia, *J. Agric. Food Chem.*, 2007, **55**, 7418–7426.
- 38 M. del Valle, *Electroanalysis*, 2010, **22**, 1539–1555.
- 39 Y. Vlasov, A. Legin, A. Rudnitskaya, C. Di Natale and A. D'Amico, *Pure Appl. Chem.*, 2005, **77**, 1965–1983.
- 40 P. Ciosek and W. Wroblewski, *Analyst*, 2007, **132**, 963–978.
- 41 S. M. Scott, D. James and Z. Ali, *Microchim. Acta*, 2006, **156**, 183–207.
- 42 P. Ciosek, Z. Brzozka and W. Wroblewski, *Sens. Actuators, B*, 2004, **103**, 76–83.
- 43 J. M. Gutiérrez, A. Gútes, F. Céspedes, M. del Valle and R. Muñoz, *Talanta*, 2008, **76**, 373–381.
- 44 Y. N. Ni and S. Kokot, *Anal. Chim. Acta*, 2008, **626**, 130–146.
- 45 D. Ebrahimi, E. Chow, J. J. Gooding and D. B. Hibbert, *Analyst*, 2008, **133**, 1090–1096.
- 46 L. Moreno-Barón, R. Cartas, A. Merkoçi, S. Alegret, J. M. Gutiérrez, L. Leija, P. R. Hernández, R. Muñoz and M. del Valle, *Anal. Lett.*, 2005, **38**, 2189–2206.
- 47 L. Zhang, Y. Z. Liang, J. H. Jiang, R. Q. Yu and K. T. Fang, *Anal. Chim. Acta*, 1998, **370**, 65–77.
- 48 F. Céspedes and S. Alegret, *TrAC, Trends Anal. Chem.*, 2000, **19**, 276–285.
- 49 R. Olivé-Monllau, M. Baeza, J. Bartrolí and F. Céspedes, *Electroanalysis*, 2009, **21**, 931–938.
- 50 X. Cetó, J. M. Gutiérrez, L. Moreno-Barón, S. Alegret and M. del Valle, *Electroanalysis*, 2011, **23**, 72–78.
- 51 L. Moreno-Barón, R. Cartas, A. Merkoçi, S. Alegret, M. del Valle, L. Leija, P. R. Hernandez and R. Muñoz, *Sens. Actuators, B*, 2006, **113**, 487–499.
- 52 J. R. M. Smits, W. J. Melssen, L. M. C. Buydens and G. Kateman, *Chemom. Intell. Lab. Syst.*, 1994, **22**, 165–189.
- 53 X. B. Zou, J. W. Zhao, M. J. W. Povey, M. Holmes and H. P. Mao, *Anal. Chim. Acta*, 2010, **667**, 14–32.
- 54 R. M. Balabin and S. V. Smirnov, *Anal. Chim. Acta*, 2011, **692**, 63–72.
- 55 Z. Boger, *Anal. Chim. Acta*, 2003, **490**, 31–40.
- 56 W. Wu and D. L. Massart, *Chemom. Intell. Lab. Syst.*, 1996, **35**, 127–135.

Supporting Information

Resolution of Phenolic Antioxidant Mixtures employing a Voltammetric Bio-Electronic Tongue

Xavier Cetó ^a, Francisco Céspedes ^a, Maria Isabel Pividori ^a, Juan Manuel Gutiérrez ^b
and Manel del Valle ^{a,*}

^a *Sensors and Biosensors Group, Department of Chemistry, Universitat Autònoma de
Barcelona, Edifici Cn, 08193 Bellaterra, Barcelona, SPAIN*

^b *Bioelectronics Section, Department of Electrical Engineering, CINVESTAV, 07360
Mexico D.F., Mexico*

This supplementary information gives detailed information of the formulation of the biocomposites of the BioET array and their voltammetric responses towards individual solutions of the phenolic compounds tested (Catechol, Caffeic acid and (±)-Catechin). Also detailed results obtained during the ANN optimization process for the RMSE values and the correlation coefficients and slopes calculated from the comparison graph between expected and found concentrations for the different configurations tested. Residual values were calculated between expected and found concentration values, for each sample (i) and for each of the three analytes (j) considered according to equation 1;

$$RMSE = \sqrt{\frac{\sum_{ij} (c_{ij} - \hat{c}_{ij})^2}{3n-1}} \quad (1)$$

* E-mail: manel.delvalle@uab.cat; tel: +34 93 5811017; fax: +34 93 5812379

Table S1. Formulation of the biocomposites used for preparing the bioelectronic tongue.

| Sensor | Epotek H77 [*] | Graphite | Component added |
|--|-------------------------|----------|---|
| (A) Graphite-epoxy sensor | 85% | 15% | - |
| (B) Tyrosinase biosensor | 83% | 15% | 2% Tyrosinase from mushroom (EC 1.14.18.1, 4276 U/mg) |
| (C) Laccase biosensor | 83% | 15% | 2% Laccase from <i>Trametes versicolor</i> (EC 1.10.3.2, 21 U/mg) |
| (D) Copper nanoparticles modified sensor | 83% | 15% | 2% Copper nanoparticles (50 nm) |

^{*}Previously mixing the resin (part A) with its corresponding hardener (part B) in a ratio 20:3 (w/w)

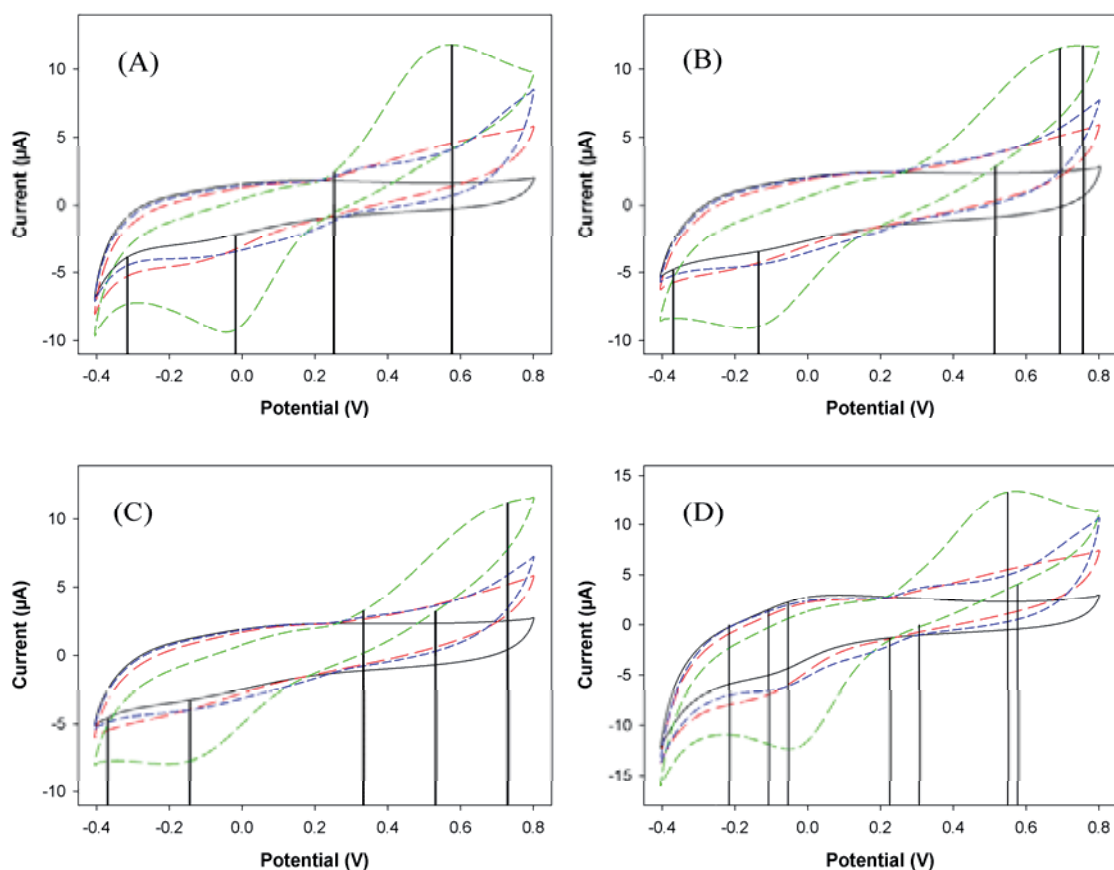


Figure S1. Example of the voltammograms obtained for the same standard solutions of each of the three phenols and buffer solution with the four sensors used. (—) Buffer solution, (---) 87.7 μM Caffeic acid solution, (---) 151.2 μM Catechol solution and (---) 96.2 μM (\pm)-Catechin solution. Sensors: (A) Graphite-epoxy sensor, (B) Tyrosinase biosensor, (C) Laccase biosensor and (D) Copper nano-particle modified sensor.

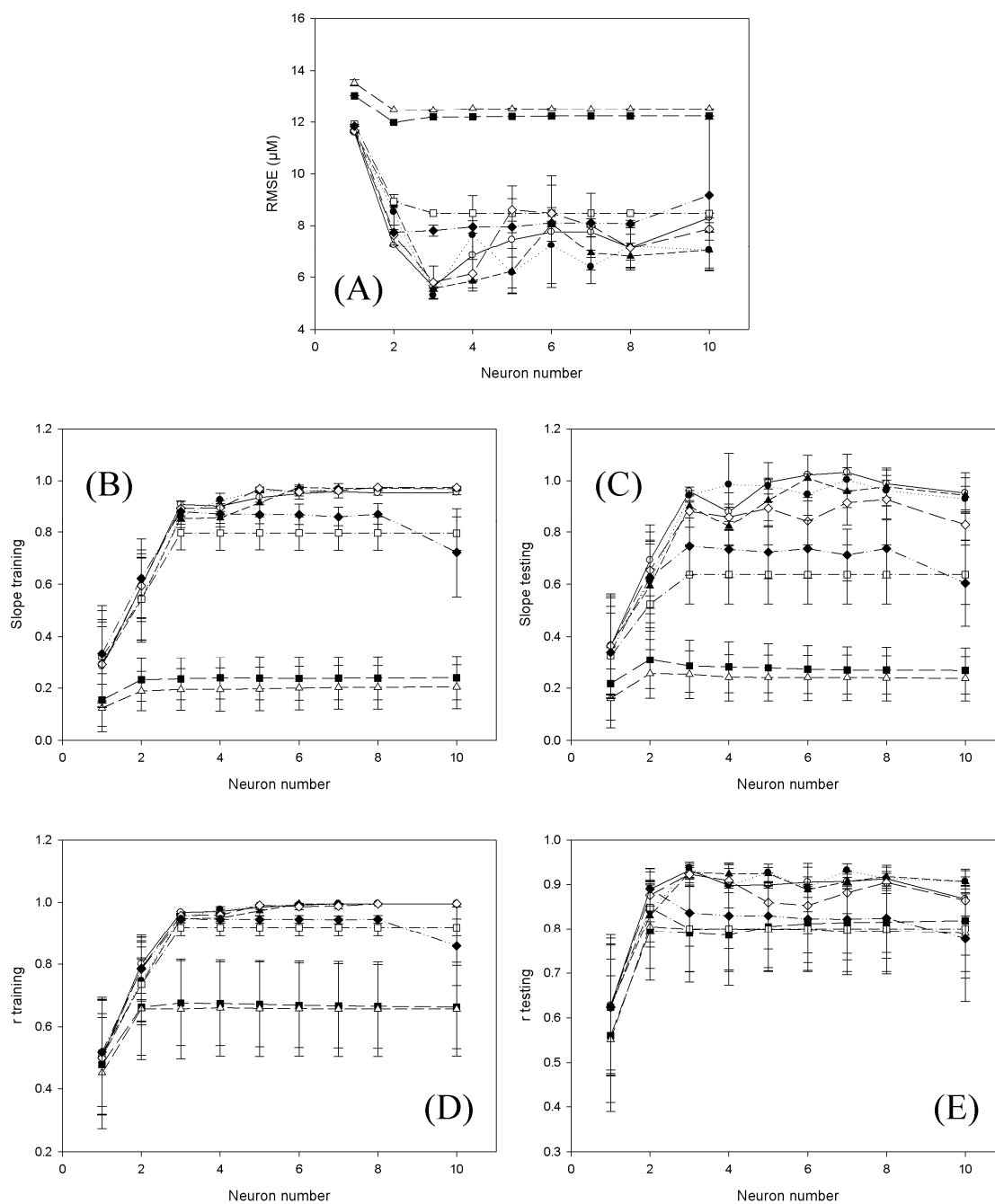


Figure S2. Detailed results of the ANN optimization. Obtained RMSEs (A), slopes (B, C) and correlation coefficient values (D, E) of obtained vs expected comparison graphs (both for training and testing sets respectively) are plotted against different transfer function combinations and number of neurons in the hidden layer-output layer. L:logsig, P:purelin, S:satlins and T:tansig; T-T (—○—), T-L (—■—), T-P (---▲---), L-L (---△---), L-P (.....●.....), P-P (---□---), L-S (—◇—), P-S (---◆---).

Article 3

Determination of total polyphenol index in wines employing a voltammetric electronic tongue

X. Cetó, J. M. Gutiérrez, M. Gutiérrez, F. Céspedes, J. Capdevila, S. Mínguez, C. Jiménez-Jorquera and M. del Valle

Analytica Chimica Acta, **732**(0): 172-179



Determination of total polyphenol index in wines employing a voltammetric electronic tongue

Xavier Cetó^a, Juan Manuel Gutiérrez^b, Manuel Gutiérrez^c, Francisco Céspedes^a, Josefina Capdevila^d, Santiago Mínguez^d, Cecilia Jiménez-Jorquera^c, Manel del Valle^{a,*}

^a Sensors and Biosensors Group, Department of Chemistry, Universitat Autònoma de Barcelona, Edifici Cn, 08193 Bellaterra, Spain

^b Bioelectronics Section, Department of Electrical Engineering, CINVESTAV, 07360 Mexico D.F., Mexico

^c Instituto de Microelectrónica de Barcelona (IMB-CNM), CSIC, 08193 Bellaterra, Spain

^d Estació de Viticultura i Enologia, INCAVI, Vilafranca del Penedès, Spain

ARTICLE INFO

Article history:

Received 1 September 2011

Received in revised form 18 January 2012

Accepted 13 February 2012

Available online 22 February 2012

Keywords:

Voltammetric sensors

Electronic tongue

Artificial neural network

Wine analysis

Folin–Ciocalteu

Polyphenol

ABSTRACT

This work reports the application of a voltammetric electronic tongue system (ET) made from an array of modified graphite-epoxy composites plus a gold microelectrode in the qualitative and quantitative analysis of polyphenols found in wine. Wine samples were analyzed using cyclic voltammetry without any sample pretreatment. The obtained responses were preprocessed employing discrete wavelet transform (DWT) in order to compress and extract significant features from the voltammetric signals, and the obtained approximation coefficients fed a multivariate calibration method (artificial neural network-ANN-or partial least squares-PLS-) which accomplished the quantification of total polyphenol content. External test subset samples results were compared with the ones obtained with the Folin–Ciocalteu (FC) method and UV absorbance polyphenol index (I_{280}) as reference values, with highly significant correlation coefficients of 0.979 and 0.963 in the range from 50 to 2400 mg L⁻¹ gallic acid equivalents, respectively. In a separate experiment, qualitative discrimination of different polyphenols found in wine was also assessed by principal component analysis (PCA).

© 2012 Elsevier B.V. All rights reserved.

1. Introduction

Wine analysis is an important analytical field, with emphasis placed on new methodologies for its characterization and elaboration control [1]. One important wine parameter, determining some organoleptic and sensorial properties, is its polyphenol content [2]. Phenolics in plants may act as phytoalexins, antifeedants, and attractants for pollinators, contributors to plant pigmentation, antioxidants and protective agents against UV light, among others [3]. Meanwhile in food, phenolics may contribute to the bitterness, astringency, color, flavor, odor and oxidative stability of food. In addition to their health-protecting capacity and some properties other than nutritional of plants, phenolics are of great importance to both consumers and producers.

In grapes, synthesis of polyphenols is induced by factors such as grape cultivar, developmental stage of the berry and maturation, climatology and UV radiation and viticultural practices; it can also be affected by fungal infection (*Botrytis cinerea*) and injuries [4–6]. In wine, phenolics have an important role in pigmentation (in red,

but also in white wines), aging, oxygen-depleting compounds and the bitter and astringent features, which are determinants for the wine taste and character.

Several methods to quantify total phenols and polyphenols have been described in the literature [7]. The Folin–Ciocalteu (FC) method is widely employed in the wine industry [8]. This spectrophotometric method measures the sample reducing capacity. As a faster alternative to this method, the use of what is known as polyphenol index (I_{280}) is now arising [9]; in this case wine absorbance is measured directly at 280 nm and straightly correlated with phenolic content. Another spectrophotometric method widely used is the reaction with 4-aminoantipyrine [10], a generic reaction for phenols. On the one hand, these methods yield a total phenol content value, and therefore do not allow for the discrimination between individual constituents. However there are some spectrophotometric methods, developed for quantification of phenolics in plants, that bring some specificity; these assays are based on differential reactivity principles and are used to determine different structural groups concomitant in phenolic compounds [11]. On the other hand, there is also the use of chromatographic techniques such as HPLC [12] or GC [13], which are able to perform individual determination of phenolic compounds; however these methods need complex and time-consuming sample pre-treatment

* Corresponding author. Tel.: +34 93 5811017; fax: +34 93 5812477.

E-mail address: manel.delvalle@uab.cat (M. del Valle).

procedures, and are not suitable for on-site analyses. Nowadays, biosensors are arising as an alternative to the traditional techniques given their low cost and their ease of use to carry out on field analyses. Therefore, biosensors represent an attractive alternative for the detection of polyphenolic compounds. To this end, biosensors have been developed incorporating enzymes such as laccase [14], tyrosinase [15] or peroxidase [16], and even by the coimmobilization of two enzymes in the same biosensor [17].

The present work reports the application of an electronic tongue (ET) in the analysis of polyphenols. The ET is a recent trend from the sensory field, which entails the use of an array of sensors capable of giving a wide and complete response of the analyzed species, plus a chemometric processing tool able to interpret the chemical signals and extract meaningful data from the complex readings [18,19]. Although there are a number of papers related to ETs and wine, they have been previously applied to distinguish different wine varieties and tastes [20–22], to quantify some analytical parameters of wines [23–25] or even to detect wine adulterations [26,27]. Potentiometric sensors have been used as well, and their signal correlated to polyphenol content of wines [28].

The ET is, in this context, the coupling between sensors and the use of chemometrics, such as artificial neural networks (ANNs); these advanced signal processing variants allow the interpretation, modelling and calibration of complex analytical signals [29,30]. Despite the great advantages provided by the use of ANNs, when voltammetric sensors are used, the high complexity of the generated data matrix makes their treatment extremely difficult. The straightforward solution is the use of multiway processing methods (samples \times sensors \times polarization potential), but the complexity of this technique is also critical [31,32]. One solution when dealing with a set of voltammograms is to employ a preprocessing stage for data reduction; this option permits an advantageous reduction in training time, to avoid redundancy in input data and to obtain a model with better generalization ability. This compression stage may be achieved by the use of methods such as principal component analysis (PCA), “kernels” [33] or discrete wavelet transform (DWT) [34]. In this sense, the latter is particularly useful thanks to its ability to compress and denoise data.

Here we report a multidimensional ET aimed at the determination of the total polyphenol content, formed by an array of six voltammetric sensors: a Au microelectrode plus five epoxy-graphite composite sensors, four of them bulk-modified with cobalt phthalocyanine, different nanoparticles (Pt and Cu) and polypyrrole powder (a conducting polymer). The array is designed in this manner to produce differentiated catalytic responses towards specific phenolic compounds present in wine [27]. The obtained responses were preprocessed employing DWT in order to extract the significant information present and the resulting approximation coefficients fed different multivariate calibration models (ANN or PLS-2 models), specially trained to predict polyphenol indexes, in what might be considered a bioinspired analytical method using artificial intelligence tools.

2. Experimental

2.1. Samples under study

A total of 20 wine samples of different varieties from 2008 and 2009 vintage were analyzed; all of them were from the Penedès region (Catalonia, Spain). Samples were selected as to obtain a set with sufficiently differentiated total polyphenol indexes and grape varieties, with values ranging from 54 to 2374 mg L⁻¹ gallic acid equivalents. Table 1 summarizes detailed information about the wines used.

Table 1
Detailed information of the wine samples under study.

| Sample | Grape variety | FC Index (mg L ⁻¹) | I ₂₈₀ (arb. unit) |
|--------|--------------------|--------------------------------|------------------------------|
| S1 | Trepat | 149 | 7.53 |
| S2 | Trepat | 54 | 3.78 |
| S3 | Ull de llebre | 1199 | 27.25 |
| S4 | Picapoll negre | 933 | 31.28 |
| S5 | Marselan | 1598 | 30.64 |
| S6 | Ull de llebre | 1239 | 37.39 |
| S7 | Cabernet sauvignon | 1467 | 42.47 |
| S8 | Merlot | 1583 | 41.08 |
| S9 | Cabernet sauvignon | 1366 | 46.97 |
| S10 | Petit verdot | 1645 | 45.72 |
| S11 | Cabernet sauvignon | 1468 | 46.27 |
| S12 | Cabernet sauvignon | 1392 | 46.44 |
| S13 | Malbec | 1351 | 48.51 |
| S14 | Sumoll | 1613 | 59.61 |
| S15 | Merlot | 1358 | 47.46 |
| S16 | Merlot | 1355 | 46.10 |
| S17 | Cabernet sauvignon | 1913 | 58.58 |
| S18 | Cabernet sauvignon | 2054 | 63.73 |
| S19 | Petit verdot | 2153 | 69.80 |
| S20 | Caladoc | 2374 | 72.91 |

2.2. Reagents and solutions

All reagents used were analytical reagent grade and all solutions were prepared using deionized water from a Milli-Q system (Millipore, Billerica, MA, USA). Gallic acid, catechin, *p*-coumaric acid, caffeic acid, catechol, *m*-cresol, copper and platinum nanoparticles (<50 nm), polyaniline and polypyrrole were purchased from Sigma–Aldrich (St. Louis, MO, USA). KCl was purchased from Merck KGaA (Darmstadt, Germany). Folin–Ciocalteu’s reagent and sodium carbonate were purchased from Panreac Química (Barcelona, Spain). A solution containing 0.1 M KNO₃ (Fluka) was used to activate the gold (Au) microelectrode.

2.3. Determination of polyphenol content by Folin–Ciocalteu method and spectrophotometric measurements

For comparative purposes, the Folin–Ciocalteu index of the wines was also determined spectrophotometrically [35]. The FC test was carried out according to the following procedure: 1 mL of sample (wines were diluted 1:100 or 1:50), 6 mL of deionized water, 0.5 mL of Folin–Ciocalteu reagent and 2 mL of a 20% sodium carbonate solution were added in this order to a 10 mL beaker and diluted to volume with deionized water. The resulting solution was stirred and allowed to react for half an hour at room temperature in darkness. The absorbance was then read at 760 nm using a spectrophotometer PerkinElmer Lambda 20 UV/VIS (MA, USA). Total phenolic content, expressed in gallic acid equivalents, was evaluated from the absorbance value by interpolation into the calibration plot obtained with gallic acid standard solutions, multiplying the resulting value by 10 and by the proper dilution factor. Different dilution factors were applied given when carrying out spectrophotometric measurements absorbance value should be around 0.3 [35].

For the determination of the total content of polyphenolic compounds in wines the polyphenol index I₂₈₀ was also considered: wine was diluted with water (1:100 or 1:50) and the absorbance was measured directly at 280 nm. The value of I₂₈₀ for each sample was given as the absorbance multiplied by the proper dilution rate. In both methods the blank solution was a hydro-alcoholic solution (12%, v/v ethanol) of tartaric acid 3 g L⁻¹.

2.4. Electrochemical measurements

Samples were analysed with two types of amperometric sensors: an array of 5 graphite-epoxy voltammetric sensors made with

different modifiers added to the bulk conductive mixture, selected according with previous experiments in our laboratory [27,36], and a microfabricated Au microelectrode [37,38]. Electroanalytical experiments were carried out at room temperature (25 °C) under quiescent conditions, and without any pretreatment or dilution of the sample. Electrodes were cycled in saline solution in order to get stable voltammetric responses before performing the measurements with real samples.

2.4.1. Composite electrodes

Working electrodes were prepared following the conventional methodology previously established in our laboratories [39]. A resin EpoTek H77 (Epoxy Technology, Billerica, MA, USA) and its corresponding hardener compound were mixed in the ratio 20:3 (w/w); afterwards a 15% of graphite (w/w) and a 2% of the modifier (w/w) were added to the previous mixture before hardening, obtaining the composite. Then, it was manually homogenized for 60 min, and afterwards, the paste was allowed to harden for three days at 80 °C. Finally, the electrode surface was polished with different sandpapers of decreasing grain size, with a final electrode area of 28 mm².

In this manner, an array of 5 different graphite-epoxy voltammetric sensors was prepared using bare graphite C and adding different modifiers such as cobalt phthalocyanine, conducting polymer as polypyrrole and nanoparticles of copper and platinum to the bulk mixture – one component per electrode.

2.4.2. Gold microelectrode

Also a conventional Au microelectrode fabricated according to standard photolithographic techniques was used [40]. The Au microelectrode was firstly chemically cleaned, successively with ethanol 96%, H₂SO₄ 6.0 M and deionized water. Next, an electrochemical activation was carried out in 0.1 M KNO₃, where the electrode was cycled from +0.8 V to –2.2 V for at least 20 times.

2.4.3. Amperometric measurements

The amperometric measurement cell was formed by the 6-sensor voltammetric array and a reference double junction Ag/AgCl electrode (Thermo Orion 900200, Beverly, MA, USA) plus a commercial platinum counter electrode (Model 52–67, Crison Instruments, Barcelona, Spain). Cyclic Voltammetry measurements were taken using a 6-channel AUTOLAB PGSTAT20 (Ecochemie, Netherlands), in a multichannel configuration, using GPES Multichannel 4.7 software package.

Potential was cycled between –1.0 V and 1.3 V vs. Ag/AgCl (–0.5 V and 1.6 V for the Au microelectrode), with a scan rate of 100 mV s^{–1} and a step potential of 9 mV. Apart, all experiments were carried out without performing any physical surface regeneration of the working electrodes. In order to prevent the accumulative effect of impurities on the working electrode surfaces, an electrochemical cleaning stage was done between each measurement applying a conditioning potential of +1.5 V for a duration of 40 s after each experiment, in a cell containing 25 ml of distilled water [36].

2.5. Spiked samples

In addition to the wine samples summarized in Table 1, some spiked samples were prepared in order to assess the discrimination ability shown by the sensor array. In this sense, the same reference wine was spiked with different quantities of a stock solution of certain polyphenols. The polyphenols considered were selected according to major polyphenols found in wine [41,42]; gallic acid, catechin, *p*-coumaric acid, caffeic acid, catechol and

m-Cresol were the ones tested. It must be reckoned that average total polyphenol content measured by the Folin method is higher than the individual content of each polyphenol [41,43]. For example, total polyphenol content is around 2160 mg L^{–1} for red wine, 820 mg L^{–1} in rosé wine and 320 mg L^{–1} for white wines; however individual polyphenol concentration is much lower with values under 300 mg L^{–1} for red wine and 50 mg L^{–1} for white wines [43].

In this manner, a rosé wine was chosen as the one to which phenolic compounds would be added given their lower polyphenolic content. Then, from it, 35 different samples were prepared based on seven groups (one for each of the 6 polyphenols under study plus another for the unspiked wine). To confirm that differentiation between compounds was not performed due to different amounts of phenolic compounds being added, in all the cases the same amount was used. Thus, for each class, 5 μmol of polyphenolic compound were added to 25 mL of wine, which represents approximately an increase of 36 mg L^{–1} (200 μM).

2.6. Data processing

Chemometric processing was done by specific routines in MATLAB 7.0 (MathWorks, Natick, MA) written by the authors, using Neural Network and Wavelet Toolboxes (v.4.0). Partial least squares (PLS) regression was done employing The Unscrambler (CAMO Software AS, Oslo, Norway) informatics package. Sigmaplot 2000 (Systat Software Inc, California, USA) was used for graphic representations of data and results.

The whole cyclic voltammograms obtained from each sensor were included in the data processing stage. In order to reduce the multidimensional data matrix generated in each measurement, a preprocessing stage employing the discrete wavelet transform was used [44]. In this way, and using the proposed sensor array, the corresponding compressed voltammograms were processed employing either artificial neural networks or partial least squares (PLS-2) models. Both allowed the carrying out of the quantification of the FC index and the *I*₂₈₀ values.

In order to find the appropriate ANN model, significant effort is needed to optimize the configuration details that determine its operation. Normally, this is a trial-and-error process, where several parameters (training algorithms, number of hidden layers, transfer functions, etc.) are fine-tuned in order to find the best configuration to optimize the performance of the model.

Given the complexity of the data set and for comparison purposes, also a linear model like PLS-2 was evaluated; PLS-2 was chosen given its ability to model several variables together, unlike PLS-1 or PCR. Therefore the obtained model would be equivalent to the one built with ANN. In this case, the parameter that must be taken into account for its optimization are the number of latent variables used to build the PLS-2 model; to optimize this parameter, the lower value of the total normalized root mean square error (NRMSE) reached by the model for predicted values was used. Also, as done in the case of the ANN, voltammetric signals were preprocessed employing DWT before employing PLS due to this data reduction improves model prediction and generalization abilities [45].

Given the reduced size of the data set, a jack-knife method was used [46,47]. The jack-knife method was originally suggested in statistics as a general approach for testing hypotheses and calculating confidence intervals in situations where apparently no better methods could be used. With this procedure, standard errors are calculated from different resampling with random distribution and repeating the modelling stage. In this manner, it may be used either leaving one sample, or even many samples, out each time [47].

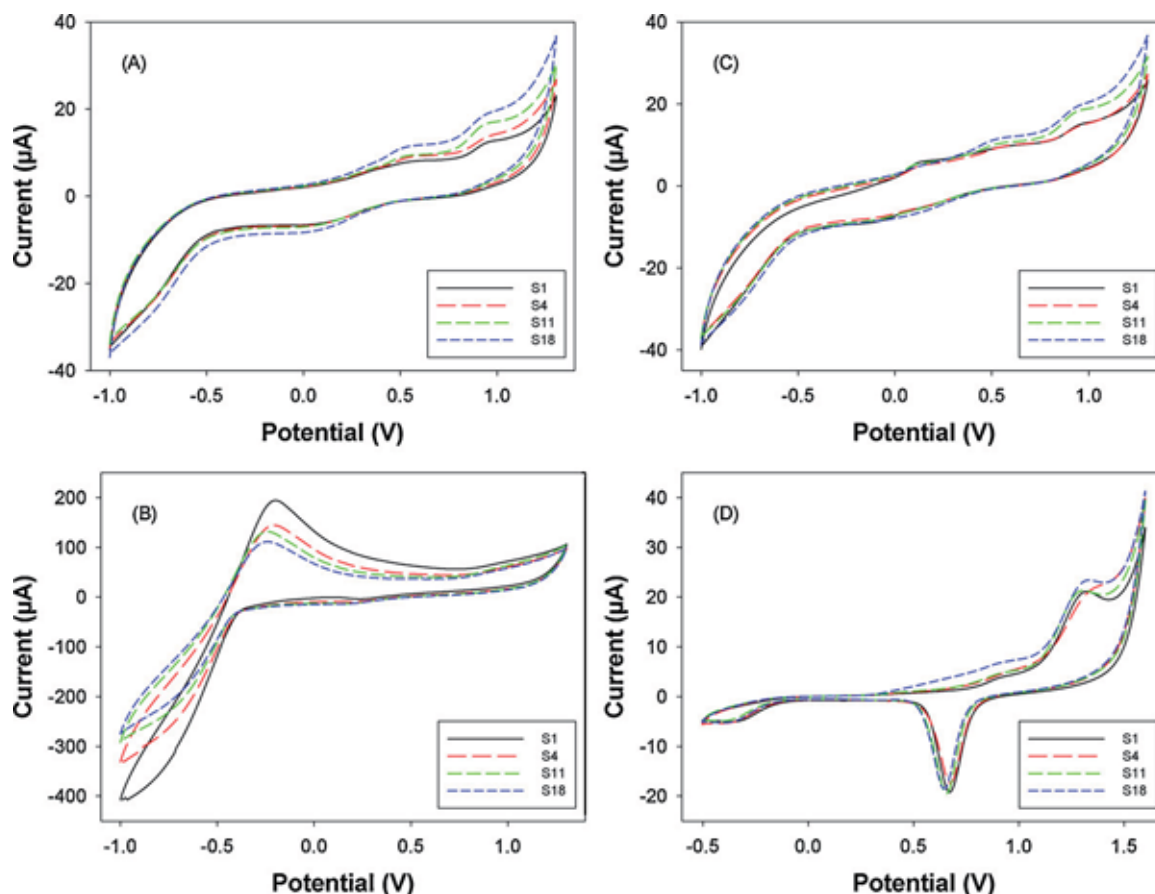


Fig. 1. Some cyclic voltammograms obtained with the sensor array for some of the wine samples with different Folin–Ciocalteu index: (S1) 149 mg L⁻¹, (S4) 933 mg L⁻¹, (S11) 1468 mg L⁻¹ and (S18) 2054 mg L⁻¹ gallic acid equivalents. Also, signals from different sensors are shown: (A) Graphite-epoxy sensor, (B) Platinum nanoparticle sensor, (C) Copper nanoparticle sensor and (D) Au microelectrode.

In our case, it was applied using 15 samples for the training process and leaving 5 samples out each time in order to evaluate model's response; then, this subdivision of the original data set was repeated 25 times leaving 5 different samples out each time, which were selected randomly. Once all the responses from the models were obtained, data was grouped depending if it was intervening in the training process or used in the external test subset.

Also, principal component analysis (PCA) was used to visualize the discrimination capability towards individual polyphenols derived from the voltammetric signals acquired, and grouped with a cluster analysis tool to build a preliminary recognition model. PCA allows the projection of the information carried by the original variables onto a smaller number of underlying ("latent") variables called principal components (PCs) with new coordinates called scores, obtained after data transformation. Then by plotting the PCs, one can view interrelationships between different variables, and detect and interpret sample patterns, groupings, similarities or differences [27]. Moreover, PCA is a useful method to reduce the dimensionality of large data sets, such as those from voltammetric sensor arrays.

Furthermore, cluster analysis tools allow constructing clusters from data using pairwise distance between observations (usually Euclidean distance), as linkage to construct the hierarchical cluster tree [48]. Thus, the set of objects are assigned into groups so that the objects in the same cluster are more similar (in some sense or another) to each other than to those in other clusters. Clustering is a main task of explorative data mining, and a common technique for statistical data analysis used in many fields, including pattern recognition.

3. Results and discussion

3.1. Voltammetric responses and feature extraction

Different voltammetric responses are observed for each kind of sensor, as shown in Fig. 1. These signals presumably contribute in different manners with the data needed for model quantification. Differentiated signals are obtained for each type of sensor used. Catalytic oxidative signals seem to originate from the sensors using metal nanoparticles, which may be due to a catalytic oxidation of saccharides and/or polyphenols. Similarly, sensors modified with conducting polymers bring new information with completely different waveforms. Gold electrode would provide the most generic redox behaviour of the sample. At the same time, it can be seen how currents increase in the same way as FC index increases and with different behaviour for each sensor; meanwhile each sensor shows its distinctive profile, generating very rich data that is very useful as the departure point. For developing an ET, the first necessary condition is to have analytical signals responding to the phenomena to which the objective is aimed, with variability among them and the different sensors forming the sensor array. As demonstrated, bulk modification of voltammetric electrodes is an easy way to enrich the cross-response of the sensor array to different aspects of the solution under scrutiny [27]. However, the extreme complexity of the generated signals (the set of voltammograms) hinders the processing step. As already commented, all these data is pretreated using a compression step, required to gain advantages in training time, to avoid redundancy in input data and to obtain a model with better generalization ability. An additional reason is the need

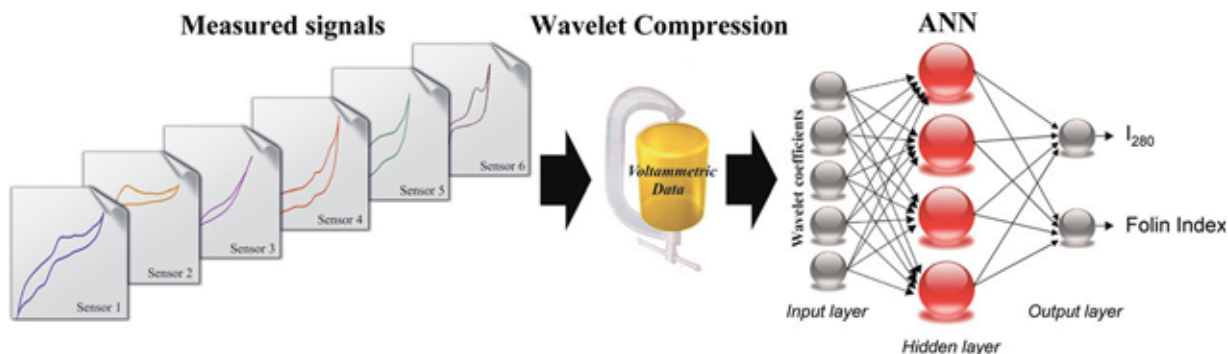


Fig. 2. Processing scheme for building the quantification model. After optimization of the ANN, the final model had 84 approximation coefficients obtained from wavelet transform of the sensor signals in the input layer, 10 neurons and *logsig* transfer function in the hidden layer and 2 neurons and *purelin* transfer function in the output layer.

to extract reduced and significant information compatible with the ANN structure [44]. This was accomplished by the use of the DWT.

3.2. Estimation of polyphenol indexes

A DWT preprocessing stage was performed employing Daubechies wavelet and a sixth decomposition level, the best choice from preliminary tests and previous experience [34,44]. The DWT allowed compressing the original data set information up to 97.27% without any loss of relevant information. From the proposed 6-sensor array, the corresponding voltammograms were compressed, and the obtained coefficients were fed into multivariate calibration methods in order to predict the total polyphenol content in wines. In this sense, two different methods were evaluated: an ANN as a non-linear data modelling tool and PLS-2 as a linear one. A simplified scheme of the procedure followed for data treatment and ANN architecture can be observed in Fig. 2.

After a systematic evaluation of topologies, the final DWT-ANN architecture model had 84 input neurons (corresponding to the 14 wavelet approximation coefficients obtained from wavelet analysis of each of the 6 sensor signals), 10 neurons and *logsig* function in the hidden layer and 2 output neurons and *purelin* function in the output layer corresponding to FC and I_{280} indexes. Although, individual ANN models may also be developed to separately predict these two indexes. Bayesian regularization was used to train the network, this algorithm has the advantage of avoiding

overfitting without the need of an internal validation subset [44], and consequently this precaution is not performed.

For the optimization of PLS model only two considerations were taken into account: the differences obtained when using the 84 DWT coefficients instead of the raw voltammetric data and the number of latent variables used to build the model. Despite PLS has no need of a preprocessing stage, it was found that better models were obtained when this was performed (total NRMSE improved from 0.31 to 0.20 for predicted values). Thus, the final model was a DWT-PLS2 with 7 latent variables, which has a total explained variance ca. 95.2%.

After building the models, which were evaluated for training with 75% of the data and tested with the remaining 25%, a jack-knife method was employed to visualize the dependence of predictions from the specific subdivision of data. Then, in order to characterize the accuracy of the identification models and obtain unbiased data, train/test data division was repeated randomly 25 times using a five-fold cross validation process; this precaution ensured that performance indicators were independent of the subsets used.

Comparison graphs for each model were built grouping the replicas for each individual sample, differentiating when it was intervening in the training process and when used as external test. The predicted indexes were then plotted against the expected ones and fitted with linear least-squares regression. To give the same weight to all points independently of the phenol index value, weighed regression was used. The obtained results for the ANN model can be seen on Fig. 3 for the Folin-Ciocalteu Index and Fig. 4,

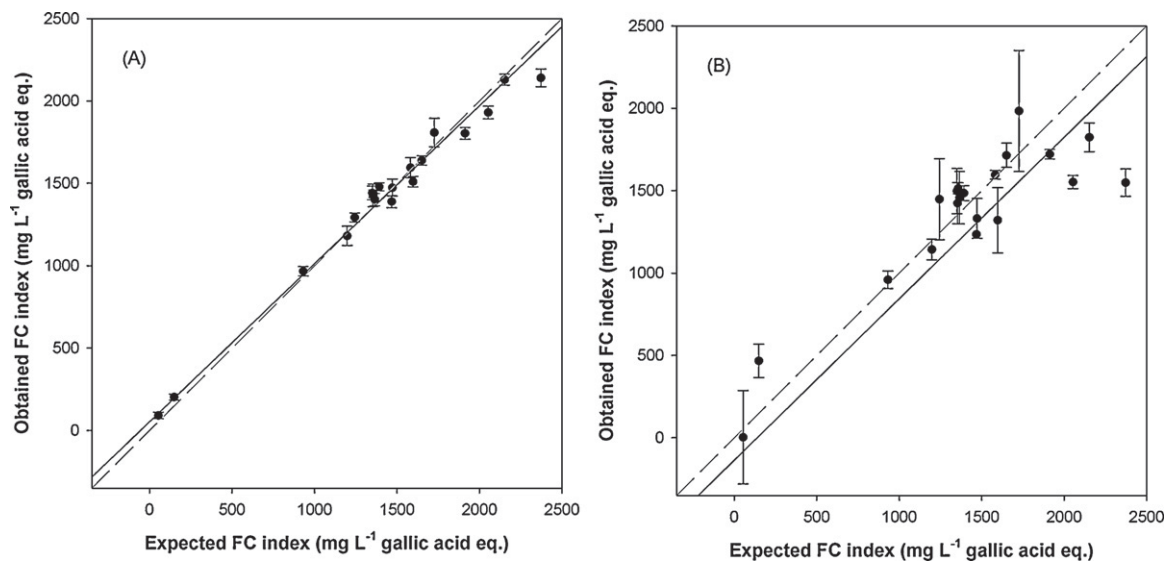


Fig. 3. Modelling ability of the optimized ANN. Comparison graph of expected vs. obtained concentrations for Folin-Ciocalteu indexes, both for (A) training and (B) testing subset. Dashed line corresponds to theoretical diagonal line.

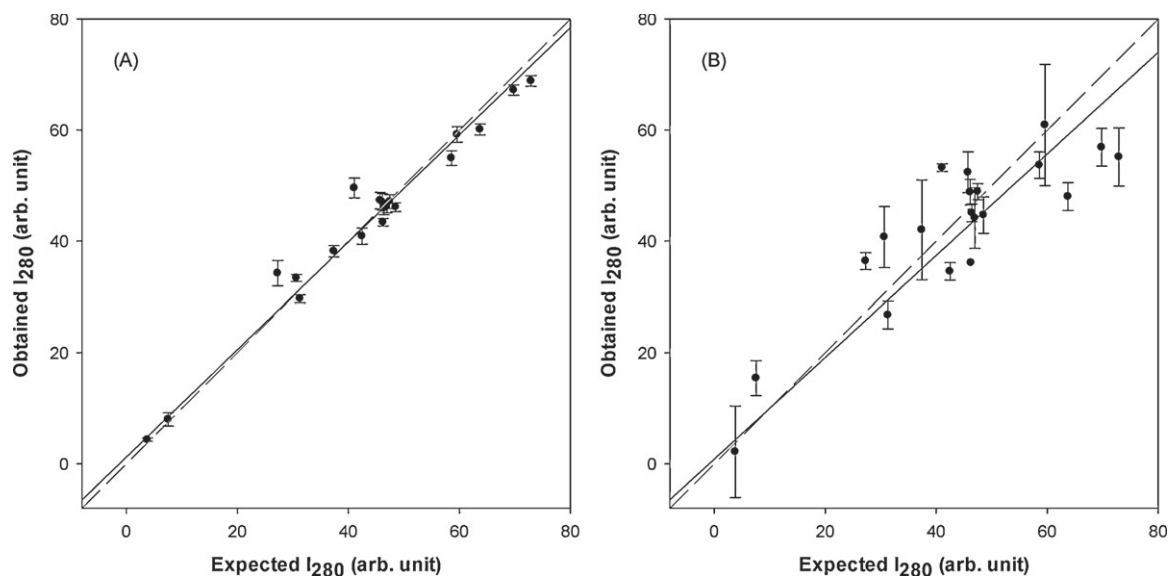


Fig. 4. Modelling ability of the optimized ANN. Comparison graph of expected vs. obtained concentrations for polyphenol indexes (I_{280}), both for (A) training and (B) testing subset. Dashed line corresponds to theoretical diagonal line.

for the I_{280} index. In the same way, results from the DWT-PLS-2 model can be seen in Fig. 5 for the FC index, and Fig. 6 for the I_{280} index. In both cases, an estimation of prediction errors (intervals calculated at the 95% confidence level) corresponding to the average of the 25 jack-knife calculations, taken from the dispersion of the replicas, can be also visualized.

As usual in multivariate calibration methods models, training stage behaviour shows better agreement and reduced errors than the testing stage. The general behaviour is also good in both cases, although individual prediction errors are largely increased, especially for certain samples that were more dependant on the specific data used in training. Despite the good trend in both cases, larger uncertainties were obtained with the DWT-PLS-2, thus meaning that the ANN model is slightly more robust. Regardless of this, Table 2 presents the results of the weighed regressions, where the good correlation coefficients and the small associated errors stand out (intervals calculated at the 95% confidence level). The obtained comparison results are close to the ideal values, with intercepts

near to 0 and slopes and correlation coefficients around 1, meaning that there are no significant differences between the values predicted by the multivariate calibration methods and the reference ones.

3.3. Classification of polyphenols

Employing the same voltammetric array and following equivalent procedure for the data treatment, spiked wine samples with typical polyphenolic compounds present in wine, as described in Section 2.5, were measured as before but at random. An extract of the results is shown on Fig. 7, corresponding to the polypyrrole sensor; this sensor was selected for being the one where the most clearly differentiated cathodic response is obtained for each compound. Then the information obtained from the complete set of voltammograms was evaluated using PCA and groups were formed using cluster analysis tools.

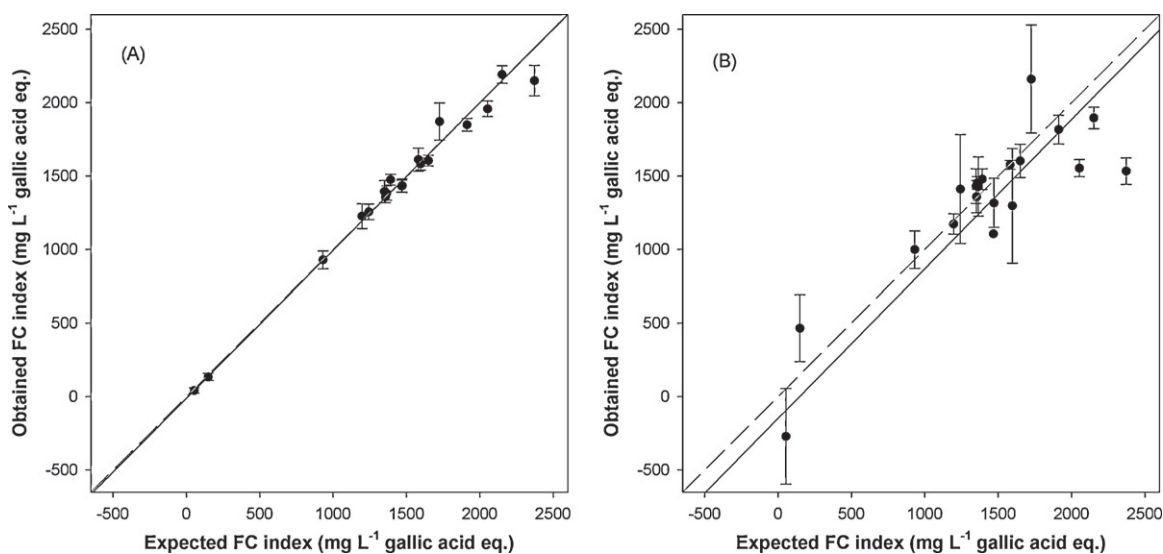


Fig. 5. Modelling ability of the PLS model. Comparison graph of expected vs. obtained concentrations for Folin-Ciocalteu indexes, both for (A) training and (B) testing subset. Dashed line corresponds to theoretical diagonal line.

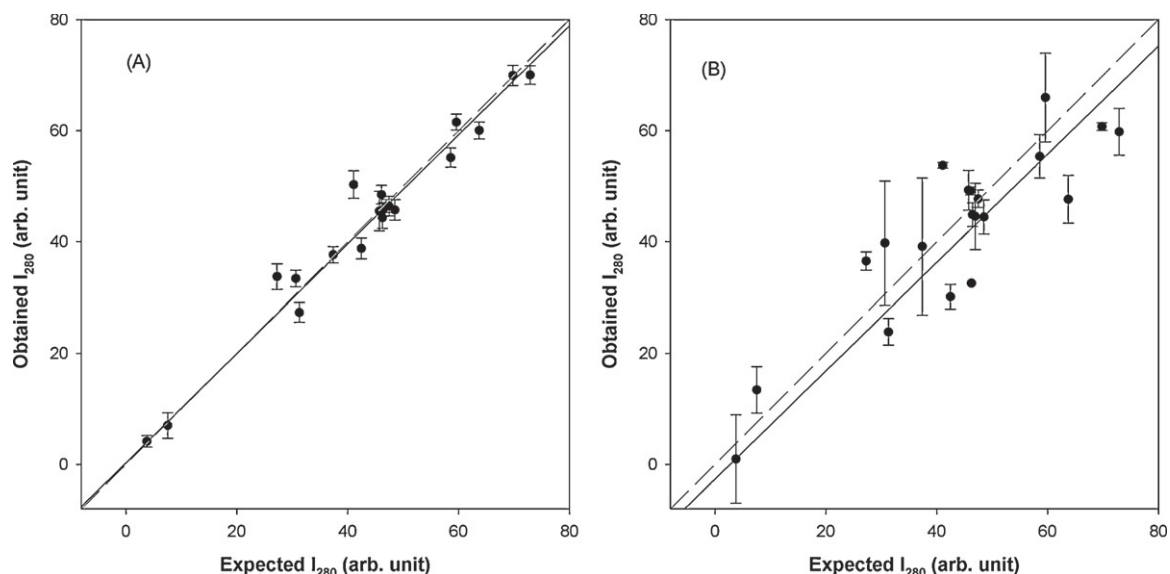


Fig. 6. Modelling ability of the PLS model. Comparison graph of expected vs. obtained concentrations for polyphenol indexes (I_{280}), both for (A) training and (B) testing subset. Dashed line corresponds to theoretical diagonal line.

Table 2
Fitted regression lines of the comparison between obtained vs. expected results provided by the proposed ET, both for Folin–Ciocalteu and I_{280} indexes, averaging each value per sample as pertaining to the training or external test subsets for the 25 replicated calculations.

| Folin–Ciocalteu index | | | | | | | | |
|-----------------------|-------------|-------------------|------------------|------|-------------|-------------------|------------------|------|
| ANN | | | | | PLS-2 | | | |
| | Correlation | Slope | Intercept | RMSE | Correlation | Slope | Intercept | RMSE |
| Train subset | 0.998 | 0.959 ± 0.029 | 53.0 ± 30.6 | 89 | 0.999 | 1.006 ± 0.020 | -11.8 ± 3.6 | 92 |
| Test subset | 0.979 | 0.952 ± 0.093 | -43.6 ± 28.7 | 366 | 0.932 | 1.017 ± 0.184 | -149 ± 162 | 404 |
| I_{280} Index | | | | | | | | |
| ANN | | | | | PLS-2 | | | |
| | Correlation | Slope | Intercept | RMSE | Correlation | Slope | Intercept | RMSE |
| Train subset | 0.994 | 0.965 ± 0.049 | 1.24 ± 1.68 | 3.5 | 0.995 | 0.983 ± 0.047 | 0.30 ± 0.71 | 3.8 |
| Test subset | 0.963 | 0.914 ± 0.119 | 0.87 ± 3.77 | 10.1 | 0.968 | 0.974 ± 0.118 | -2.64 ± 0.69 | 9.5 |

The PCA analysis was done, and with the two first PCs, the explained variance accumulated was ca. 95%; this large value means that nearly all the variance contained in the original data is explained by just using the first new coordinates. By plotting them,

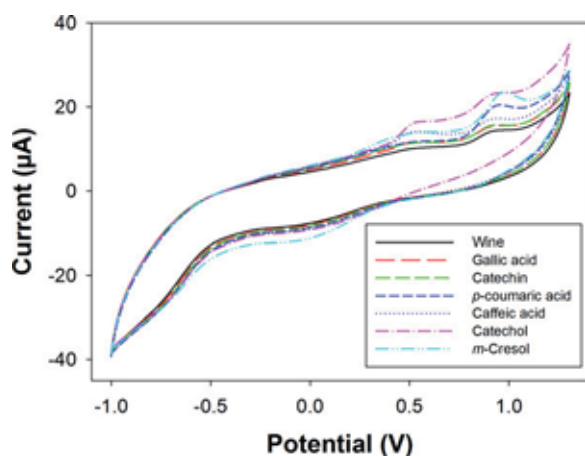


Fig. 7. Example of cyclic voltammograms obtained with polypyrrole sensor (one of the sensors from the sensor array) for some of the spiked samples.

different clusters were obtained as can be seen in Fig. 8; patterns in the figure show evidence that wine samples are grouped based on which polyphenol was added. These well established clusters clearly separate the main classes of samples corresponding to: (I) Wine, (II) Gallic acid, (III) Catechin, (IV) *p*-Coumaric acid, (V) Caffeic acid, (VI) Catechol and (VII) *m*-Cresol.

The fact that gallic acid and catechin clusters were so close to one another, and also to the wine cluster, is due to the fact that these two compounds are the ones present in higher concentrations in wine. As expected, *p*-Coumaric and caffeic acids clusters were more separated from the wine cluster and in close proximity to each other, given both are hydroxycinnamic acids differing only by one alcohol group. Finally, *m*-Cresol and Catechol were clearly far away from the rest of clusters and mainly represented by PC2. The position of these clusters could be explained by the fact that these components are found in lesser proportions in wine and their chemical structure is quite different from the others.

In this interpretation of polyphenol identification, we could see that individual groups were obtained for each phenolic compound; moreover, compounds with structural similarity arranged closely together, meaning there is some background relationship. Although the good results obtained with those spiked samples, it should be taken into account that the quantification of individual polyphenols

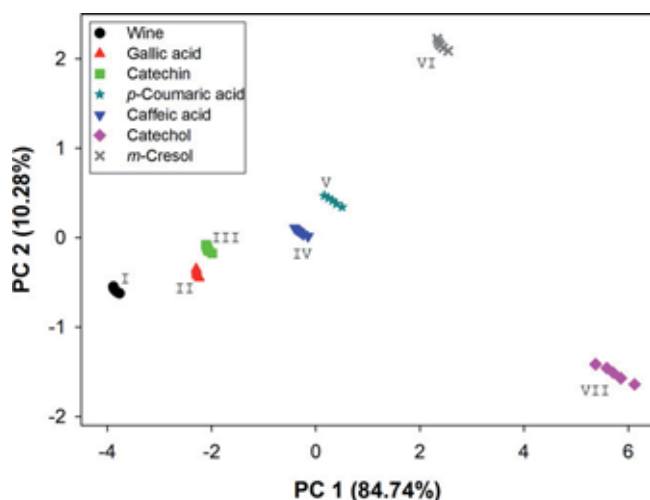


Fig. 8. Score plot of the first two components obtained after PCA of the spiked samples. A total of 35 samples were analysed. As can be seen, clear discrimination is obtained for the different polyphenols: (I) Wine, (II) Gallic acid, (III) Catechin, (IV) *p*-Coumaric acid, (V) Caffeic acid, (VI) Catechol and (VII) *m*-Cresol.

in a complex sample such as wine is very difficult, but the results obtained suggest this possibility. In essence, the study demonstrates the capabilities of the proposed ET, both in the quantification of total polyphenol content and in the differentiation of different polyphenols found in wine.

4. Conclusions

An electronic tongue based on voltammetric sensors with different modifiers (metallic nano-sized particles, conducting polymers and cobalt phthalocyanine) was developed in order to create a tool capable of quantifying total polyphenol content in wines. Complex voltammetric data required the use of a preprocessing stage that was achieved by the use of DWT which provides a good compression of data preserving relevant information. The use of ANN or PLS-2 allowed the prediction of phenolic content indexes obtained with two different reference methods (Folin–Ciocalteu and I_{280} indexes). In a more qualitative application of the technique, we have been also able to differentiate between different polyphenols found in wine samples by the use of PCA, clearly discriminating them in an application comparable to much more complex analytical techniques such as HPLC.

Finally, the superior performance of ET combined with the use of chemometric tools such as ANN in tasks of multivariate calibration or pattern recognition has been again demonstrated, presenting the proposed ET an alternative to traditional methods for polyphenols quantification and even distinction of different polyphenols present in wine.

Acknowledgments

Financial support for this work was provided by Spanish Ministry of Science and Innovation MCINN (Madrid) through the projects TEC2007-68012-C03 and CTQ2010-17099, by the Mexican National Council of Science and Technology CONACYT (Mexico) through the project 154243 and by the Catalanian program ICREA Academia. X. Cetó thanks the support of *Departament d'Innovació, Universitats i Empresa de la Generalitat de Catalunya* for the predoctoral grant.

References

- [1] I.S. Arvanitoyannis, M.N. Katsota, E.P. Psarra, E.H. Soufleros, S. Kallithraka, *Trends Food Sci. Technol.* 10 (1999) 321.
- [2] P. Ribéreau-Gayon, Y. Glories, A. Maujean, D. Dubourdieu, *Handbook of Enology*, John Wiley & Sons, Ltd., Chichester (UK), 2006, pp. 141.
- [3] F. Shahidi, M. Naczk, *Phenolics in Food and Nutraceuticals: Sources Applications and Health Effects*, CRC, Press, FL, 2003.
- [4] L. Bavaresco, D. Petegolli, E. Cantù, M. Fregoni, G. Chiusa, M. Trevisan, *Vitis* 36 (1997) 77.
- [5] A. Versari, G.P. Parpinello, G.B. Tornielli, R. Ferrarini, C. Giulivo, *J. Agric. Food Chem.* 49 (2001) 5531.
- [6] E.H. Siemann, L.L. Creasy, *Am. J. Enol. Vitic.* 43 (1992) 49.
- [7] I. Ignat, I. Volf, V.I. Popa, *Food Chem.* 126 (2011) 1821.
- [8] V.L. Singleton, J.A. Rossi, *Amer. J. Enol. Vitic.* 16 (1965) 144.
- [9] E. Mataix, M.D.L. de Castro, *Analyst* 126 (2001) 251.
- [10] M. Fayyad, M. Tutunji, L. Abdel-Nour, *Dirasat Nat. Sci.* 10 (1983) 107.
- [11] M. Naczk, F. Shahidi, *J. Pharmaceut. Biomed.* 41 (2006) 1523.
- [12] H. Sakakibara, Y. Honda, S. Nakagawa, H. Ashida, K. Kanazawa, *J. Agric. Food Chem.* 51 (2002) 571.
- [13] G.J. Soleas, E.P. Diamandis, A. Karumanchiri, D.M. Goldberg, *Anal. Chem.* 69 (1997) 4405.
- [14] D.M.A. Gil, M.J.F. Rebelo, *Eur. Food Res. Technol.* 231 (2010) 303.
- [15] V.C. Sanz, M.L. Mena, A. Gonzalez-Cortes, P. Yanez-Sedeno, J.M. Pingarron, *Anal. Chim. Acta* 528 (2005) 1.
- [16] R. Solna, P. Skladal, *Electroanalysis* 17 (2005) 2137.
- [17] M.R. Montareali, L. Della Seta, W. Vastarella, R. Pilloton, *J. Mol. Catal. B-Enzym.* 64 (2010) 189.
- [18] M. del Valle, *Electroanalysis* 22 (2010) 1539.
- [19] Y. Vlasov, A. Legin, A. Rudnitskaya, C. Di Natale, A. D'Amico, *Pure Appl. Chem.* 77 (2005) 1965.
- [20] M.L. Rodríguez-Méndez, V. Parra, C. Apetrei, S. Villanueva, M. Gay, N. Prieto, J. Martínez, J.A. de Saja, *Microchim. Acta* 163 (2008) 23.
- [21] S. Buratti, S. Benedetti, M. Scampicchio, E.C. Pangerod, *Anal. Chim. Acta* 525 (2004) 133.
- [22] L. Pigani, G. Foca, K. Ionescu, V. Martina, A. Ulrici, F. Terzi, M. Vignali, C. Zanardi, R. Seeber, *Anal. Chim. Acta* 614 (2008) 213.
- [23] A. Legin, A. Rudnitskaya, L. Lvova, Y. Vlasov, C. Di Natale, A. D'Amico, *Anal. Chim. Acta* 484 (2003) 33.
- [24] M. Gay, C. Apetrei, I. Nevares, M. del Alamo, J. Zurro, N. Prieto, J.A. De Saja, M.L. Rodríguez-Méndez, *Electrochim. Acta* 55 (2010) 6782.
- [25] X. Cetó, F. Céspedes, M.I. Pividori, J.M. Gutiérrez, M. del Valle, *Analyst* 137 (2012) 349.
- [26] V. Parra, Á.A. Arrieta, J.-A. Fernández-Escudero, M.L. Rodríguez-Méndez, J.A. De Saja, *Sens. Actuators B-Chem.* 118 (2006) 448.
- [27] J.M. Gutiérrez, L. Moreno-Barón, M.I. Pividori, S. Alegret, M. del Valle, *Microchim. Acta* 169 (2010) 261.
- [28] G. Verrelli, L. Lvova, R. Paolesse, C. Di Natale, A. D'Amico, *Sensors* 7 (2007) 2750.
- [29] F. Despagne, D. Luc Massart, *Analyst* 123 (1998) 157R.
- [30] Y.N. Ni, S. Kokot, *Anal. Chim. Acta* 626 (2008) 130.
- [31] D. Ebrahimi, E. Chow, J.J. Gooding, D.B. Hibbert, *Analyst* 133 (2008) 1090.
- [32] R. Cartas, A. Mimendia, A. Legin, M. del Valle, *Electroanalysis* 23 (2011) 953.
- [33] R. Gutiérrez-Osuna, H.T. Nagle, *IEEE Sys. Man Cybern.* 29 (1999) 626.
- [34] L. Moreno-Barón, R. Cartas, A. Merkoçi, S. Alegret, M. del Valle, L. Leija, P.R. Hernandez, R. Muñoz, *Sens. Actuators B-Chem.* 113 (2006) 487.
- [35] Commission Regulation (EC), *Official Journal L 272*, Publication Office of the European Union, 17 September 1990, p. 0001.
- [36] X. Cetó, J.M. Gutiérrez, L. Moreno-Barón, S. Alegret, M. del Valle, *Electroanalysis* 23 (2011) 72.
- [37] M. Gutiérrez, A. Llobera, J. Vila-Planas, F. Capdevila, S. Demming, S. Buttgenbach, S. Mínguez, C. Jiménez-Jorquera, *Analyst* 135 (2010) 1718.
- [38] M. Gutiérrez, C. Domingo, J. Vila-Planas, A. Ipatov, F. Capdevila, S. Demming, S. Buttgenbach, A. Llobera, C. Jiménez-Jorquera, *Sens. Actuators B-Chem.* 156 (2011) 695.
- [39] F. Céspedes, E. Martínez-Fàbregas, S. Alegret, *TrAC-Trends Anal. Chem.* 15 (1996) 296.
- [40] J. Orozco, G. Suarez, C. Fernandez-Sanchez, C. McNeil, C. Jimenez-Jorquera, *Electrochim. Acta* 53 (2007) 729.
- [41] V. Rastija, G. Srećnik, S. Marica Medic, *Food Chem.* 115 (2009) 54.
- [42] N. Paixao, R. Perestrelo, J.C. Marques, J.S. Camara, *Food Chem.* 105 (2007) 204.
- [43] E.N. Frankel, A.L. Waterhouse, P.L. Teissedre, *J. Agric. Food Chem.* 43 (1995) 890.
- [44] L. Moreno-Barón, R. Cartas, A. Merkoçi, S. Alegret, J.M. Gutiérrez, L. Leija, P.R. Hernández, R. Muñoz, M. del Valle, *Anal. Lett.* 38 (2005) 2189.
- [45] J.M. Palacios-Santander, L.M. Cubillana-Aguilera, M. Cocchi, A. Ulrici, I. Naranjo-Rodríguez, R. Seeber, J.L. Hidalgo-Hidalgo de Cisneros, *Chemom. Intell. Lab. Syst.* 91 (2008) 110.
- [46] J.W. Tukey, *Ann. Math. Stat.* 29 (1958) 614.
- [47] S.D. Peddada, in: C.R. Rao (Ed.), *Handbook of Statistics*, Elsevier, North-Holland, New York, 1993, pp. 723.
- [48] M. Marina, *J. Multivariate Anal.* 98 (2007) 873.

Article 4

Bioelectronic Tongue for the quantification of total polyphenol content in wine

X. Cetó, F. Céspedes and M. del Valle

Talanta, **99**(0): 554-551



BioElectronic Tongue for the quantification of total polyphenol content in wine

Xavier Cetó, Francisco Céspedes, Manel del Valle*

Department of Chemistry, Sensors and Biosensors Group, Universitat Autònoma de Barcelona, Edifici Cn, 08193 Bellaterra, Barcelona, Spain

ARTICLE INFO

Article history:

Received 27 February 2012
Received in revised form
13 June 2012
Accepted 15 June 2012
Available online 20 June 2012

Keywords:

Electronic Tongue
Artificial neural network
Polyphenols
Folin–Ciocalteu
Biosensor
Wine

ABSTRACT

This work reports the application of a BioElectronic Tongue (BioET) in the estimation of polyphenol content in wine. The approach used an array of enzyme biosensors capable of giving a wide and complete response of the analyzed species, plus a chemometric processing tool able to interpret the chemical signals and extract meaningful data from the complex readings. In our case, the proposed BioET was formed by an array of four voltammetric enzymatic biosensors based on epoxy-graphite composites, one blank electrode and the other three bulk-modified with tyrosinase and laccase on one side, and copper nanoparticles on the other; these modifiers were used in order to incorporate differentiated or catalytic response to different polyphenols present in wine and aimed to the determination of its total polyphenol content value. The obtained voltammetric responses were pre-processed employing the Fast Fourier Transform (FFT); this was used to compress the relevant information whereas the obtained coefficients fed an Artificial Neural Network (ANN) model that accomplished the quantification of total polyphenol content. For comparison purposes, obtained polyphenol content was compared against the one assessed by two different reference methods: Folin–Ciocalteu and UV polyphenol index (I_{280}); good prediction ability was attained with correlation coefficients higher than 0.949 when comparing against reference methods. Qualitative discrimination of individual polyphenols found in wine was also assessed by means of Principal Component Analysis which allowed the discrimination of the individual polyphenols under study.

© 2012 Elsevier B.V. All rights reserved.

1. Introduction

Analytical methods with high sensitivity, good selectivity and fast response are needed to meet new challenges in environmental contamination, food safety, and public health. In this sense, winemaking industry is focusing much attention in fast-response and low-cost methods for wine characterization during its elaboration. From the analytical point of view, as the sample, wine is a complex mixture of diverse substances that exhibit biological activity and considerable influence on taste and other properties [1]. Among such compounds, polyphenols are ones of the most important given their antioxidant properties and their effect in wine sensorial features, in particular, colour, body and astringency. The health benefits of phenolic compounds derive from their antioxidant activity acting as free radical scavengers and inhibitors of lipoprotein oxidation [2]; these effects provide a protective effect against aging pathologies like cardiovascular diseases or cancers mutation. Therefore, there is a clear interest to evaluate wine polyphenolic content. Usually, its global content

(expressed as equivalents of gallic acid) is around 320 mg L^{-1} for white wine, 820 mg L^{-1} for rosé wine and 2160 mg L^{-1} for red wine [3].

Several methods to quantify polyphenols have been described in the literature [4]. Basically, those methods may be divided into two large groups. On the one hand, there are generic methods such as Folin–Ciocalteu (FC) or polyphenol index (I_{280}), which are widely employed in the wine industry and yield a total polyphenol content value [5,6]. FC method measures the sample reducing capacity and I_{280} index is a measure of the sample absorbance at 280 nm, with different UV contributions that do not necessarily reflect an absolute measurement of the total phenolic content [7]. On the other hand, there are the chromatographic techniques such as High-Performance Liquid Chromatography (HPLC) [8] and Gas Chromatography (GC) [9], which are able to perform the individual determination of polyphenolic compounds. Main disadvantages of these methods are that they need complex and time-consuming sample pre-treatment procedures, and are not suited for on site analysis.

Biosensors provide an interesting alternative to the determination of polyphenols thanks to their low cost and because they can be easily used to carry out on-field analyses [10]. An easy way for biosensor development is the use of bulk-modified composites

* Corresponding autor. Tel.: +34 93 5811017; fax: +34 93 5812379.
E-mail address: manel.delvalle@uab.es (M. del Valle).

[11], which allows the immobilization of enzymes or other biomolecules at the electrode. The applicability of electrochemical biosensors to the analysis of antioxidant compounds, including polyphenols, is promising and there is a growing interest in the development of such devices. Nevertheless, it is deemed that further work is required to avoid and/or take into account the interference problem [12]. Is for this reason that the combination of biosensors with Artificial Neural Networks (ANNs) modelling may represent an alternative to classical methods, taking benefit of the advantages of both parts. On one hand we have the specificity and selectivity of biosensors [13], and on the other the use of ANNs modelling to derive meaning from complex or imprecise data [14]; for example in multivariate calibration or pattern recognition [15]. This coupling, known as BioElectronic Tongue (BioET) may be postulated as a tool combining ANN to solve interference problems from biosensors, and biosensors as the tool that solves the selectivity problem from the Electronic Tongue (ET). This is the most recent variant of ETs, when biosensors are incorporated in the sensor array. This approach has already been applied towards the analysis of wines [16], and more specifically to the analysis of polyphenols and resolution of their mixtures [17–19]. Apart from BioETs, conventional ET systems based either on the use of voltammetric modified sensors [20] or potentiometric sensors [21,22] have been also used for estimation of polyphenols in wine.

ETs are a new trend from the sensory field, which can be defined as a multisensor system with marked mix-response, capable of giving a wide and complete response towards the analyzed species; ETs use advanced mathematical procedures for signal processing based on pattern recognition and/or multivariate analysis, able to extract meaningful data from the complex readings [14,15]. A known problem with some ETs is the large dimensionality of the generated data; for example, this occurs when a complete voltammogram is recorded for each sensor forming the array and data needs to be processed with ANN. In these cases, departure information is too complex and needs to be pre-processed. This step is needed to reduce the complexity of the input signal preserving the relevant information. Usually, this has been accomplished by the use of Principal Component Analysis (PCA), “kernel” functions [23], Discrete Wavelet Transform (DWT) [24] or even Fast Fourier Transform (FFT) [25].

The aim of this work is to present a BioET analysis system formed by an array of voltammetric enzymatic biosensors, e.g., employing tyrosinase or laccase, incorporated on bulk-modified biocomposites, for the analysis of polyphenols in wine. The combined electrochemical responses obtained from the set of biosensors' voltammograms were pre-processed employing FFT in order to extract the significant information and compress the departure information. Next, the obtained coefficients fed an ANN model specially trained to predict total polyphenol content. Furthermore, distinction of individual polyphenols present in wine was evaluated with PCA and classification was assessed with a PCA-ANN as the pattern recognition algorithm carrying out the identification of the different chemical compounds.

2. Experimental

2.1. Reagents and solutions

All reagents used were analytical grade and all solutions were prepared using deionised water from a Milli-Q system (Millipore, Billerica, MA, USA). Tyrosinase from mushroom (EC 1.14.18.1, 4276 U mg⁻¹), Laccase from *Trametes versicolor* (EC 1.10.3.2, 21 U mg⁻¹), copper nanoparticles (50 nm), gallic acid, (±)-catechin, *p*-coumaric acid, caffeic acid, catechol phenol, *m*-cresol,

ferulic acid, chlorogenic acid and quercetin were purchased from Sigma-Aldrich (St. Louis, MO, USA). KCl was purchased from Merck KGaA (Darmstadt, Germany). Folin–Ciocalteu's reagent and sodium carbonate were purchased from Panreac Química (Barcelona, Spain).

2.2. Wine samples

2.2.1. Evaluation of total polyphenol content

For this purpose, a total set of 29 wine samples of different varieties and from different regions of Spain where acquired at the local supermarket and analyzed. Samples were chosen in order to obtain a set with sufficiently differentiated total polyphenol content and grape varieties, i.e., including the most typical ones of each region; e.g., *merlot* and *trepal* from *Penedès* region (*Catalonia*), *tempranillo* from *Rioja*, *malbec* from *Castilla* or *petit verdot* from *Andalusia*, all of them from recent harvests on 2008, 2009 and 2010.

2.2.2. Discrimination of individual polyphenols

Nowadays, there is also an increasing demand for highly sensitive and selective analytical methods for the determination of individual polyphenols [4,26]. In this way, despite the great amount of research in the field, the separation and quantification of different polyphenols remain difficult, especially the simultaneous determination of those from different chemical subgroups.

In order to assess the ability of the BioET to discriminate between different polyphenols present in wine, a first qualitative attempt was carried out. For this, some spiked samples were prepared adding different quantities of certain polyphenols to wine samples with low initial content. Polyphenols considered were selected according to their expected presence in wine [27,28]; being gallic acid, (±)-catechin, *p*-coumaric acid, caffeic acid, catechol, phenol, *m*-cresol, ferulic acid, chlorogenic acid and quercetin the compounds selected for the study.

In this manner, 5 separate added samples for each polyphenol tested plus five samples of non-spiked wine were prepared, which means a total of 55 samples divided into 11 classes. To confirm that discrimination between compounds was not due to different amount of polyphenolic compound being added, in all the cases 5 μmol of each compound were added to 25 mL of wine, which represents approximately an increase of 36 mg L⁻¹ (200 μM). This small amount was chosen given individual polyphenol concentration is much lower than average total polyphenol content [3]. Apart, solutions were analyzed in random order to be sure their grouping its not a consequence of the order in which samples were analyzed.

2.3. Apparatus

Electrochemical measurements were carried out at room temperature (25 °C) under quiescent condition, in a multichannel electrode configuration using a multipointentiostat AUTOLAB PGSTAT20 (Ecochemie, Netherlands). The voltammetric cell was formed by the array of four composite working electrodes plus a double junction electrode Ag/AgCl Orion 900200 (Thermo Electron Corporation, Beverly, MA, USA) and a platinum-based electrode (Crison 52–67, Barcelona, Spain) used as reference and auxiliary electrodes, respectively.

Spectrophotometric measurements were carried out using a Spectronic Helios Epsilon spectrophotometer from Thermo Electron Scientific Instruments LLC (Madison, WI USA).

2.4. Electrode fabrication

Based on previous studies with polyphenols in our laboratory [18], an array of four working electrodes were prepared following the conventional biocomposite methodology in our laboratories [29]. First, resin EpoTek H77 (Epoxy Technology, Billerica, MA, USA) and its corresponding hardener compound were mixed in the ratio 20:3 (w/w). Then, each electrode was prepared adding a 15% of graphite (w/w) and a 2% of either the enzyme (tyrosinase or laccase) or the modifier (w/w) (copper nanoparticles) into the epoxy resin before hardening—one electrode with each modifier plus a blank electrode without any modifier. Afterwards, the biocomposite was manually homogenised for 60 min. Finally, the biocomposite paste electrode was allowed to harden during 7 day at 40 °C. Electrode surface was then polished with different sandpapers of decreasing grain size, with a final electrode area of 28 mm².

Hence, the BioET array will be formed by two enzymes, which belong to the class of copper containing oxidases, and that catalyze the oxidation of the phenolic compounds into their quinones, which are directly measurable on an amperometric sensor surface [10]. Also, copper-nanoparticles were chosen given the fact that both tyrosinase and laccase have a copper centre in its active site. Then it was thought that some catalytic effect could be derived; a fact finally proved by sensor's response [18].

2.5. Procedures

2.5.1. Electrochemical measurements

The electrodes were cycled for 3–5 times in buffer solution in order to get stable voltammetric responses before performing the measurements with real samples. Potential was cycled between –0.4 V and +0.8 V vs Ag/AgCl, with a scan rate of 100 mV s⁻¹ and a step potential of 9 mV. No pre-treatment or dilution of sample was performed when measuring wine samples; apart, all experiments were carried out without performing any physical surface regeneration of the working electrodes. In order to prevent the accumulative effect of impurities on the working electrode surfaces, an electrochemical cleaning stage was done between each measurement applying a conditioning potential of +1.0 V for 40 s after each experiment, in a cell containing 10 mL of distilled water [30]. This step ensures that electrode fouling is minimized, recovering the original signal (baseline) after each measurement.

2.5.2. Spectrophotometric measurements

For comparison purposes, polyphenolic content of wines was also assessed spectrophotometrically with two different methods: Folin–Ciocalteu index (FC) and UV Polyphenol Index (I_{280}).

The first one is a colorimetric assay measuring the amount of phenol needed to inhibit the oxidation of the Folin–Ciocalteu reagent (a mixture of phosphomolybdate and phosphotungstate, which are reduced to the respective oxides). As a drawback, this reagent does not measure total phenols uniquely and will react with other reducing substances present. Therefore, it measures the total reducing capacity of a sample, not just the level of phenolic compounds. The second one is a direct measurement of the absorbance at 280 nm. The relation between I_{280} index and phenolic concentration is due to the fact that all phenolic compounds absorb UV light, and even more, all of them have some absorbance at 280 nm. One problem with this method is that each class of phenolic substances has a different absorptivity (extinction coefficient, ϵ) at 280 nm [6]. Thus, the results cannot be related to any specific standard and are reported directly in absorbance units or arbitrary units (arb. unit). Despite this method is less sensitive and more inespecific, its usage has grown in the last years given its simplicity and low cost.

The Folin–Ciocalteu test was carried out according to the established procedure for wine analysis [31]. First, 200 μ L of sample (wines were previously diluted 1:50), 1300 μ L of deionized water, 100 μ L of Folin–Ciocalteu reagent and 400 μ L of a 20% sodium carbonate solution were mixed into an Eppendorf tube. Afterwards, the resulting solution was allowed to react for 30 min in darkness at room temperature (25 °C), and finally its absorbance was read directly at 760 nm. The total phenolic content (FC index), expressed in mg L⁻¹ equivalents of gallic acid, was evaluated from the absorbance value by interpolation into the calibration plot obtained with gallic acid standard solutions, multiplying the resulting value by the proper dilution rate.

Polyphenol index (I_{280}) was also considered [31]. For this, wine was first diluted with deionized water (1:50) and then absorbance was measured directly at 280 nm. The value of I_{280} for each sample was given as the absorbance multiplied by the proper dilution rate.

2.6. Data processing

Chemometric processing was done by specific routines in MATLAB 7.1 (MathWorks, Natick, MA) written by the authors, and using Neural Network Toolbox (v.4.0.6). Sigmaplot 2000 (Systat Software Inc, California, USA) was used for graphic representations of data and results.

For each sample, one voltammogram was recorded for each sensor from the BioET array. In order to reduce the high dimensionality of the recorded signals (samples \times sensors \times potentials), a preprocessing stage was required. For this, FFT was used as feature extraction tool to compress the information from the original signals and extract meaningful data from the readings [32]. FFT preprocessing consists in the decomposition of the signal using sine/cosine function pair at different frequencies, calculating a coefficient for each one taking into account its contribution to the original signal [33]. In this manner, as many coefficients as points in the original signal are obtained, where most relevant information is kept by the first coefficients, ergo good signal reconstruction could be achieved just preserving the first coefficients and discarding the rest. Then, the obtained coefficients fed the ANN model which was used for the quantification of total polyphenol content. Similarly, a classification procedure was developed from a PCA pretreatment and an ANN classifier in order to discriminate between the different polyphenols added to wine samples.

The used ANN training algorithm was Bayesian regularization. This algorithm has the particularity that it avoids overfitting without the need to monitor the fitness degree of an internal validation subset [34], then this precaution is not performed. To evaluate the BioET response a jack-knife method was used given the reduced size of the data set [35,36]. This method has the advantage that avoids dependence of predictions from the specific subdivision of data. In this way, train/test data subdivision is repeated randomly k times, evaluating model's response for test validation subset each time; afterwards all data is grouped depending if it was intervening in the training process or used in the external test subset, building the response model, which in addition allows to calculate confidence intervals for the predictions.

3. Results and discussion

3.1. Voltammetric responses

Examples of the different obtained signals coming from the BioET array for different wines are shown in Fig. 1. A clear voltammetric response is obtained for all the sensors, with currents

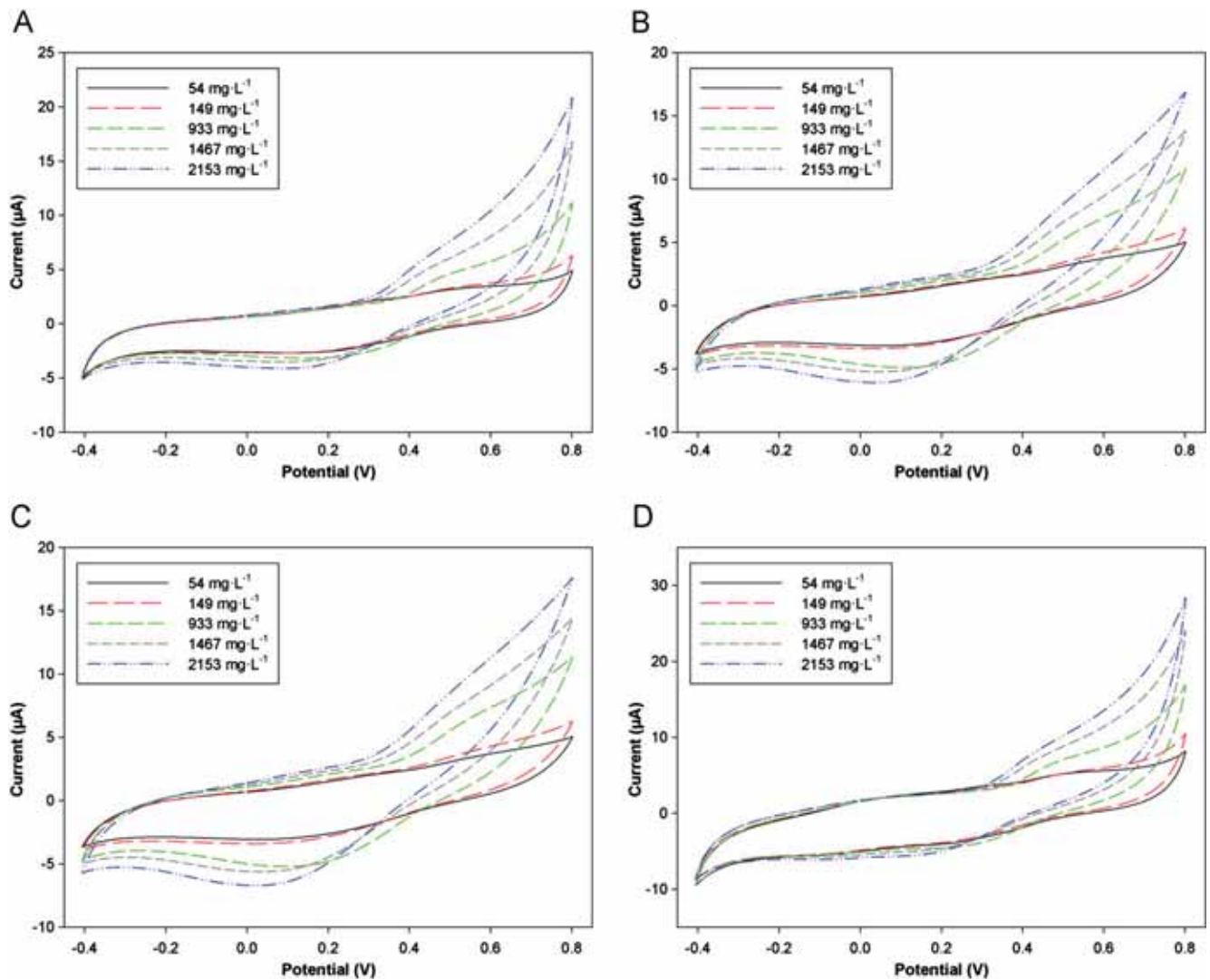


Fig. 1. Example of the different voltammograms obtained for certain arbitrary wine samples. Folin indexes are expressed as equivalents of gallic acid. Also signals provided by different (bio)sensors are shown: (A) graphite-epoxy sensor, (B) tyrosinase biosensor, (C) laccase biosensor and (D) copper nano-particle modified sensor.

monotonously increasing as FC index increases and with differentiated behaviour for each sensor; but as can be seen, higher currents are obtained for the biosensors containing tyrosinase (Fig. 1B) and laccase (Fig. 1C) enzymes, especially in the reduction region close to 0 V where obtained net response is ca. 3 times higher. Also, some degree of catalytic effect is obtained from the copper nanoparticles, a fact somehow explained given both tyrosinase and laccase are copper-containing redox enzymes. These enzymes, both having two copper atoms (three in the case of laccase) within their active site, interact with dioxygen to form a highly reactive chemical intermediate that then oxidizes the substrate; copper nanoparticles may therefore have a similar catalytic effect [18].

3.2. FFT compression

To fully exploit all the information obtained from each voltammogram and not overdosing the associated ANN, a compression step is required given the extreme complexity of the considered departure signals (268 current values \times 4 sensors \times 29 samples). This step may also help to gain advantages in training time, to avoid redundancy in input data and to obtain an ANN model with better generalization ability [30].

This compression step was achieved in this work using the Fast Fourier Transform (FFT).

In order to quantify the reconstruction degree, the coefficient of determination (R^2) between reconstructed and original signal and a comparison factor named f_c that considers the area under both signals when superimposed were used [24]. The f_c is defined as the ratio of the area intersected by both curves to the total area under both curves. From the set theory, with A and B as the areas under each curve, the f_c can be expressed as $f_c = (A \cap B) / (A \cup B)$. This factor ranges from 0 to 1 depending on signals similarity; it values 0 when two signals have nothing in common and increases its value as similarity does. In this sense, f_c computes similarity in a way related to a correlation coefficient R , but being more sensitive to small differences.

As can be seen in Fig. 2(A), by increasing the number of coefficients we preserve a better representation of the voltammetric signal, but implying a higher complexity for the ANN model. Then, a compromise is needed; we selected the number of coefficients by checking the correlation of the reconstructed vs. original signal representation (Fig. 2B). Both R^2 and f_c had the same trend, but the factor f_c presented better discrimination capability than R ; e.g., with only 8 Fourier coefficients, where signal reconstruction is still poor, the obtained value for R^2 was 0.9974, while f_c value was 0.9020. Therefore, the optimal number

of coefficients was selected as the first time f_c surpassed 0.98. This was achieved with the first 32 Fourier coefficients, which allowed a compression of the original information up to 88.1% without any loss of significant information.

Hence, it could be seen the good performance of FFT when applied to voltammetric data, providing similar or even better

signal compression/reconstruction ratio than DWT. This is mainly due to the recorded signal, which has some sinusoidal trend, making FFT a very suitable method in these cases. As also happens with Wavelet pre-processing (DWT), both methods are very effective in cases where complex and highly overlapped responses are obtained, given obtained coefficients are a representation of the entire voltammetric signal which in addition is denoised.

3.3. Quantification of total polyphenol indexes

After pre-processing the recorded voltammograms with FFT, the obtained coefficients fed an ANN model in order to predict the total polyphenol index in wines. In order to find the appropriate ANN model, significant effort is needed to optimize the configuration details that determine its operation. Normally, this is a trial-and-error process, where several parameters (training algorithms, number of hidden layers, transfer functions, etc.) are fine-tuned in order to find the best configuration that optimizes the performance of the model. A simplified scheme of the procedure followed for the measurement and data treatment could be seen in Fig. 3.

After some preliminary tests, the final architecture of the ANN model had 128 neurons (4 sensors \times 32 coeffs. obtained from the FFT analysis) in the input layer, 6 neurons and *logsig* transfer function in the hidden layer and two neurons and *tansig* transfer function in the output layer, providing the two phenol indexes considered.

Accuracy of the generated model was evaluated employing a jack-knife method to avoid dependence of predictions from the specific subdivision of data, training with 80% of the data (23 samples) and tested with the remaining 20% (6 samples). In this way, train/test data subdivision was repeated randomly 29 times (as many times as samples, similarly to *k*-fold method) in order to ensure that model's accuracy is good enough and BioET performance does not depend on the specific subsets used. Then, once all responses from all the constructed models were obtained, predicted values by each model were grouped depending if they were used in the training process or in the testing subset (again, similarly to *k*-fold method). Finally, average values for each sample were calculated, allowing us to calculate model uncertainties and obtain unbiased data [36].

Comparison graphs of predicted vs. expected concentration for FC (Fig. 4) and I_{280} (Fig. 5) indexes were built, both for train and test subsets, to check the prediction ability of the obtained ANN model. It may be seen that a satisfactory trend is obtained for both indexes, with regression lines almost indistinguishable from the theoretical ones. Also, as usual in ANN models, lower dispersion and uncertainties are obtained for the training subsets.

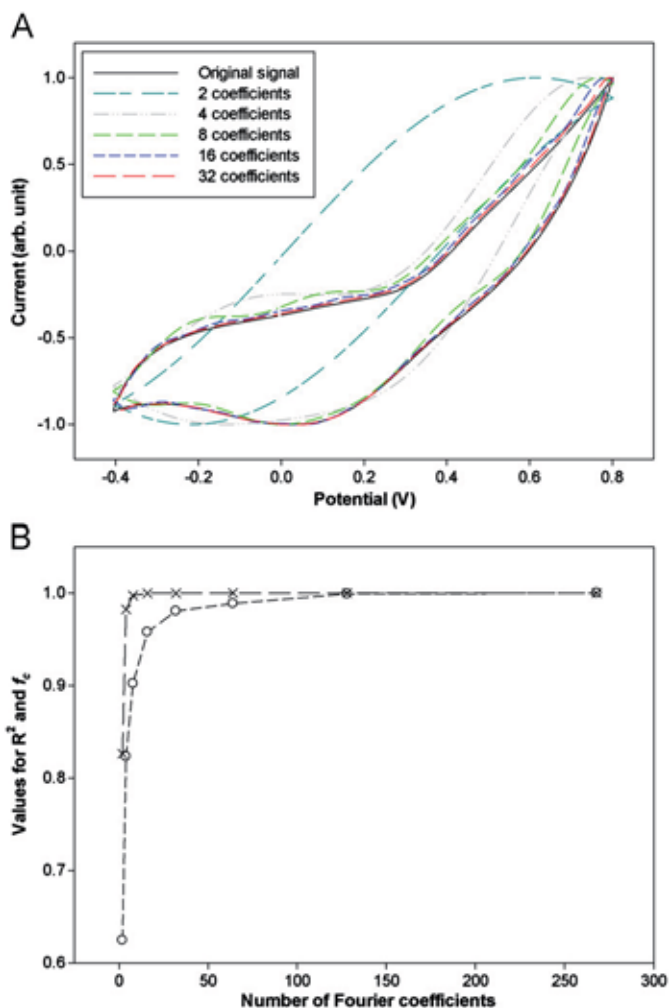


Fig. 2. (A) Reconstruction of the voltammetric signal according to the number of Fourier coefficients employed, tyrosinase biosensor shown as example. (B) Representation of the coefficient of determination (R^2 , \times) and f_c (\circ) against the number of Fourier coefficients used from comparing the raw voltammograms with the reconstructed signals.

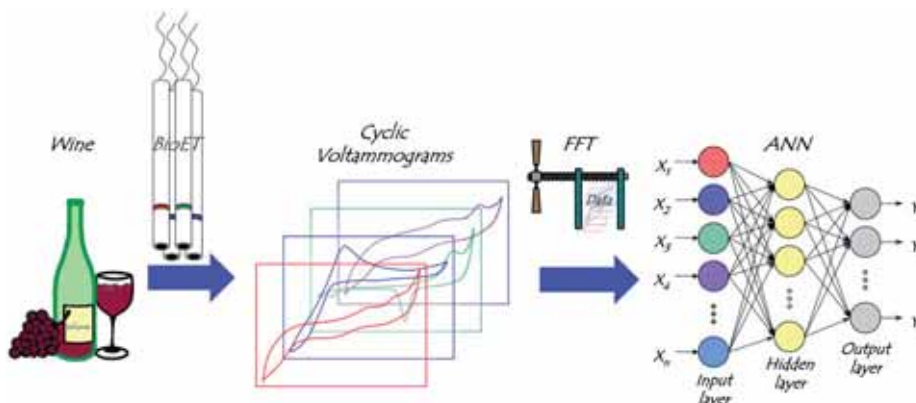


Fig. 3. Voltammetric BioET approach. Voltammetric signals are compressed employing FFT and the obtained coefficients are taken as input in the ANN. Appropriate weights and biases are applied by the learning algorithm until the targets are reached within the established error.

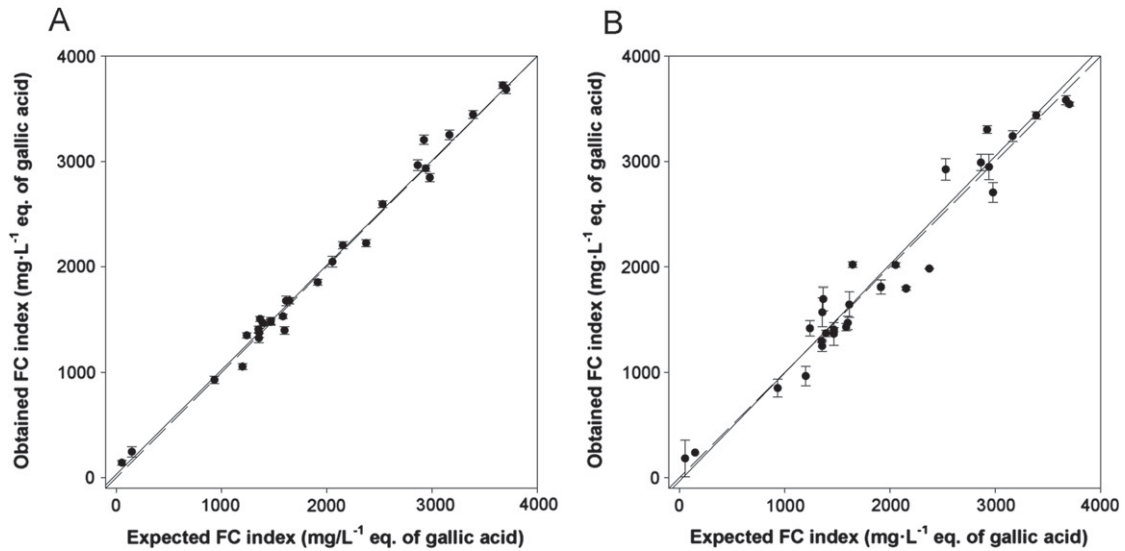


Fig. 4. Modelling ability of the optimized FFT-ANN. Sets adjustments of expected vs. obtained concentrations for Folin-Ciocalteu index, both for (A) training and (B) testing subsets. Dashed line corresponds to theoretical diagonal line. Results provided correspond to the average of the values obtained for each sample after 29 repeated calculations, done with random division of samples for train/test subsets each time. Uncertainties calculated at the 95% confidence level.

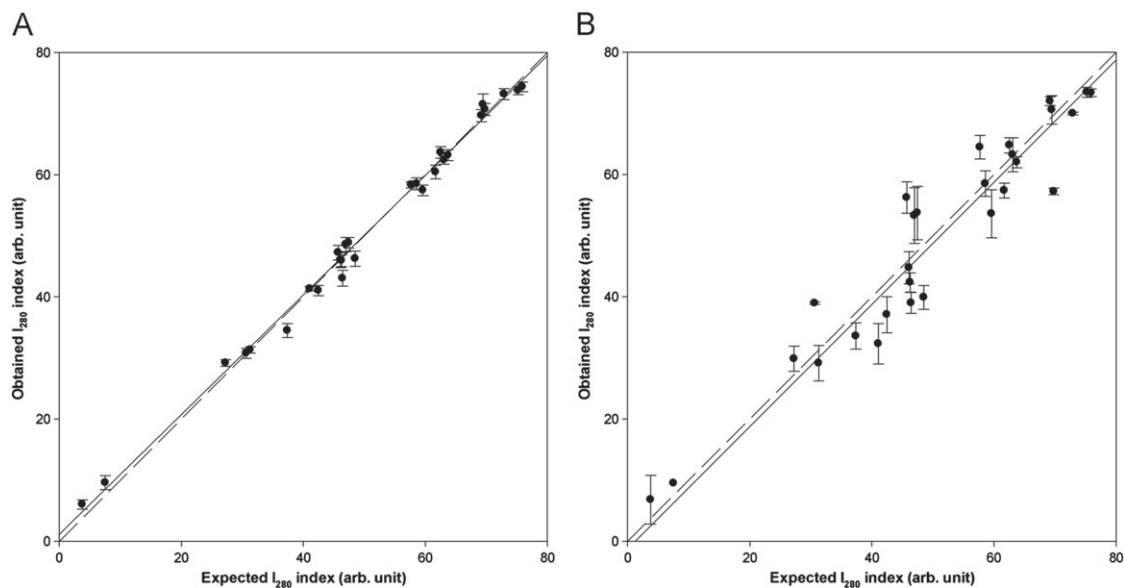


Fig. 5. Modelling ability of the optimized FFT-ANN. Sets adjustments of expected vs. obtained values for I_{280} index, both for (A) training and (B) testing subsets. Dashed line corresponds to theoretical diagonal line. Results provided correspond to the average of the values obtained for each sample after 29 repeated calculations, done with random division of samples for train/test subsets each time. Uncertainties calculated at the 95% confidence level.

Table 1 summarizes the obtained regression parameters. As expected from the comparison graphs, a good linear trend is attained for all the cases, but with better correlation coefficients in the training subsets due to the lower dispersion. Despite this, the results obtained for both subsets are close to the ideal values, with intercepts close to 0 and slopes and correlation coefficients practically 1.

As an additional verification of the proposed approach, a Student's paired samples t -test was performed between both reference methods and the BioET methodology. Obtained experimental t values were 0.025 and 0.127 for the comparison against FC and I_{280} , respectively, while the critical tabulated t value with 95% confidence level and 28 degrees of freedom was 2.048. Therefore, from the comparison graphs and these t -test results it could be concluded that there are no significant differences

Table 1

Results of the fitted regression lines for the comparison between obtained vs. reference values for the BioET, both for Folin-Ciocalteu and I_{280} indexes. Results provided correspond to the average of the values obtained for each sample after 29 repeated calculations, done with random division of samples for train/test subsets each time. Uncertainties calculated at the 95% confidence level.

| | Correlation | Slope | Intercept | RMSE (mg L ⁻¹) |
|--|-------------|---------------|--------------|----------------------------|
| Folin-Ciocalteu index (54 to 3705 mg L ⁻¹ gallic acid) | | | | |
| Train subset | 0.994 | 0.993 ± 0.008 | 29.5 ± 17.4 | 102 |
| Test subset | 0.978 | 1.027 ± 0.033 | -31.6 ± 74.0 | 205 |
| | Correlation | Slope | Intercept | RMSE (arb. unit) |
| I_{280} index (3.78 to 75.93 arbitrary units) | | | | |
| Train subset | 0.995 | 0.979 ± 0.007 | 1.15 ± 0.39 | 1.85 |
| Test subset | 0.949 | 0.999 ± 0.046 | -1.23 ± 2.57 | 5.17 |

between the BioET predicted values and the ones obtained with the reference methods.

3.4. Discrimination and classification of individual polyphenols

In this qualitative approach, spiked wine samples with reference polyphenols were analyzed in random order with the BioET array, following equivalent measuring procedure as before. Then, the obtained responses were processed employing PCA analysis and grouped using cluster analysis tools, which allowed to summarize almost all variance contained in the departure information onto a fewer number of directions (the PCs) with new coordinates called scores, obtained after data transformation. Thus, a preliminary recognition was attained which allowed the visually distinction of the formed groups.

As can be seen in Fig. 6, with the two first PCs, the explained variance accumulated was ca. 98% and a clear distinction between clusters is obtained; this large value means that nearly all the variance contained in the original information is represented by the two new coordinates. Patterns in the figure evidence that samples are grouped based on which polyphenol was added. These well established clusters clearly separate the main classes of samples corresponding to: (I) wine, (II) gallic acid, (III) (\pm)-catechin, (IV) *p*-coumaric acid, (V) caffeic acid, (VI) catechol, (VII) phenol, (VIII) *m*-cresol, (IX) ferulic acid, (X) chlorogenic acid and (XI) quercetin. This satisfactory separation of classes validates on its respect, the proper selection of (bio)sensors to form the BioET.

Again, an ANN was trained to act as a classifier of the above stated groups. Unlike the previous case, where the responses obtained from the model were quantitative variables, in this case the output of the ANN model was formed by binary predictors (1/0) for each of the classes. As in the previous case, the ANN configuration needs to be optimized. Then after some preliminary tests, the final ANN architecture model had 3 neurons (corresponding to the first three components of the PCA, with explained variance accumulated ca. 99%) in the input layer, 7 neurons in the hidden layer and 11 binary (1/0) neurons in the output layer (one for each class) with *logsig* transfer function in the three layers. The backpropagation algorithm was used to train the

network and the expected output error was programmed to reach a value of 0.01.

PCA-ANN model was trained with 60% of the data (33 samples) and confusion matrix was performed using the information of the testing set (remaining 40% of the data; 22 samples) in order to characterize the accuracy of the identification model and obtain unbiased data. Unlike the previous case, a *k*-fold strategy was used instead of jack-knife data division, given in this case it was dealing with a qualitative approach, which means an easier modelling situation and with less statistical parameters; besides that, five replicates for each sample were measured.

From the classification results, the corresponding confusion matrix was built. Correct classification for all the classes was obtained (i.e., a classification rate of 100% for each of the groups), as could be explained from the direct visualization of the PCA analysis. The percentage of correct classifications was estimated, from individual sample calculation in the test subset, as 100%. The efficiency of the classification obtained was also evaluated according to its sensitivity, i.e., the percentage of objects of each class identified by the classifier model, and to its specificity, the percentage of objects from different classes correctly rejected by the classifier model. The value of sensitivity, averaged for the classes considered, was 100%, and that of specificity was 100%.

4. Conclusions

In summary, a BioElectronic Tongue based on voltammetric enzyme-modified biosensors with different modifiers (tyrosinase and laccase on one side and copper nanoparticles on the other) has been applied in wine analysis in order to create a tool capable of quantifying total polyphenol content and discriminate individual polyphenolic compounds. Pre-processing tools such as FFT (an approach rarely used in literature with voltammetric data) could provide reasonable compression of data preserving relevant information, while the use of ANN allowed us to predict phenolic content index obtained with two different reference methods (Folin–Ciocalteu and I_{280} indexes). Finally, PCA allowed the discrimination of individual polyphenolic compounds.

With this and the previous results, proposed BioET has demonstrated its powerful applications in winemaking industry in the analysis of polyphenols found in wine, both in the quantitative analysis of total polyphenol content and in the resolution and quantification of phenolic mixtures, and also in the qualitative discrimination and classification of polyphenolic compounds. Its performance characteristics may satisfy food industry requirements of precision, rapidity, sensitivity, simplicity and low cost required to be considered as a useful analytical tool.

Furthermore, it represents an alternative to Folin–Ciocalteu method reducing considerably analysis time (from 30 min to ca. 3 min), avoiding the sample pretreatment (proper dilution factor) and the use of reagents (Folin–Ciocalteu and sodium carbonate). At the same time and after proper training the BioET, it is possible to perform a discrimination of individual phenolic compounds in an application more comparable to HPLC.

Acknowledgments

Financial support for this work was provided by Spanish Ministry of Science and Innovation, MCINN (Madrid) through the project CTQ2010–17099 and by program ICREA Academia. X. Cetó thanks the support of Dept. d'Innovació, Universitat i Empresa de la Generalitat de Catalunya for the predoctoral grant.

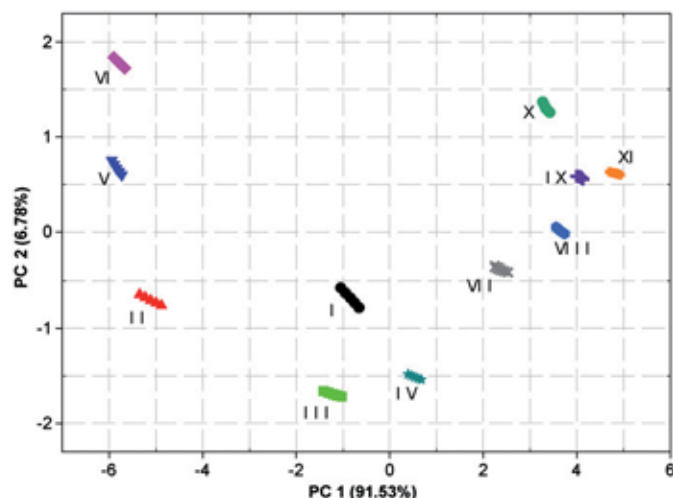


Fig. 6. Score plot of the first two components obtained after PCA analysis of the spiked wine samples. A total of 55 spiked wine samples were analysed. As can be seen, clear discrimination is obtained for the different polyphenols considered: (I) wine, (II) gallic acid, (III) (\pm)-catechin, (IV) *p*-coumaric acid, (V) caffeic acid, (VI) catechol, (VII) phenol, (VIII) *m*-cresol, (IX) ferulic acid, (X) chlorogenic acid and (XI) quercetin.

References

- [1] P. Ribéreau-Gayon, Y. Glories, A. Maujean, D. Dubourdieu, Handbook of Enology, Volume 2, The Chemistry of Wine Stabilization and Treatments, second ed., Wiley-Blackwell, London, 2000.
- [2] C. Sanchez-Moreno, J.A. Larrauri, F. Saura-Calixto, Food Res. Int. 32 (1999) 407–412.
- [3] E.N. Frankel, A.L. Waterhouse, P.L. Teissedre, J. Agr. Food Chem. 43 (1995) 890–894.
- [4] I. Ignat, I. Volf, V.I. Popa, Food Chem. 126 (2011) 1821–1835.
- [5] V.L. Singleton, J.A. Rossi, Amer. J. Enol. Viticult. 16 (1965) 144–158.
- [6] A.L. Waterhouse, Determination of Total Phenolics, in: Current Protocols in Food Analytical Chemistry, John Wiley & Sons, Inc., 2001.
- [7] D. De Beer, J.F. Harbertson, P.A. Kilmartin, V. Roginsky, T. Barsukova, D.O. Adams, A.L. Waterhouse, Am. J. Enol. Vitic. 55 (2004) 389–400.
- [8] H. Sakakibara, Y. Honda, S. Nakagawa, H. Ashida, K. Kanazawa, J. Agr. Food Chem. 51 (2002) 571–581.
- [9] P. Viñas, N. Campillo, N. Martínez-Castillo, M. Hernández-Córdoba, J. Chromatogr. A 1216 (2009) 1279–1284.
- [10] R. Solna, P. Skladal, Electroanalysis 17 (2005) 2137–2146.
- [11] F. Céspedes, E. Martínez-Fàbregas, S. Alegret, TrAC-Trend. Anal. Chem. 15 (1996) 296–304.
- [12] V.L. Singleton, R. Orthofer, R.M. Lamuela-Raventós, Analysis of Total Phenols and Other Oxidation Substrates and Antioxidants by Means of Folin-Ciocalteu Reagent, in: Oxidants and Antioxidants, Pt A, Academic Press Inc., San Diego, 1999, pp. 152–178.
- [13] S.P. Mohanty, E. Kougiannos, IEEE Potentials 25 (2006) 35–40.
- [14] M. del Valle, Electroanalysis 22 (2010) 1539–1555.
- [15] Y. Vlasov, A. Legin, A. Rudnitskaya, C. Di Natale, A. D'Amico, Pure Appl. Chem. 77 (2005) 1965–1983.
- [16] J. Zeravik, A. Hlavacek, K. Lacina, P. Skladal, Electroanalysis 21 (2009) 2509–2520.
- [17] A. Gutiérrez, F. Céspedes, S. Alegret, M. del Valle, Biosens. Bioelec. 20 (2005) 1668–1673.
- [18] X. Cetó, F. Céspedes, M.I. Pividori, J.M. Gutiérrez, M. del Valle, Analyst 137 (2012) 349–356.
- [19] A. Gutiérrez, A.B. Ibáñez, F. Céspedes, S. Alegret, M. del Valle, Anal. Bioanal. Chem. 382 (2005) 471–476.
- [20] M. Gay, C. Apetrei, I. Nevaes, M. del Alamo, J. Zurro, N. Prieto, J.A. De Saja, M.L. Rodríguez-Mendez, Electrochim. Acta 55 (2010) 6782–6788.
- [21] D. Kirsanov, O. Mednova, V. Vietoris, P.A. Kilmartin, A. Legin, Talanta 90 (2012) 109–116.
- [22] A. Rudnitskaya, H. Nieuwoudt, N. Muller, A. Legin, M. du Toit, F. Bauer, Anal. Bioanal. Chem. 397 (2010) 3051–3060.
- [23] R. Gutiérrez-Osuna, H.T. Nagle, IEEE Trans. Syst. Man Cybern. Part B 29 (1999) 626–632.
- [24] L. Moreno-Barón, R. Cartas, A. Merkoçi, S. Alegret, M. del Valle, L. Leija, P.R. Hernandez, R. Muñoz, Sens. Actuators, B 113 (2006) 487–499.
- [25] R.A. DePalma, S.P. Perone, Anal. Chem. 51 (1979) 829–832.
- [26] A.S. Arribas, M. Martínez-Fernández, M. Chicharro, TrAC-Trend. Anal. Chem. 34 (2012) 78–96.
- [27] P. Ribéreau-Gayon, Y. Glories, A. Maujean, D. Dubourdieu, Phenolic Compounds, in: Handbook of Enology, John Wiley & Sons, Ltd, Chichester (UK), 2006, pp. 141–203.
- [28] J. Garrido, F. Borges, Food Res. Int. 44 (2011) 3134–3148.
- [29] F. Céspedes, S. Alegret, TrAC-Trend. Anal. Chem. 19 (2000) 276–285.
- [30] X. Cetó, J.M. Gutiérrez, L. Moreno-Barón, S. Alegret, M. del Valle, Electroanalysis 23 (2011) 72–78.
- [31] Commission Regulation (EEC), in: Official Journal L 272, Publication Office of the European Union, 17 September 1990, pp. 0001–0192.
- [32] D. Calvo, A. Duran, M. del Valle, Sens. Actuators, B 131 (2008) 77–84.
- [33] V. Tchistiakov, C. Ruckebusch, L. Duponchel, J.-P. Huvenne, P. Legend, Chemometr. Intell. Lab. 54 (2000) 93–106.
- [34] L. Moreno-Barón, R. Cartas, A. Merkoçi, S. Alegret, J.M. Gutiérrez, L. Leija, P.R. Hernández, R. Muñoz, M. del Valle, Anal. Lett. 38 (2005) 2189–2206.
- [35] J. Riu, R. Bro, Chemometr. Intell. Lab. 65 (2003) 35–49.
- [36] J. Shao, D. Tu, The Jackknife and Bootstrap, Springer, New York etc, 1995.

Article 5

Assessment of individual polyphenol content in beer by means of a voltammetric
BioElectronic Tongue

X. Cetó, F. Céspedes and M. del Valle

Electroanalysis, **25**(1): 68-76

Assessment of Individual Polyphenol Content in Beer by Means of a Voltammetric BioElectronic Tongue

Xavier Cetó, Francisco Céspedes, Manel del Valle*

Sensors and Biosensors Group, Department of Chemistry, Universitat Autònoma de Barcelona, Edifici Cn, 08193 Bellaterra, Barcelona, Spain

tel: +34 93 5811017; fax: +34 93 5812477

*e-mail: manel.delvalle@uab.cat

Received: June 18, 2012

Accepted: July 10, 2012

Published online: September 5, 2012

Abstract

This work describes the simultaneous determination of three major phenolic compounds found in beer, namely ferulic, gallic and sinapic acids, by means of a BioElectronic Tongue (BioET). The proposed BioET is formed by an array of four voltammetric biosensors, with marked mix-response towards the involved compounds, plus an Artificial Neural Network (ANN) model able to extract meaningful data from the complex readings, overcoming signal overlapping. The obtained responses were first preprocessed employing the *windowed slicing integral* method to reduce data dimension and the obtained coefficients fed the ANN model which carried out the quantification. Finally, its performance was successfully assessed with some external test and spiked beer samples.

Keywords: Electronic tongue, Artificial neural network, Voltammetric biosensors, Polyphenols, Beer

DOI: 10.1002/elan.201200299

Supporting Information for this article is available on the WWW under <http://dx.doi.org/10.1002/elan.201200299>.

1 Introduction

Beer contains a wide variety of polyphenols of different classes that act as natural antioxidants. However, content values for each of the single polyphenols are rather low. Nevertheless polyphenols play an important role in beer flavour (bitterness, astringency and harshness), colour and stability. Therefore, control of phenolic compounds is of clear interest in the field.

There are several methods reported in the literature to individually quantify phenolic compounds. Usually, separation is achieved by the use of chromatographic techniques such as HPLC [1] or GC [2] which are able to perform individual determination of phenolic compounds. Also, there are some spectrophotometric methods, developed for quantification of phenolics in plants, that bring some specificity. These assays are based on differential reactivity principles and are used to determine different structural groups concomitant in phenolic compounds [3]. However these procedures usually require additional preparative steps, the use of heavy, dedicated laboratory instruments and are not suitable for on-site analysis. In addition, due to similar chemical behaviour, the resolution could be incomplete or not always well performed, getting some signal overlapping and making it difficult to quantify the compounds separately.

Alternatively, there are other methods, such as the Folin–Ciocalteu index [4], that yield a total polyphenol content value. This method is the most widely used one,

and measures the sample reducing capacity, although this global index can lead to an overestimation of the total polyphenol content.

In this fashion, the total polyphenol content in beer, as measured by the Folin–Ciocalteu method, varies between 120 and 520 mg/L, depending on the beer type. However, when comparing the total polyphenol value measured by the later (280 mg/L for regular beer) with the total polyphenol amount calculated from the sum of the individual polyphenols measured by chromatographic methods (40.4 mg/L for regular beer), the sum appears to be inferior [5]. This could be explained in one part by the lack of data on beer polyphenols in the literature, and thus in the food composition table. Also, it can be justified by the difficulty to measure certain polyphenols by chromatographic methods. Additionally, the Folin–Ciocalteu assay has the drawback that it can lead to an overestimation of the final value, as it may include the response of other non-phenolic reducing compounds. In this way, the most abundant polyphenols found in regular beer are [6] ferulic acid, gallic acid, vanillic acid, *p*-coumaric acid and sinapic acid, which concentration may vary depending on the type of beer.

Nowadays biosensors are arising as an alternative to traditional laboratory techniques given their low cost and their ease of using them to carry out on field analyses. Focusing on the detection of phenolic compounds, different enzymatic biosensors have been developed by means of laccase [7], tyrosinase, [8] or peroxidase enzymes [9] and

even by the co-immobilization of two enzymes in the same biosensor [10]. On that account, the applicability of biosensors to the analysis of antioxidant compounds is promising and represents an attractive alternative for their detection.

These systems employ enzymes to specifically oxidise the phenolic compounds into their quinones, which are directly measurable on an amperometric sensor surface. However, the voltammetric signals produced by these electrochemical methods correspond to the global overlapped voltammogram arising from all the different phenols present in the sample. In this sense, the use of chemometric tools such as Artificial Neural Networks (ANN) has helped to overcome this limitation [11,12]. This coupling, known as BioElectronic Tongue (BioET), consist in the combination of an array of biosensors with marked mix-response towards the desired species plus a chemometric processing tool able to interpret and extract meaningful data from the complex readings [13]. This approach has already been applied to the analysis of wines [12] and beers [14], and more specifically to the analysis of polyphenols and the resolution of their mixtures [15,16].

The main drawback of this approach is the large dimensionality of the generated data that hinders its treatment. That is when a complete voltammogram is recorded for each sensor that forms the array and data needs to be processed with ANN. In these situations, departure information is too complex and needs to be preprocessed. This feature extraction step is needed to reduce the complexity of the input signal meanwhile preserving the relevant information. Usually, this could be accomplished by the use of Principal Component Analysis (PCA) [17], "kernel" functions [18], Discrete Wavelet Transform (DWT) [19] or even Fast Fourier Transform (FFT) [20].

In this work we report the application of a voltammetric BioET towards the simultaneous quantification of three major polyphenols found in beers. The combined electrochemical responses obtained from the set of biosensors were preprocessed employing the *windowed slicing integral* in order to extract the significant information and compress the departure data. Then the obtained coefficients fed an ANN model specially trained to predict individual polyphenols present in beer.

2 Experimental

2.1 Reagents and Solutions

All reagents used were analytical reagent grade and all solutions were prepared using deionised water from a Milli-Q system (Millipore, Billerica, MA, USA). Tyrosinase from mushroom (EC 1.14.18.1, 4276 U mg⁻¹), laccase from *Trametes versicolor* (EC 1.10.3.2, 21 U mg⁻¹), copper nanoparticles (<50 nm), ferulic acid, gallic acid, sinapic acid, potassium dihydrogenphosphate and potassium monohydrogenphosphate were purchased from

Sigma-Aldrich (St. Louis, MO, USA). KCl was purchased from Merck KGaA (Darmstadt, Germany).

2.2 Samples Under Study

2.2.1 Standard Solutions

Firstly, the discrimination of the three phenolic mixtures was attempted in buffered solutions (0.1 M phosphate buffer at pH 6.50 and 0.1 M KCl) in order to prove the capabilities of the proposed approach. For this, a total set of 42 samples were manually prepared with a concentration range for the three species from 0 to 200 μM for each polyphenol. The set of samples was divided into two data subsets: a training subset formed by 27 samples (64%), which were distributed in a cubic design [21] and used to establish the response model; plus 15 additional samples (36%) for the testing subset, randomly distributed along the experimental domain, and used to evaluate the model predictive response.

2.2.2 Beer Spiked Samples

Once we had trained and tested the BioET system using standard solutions, the next step was to apply it to the resolution of these three phenolic mixtures in real beer samples. For this, 15 extra samples were prepared adding different quantities of a stock solution of each polyphenol to a lager beer (the same one was used for the preparation of all the spiked samples). As the ones prepared for the external test set, values of added concentration was ranged and distributed randomly along the experimental domain

2.3 BioElectronic Tongue

2.3.1 Electrodes and Measurement Cell

Working electrodes were prepared following the conventional methodology previously established in our laboratories [22]. The electrode fabrication begins with the preparation of the composite paste. For this, the resin EpoTek H77 (Epoxy Technology, Billerica, MA, USA) and its corresponding hardener compound were mixed with a the ratio of 20:3 (w/w). Afterwards 15% (w/w) of graphite (50 μm BDH Laboratory Supplies) and 2% (w/w) of the modifier (either the enzyme or the catalyst) were added to the previous mixture before hardening. Then, it was manually homogenized for 60 min, and afterwards, the paste was allowed to harden for seven days at 40°C. Finally, the electrode surface was polished with different sandpapers of decreasing grain size, with a final electrode area of 28 mm².

In this manner, an array of 4 different graphite-epoxy voltammetric sensors was prepared using bare graphite C and adding different modifiers such as tyrosinase, laccase and copper nanoparticles to the bulk mixture – one com-

ponent per electrode plus a blank electrode without any modifier.

The amperometric measurement cell was formed by the 4-sensor voltammetric array plus a reference double junction Ag/AgCl electrode (Thermo Orion 900200, Beverly, MA, USA) and a commercial platinum counter electrode (Model 52-67, Crison Instruments, Barcelona, Spain). Cyclic voltammetry measurements were carried out using a 6-channel AUTOLAB PGSTAT20 (Ecochemie, Netherlands) in a multichannel electrode configuration, controlled by GPES Multichannel 4.7 software package.

2.3.2 Voltammetric Measurements

Electrochemical measurements were performed at room temperature (25 °C) under quiescent condition. Firstly, prior to sample measurements, the electrodes were cycled in saline solution in order to get stable voltammetric responses.

For the measurements, the potential was cycled between -0.4 V and $+0.8$ V vs. Ag/AgCl, with a scan rate of 100 mVs^{-1} and a step potential of 9 mV. All experiments were carried out without performing any physical surface regeneration of the working electrodes. In order to prevent the accumulative effect of impurities on the working electrode surfaces, an electrochemical cleaning stage was performed between each measurement applying a conditioning potential of $+1.0$ V for 40 s after each experiment, in a cell containing 10 mL of distilled water [23].

2.4 Data Processing

The whole set of samples was measured altogether in random order, and divided into three subsets: training subset, testing subset and real samples subset. For each sample, four voltammograms were obtained, i.e. one voltammogram for each of the sensors from the BioET. Hence, in order to reduce the high dimensionality of the recorded signals (samples \times sensors \times potentials), a preprocessing stage of feature extraction was required.

This variable reduction stage was achieved by means of the *windowed slicing integral* method [18]. For this, the voltammogram is divided into k sections and the area under each section is used as input data point to the modelling stage (Figure 1). The idea behind this compression technique is to capture somehow the information about the voltammetric waveform characteristics. The number of sections (k) into which the voltammogram is divided is arbitrary and it depends on the response profile of the data. On the one side, by increasing k , the original information of the voltammetric data is best reproduced, but at expenses of the compression ratio. On the other hand, when k value is close to the number of points in the original signal, the obtained coefficients are just a scaled value of the original signal; that is, the original value multiplied

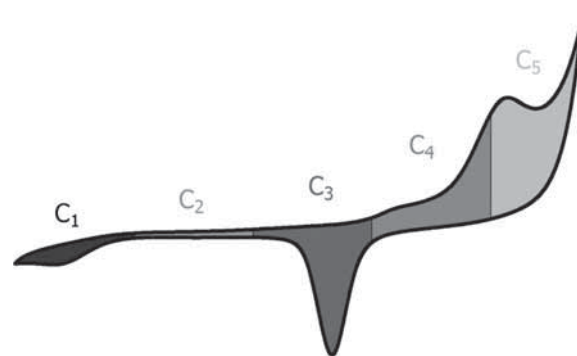


Fig. 1. Preprocessing scheme of the *windowed slicing integral* method followed by the compression and extraction of coefficients.

by a constant factor. Hence, a compromise between the compression ratio and signal reproduction must be taken.

After this, the obtained coefficients fed the ANN model which was used for the resolution and quantification of the three phenolic mixtures. In this way, an exhaustive study in the number of sections k and in the ANN architecture and configuration was done in order to optimize the separate quantification of the three phenols considered.

All computations and chemometric processing were carried out by specific routines written by the authors using MATLAB 7.1 (MathWorks, Natick, MA) and its Neural Network Toolboxes (v4.0.6). Sigmaplot 2000 (Systat Software Inc, California, USA) was used for graphic representations of data and results.

3 Results and Discussion

3.1 Voltammetric Responses

Following the described procedure, the set of samples was measured employing the BioET array, obtaining a whole cyclic voltammogram for each of the sensors from the array. Figure 2 shows some of the obtained voltammograms for the different phenolic mixtures. As can be seen, complex and highly overlapped signals are obtained along the whole voltammogram, with differentiated signals obtained for the different kinds of sensors; these changes were even more noticeable in the case of real samples analysis [20]. This situation, where the different sensors display marked distinct features for the different samples is an ideal departure point for building a BioET application.

As a general trend, two processes are observed for all the sensors corresponding to the oxidation of the corresponding phenol to its quinone form, and the reduction of the quinone to the polyphenolic form. Even with the overlapped signals obtained for the mixtures, some peaks that could be attributed to the oxidation of the $-\text{OCH}_3$ groups of the phenols are observed (Figure 3). Thus, as expected two peaks are obtained for sinapic acid, one for

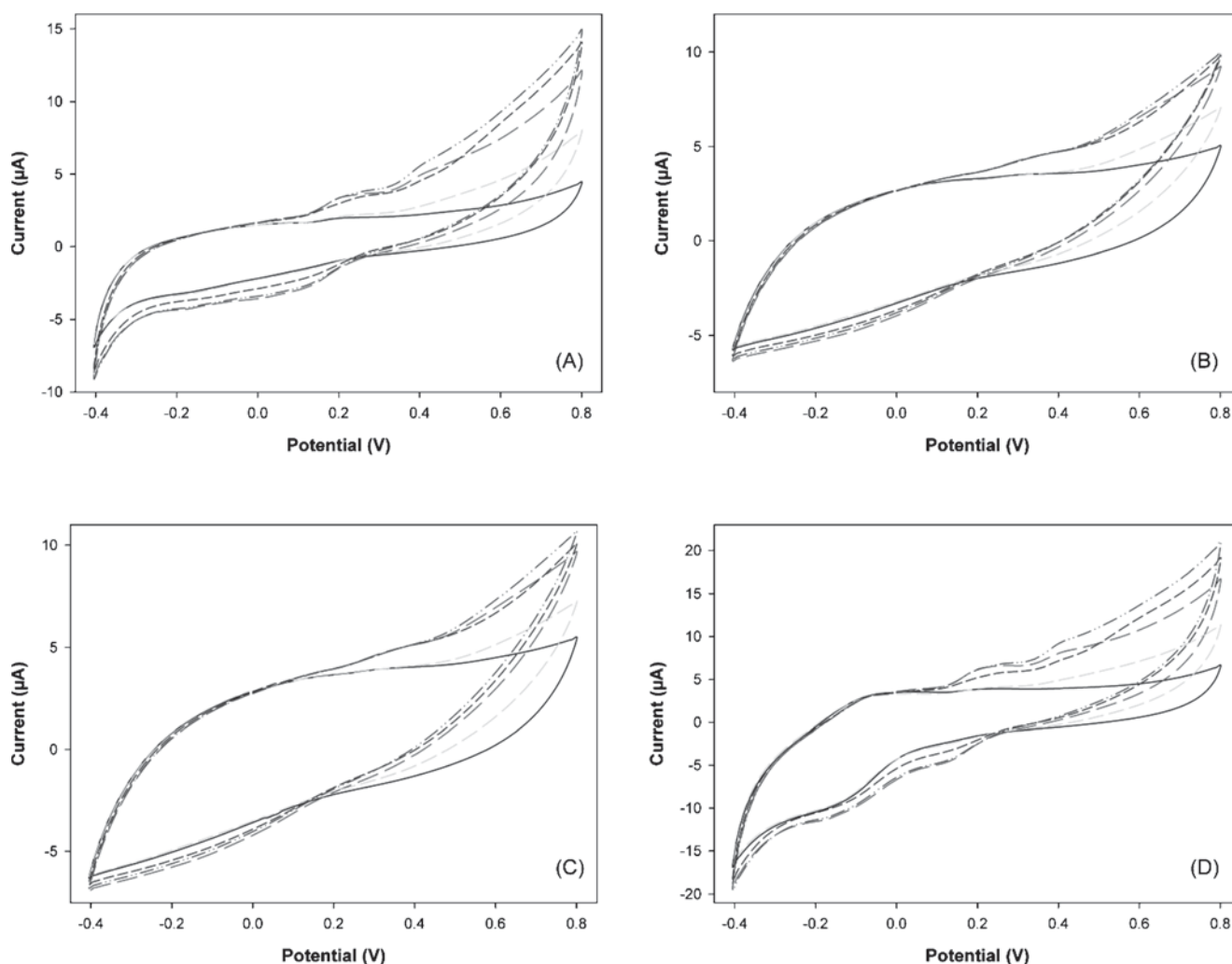


Fig. 2. Example of the different voltammograms obtained for certain arbitrary mixtures of the three phenolic compounds. Concentrations for each phenol are: (—) 0, 15.4, 61.6 μM ; (---) 84.8, 53.6, 176.8 μM ; (----) 153.6, 46.4, 153.6 μM ; (-----) 184.8, 138.4, 0 μM and (-·-·-·) 166.4, 100.8, 141.6 μM for ferulic, gallic and sinapic acids, respectively. Also signals provided by the different sensors are shown: (A) Graphite-epoxy sensor, (B) Tyrosinase biosensor, (C) Laccase biosensor and (D) Copper nanoparticle modified sensor.

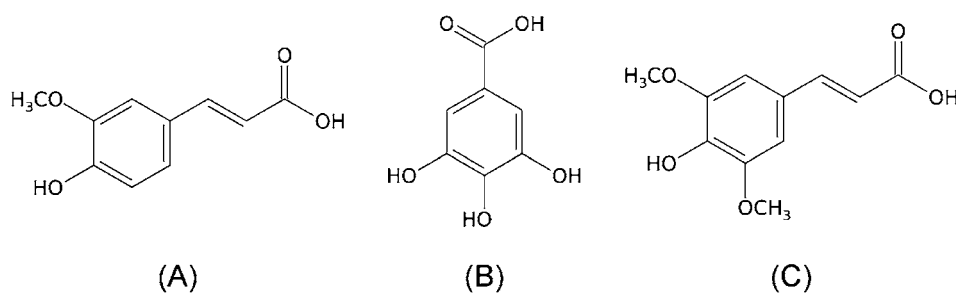


Fig. 3. Chemical structures of the polyphenols under study: (A) ferulic acid, (B) gallic acid and (C) sinapic acid, selected as the most abundant in beers.

ferulic acid and no peaks are observed in the case of gallic acid, as it does not have any methoxy group.

The identification of these peaks was achieved after the characterization of the BioET towards standard solutions of each of the polyphenols. To confirm this differentiated behaviour, and as a previous step prior to the construc-

tion of the quantitative model, raw voltammetric responses of those standards were analyzed employing PCA. As shown in Figure 4, clear discrimination was achieved for the three compounds. Analyzing the plot more deeply, it could be seen that samples are sorted along PC1 depending on their number of methoxy groups, with gallic acid

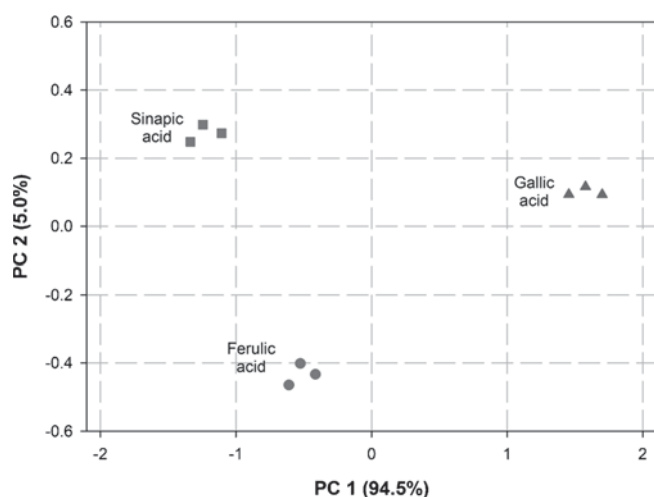


Fig. 4. Score plot of the first two components obtained after PCA analysis of the different polyphenols standards (50 μ M solutions).

on the right far from the others, and ferulic and sinapic acids close to each other (both containing one and two groups, respectively), mainly separated by PC2.

In the case of biosensors (Figure 2B and 2C), less significant features were shown in the voltammetric response, given the enzymes catalyze similarly the oxidation of the three phenolic compounds, but with differentiated selectivity towards each one [9]. Some oxidative response is still observed at higher potentials ($E > 0.5$ V) mainly due to direct oxidation onto the graphite electrode surface, but with less discriminated response in the intermediate oxidation zone due to the enzyme already catalyzes the oxidation of all the compounds, hence no such differences as the ones obtained for the oxidation of the methoxy groups on the other electrodes are observed. In this sense, most relevant information along the voltammetric signal, in the case of tyrosinase and laccase biosensors, is found in the cathodic part of the CV where the reduction of the generated quinones by the enzymes takes place. Although in the case of synthetic samples, the benefits on the improvement of the response due to the use of enzymatic biosensors might not be so clear; there is a huge improvement on the response when dealing with real samples (especially in the reduction region close to 0 V) [20] due to the biosensor specificity provided by the use of enzymes.

3.2 Construction of the ANN Model

Prior to building the quantification model, and in order to gain advantages in the modelling stage and to fully exploit all the information obtained from each voltammogram, a preprocessing compression step was used to reduce the high dimensionality of the recorded signals (samples \times sensors \times potentials) while preserving the relevant information. In addition to data reduction, compression is intended for extracting significant features from

the departure information, besides eliminating of irrelevant content, such as noise or redundancies [24]. That is, in addition to size reduction, advantages are gained in training time and reduction of memory needs; moreover, this compression helps to avoid redundancy in input data, to obtain a model with better modelling accuracy and higher robustness plus better generalization ability and easier operation and understanding [25]. Lastly, an additional reason is the need to use reduced and significant information compatible with the ANN structure, which must have a reasonable size and complexity at its inputs. As stated, in our case this variable reduction was achieved by means of the *windowed slicing integral* method.

The main drawback of this method is the difficulty to predict in advance the optimal number of sections prior to building the multivariate calibration model, in our case, the ANN model. Hence, the number of coefficients that must be used as input was optimized together with the ANN architecture, varying the number of sections (k).

The first step in building the ANN model is selecting the topology of the network. This requires a significant effort due to the difficulties in predicting the optimum ANN configuration in advance. A trial-and-error process is needed, where several parameters (training algorithms, number of hidden layers, transfer functions, etc.) are fine-tuned in order to find the best configuration to optimize the performance of the neural network model [25]. In this way, an exhaustive search of the ANN architecture was performed varying the number of sections (k) from 1 to 20, the number of neurons in the hidden layer (between 1 and 12) and evaluating the combination of different transfer functions (*tansig*, *logsig*, *purelin* and *satlin*) in both hidden and output layers.

For the selection of the best ANN model, the Root Mean Square Error (*RMSE*) for each one of the configurations was calculated and an overall comparison was performed. Also, comparison graphs of predicted vs. expected concentrations for the three determined phenols were built to check the prediction ability of the ANN. In this manner, the optimum model would have a small *RMSE*; and also slope, intercept and correlation coefficient of the comparison close to the ideal values (1, 0 and 1, respectively).

Then, after an extensive study varying its configurations, the final ANN architecture model had 44 neurons (4 sensors \times 11 coefficients obtained from the *windowed slicing integral* method) in the input layer, 5 neurons and *tansig* transfer function in the hidden layer and 3 neurons and *purelin* transfer function in the output layer (one for each polyphenol).

It should be noticed, that the use of *windowed slicing integral* method allowed a compression of the original information up to 95.9% without any loss of significant information (11 coeffs. vs. originally 268 current values in each voltammogram). This is a remarkable value taking into account the simplicity of the method, both from the theoretical point of view as well as from the calculation

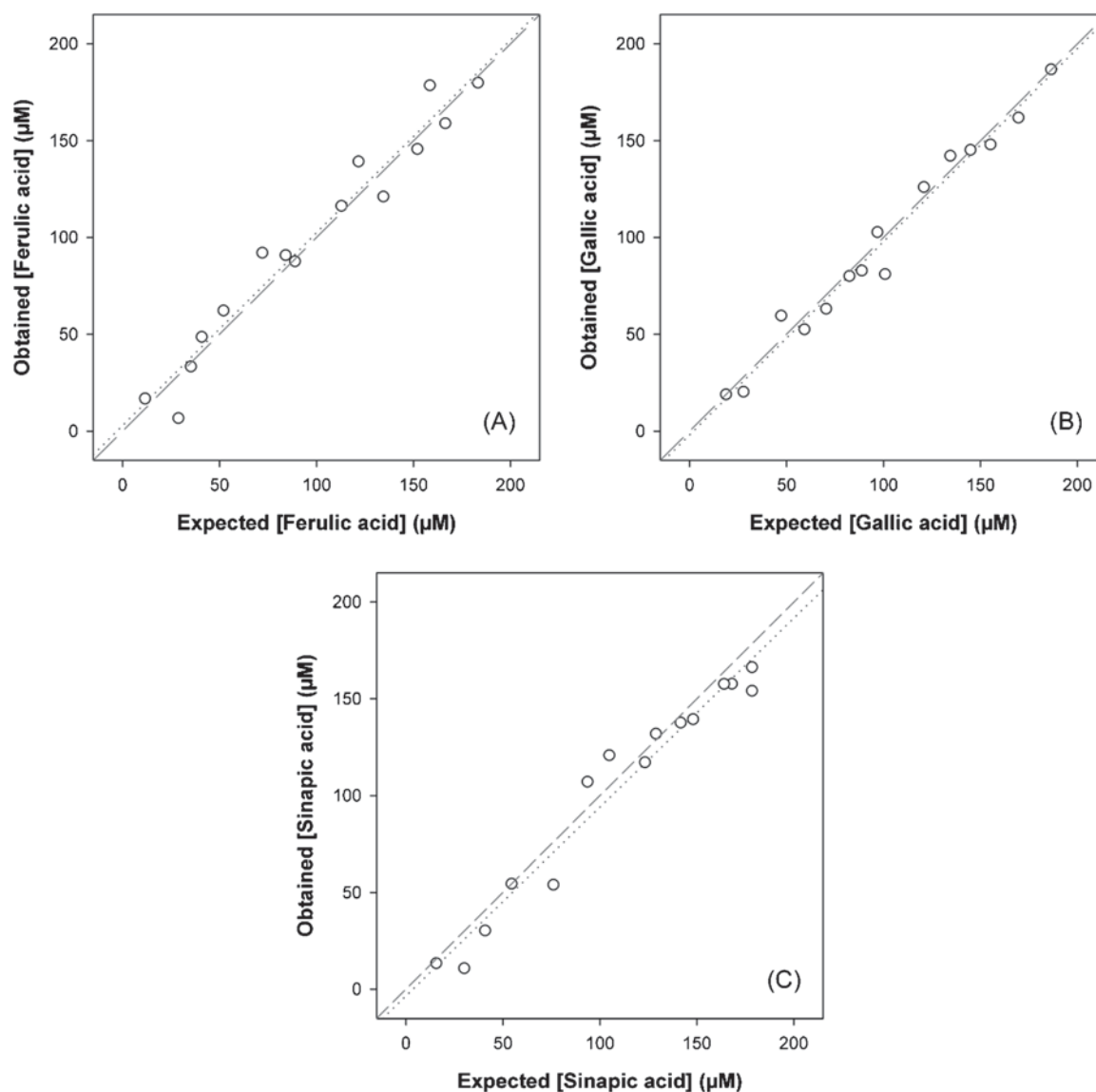


Fig. 5. Modelling ability of the optimized ANN for the external test subset of samples. Set adjustments of obtained vs. expected concentrations for (A) ferulic, (B) gallic and (C) sinapic acids. Dashed line corresponds to theoretical diagonal line.

requirements. Detailed information of the results obtained when varying the number of the sections (k) using the *windowed slicing integral* method can be found in Supporting Information. As a general trend, and as could be expected, by increasing the number of coefficients extracted, which means and increase of raw data representation, results tend to improve with huge improvements at the beginning and lower differences at higher k levels.

3.3 Quantification of the Three Phenolic Mixtures

3.3.1 Standard Solutions

After building the ANN model, and in order to prove its capabilities, its performance was first evaluated towards samples in the external test subset. Figure 5 shows the

testing performances for the three compounds with the selected ANN model, where its modelling and prediction capabilities are demonstrated. It must be remarked that the external test subset data is not employed at all for the modelling, so its goodness of fit is a measure of the accomplished modelling performance. In all the cases, very good correlations were obtained, with regression lines almost indistinguishable from the theoretical ones.

Table 1 summarizes the obtained regression parameters. As expected from the comparison graphs, a good linear trend is attained for all the cases, but with better correlation coefficients in the training subsets due to the lower dispersion. Despite this, the results obtained for both subsets are close to the ideal values, with intercepts close to 0 and slopes and correlation coefficients practically 1 (r values > 0.977).

Table 1. Results of the fitted regression lines for the comparison between obtained vs. expected values for the BioET, both for the training and testing subsets of samples and the three considered phenols (intervals calculated at the 95 % confidence level).

| Phenol | Correlation | Slope | Intercept (μM) | RMSE (μM) | Total RMSE (μM) |
|---------------------------|-------------|-------------------|-----------------------------|------------------------|------------------------------|
| Training subset | | | | | |
| Ferulic acid | 0.994 | 0.973 ± 0.044 | 2.3 ± 5.1 | 8.82 | 8.00 |
| Gallic acid | 0.998 | 0.988 ± 0.029 | 0.7 ± 3.3 | 5.75 | |
| Sinapic acid | 0.994 | 0.972 ± 0.045 | 3.3 ± 5.2 | 9.02 | |
| Testing subset | | | | | |
| Ferulic acid | 0.977 | 0.995 ± 0.131 | 2.9 ± 14.4 | 11.97 | 11.10 |
| Gallic acid | 0.988 | 0.998 ± 0.094 | -1.9 ± 10.5 | 8.03 | |
| Sinapic acid | 0.978 | 0.976 ± 0.123 | -3.5 ± 15.1 | 12.71 | |
| Real sample subset | | | | | |
| Ferulic acid | 0.986 | 0.986 ± 0.098 | 3.2 ± 11.6 | 9.21 | 10.46 |
| Gallic acid | 0.981 | 1.001 ± 0.119 | 2.4 ± 12.9 | 10.07 | |
| Sinapic acid | 0.976 | 0.994 ± 0.134 | 4.6 ± 15.3 | 11.92 | |

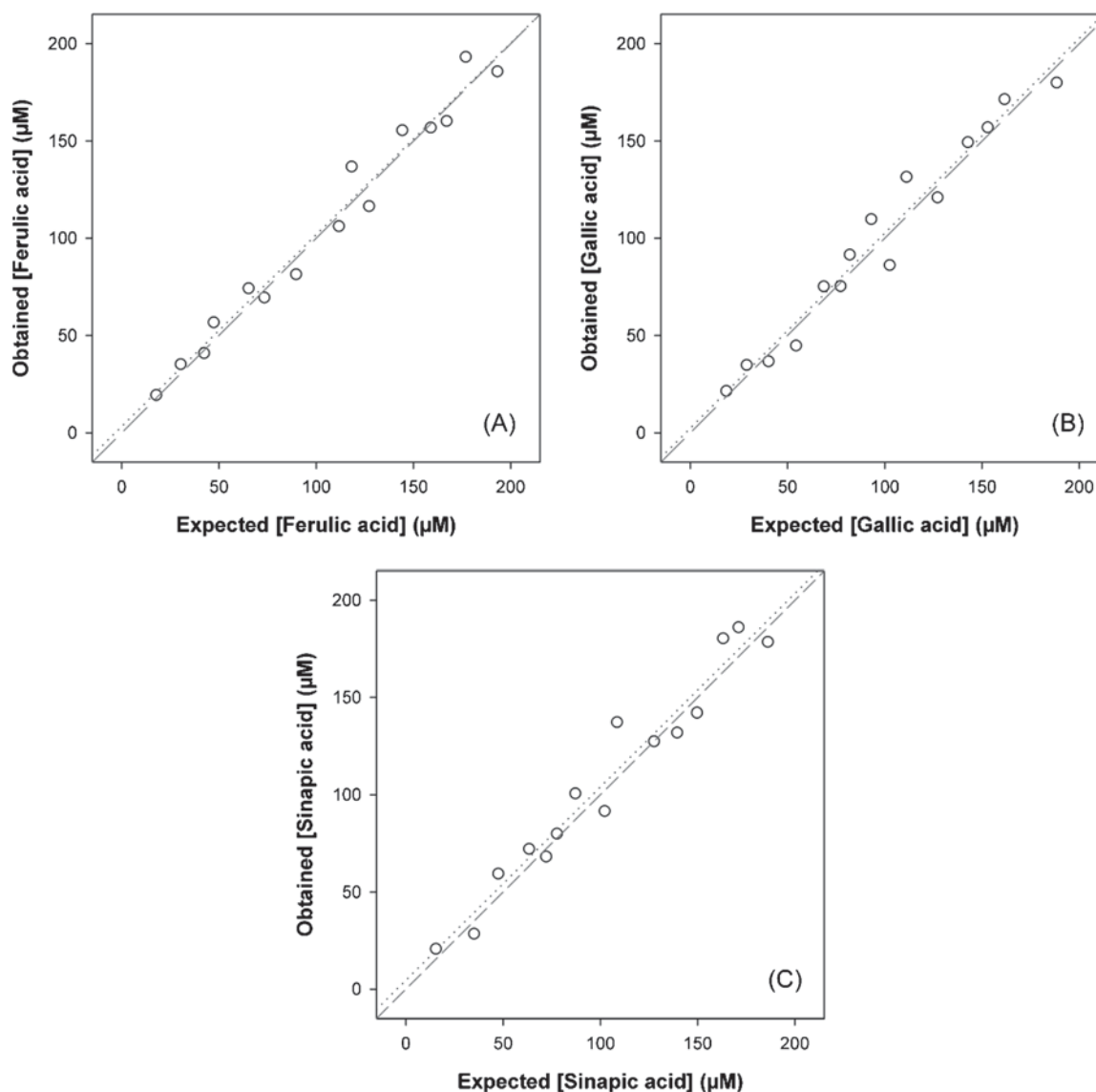


Fig. 6. Modelling ability of the optimized ANN for spiked beer samples. Set adjustments of obtained vs. expected concentrations for (A) ferulic, (B) gallic and (C) sinapic acids. Dashed line corresponds to theoretical diagonal line.

Table 2. Detailed results obtained with the BioET approach for the spiked beer samples against added concentrations of the three phenolics considered. Recovery yield is also expressed for each sample based on found vs. added concentrations.

| Beer | Ferulic acid concentration | | | Gallic acid concentration | | | Sinapic acid concentration | | |
|------|----------------------------|-------------------------|--------------------|---------------------------|-------------------------|--------------------|----------------------------|-------------------------|--------------------|
| | Added (μM) | BioET (μM) | Recovery yield (%) | Added (μM) | BioET (μM) | Recovery yield (%) | Added (μM) | BioET (μM) | Recovery yield (%) |
| 1 | 111.64 | 106.20 | 95.1 | 54.38 | 44.84 | 82.5 | 72.06 | 68.11 | 94.5 |
| 2 | 158.91 | 156.90 | 98.7 | 81.97 | 91.71 | 111.9 | 139.37 | 132.01 | 94.7 |
| 3 | 89.64 | 81.51 | 90.9 | 142.66 | 149.39 | 104.7 | 149.67 | 142.17 | 95.0 |
| 4 | 193.13 | 185.70 | 96.2 | 161.57 | 171.57 | 106.2 | 87.11 | 100.69 | 115.6 |
| 5 | 127.13 | 116.49 | 91.6 | 77.24 | 75.42 | 97.7 | 47.51 | 59.48 | 125.2 |
| 6 | 167.06 | 160.29 | 96.0 | 102.46 | 86.26 | 84.2 | 108.49 | 137.38 | 126.6 |
| 7 | 42.38 | 40.95 | 96.6 | 40.20 | 36.70 | 91.3 | 77.61 | 79.93 | 103.0 |
| 8 | 118.16 | 136.89 | 115.8 | 111.13 | 131.52 | 118.4 | 127.50 | 127.53 | 100.0 |
| 9 | 17.72 | 19.36 | 109.2 | 18.52 | 21.47 | 115.9 | 63.35 | 72.12 | 113.8 |
| 10 | 73.34 | 69.56 | 94.8 | 188.37 | 179.88 | 95.5 | 171.05 | 186.10 | 108.8 |
| 11 | 176.84 | 193.13 | 109.2 | 127.09 | 120.95 | 95.2 | 15.44 | 20.80 | 134.7 |
| 12 | 144.24 | 155.59 | 107.9 | 28.96 | 34.81 | 120.2 | 35.04 | 28.62 | 81.7 |
| 13 | 47.27 | 56.76 | 120.1 | 68.57 | 75.27 | 109.8 | 102.15 | 91.65 | 89.7 |
| 14 | 30.36 | 35.22 | 116.0 | 93.00 | 109.80 | 118.1 | 186.10 | 178.56 | 96.0 |
| 15 | 65.19 | 74.26 | 113.9 | 152.90 | 157.09 | 102.7 | 163.13 | 180.38 | 110.6 |

3.3.2 Spiked Beer Samples

Once that the BioET model behaviour towards the resolution of these phenolic compounds was proven with standard solutions, and in order to test its applicability to real samples, some beer samples were spiked with variable amounts of the three phenolic compounds considered, with additions randomly distributed in the range of the experimental design.

Employing the same ANN architecture with leave-one-out validation method, forced by the reduced number of samples and to ensure that each sample in the set is used in the validation step, the system was retrained. Then, prediction capability of the ANN model for beer samples was evaluated. As before, comparison graphs of predicted vs. expected concentrations for the three determined phenols were built (Figure 6). As can be seen, a good trend is obtained for the three phenolics, with regression lines almost indistinguishable from the theoretical ones. Again, obtained regression parameters can be found in Table 1 (r values > 0.976). As expected from the plot, the uncertainty intervals included the ideal slopes of 1 as well as the 0 intercepts in all the cases, demonstrating the goodness of fit.

In addition, the recovery yield of the added theoretical concentration was calculated for each phenolic compound and each spiked sample, which results are summarized in Table 2. As can be seen, obtained results are satisfactory, with a good recovery yield in general (average values of 103%, 103% and 106% for ferulic, gallic and sinapic acids, respectively), illustrating the value of the approach.

4 Conclusions

The quantitative resolution at low concentration level of mixtures of major polyphenols found in beers has been

achieved by means of a BioET using an array of voltammetric biosensors. For this, obtained responses of the BioET array were firstly preprocessed employing a *windowed slicing integral*, which provided a good compression of data while preserving relevant information. Then, the obtained coefficients fed an ANN model that allowed the quantification of the three phenolics, overcoming the signal overlapping and providing the resolution of the mixtures. Particularly, the developed BioET has been evaluated for the resolution of ferulic, gallic and sinapic acids in aqueous and beer sample matrices, with good results in an application comparable to more complex analytical techniques such as HPLC.

In addition, the use of *windowed slicing integral* has been firstly proposed and successfully used as feature extraction tool in voltammetric data with a good compression degree and faithful representation of data prior to ANN modelling.

Acknowledgments

Financial support for this work was provided by the *Spanish Ministry of Science and Innovation, MCINN* (Madrid) through Project CTQ2010-17099 and by Program ICREA Academia *Generalitat de Catalunya*. X. Cetó thanks the support of *Dept. d'Innovació, Universitats i Empresa de la Generalitat de Catalunya* for the predoctoral grant.

References

- [1] H. Sakakibara, Y. Honda, S. Nakagawa, H. Ashida, K. Kanazawa, *J. Agric. Food Chem.* **2002**, *51*, 571.
- [2] G. J. Soleas, E. P. Diamandis, A. Karumanchiri, D. M. Goldberg, *Anal. Chem.* **1997**, *69*, 4405.
- [3] M. Naczka, F. Shahidi, *J. Pharmaceut. Biomed.* **2006**, *41*, 1523.

- [4] V. L. Singleton, J. A. Rossi, *Am. J. Enol. Viticult.* **1965**, *16*, 144.
- [5] M. Dvořáková, M. Douanier, M. Jurková, V. Kellner, P. Dostálek, *J. Inst. Brew.* **2008**, *114*, 150.
- [6] M. Nardini, A. Ghiselli, *Food Chem.* **2004**, *84*, 137.
- [7] D. M. A. Gil, M. J. F. Rebelo, *Eur. Food Res. Technol.* **2010**, *231*, 303.
- [8] V. C. Sanz, M. L. Mena, A. González-Cortés, P. Yáñez-Sedeño, J. M. Pingarrón, *Anal. Chim. Acta* **2005**, *528*, 1.
- [9] R. Solná, P. Skládal, *Electroanalysis* **2005**, *17*, 2137.
- [10] M. Elkaoutit, I. Naranjo-Rodríguez, K. R. Temsamani, M. D. La Vega, J. L. H. De Cisneros, *J. Agric. Food Chem.* **2007**, *55*, 8011.
- [11] M. del Valle, *Electroanalysis* **2010**, *22*, 1539.
- [12] J. Zeravik, A. Hlavacek, K. Lacina, P. Skládal, *Electroanalysis* **2009**, *21*, 2509.
- [13] Y. Vlasov, A. Legin, A. Rudnitskaya, C. Di Natale, A. D'Amico, *Pure Appl. Chem.* **2005**, *77*, 1965.
- [14] M. Ghasemi-Varnamkhasti, M. L. Rodríguez-Méndez, S. S. Mohtasebi, C. Apetrei, J. Lozano, H. Ahmadi, S. H. Razavi, J. Antonio de Saja, *Food Control* **2012**, *25*, 216.
- [15] A. Gútes, F. Céspedes, S. Alegret, M. del Valle, *Biosens. Bioelectron.* **2005**, *20*, 1668.
- [16] X. Cetó, F. Céspedes, M. I. Pividori, J. M. Gutiérrez, M. del Valle, *Analyst* **2012**, *137*, 349.
- [17] R. M. de Carvalho, C. Mello, L. T. Kubota, *Anal. Chim. Acta* **2000**, *420*, 109.
- [18] R. Gutiérrez-Osuna, H. T. Nagle, *IEEE T. Syst. Man Cy. B* **1999**, *29*, 626.
- [19] L. Moreno-Barón, R. Cartas, A. Merkoçi, S. Alegret, M. del Valle, L. Leija, P. R. Hernandez, R. Muñoz, *Sens. Actuators B, Chem.* **2006**, *113*, 487.
- [20] X. Cetó, F. Céspedes, M. del Valle, *Talanta* **2012**, dx.doi.org/10.1016/j.talanta.2012.06.031, in press.
- [21] L. Zhang, Y. Z. Liang, J. H. Jiang, R. Q. Yu, K. T. Fang, *Anal. Chim. Acta* **1998**, *370*, 65.
- [22] F. Céspedes, E. Martínez-Fàbregas, S. Alegret, *TrAC – Trends Anal. Chem.* **1996**, *15*, 296.
- [23] X. Cetó, J. M. Gutiérrez, L. Moreno-Barón, S. Alegret, M. del Valle, *Electroanalysis* **2011**, *23*, 72.
- [24] L. Moreno-Barón, R. Cartas, A. Merkoçi, S. Alegret, J. M. Gutiérrez, L. Leija, P. R. Hernández, R. Muñoz, M. del Valle, *Anal. Lett.* **2005**, *38*, 2189.
- [25] F. Despagne, D. Luc Massart, *Analyst* **1998**, *123*, 157R.

Assessment of Individual Polyphenol Content in Beer by Means of a Voltammetric BioElectronic Tongue

Xavier Cetó, Francisco Céspedes and Manel del Valle*

Sensors and Biosensors Group, Department of Chemistry, Universitat Autònoma de Barcelona, Edifici Cn, 08193 Bellaterra, Barcelona, SPAIN

Supporting information

This supplementary information gives detailed results obtained during the ANN optimization process for the RMSE values and the correlation coefficients and slopes calculated from the comparison graph between expected and found concentrations, for some of the different configurations tested just to illustrate the effect of the number of sections used with the *windowed slicing integral* method.

Residual values were calculated between expected and found concentration values, for each sample (i) and for each of the three analytes (j) considered according to equation 1. Also, for easier interpretation of the results and to simplify the selection of the best model, the difference of the optimal values for the correlation coefficients (Δr) and the slopes (Δm) from the comparison graphs of the 3 analytes were calculated according to equations 2 and 3, respectively. Hence, given ideal values for each parameter are 1, both Δr and Δm will be ranged between 3 and 0; they value 3 when there is no prediction capability by the model and decrease as model behaviour improves.

$$RMSE = \sqrt{\frac{\sum_{ij} (c_{ij} - \hat{c}_{ij})^2}{3n-1}} \quad (1)$$

$$\Delta r = \sum_j |1 - r_j| \quad (2)$$

$$\Delta m = \sum_j |1 - m_j| \quad (3)$$

* E-mail: manel.delvalle@uab.cat; tel: +34 93 5811017; fax: +34 93 5812379

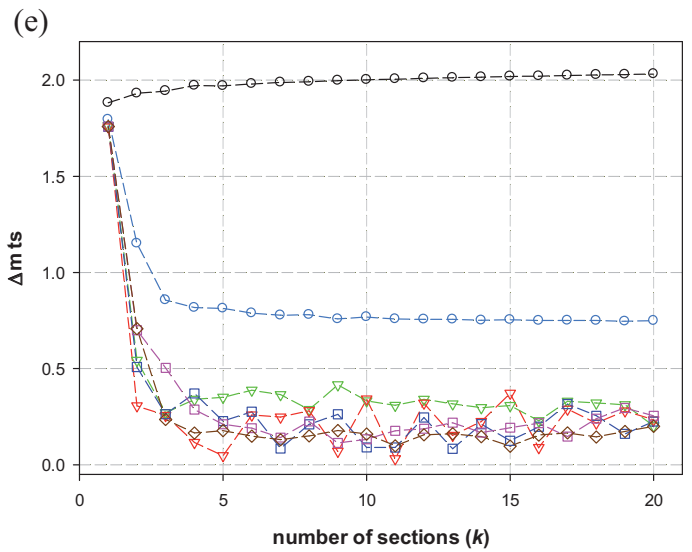
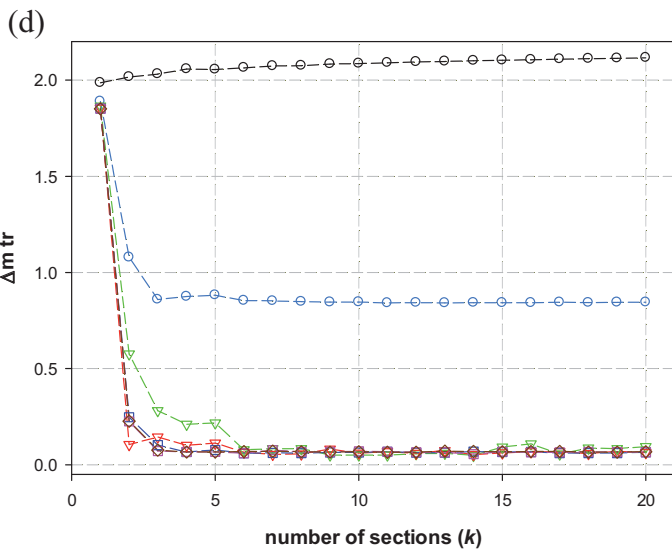
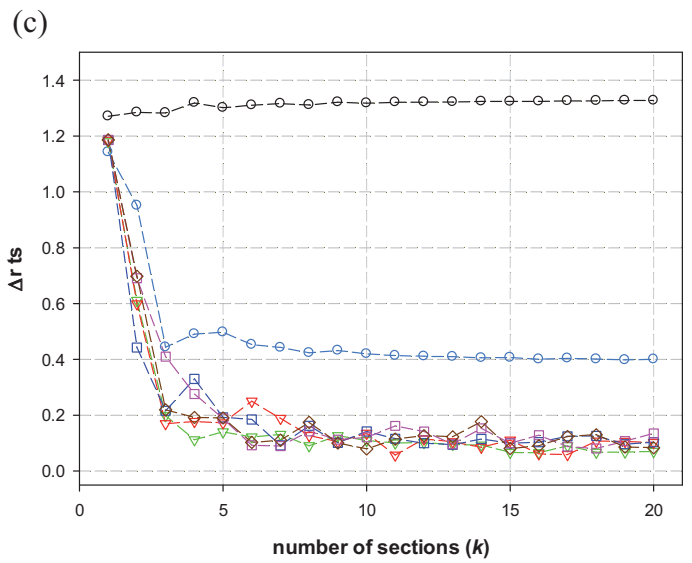
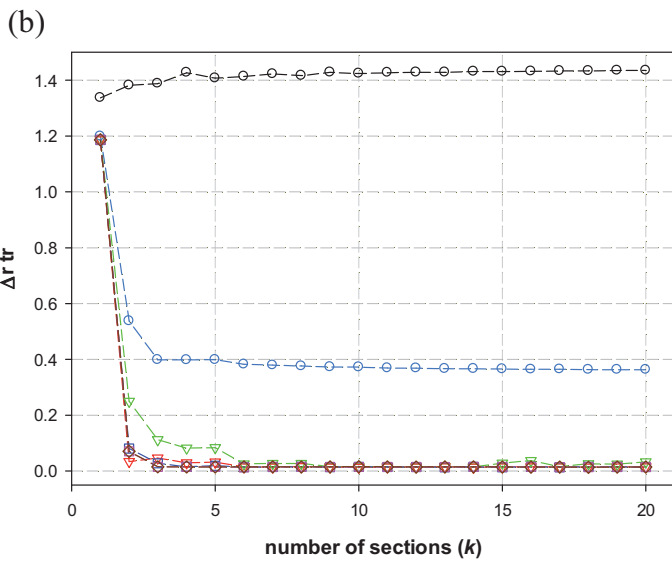
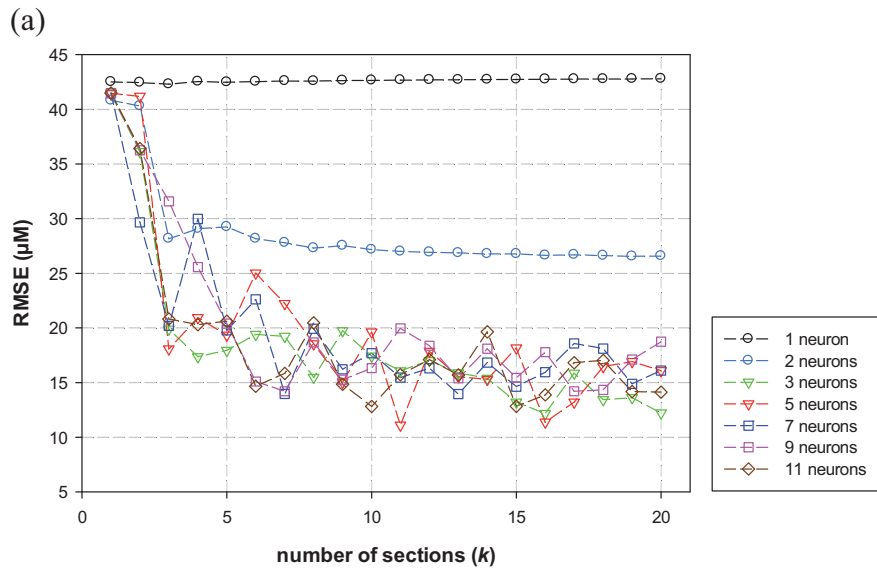


Figure S1. Detailed results of the ANN optimization. Obtained RMSEs (a), slopes (b, c) and correlation coefficient values (d, e) of obtained vs expected comparison graphs (both for training (tr) and testing subsets (ts), respectively) are plotted against different number of sections in the *windowed slicing integral* method. Results shown correspond to different ANN architectures, varying both the number of neurons in the input (number of coefficients) and hidden layers; but all with *tansig* and *purelin* transfer functions between input and hidden layer and between hidden and output layer, respectively.

Article 6

Chapter 6: BioElectronic Tongues: When the sensor array incorporates biosensors

M. del Valle, X. Cetó and M. Gutiérrez

in *Multisensor systems for chemical analysis - materials and sensors* by

L. Lvova, D. Kirsanov, C. Di Natale and A. Legin

Pan Stanford Publishing Pte Ltd. ISBN: 9789814411158

1.1 INTRODUCTION

Most of the developments gathered in the field of sensors and biosensors are aimed to new designs, configurations or implementations towards the ideal sensor showing perfect selectivity. Opposite to the search of more selective sensors, there has been a trend from the late 90s of using sensors in array mode in order to obtain some added value in the generation of analytical information. Some variants of this trend are the use of extra sensors for detection of malfunctioning episodes, parallel multi-determination schemes to accelerate sample throughput, or the array with cross-sensitivity features combined with pattern recognition tools, which has led to what is referred to as electronic noses and tongues. According to the agreed IUPAC definition [1], an electronic tongue is "a multisensor system, which consists of a number of low-selective sensors and uses advanced mathematical procedures for signal processing based on Pattern Recognition and/or Multivariate data analysis – Artificial Neural Networks (ANNs), Principal Component Analysis (PCA), etc". The underlying motivation of electronic tongues (ETs) is then different from the general trend; this is, to use low selectivity sensors or with cross response features, a prerequisite for the development of these biomimetic systems. In the aforementioned systems, each sensor uses one element to differentiate the response. This can be different ionophores originating the potentiometric response, different metal electrodes to induce different redox behaviour, or different catalysts that show different voltammetric response (see Figure 6.1). Many research papers are found along the ten-year history of ETs that use sensors of different types, but a new class is emerging, named the bioelectronic tongue (BioET); this one is characterized by including one or several biosensors into the sensor array. To be considered a biosensor, the sensor's recognition element must be of biological origin, namely enzymes, antigens, antibodies, nucleic acids, receptor proteins, cells, or even tissues. Our research group was pioneering in this variant, developing a BioET with voltammetric biosensors for glucose [2], another BioET with voltammetric biosensors for phenolic compounds [3] or, also, a BioET employing potentiometric biosensors for determination of urea [4]. In the literature, it is not usual to find BioETs, as they require the conjunction of very diverse knowledge areas: biochemistry, analytical chemistry, instrumentation and chemometrics.

Chapter 6

BioElectronic Tongues: When the Sensor Array Incorporates Biosensors

Manel del Valle^{1,*}, Xavier Cetó¹, Manuel Gutiérrez-Capitán²

¹Sensors and Biosensors Group, Department of Chemistry, Universitat Autònoma de Barcelona, Edifici Cn, 08193 Bellaterra, Spain

²Instituto de Microelectrónica de Barcelona (IMB-CNM), CSIC, 08193 Bellaterra, Spain

manel.delvalle@uab.es

One of the recent advances in the design of electronic tongues has been the incorporation of biosensors, in order to tackle new application fields or to improve existing ones. These bioelectronic tongues, as they have been named, are only distinguished from conventional ones in the incorporation of one or several biosensors into the sensor array, normally sharing the same transduction principle to facilitate compatibility. After reviewing existing examples from the literature, this chapter will deepen in two case studies, one with potentiometric sensors and the other with devices of the voltammetric type. The first is designed having in mind the monitoring of the haemodialysis process, where urea and creatinine biosensors constructed on the basis of the proper enzymes and ammonium ion selective electrodes are complemented with additional sensors to alkaline ions. The second case, aimed to the detection of phenolic antioxidant compounds, uses an array of voltammetric biosensors incorporating different phenol-oxidizing enzymes to generate the cross-sensitive responses. In both cases, the data processing tool used has been Artificial Neural Networks, able to provide a precise response model to estimate concentrations of involved species.

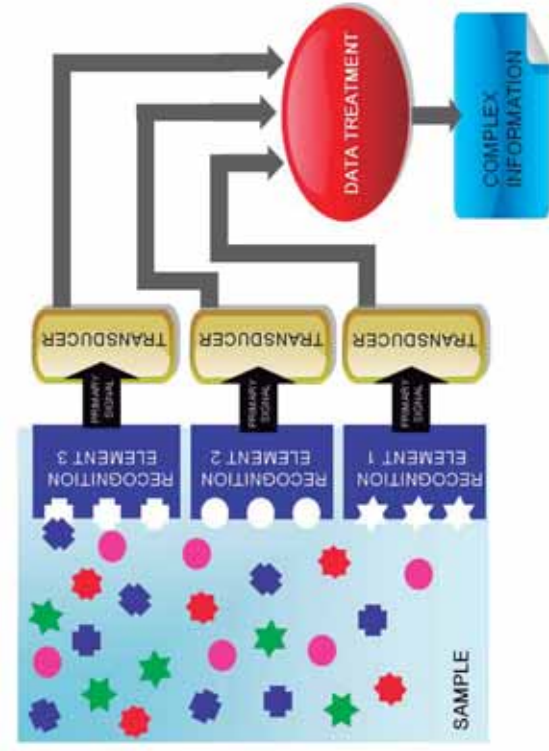


Figure 6.1. The use of the biosensor array as key concept in the development of (bio)electronic tongues.

1.1.1 Building of a BioElectronic Tongue

An electronic tongue can be defined as an analytical system applied to liquid analysis formed by a sensor array in order to generate multidimensional information, plus a chemometric processing tool to extract meaning from these complex data. Recent reviews can be consulted in the literature to check the extension of variants that have been employed up to now to develop these systems [5]. Although there are electronic tongues described employing optical sensors or piezoelectric (mass) sensors, most of the described works employ sensors of the electrochemical type [6]. Among these, many exploit sensors of potentiometric [7, 8] or voltammetric type [9], although systems employing impedimetric sensors are also significant [10]. To respond to the expectation in the sensor community, recent special issues of significant journals in the analytical chemistry field have appeared dealing with the topic [11, 12].

One would expect that biosensors don't need to be used in array mode, as they usually involve very selective, almost specific determinations. But, although of the general thought, biosensors use can present response to a group of specific substances, e.g., sugars or pesticides, or may show interfering effects by tertiary substances. If one creates a set of biosensors, where each biosensor recognizes one or a group of analytes or compensates for the presence of interferences, then with the help of advanced statistical methods, very accurate characterization of the sample can be obtained.

One of the precautions to be taken for the correct operation of the BioET, is that collinearity must be absent; this is, every element forming the array must originate a signal which is not linearly dependent to the rest of biosensors. This can be accomplished through the use of different recognition elements, the use of biological elements of equivalent function but different origin (e.g. the use of equivalent enzymes but from different animals to show different response) or by affecting the response of the biological element, e.g. like the use of chemical modifiers, i.e. the use of redox mediators, or also with the use of physical modifiers, e.g. the use of metallic catalysts.

Regarding the different data processing tools used with electronic tongue systems, although being one of their chief parts, a detailed description is out the scope of this chapter and they will only be mentioned as used; a proper exposition can be consulted in the review by Richards et al. [13]. It can be said, anyhow, that data processing details are mostly equivalent between the use of ETs and electronic noses [14], i.e. the equivalent analytical systems but using gas sensors and applied to gas media.

Basically, the lay reader needs to know that certain processing algorithms are more disposed to identification/classification purposes, such as Principal Component Analysis (PCA), and are therefore used for qualitative applications. Other procedures, for example regression using Partial Least Squares (PLS), are specially devised for quantification purposes, and mainly used for multidetermination applications [15]. The mentioned chemometric variants are conventional pattern recognition techniques, mainly linear in nature, which can be somehow limited if the sensors considered show nonlinear response trends. To improve results, researchers have proposed the use of Artificial Neural Networks (ANNs), which is a massively parallel computing technique, specially

suitable to non-linear sensor responses and very much related to human pattern recognition. Concerning the experience in our laboratory, we are most in favour of the use of ANNs, as these are very powerful modelling tools, amenable to both qualitative and quantitative applications [16].

Figure 6.2 sketches the operation of an ET system based on an ANN model. In this case, it develops a multidetermination application, i.e. to simultaneously yield the concentrations of a number of species from the readings provided by a sensor array. Measures are directly fed to an input layer of elements that distribute information to a number of processing elements in the hidden layer. The operation is directly inspired by the biological brain, as each previous element acts as a neuron; then, it is activated if enough information is introduced into it. The sought information (in this case the concentrations of species A and B) is obtained from an output layer, also processing its entering information as before. To obtain the proper response model, one must obtain the proper weight connections distributing the information, in a process known as backpropagation, and also the way to activate the output of each neuron, i.e. the transfer function. To obtain the weights, a mathematical optimization procedure is followed, in which known samples are first presented to the system, and a learning algorithm is applied (this is the training phase of the model building). Other details of the final configuration are the number and layers of neurons used, also to be obtained by trial and error. Once the response has been built, the recommended method is to use new data to verify its performance, as any other validation should be fouled by previously known information. For this purpose, new samples are presented to the system, and the obtained responses are compared to those expected (testing phase). A usual procedure to check this data is to build obtained/vs./expected comparison graphs, and use the correlation coefficient, slope and intercept of the linear regression of the data as indicators of the goodness of the modelling. For each modelled species, a sufficiently significant correlation should be obtained, and the general trend should not show any bias. As commented, the use of ANNs will provide improved results in comparison with other tools, especially if there are non-linearities in the biosensors response. One of its drawbacks is the need for a sufficiently large number of samples, so the weight coefficients (the training) can be obtained with enough precision, and the final results are not conditioned by which samples are designated for the training phase

and which others for testing. This drawback has been counterbalanced with the use of automated flow systems capable of preparing without intervention the needed sets of samples [17]. For a detailed description of the use of ANNs to build multidetermination response models, the reader can refer to [18]. In [19], the use of the shareware software EasyNNplus has been shown for that purpose.

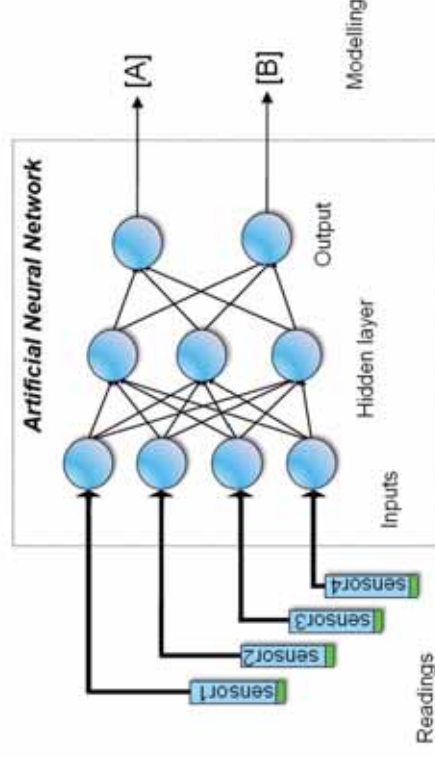


Figure 6.2. The electronic tongue concept applied for multi-determination employing a neural network model.

1.1.2 BioElectronic Tongues in the literature

Several designs of bioelectronic tongues can be found in the literature. The multisensor approach, to be viable must bring some added value, i.e. some new information, different to the individual concentrations yielded by each single sensor. This can be the imaging of a concentration profile from a multichannel portrait, or some error prevention ability gained along the process. The potential of these new tools lies in the new types of analytical information obtained, different from the simple sum of individual signals; examples of these are the correction of a matrix effect, the obtaining of some characteristic not directly attainable through a chemical composition, like the identification of a sample variety or the correlation with a perception from a panel of sensory experts.

The type of applications described can be classified, first, in qualitative type, like identification of species, classification of sample varieties, or recognition of an adulteration episode; conversely, the application can be also of quantitative type, normally the multidetermination of a set of chemical species, an interesting objective for process control. A different, more bioinspired trend is the artificial taste concept [20] devised with the goal of performing automated taste perception, especially in the industrial field. This idea, initiated by the works of Prof. Toko in Japan [21], consists in developing a sensor array inspired in the tongue's papillae and/or responding to the basic taste types (sweet, sour, salty and bitter, plus the far-east concept umami, which meaning is delicious); the ultimate goal is to mimic analogues of human taste assessment in cases where a human expert is not feasible, like in automatic control, poisonous/extreme condition samples, etc.

The first analytical system employing an array of biosensors was described by Magalhaes in an application to determine creatinine in urine samples [22]. The approach used two identical creatinine potentiometric biosensors employing the enzyme creatinine iminohydrolase and four PVC-membrane ion selective electrodes (ISEs) for K^+ , Na^+ , NH_4^+ and Ca^{2+} , intended to compensate any interference or matrix effect. The setup was calibrated employing a multivariate response model based on PLS, and the results validated against the standard Jaffé method.

In our laboratory, the first potentiometric bioelectronic tongue was designed employing an array formed by two urea biosensors based on the enzyme urease, covalently linked to carboxylated PVC, plus ISEs for H^+ , K^+ , Na^+ , NH_4^+ and cationic generic response [4]. Response models were built employing PLS and ANN, with slightly better function of the latter. With this array, it was possible to determine urea in clinical samples without the need to separate endogenous ammonium or the interfering ions sodium and potassium. In the same progress line, the system was enlarged with an additional potentiometric biosensor employing creatinine iminohydrolase to determine urea, creatinine and alkaline interferents [23]. Such a bioelectronic tongue is ready at the moment for its clinical application in studies of renal function or the monitoring of haemodialysis processes.

Next, it must be presented significant examples of the use of voltammetric biosensors. The experience of our research group in the

development of biosensors has allowed the immobilization of the biological component on the surface of electrochemical transducers and also to incorporate it in the bulk of composite materials, showing a large versatility, especially for enzyme-based biosensors.

In a significant contribution from our laboratory, Gutiérrez prepared a voltammetric bioelectronic tongue grouping different glucose biosensors, formed by epoxy-graphite biocomposites with glucose oxidase enzyme and different metal catalysts (Pt, Pd and Au-Pd) to promote a differentiated response [2]. The system was successfully applied to the simultaneous determination of glucose and ascorbic acid in an attempt to determine simultaneously glucose plus its typical interfering species in biological fluids without any separation stage or barrier membrane. In this case, the sensor array used the same enzyme, but the response in the array could be differentiated through the use of the different metal catalysts. The metal catalysts improved the biosensor response by decreasing the oxidation potential for hydrogen peroxide from the enzymatic reaction. This electronic tongue was applied to measure both components in juice samples.

A more classical approach is the use of separate channels for each analyte. For example, there is the work of Moser et al. who describes a flow system with an enzymatic array to determine glutamine, glucose, lactate, and glutamate in food products [24]. However this case does not get any advantage from the cross response to counterbalance interfering or matrix effects. In the collaborative work of Tønning [25], a screen printed enzyme array employing tyrosinase, horseradish peroxidase, acetyl cholinesterase and butyryl cholinesterase was applied in environmental analysis, to identify by PCA different water qualities, like untreated, alarm, alert and normal.

A number of works in the literature involve the determination of phenols, and also polyphenols. In an interesting case study [26], Solná and Skládal attempted the determination of various phenolic compounds combining the responses of a multiple screen-printed sensor array, which employed amperometric biosensors made using laccase, tyrosinase and peroxidase enzymes. The amperometric flow-injection determination of phenolic compounds at discrete applied potentials and using the multichannel biosensor was demonstrated at the nM level, although the multivariate approach with combination of the multiple responses was not fully developed. A contribution more in consonance

with the B approach was the determination of mixtures of phenolic compounds (phenol, catechol and 3- cresol) from the overlapping voltammogram obtained with the tyrosinase enzyme biosensor [3]. ANN was the tool used to resolve the mixture at the mM level. Torrecilla and co-workers derived a similar work afterwards to determine phenolic content in olive oil mill wastewater, in which the main difference was the use of the laccase enzyme for building the biosensor [27].

In an interesting approach, a BioET was devised employing three enzymatic biosensors based on tyrosinase and three different phthalocyanines as electron mediators, to evaluate the changes that occur during the aging of beers. For this purpose, alcoholic and non alcoholic beers, were analyzed using cyclic voltammetry, showing significant changes during the aging process. The features extracted from the cyclic voltammograms were used to perform Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) in order to discriminate between their types and age [28].

In a further attempt to use the phenol-responsive enzymes, the co-immobilization of laccase and tyrosinase on the same electrode using glutaraldehyde as a crosslinker and Nafion- ion exchanger as a protective additive provided a biosensor with an integrated response of the two enzymes to five individual polyphenols. With this system, a good correlation between the estimated total phenol index in beer samples and that obtained using the Folin-Ciocalteu reagent was obtained [29].

In our laboratory, the last approach consisted of developing a voltammetric BioET formed by an array of epoxy-graphite biosensors, bulk-modified with different redox enzymes, laccase and tyrosinase, aimed to the simultaneous determination of different polyphenols in the wine matrix. Departure information was the set of voltammograms generated with the biosensor array and Artificial Neural Networks (ANNs) were used for the extraction and quantification of phenolic compounds. The system was applied to resolve ternary mixtures of catechol, caffeic acid and catechin [30].

The proposal of BioETs has also been extended to the use of inhibition-based biosensors, such as those employed for the determination of pesticides; the principle here, is to use enzymes from different biological origins, that may show different degrees of inhibition to a set of substances. A first biosensor array was constructed, in a

collaborative work taking part our laboratory, employing acetylcholinesterase enzymes from different origins in order to obtain a differentiated inhibitory effect from different pesticides [31]. The array employed the wild enzyme type from electric eel (EE) and two different genetically modified enzymes (B1 and B394). In this way, an inhibitory BioET was constructed to resolve dichlorvos and carbofuran mixtures at the nM level. The signal employed was the % of signal inhibited for the different assays, done with equivalent-response, disposable, screen-printed amperometric biosensors, given that the pesticide inhibition is irreversible. Recently, a similar strategy was attempted but employing an automated FIA system, in which dichlorvos and methylparaoxon pesticide mixtures were resolved [32]. These initial results have been extended to new pesticides, in this case the ternary mixture of chlorpyrifos oxon, chlorfenvinfos and azinphos methyl-oxon [33].

Similarly to the acetylcholinesterase inhibition, the system of sarcosine oxidase has been explored for the determination of carboxylic acids [34]. This enzyme was inhibited by citric, malic, succinic, acetic and formic acids while tartaric and lactic acid did not provide any inhibition responses. This response features may be used as a starting point to monitor mixtures of carboxylic acids, for example in wine production.

A fascinating idea is the extension of the electronic tongue principles to multiplex detection with arrays of biosensors based on antibodies or DNA. An especially interesting feature of this is that the data treatment option can be used to solve a defective specificity [35]. In any case, it is evident that there is a trend in the food and clinical diagnostic systems to adopt multichannel (multiplex) biosensing, that takes its inspiration from DNA or peptide chips, and now utilise electrochemical transduction for the sake of simplicity and portability [36]. From the works in the literature there is the interesting work of Wilson and Nie, who described a multiplexed electrochemical immunosensor for the determination of seven tumour markers using a single chip [37]. The simultaneous determination of carcinoembryonic antigen, α -fetoprotein, β -human chorionic gonadotropin, ferritin, or carbohydrate antigens 15-3, 125 or 19-9 was demonstrated at the ng/mL level, although the system did not employ any cross-response feature. A more advanced system [38], able to correct the cross-talk between assays employing the cross reactivity profiles of eight antibodies for androgenic anabolic steroids permitted the determination of four of these (androstandione,

methylboldenone, progesterone and stanozolol) in the 0.1 to 300 nM range, in this case following a rather standard ELISA.

Analogously, it is motivating the prompting of the design of an electrochemically addressed system for multiplex sensing of DNA sequences [39], a tool that may be the basis for systems aimed at clinical diagnostics or food control [40]. In our laboratory, an innovative proposal to detect two model genes from a single electrode has been already described [41]. The multiplex biosensing concept was implemented with an impedimetric genosensor scheme in which two DNA probes were immobilized on an electrode surface and hybridization experiments were monitored by following its electrochemical impedance spectrum. The direct observation did not allow discernment if the first, second or both DNA targets were present in the sample. The application of the electronic tongue concept, where an ANN response model was built from the complex impedance spectra, permitted the correct identification of all the training cases in a leave-one-out scheme. Equivalent schemes may be also developed for protein detection by electrically addressed preparation of aptamer-modified arrayed electrodes [42], demonstrated in the model work with the detection of human IgE.

From the scope of the use of biosensor arrays, as shown in the literature, this chapter will now present two basic applications, one employing potentiometric sensors and the other voltammetric sensors. The first example will describe the construction of urea and creatinine biosensors based on urease and creatinine deiminase, respectively, covalently immobilized onto ammonium selective electrodes, and its coupling in an array with ISEs sensitive to ammonium, potassium and sodium. Generic sensors to alkaline ions will be also included. The further response model based on artificial neural networks will attempt the simultaneous determination of urea, creatinine, ammonium, potassium and sodium in clinical samples.

In the second example, a BioET formed by an array of four voltammetric enzyme-modified biosensors based on bulk-modified biocomposites will be described for the analysis of polyphenols found in beers. One blank electrode, a laccase biosensor, a tyrosinase biosensor and one electrode bulk-modified with copper nanoparticles form the array; these modifiers are selected in order to incorporate differentiated or catalytic response towards different polyphenols present in beers and

aimed to their simultaneous resolution. In this case, the highly complex electrochemical responses obtained from the set of biosensors will be preprocessed employing PCA in order to extract the significant information and compress the departure information. Then, the obtained coefficients will feed an ANN model specially trained to predict major polyphenols found in beer; specifically, ferulic, gallic and sinapic acids.

1.2 BIOELECTRONIC TONGUE EMPLOYING POTENTIOMETRIC SENSORS

In order to evaluate renal function, most clinical tests are related to the measurement of the kidneys' ability to eliminate waste products from the body, usually nitrogen compounds such as urea and creatinine. On the one hand, urea is the main waste product of protein metabolism. In fact, about 95% of protein nitrogen is degraded to urea nitrogen. Urea is formed in the liver and is transported by the blood to the kidneys, where it is filtered and eliminated in the urine. On the other hand, creatinine is a marker of an amino acid pool and its generation reflects slow turnover of muscle protein. It is also transported by the blood to the kidneys. Therefore it is quite common to determine these two species in blood [43]. However, renal dysfunction will occur before a significant increase of these two compounds in blood is detected. So, for early detection of the different nephropathies, it is also necessary to calculate the mass of these two substances eliminated in the urine [44]. Nevertheless, there are few biparametric systems reported that allow the simultaneous determination of urea and creatinine. All of them, based on enzymatic methods [45-49], need to remove the interferences before the analysis.

In this study, we present a BioElectronic Tongue to analyze urine samples based on potentiometric enzymatic electrodes for urea and creatinine determination plus selective and generic ISEs for ammonium, potassium and sodium [50]. These biosensors and ISEs, based on PVC membranes, provide a cross-response towards the different species considered. The chemometric tool used to build the multivariate response model is an ANN, which is fed with the measurements from the biosensor array. The major advantage of this system is not only to determine simultaneously urea and creatinine without any pretreatment strategy, but also to quantify their most severe interferences: ammonium, potassium and sodium ions, these being of biological interest as well.

1.2.1 Fabrication of the potentiometric biosensors

The sensors used were all-solid-state ion selective electrodes (ISEs) with a solid contact made of a conductive composite. This graphite-epoxy resin composite was formed by 50% (by weight) graphite powder (50 µm particle size, BDH Laboratory Supplies, UK), 36% resin Araldite M and 14% hardener Araldite HR (Uneco, Spain). Then these components were thoroughly homogenized by mechanical stirring for 30 min and introduced into the cavity formed between the PVC tube and the copper disk. Finally, the composite was hardened at 40 °C for 24 h. The steps of the construction are depicted in the Figure 6.3.

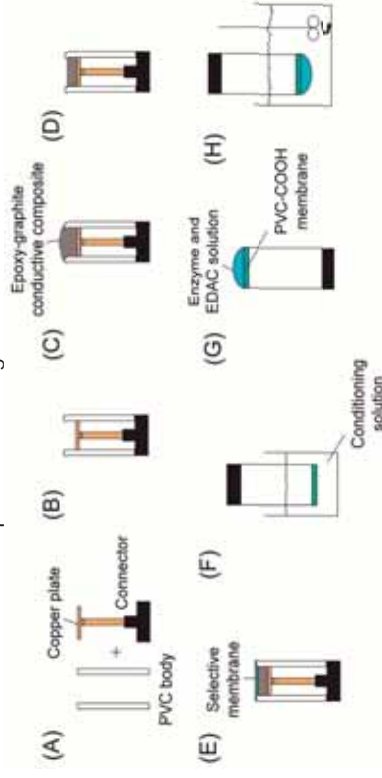


Figure 6.3. Construction of the ISEs and biosensors employed.

The PVC and PVC-COOH membranes were formed by solvent casting the sensor cocktail onto the solid contact. The formulation of the different membranes used is outlined in Table 6.1.

The biosensor array used consisted of 12 electrodes altogether: two urea biosensors, two creatinine biosensors, two ISEs for ammonium, two for potassium, two for sodium and two electrodes of generic response towards alkaline ions (Generic 1 and 2). For the fabrication of urea and creatinine biosensors, the enzymes urease (EC 3.5.1.5, 82.8 U/mg solid, Sigma) and creatinine deiminase (EC 3.5.4.21, 12.8 U/mg solid, Sigma) were used, respectively. Urease catalyses the hydrolysis of urea into ionic products: $\text{Urea} + 2\text{H}_2\text{O} + \text{H}^+ \rightarrow 2\text{NH}_4^+ + \text{HCO}_3^-$; and the enzyme creatinine deiminase also breaks creatinine: $\text{Creatinine} + \text{H}_2\text{O} + \text{H}^+ \rightarrow 1\text{-methylhydantoin} + \text{NH}_4^+$. The concentration of ammonium ions

(NH_4^+) generated as products in both bioreactions, is proportional to the concentration of the metabolite. Therefore, in this study an ammonium transducer was used as the base element for these biosensors (Figure 6.3 G and H).

Table 6. 1. Formulation of the different ion-selective membranes employed.

| Electrode | PVC | PVC-COOH | Plasticizer | Ionophore | Additive |
|------------|-------|----------|-------------|-------------------------|------------|
| Ammonium 1 | 33.0% | - | 66.0% BPA | 1.0% Nonactine | - |
| Ammonium 2 | - | 33.0% | 66.0% BPA | 1.0% Nonactine | - |
| Potassium | 31.7% | - | 66.7% DOS | 1.0% Valinomycin | 0.6%KpCIPB |
| Sodium | 21.8% | - | 70.0% NPOE | 6.0%bis(12-crown-4) | 2.2%KpCIPB |
| Generic 1 | 29.0% | - | 67.0% DOS | 4.0% Dibenzo-18-crown-6 | - |
| Generic 2 | 27.0% | - | 70.0% DBS | 3.0% Lasalocide | - |

BPA: bis(1-butylpentyl) adipate; DOS: dioctylsebacate; NPOE: o-nitrophenyloctylether; DBS: dibutyl sebacate; KpCIPB: potassium tetrakis(4-chlorophenyl) borate

The covalent immobilization of the enzymes into the ammonium PVC-COOH membrane (Ammonium 2 formulation) was performed by the carbodiimide reaction, using the reagent N-(3-dimethylaminopropyl)-N'-ethylcarbodiimide hydrochloride (EDAC, Sigma). Specifically, an aqueous solution of enzyme (700 U/ml) and EDAC (6 mg/ml) was prepared. Then, 60 µl were deposited onto the PVC-COOH membrane of an ammonium ISE, and left in a refrigerator for 12 hours. This enables the activation of the carboxylic groups of the membrane and the subsequent formation of amide bonds with the free amino groups of the enzyme in just one step. Finally, the biosensor was washed in 0.05 M TRIS buffer (pH 7.5) under stirring for at least two hours in order to remove the enzyme that may have been adsorbed. When not in use, the biosensors were stored in the refrigerator immersed in the same TRIS buffer to maintain the activity of the enzyme [51].

1.2.2 Potentiometric characterization of the biosensors

The characterization of a sensor or biosensor consists of assessing its response towards the analyte which it is selective to. In practice, the response characteristics such as sensitivity, detection limit (DL) and linear range were studied through its calibration plot. Although all sensors in the array (ammonium, potassium, sodium and those of

generic response) were calibrated to check they presented the expected behaviour, we will centre in the response for urea and creatinine.

The calibration curves were obtained by means of the method of analyte addition: the variation of potential originated by the addition of accumulated microvolumes of stock solutions in 50 mM TRIS-HCl buffer pH 7.5 was determined. TRIS buffer was chosen because it has a lower interference effect than other buffer solutions that contain alkaline ions. These interfering ions would increase the DL considerably, thus their concentration has to be controlled. In addition, pH 7.5 has been considered the most appropriate one, both for the catalytic activity of the enzymes and for the detection of ammonium ions. For each metabolite (urea and creatinine), two replicated biosensors (devices A and B) built following the same methodology and employing the same membrane composition were used. The obtained calibration parameters for the constructed biosensors are shown in Table 6.2.

Table 6.2. Formulation of the different ion-selective membranes employed.

| Sensor | Device | Sensitivity (s) (mV/dec.) | DL (s) (M) | R ² (n=28) |
|------------|--------|---------------------------|---|-----------------------|
| Urea | A | 29.8 (0.9) | $1.2 \cdot 10^{-4}$ ($5.3 \cdot 10^{-5}$) | 0.992 |
| | B | 32.2 (1.2) | $1.0 \cdot 10^{-4}$ ($3.9 \cdot 10^{-5}$) | 0.998 |
| Creatinine | A | 53.5 (2.0) | $1.4 \cdot 10^{-4}$ ($1.8 \cdot 10^{-5}$) | 0.999 |
| | B | 54.4 (1.8) | $8.5 \cdot 10^{-5}$ ($1.3 \cdot 10^{-5}$) | 0.999 |

As can be observed, while urea biosensors show a sensitivity around 30 mV/decade due to the enzyme kinetics, creatinine biosensors show a slope close to the Nernstian behaviour, about 50 mV/decade. On the other hand, the DL is quite similar for both analytes, around $1.0 \cdot 10^{-4}$ M. The linear range is around two decades of concentration, up to $1.0 \cdot 10^{-2}$ M, for the two species which is appropriate for the proposed application.

Frequent calibrations of the prepared biosensors in the TRIS buffer allowed us to estimate their lifetime. Figure 6.4 shows the variation of the sensitivity and DL of the urea biosensors during 20 days of study. As can be seen, in terms of slope and LD the loss of response was practically zero. In the case of the creatinine biosensors, the variation of these parameters is shown in Figure 6.5. The variation observed during the 27 days of study was less than 15%. Therefore, the stability of the biosensor response is appropriate to be included in the array.

Finally, in the experiments to check the response time, the method of analyte addition explained above was used and the response potentials

after each addition were recorded every second. Results are shown in Figure 6.6. Arrows in the graph indicate the exact moment in which the addition was done. As can be seen, the action of switching off the magnetic stirrer one minute after the addition allows for a great increase in the response. In any case, 95% of final response is in the range 1-3 min.

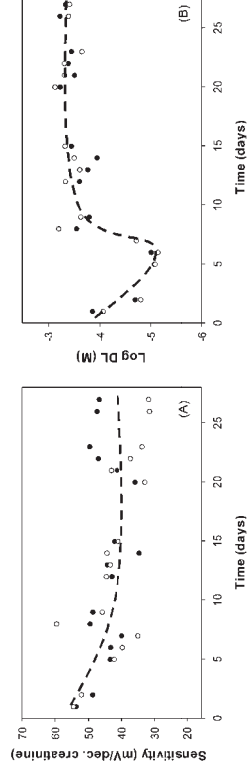


Figure 6.4. Variation in time of (A) the sensitivity and (B) the DL for the replicated urea biosensors in TRIS buffer.

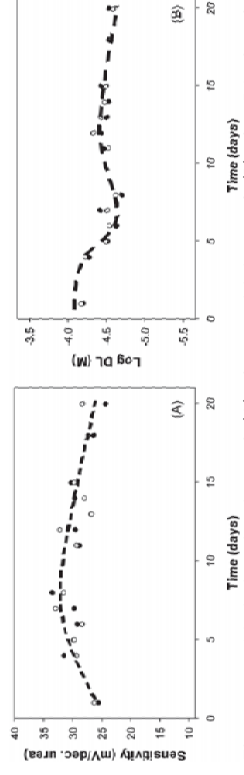


Figure 6.5. Variation in time of (A) the sensitivity and (B) the DL for the replicated creatinine biosensors in TRIS buffer.

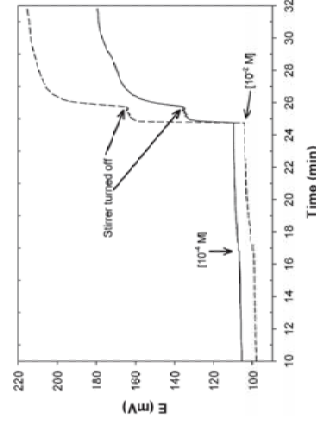


Figure 6.6. Response time for the constructed urea and creatinine biosensors.

1.2.3 Response modelling of the BioElectronic Tongue

The training of the BioElectronic Tongue was performed based on the voltage readings of the biosensor array corresponding to a set of samples defined from a statistical experimental design. The composition of the 27 solutions used for the calibration of the array was designed according to a fractional factorial design with three levels and five factors ($3^{5 \cdot 2}$) [52]. The external validation set was formed by 13 additional synthetic solutions (test-set) with their concentrations values randomly generated. The concentration range of the five analytes varies over two orders of magnitude: urea and creatinine, from 0.1 to 10 mM, ammonium, from 0.005 to 0.5 mM, and potassium and sodium, from 0.05 to 5 mM, as defined by possible physiological levels, and which supposes an extensive training space. The BioET was also applied to real urine samples collected from 3 volunteers and the results compared with reference determination procedures. To obtain a greater variability of the concentrations, 8 new spiked urine samples were prepared by adding different quantities of the five analytes.

On the other hand, for the construction of the ANN model, we initially fixed the following parameters: the number of input neurons, which was 12 (one for each sensor from the array); the number of output neurons, 5 (one for each of the modelled analytes); the transfer function for the output layer, a linear one (purelin) and a single hidden layer between the input and output ones. These selections were based on previous experience with Electronic Tongues using potentiometric sensors. The learning algorithm used was Bayesian regularization and employed as internal parameters a learning rate of 0.1 and a momentum of 0.4, selected from preliminary tests.

Therefore, we studied the 20 possible configurations that result from combining the transfer functions tansig and logsig, and a number of neurons in the hidden layer between three and twelve. Firstly, a representation of the Root Mean Squared Error (RMSE), for each of the species and the total value) versus the number of neurons in the hidden layer was performed in order to choose the best transfer function. These RMSE were obtained from the mean of three replicates of the same training configuration. As can be observed in Figure 6.7, although there is not a big difference between the RMSE of the two functions, generally smaller errors are obtained when using the tansig function.

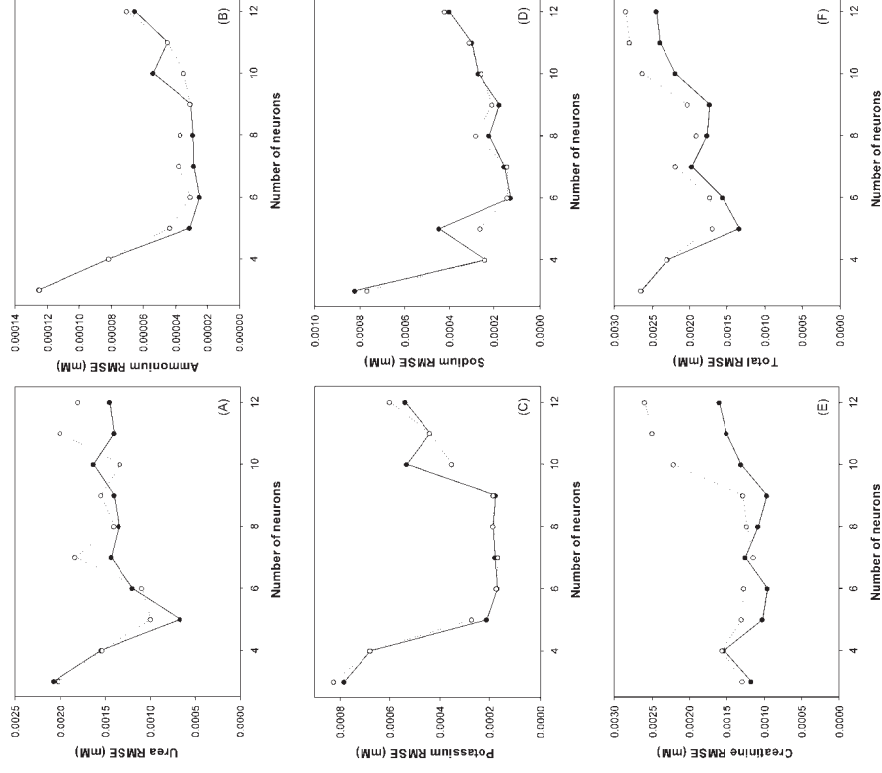


Figure 6.7. Plot of RMSE versus the number of neurons for each species: (A) urea, (B) ammonium, (C) potassium, (D) sodium, (E) creatinine and (F) total.

After selecting the transfer function, it was needed to define the number of neurons in the hidden layer. Therefore, the slope, intercept and correlation coefficient obtained from the comparison graphs between the found values and those expected for the external validation set were represented vs. the number of neurons employed in the hidden layer. These values were also obtained from the mean of the three

replicates. Thus, the choice of the best configuration was done considering closeness to the ideal values of 1.0 for slope, 0.0 for intercept and 1.0 for the correlation coefficient. As can be observed in Figure 6.8, this was achieved when 6 neurons were employed in the hidden layer.

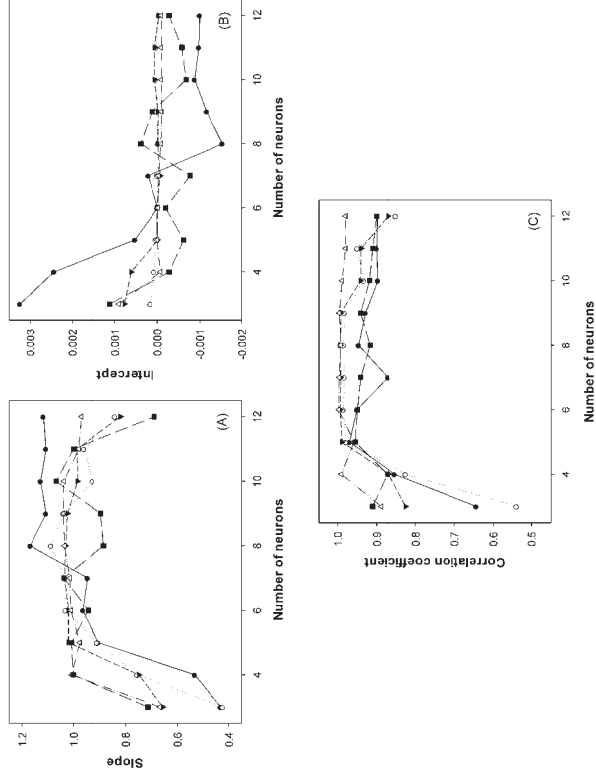


Figure 6.8. Plot of (A) slope, (B) intercept and (C) correlation coefficients versus number of neurons in the hidden layer for the external test subset.

1.2.4 Application

Performance of the final optimized model, for the external test set and for the real samples, was evaluated by comparison of found vs. expected concentrations for urea, ammonium, potassium, sodium and creatinine. A good trend was obtained for the individual prediction of species in each of the mixtures. As commented, from the comparison graphs a linear least-squares regression was fitted. In this manner, Table 6.3 presents the results of the regressions. As can be seen, the uncertainty intervals (calculated at the 95% confidence level) included the ideal comparison values of slope of 1.0 as well as 0.0 intercept in all cases.

The analysis of the three urine samples plus the eight spiked samples was performed by simple direct measurement with the biosensor array, followed by interpolation using the ANN model. Obtained results were compared with those obtained by reference methods. As before, Table 6.3 includes the comparative regression data for the five considered species. In all cases, the uncertainty intervals included the ideal slope of 1.0 as well as 0.0 intercept. Dispersion between real and spiked samples was higher than that obtained with the test set; nevertheless, the correlation coefficients were highly significant in all the cases.

Table 6.3. Regression data for the comparison of results provided by the proposed BioET for the five considered species, considering the external test-set and the real samples (intervals calculated at the 95% confidence level).

| Analyte | Testing subset | | | | Real and spiked samples | | | |
|------------|----------------|-----------------|-----------------------------|-----------------------------|-------------------------|-----------------|-----------------------------|-----------------------------|
| | Correlation | Slope | Intercept (μM) | Intercept (μM) | Correlation | Slope | Intercept (μM) | Intercept (μM) |
| Urea | 0.992 | 1.01 \pm 0.08 | -0.20 \pm 0.51 | -0.40 \pm 1.17 | 0.967 | 0.95 \pm 0.19 | -0.40 \pm 1.17 | -0.17 \pm 1.19 |
| Creatinine | 0.991 | 1.06 \pm 0.09 | -0.35 \pm 0.45 | 0.940 | 0.940 | 1.05 \pm 0.29 | 0.01 \pm 0.06 | 0.01 \pm 0.06 |
| Ammonium | 0.995 | 1.02 \pm 0.07 | 0.00 \pm 0.02 | 0.970 | 0.98 \pm 0.18 | 1.04 \pm 0.07 | 0.02 \pm 0.14 | 0.02 \pm 0.14 |
| Potassium | 0.996 | 1.02 \pm 0.06 | 0.02 \pm 0.18 | 0.995 | 1.04 \pm 0.07 | 0.99 \pm 0.10 | 0.04 \pm 0.27 | 0.04 \pm 0.27 |
| Sodium | 0.998 | 1.04 \pm 0.04 | -0.10 \pm 0.12 | 0.991 | 0.99 \pm 0.10 | 0.99 \pm 0.10 | 0.04 \pm 0.27 | 0.04 \pm 0.27 |

1.3 BIOELECTRONIC TONGUE EMPLOYING VOLTAMMETRIC SENSORS

A known problem when voltammetric sensors are used is the large dimensionality of the generated data that hinders their treatment. That is, when a complete voltammogram is recorded for each sensor from the array. This is perhaps the main reason why this approach has been rarely used in the literature; especially if ANNs are to be used, in which the departure information needs to be preprocessed.

A direct solution is the use of multiway processing methods (samples \times sensors \times polarization potential) like nPLS, but the intricacy of the theoretical background and conceptual depth of the technique is also critical [53]. An alternative when dealing with a set of voltammograms is to employ a preprocessing stage for data reduction. The main objective of this step is to reduce the complexity of the input signal preserving the relevant information and making it compatible with ANN, which in addition allows to gain advantages in training time, to avoid redundancy

in input data and to obtain a model with better generalization ability [54]. This compression stage may be achieved through the use of methods such as feature extraction [30], Principal Component Analysis (PCA) [55], “kernels” [56], Discrete Wavelet Transform (DWT) [57] or even Fast Fourier Transform (FFT) [58].

Beer contains a wide variety of polyphenols and classes. However, content values for each of the single polyphenols are rather low. Nevertheless polyphenols play an important role in beer flavour (bitterness, astringency, and harshness), colour and beer stability. Therefore, its control is of clear interest. In this sense, our BioET was developed for this purpose, thinking in the next step which will be the resolution of the same problem in real samples.

Total polyphenol content in beer, as measured by the Folin-method, varies between 120 and 520 mg/L, depending on the beer type. However, when comparing the total polyphenol value measured by Folin (280 mg/L for regular beer) with the total polyphenol calculated from the sum of the individual polyphenols measured by chromatographic methods (40.4 mg/L for regular beer), the latter appears to be inferior. This could be partially explained by the lack of data on beer polyphenols in the literature, and thus in the food composition table. Also by the difficulty to measure certain polyphenols by chromatographic methods and the reactivity of the Folin assay with non-polyphenol reducing compounds, prone to its overestimation. Being ferulic (8.4 mg/mL), gallic (5.7 mg/mL), vanillic (1.5 mg/mL), p-coumaric acid (1.4 mg/mL) and sinapic acids (0.3 mg/mL) the most abundant, as measured after hydrolysis in regular beer. To sum up, in this case, a BioET formed by an array of voltammetric enzyme-modified biosensors based on bulk-modified biocomposites for the analysis of polyphenols found in beers could be accomplished.

1.3.1 Fabrication of the amperometric biosensors

An array of 4 different graphite-epoxy voltammetric sensors and biosensors were prepared following conventional methodology previously established by our research Group [59]. Electrode fabrication begins with the soldering of a copper disk to an electric connector. Then, the connector is introduced to a PVC tube, where composite paste will be deposited. These fabrication steps are illustrated in Figure 6.9.

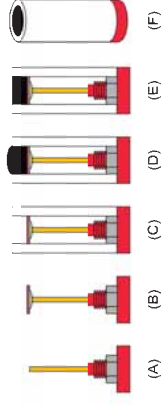


Figure 6.9. Graphite-epoxy composite electrodes construction scheme. (A) Connector. (B) Copper disk is soldered to the connector and (C) assembled into the PVC tube. (D) Preparation of the graphite-epoxy mixture, incorporation of the modifiers and deposition into the cavity. (E and F) Final aspect of the biocomposite electrode after hardening and polishing.

Afterwards, the composite paste is prepared. For this, resin EpoTek H77 (Epoxy Technology, Billerica, MA, USA) and its corresponding hardener compound were mixed in the ratio 20:3 (w/w). Then, each electrode was prepared by adding a 15% (w/w) of graphite (50 μm BDH Laboratory Supplies) and a 2% (w/w) of either the enzyme (tyrosinase or laccase) or the modifier (copper nanoparticles) into the epoxy resin before hardening – one component per electrode, and the fourth electrode without any modifier (Table 6.4) [30]. Finally, the biocomposite material was manually homogenised for 60 min, and after that it was allowed to harden for 7 days at 40°C. Electrode surface was then polished with different sandpapers of decreasing grain size, with a final electrode area of 28 mm².

Table 6.4. Formulation of the biocomposites used for preparing the BioElectronic Tongue.

| Sensor | Epotek H77 | Graphite | Modifier |
|--------|------------|----------|---|
| GEC | 85% | 15% | - |
| Tyr | 83% | 15% | 2% Tyrosinase from mushroom (EC 1.14.18.1, 4276 U/mg) |
| Lac | 83% | 15% | 2% Laccase from <i>Trametes versicolor</i> (EC 1.10.3.2, 21 U/mg) |
| Cu | 83% | 15% | 2% Copper nanoparticles (50 nm) |

^aPreviously mixing the resin (part A) with its corresponding hardener (part B) in a ratio 20:3 (w/w)

A crucial step in biosensors development is the selection of the biomolecule, which will act as the recognition element, and its type of interaction with the target molecule; i.e. biological stability, reaction type and transduction process, generation of possible interfering by-products,

etc. Also, similar considerations need to be taken into account when selecting the electrochemical modifiers for the modified sensors.

In our case, tyrosinase and laccase were chosen as the enzymes employed to detect the phenolic compounds. These enzymes belong to the class of copper containing oxidases, which catalyze the reduction of molecular oxygen by different electron donors. In those reactions the oxygen is reduced directly to water without the intermediate formation of hydrogen peroxide. To be precise, tyrosinase catalyzes the hydroxylation of monophenols to catechols, which in turn are further oxidized to o-quinones, both using molecular oxygen; then o-quinones produced in the enzymatic reaction are electrochemically reduced to o-diphenols, resulting in the recycling of the diphenols at the applied negative potential. Laccase catalyzes the oxidation of phenols giving phenoxyl radicals species which are converted to quinones in the second stage of the oxidation; then as in tyrosinase, these quinones can be electrochemically reduced again to phenols. Finally, copper-nanoparticles were chosen given that both tyrosinase and laccase are copper-containing enzymes. Then it was thought that some catalytic effect could be derived; a fact finally observed in the sensor response.

1.3.2 Voltammetric characterization of the biosensors

The biosensors response was first checked towards standard solutions of three major polyphenols found in beer (Figure 6.10). The voltammetric responses of the electrodes are illustrated in Figure 6.11. Differentiated signals were observed for the different electrodes, generating very rich data that can be useful as departure point. As a general trend, two processes are observed for all the sensors corresponding to the oxidation of the corresponding phenol to its quinone form, and the reduction of the quinone to the polyphenolic form.

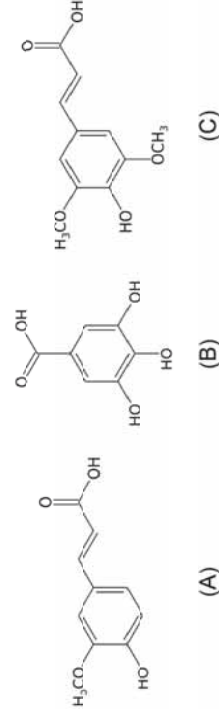


Figure 6.10. Chemical structures of the polyphenols under study: (A) ferulic acid, (B) gallic acid and (C) sinapic acid.

Moreover, some peaks that could be attributed to the oxidation of the $-OCH_3$ group are observed. Thus, as expected two peaks are obtained for sinapic acid, one for ferulic acid and no peaks are observed in the case of gallic acid which does not have any methoxy group. Chemical structures of these compounds can be seen in Figure 6.10.

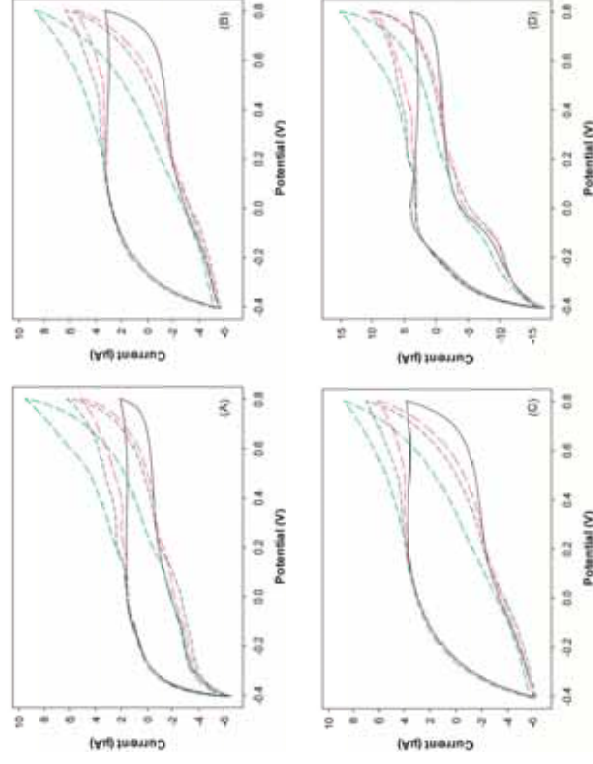


Figure 6.11. Example of the voltammograms obtained for the same standard solutions of each of the three phenols and buffer solution with the four sensors used. (—) Buffer solution, (---) 172 μ M Ferulic acid solution, (- - -) 181 μ M Gallic acid solution and (- · - ·) 171 μ M Sinapic acid solution. Sensors: (A) Graphite-epoxy sensor, (B) Tyrosinase biosensor, (C) Laccase biosensor and (D) Copper nanoparticle modified sensor.

The next step, once it was confirmed that different behaviour was obtained from the BioET array towards the different compounds, was to evaluate more deeply electrodes response for each polyphenol. Thus, standard solutions of increasing concentration of the three polyphenols were measured (Figure 6.12).

As expected, the intensity of the peaks associated with phenols increased with the concentration of the phenolic compounds. In order to

characterize electrode sensitivity for each compound, the current of the peaks associated to phenols was plotted against the concentration of each species. Although maximum response differs slightly for each compound, the same potential for each sensor was chosen when building the calibration plot like would be done in a conventional amperometry (intensity measured at a single potential). Table 6.5 summarizes the calibration curves parameters of the different biosensors.

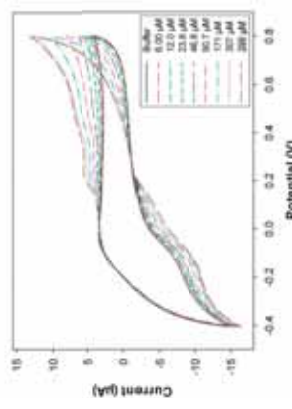


Figure 6.12. Voltammetric response of "Cu"-modified sensor towards standard solutions of increasing sinapic acid concentration.

Table 6.5. Summary of the responses of the BioET array towards the studied polyphenols.

| Sensor | Ferulic acid | | Gallic acid | | Sinapic acid | |
|--------------------|--------------|-------|-------------|-------|--------------|-------|
| | an | cat | an | cat | an | cat |
| GEC | | | | | | |
| E (V) | 0.28 | 0.08 | 0.28 | 0.08 | 0.28 | 0.08 |
| Sensitivity (µA/M) | 1756 | -2104 | 4918 | - | 2428 | -2108 |
| Tyr | | | | | | |
| E (V) | 0.40 | -0.01 | 0.40 | -0.01 | 0.40 | -0.01 |
| Sensitivity (µA/M) | 1276 | 1227 | 4633 | - | 1780 | -954 |
| Lac | | | | | | |
| E (V) | 0.40 | -0.01 | 0.40 | -0.01 | 0.40 | -0.01 |
| Sensitivity (µA/M) | 1114 | 1104 | 3630 | - | 1692 | -790 |
| Cu | | | | | | |
| E (V) | 0.31 | 0.08 | 0.31 | 0.08 | 0.31 | 0.08 |
| Sensitivity (µA/M) | 3381 | -3891 | 8555 | - | 5228 | -4234 |

The different relative responses of each phenolic compound arise from the different enzyme affinities towards the compound or in the electrochemical activity in the case of the copper modified composite. This situation, with marked mix-response for the different involved compounds and with differentiated response from the different biosensors used, is highly desirable for studies with BioET systems.

1.3.3 Response modelling of the BioElectronic Tongue

As demonstrated in the above section, the array shows complex voltammetric responses that contain global information about the compounds analyzed. Despite differentiated response in terms of specificity that can be found in conventional amperometry (intensity measured at a single potential), huge improvements of the results are obtained when using the whole curve instead of particular peaks.

With the aim of evaluating the capability of the BioET array to discriminate mixtures of the studied polyphenols, a total of 42 samples were manually prepared with a concentration range for the three species from 0 to 200 µM for each phenol. The set of samples was divided into two subsets: a training subset with 27 samples (65%), which were used to establish the response model; and a testing subset with 15 samples (35%), which were used to evaluate the models predictive response. Training samples were distributed in a cubic design [30], while testing ones were distributed randomly along the experimental domain, with the only precaution that samples must be inside the experimental design domain.

As discussed, when dealing with voltammetric sensor arrays a preprocessing step is required in order to reduce the high dimensionality of the recorded signals (samples x sensors x potentials). In this case, this preprocessing of the data (MATLAB) was done by the use of PCA. In this manner, analysis of principal components was performed on each sensor, and the obtained scores were used as the input coefficients in the neural network model.

With only the four first PCs for each sensor, the explained variance accumulated was ca. 99.9%; thus meaning that nearly all the variance contained in the original information is preserved. This allowed a compression of the original data up to 98.5% without any loss of significant information.

The first step in building the ANN model is selecting the topology of the network. This requires a significant effort of optimization due to the difficulties in predicting the optimum ANN configuration in advance. A trial-error process is needed, where several parameters (training algorithms, number of hidden layers, transfer functions, etc.) are fine-tuned in order to find the best configuration to optimize the performance of the neural network model. A simplified scheme of the procedure followed for the measurements and data treatment could be seen in Figure 6.13.

Multiple ANN architectures and topologies were assayed employing Bayesian regularization algorithms. Comparison graphs of predicted vs. expected concentrations for the three determined phenols were built to check the prediction ability of the ANN. After this step the best configuration was chosen taking into account the slope, intercept and correlation coefficient (ideal values equal to 1, 0 and 1, respectively).

Then, after an extensive study varying its configurations, the final ANN architecture model had 16 neurons (corresponding to the first four components of the PCA for each sensor, with explained variance accumulated ca. 99.9 %) in the input layer, 5 neurons and tangsig transfer function in the hidden layer and 3 neurons and purelin transfer function in the output layer (one for each polyphenol).

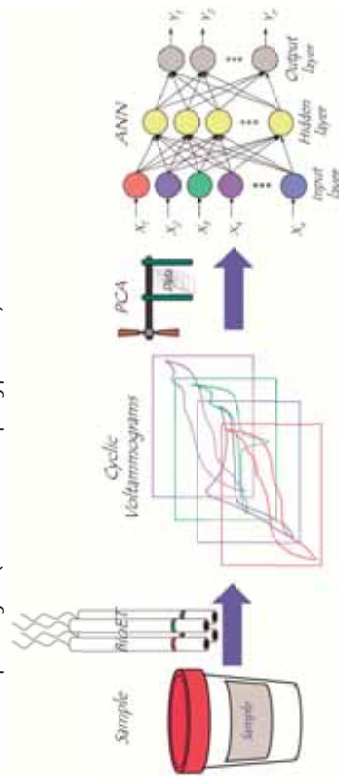


Figure 6.13. Processing scheme of the voltammetric BioET approach. After measuring with the BioET array, signals are preprocessed employing PCA. Then, obtained scores fed an ANN model which carries out the quantification of the analytes. Finally, appropriate weights and biases are applied by the learning algorithm until the targets are reached within the established error.

PCA-ANN model was trained with 65% of the data, using the remaining 35% of the data (testing subset) to characterize the accuracy of the quantification model. In this way, results obtained (calculated concentrations by the model) for the training and testing subsets were plotted against the expected ones and fitted with linear least-squares regression. The comparison slopes and intercepts, as summarized in Table 6.6, are satisfactory, with slopes and intercepts nearly 1.0 and 0.0 respectively for the three studied phenols and the different subsets of samples. The correlation coefficients between obtained and expected concentrations were also satisfactory, very close to 1.

Table 6.6. Fitted regression lines of the comparison between obtained vs. expected results provided by the proposed BioET for the training and testing subsets of samples and the three considered phenols (intervals calculated at the 95% confidence level).

| Phenol | Training subset | | | Testing subset | | |
|--------------|-----------------|-------------------|-----------------------------|----------------|-------------------|-----------------------------|
| | Correlation | Slope | Intercept (μM) | Correlation | Slope | Intercept (μM) |
| Ferulic acid | 0.994 | 0.988 ± 0.044 | 0.11 ± 5.03 | 0.974 | 1.095 ± 0.153 | -9.46 ± 16.84 |
| Gallic acid | 0.998 | 0.990 ± 0.027 | 0.94 ± 3.15 | 0.992 | 0.991 ± 0.077 | -1.11 ± 8.65 |
| Sinapic acid | 0.996 | 0.984 ± 0.035 | 1.68 ± 4.08 | 0.980 | 1.055 ± 0.127 | -6.77 ± 15.56 |

1.4 CONCLUSIONS

There is a clear demand for simpler and more efficient procedures for obtaining (bio)chemical information on the different fields. Among the progress in new analytical technologies, the use of sensors is a clear option for a fast, simple and cheap gathering of information, although sensors alone cannot solve all the existing situations. In the last decade, a new trend in the sensor field has appeared, which is to couple multidimensional sensor information with advanced computer processing strategies; this approach, known as electronic tongue, can take profit from electrochemical sensors and biosensors and derive information of different types, even tailor-size to specific requirements. In this chapter, the potential of BioETs as alternatives to standard analytical methods has been proved, in this case, with two multidetermination examples.

In the first, urea and creatinine biosensors based on covalent immobilization of urease and creatinine deiminase respectively, were prepared from ammonium transducers to be included in a sensor array. Their characterization showed adequate cross-response characteristics to complement ammonium, potassium, sodium and generic ISEs. ANNs were applied as processing tool to obtain a calibration model from the measured potentials. The developed bioelectronic tongue provided an extraordinarily simple procedure, with direct measurement, to quantify concentrations of urea and creatinine in real urine samples without the necessity of eliminating alkaline interferences. Such a system would

facilitate clinical studies of renal function or the monitoring of haemodialysis processes in real time.

In the second example, the quantitative resolution at low concentration level of mixtures of typical polyphenols present in beers was achieved by means of a voltammetric BioET. The approach combined the response from different enzymatic biosensors to complement the analytical departure information, which being more complex in this case, needed a feature extraction process employing a PCA. Resolution of signal overlapping by ANN allowed obtaining a good response model for three important phenols in the beverage field (ferulic, gallic and sinapic acids), in an application comparable to more complex analytical techniques such as HPLC.

Acknowledgments

Financial support for this work was provided by Spanish Ministry of Science and Innovation, MCINN (Madrid) through the project CTO2010-17099 and program ICREA Academia. X. Cetó thanks the support of Dept. d'Innovació, Universitat i Empresa de la Generalitat de Catalunya for the predoctoral grant.

References

1. Vlasov, Y., Legin, A., Rudnitskaya, A., Di Natale, C. and D'Amico, A., (2005). Nonspecific sensor arrays ("electronic tongue") for chemical analysis of liquids: (IUPAC technical report), *Pure Appl. Chem.*, **77**, pp. 1965-1983.
2. Gutiérrez, A., Ibañez, A. B., del Valle, M. and Cespedes, F., (2006). Automated SIA e-tongue employing a voltammetric biosensor array for the simultaneous determination of glucose and ascorbic acid, *Electroanalysis*, **18**, pp. 82-88.
3. Gutiérrez, A., Cespedes, F., Alegret, S. and del Valle, M., (2005). Determination of phenolic compounds by a polyphenol oxidase amperometric biosensor and artificial neural network analysis, *Biosens. Bioelectron.*, **20**, pp. 1668-1673.
4. Gutiérrez, M., Alegret, S. and del Valle, M., (2007). Potentiometric bioelectronic tongue for the analysis of urea and alkaline ions in clinical samples, *Biosens. Bioelectron.*, **22**, pp. 2171-2178.
5. Ciosek, P. and Wroblewski, W., (2007). Sensor arrays for liquid sensing - Electronic tongue systems, *Analyst*, **132**, pp. 963-978.
6. del Valle, M., (2010). Electronic tongues employing electrochemical sensors, *Electroanalysis*, **22**, pp. 1539-1555.

7. Vlasov, Y. G., Legin, A. V. and Rudnitskaya, A. M., (2008). Electronic Tongue: Chemical Sensor Systems for Analysis of Aquatic Media, *Russ. J. Gen. Chem.*, **78**, pp. 2532-2544.
8. Bratov, A., Abramova, N. and Ipatov, A., (2010). Recent trends in potentiometric sensor arrays-A review, *Anal. Chim. Acta*, **678**, pp. 149-159.
9. Winquist, F., (2008). Voltammetric electronic tongues - Basic principles and applications, *Microchim. Acta*, **163**, pp. 3-10.
10. Riul Jr, A., Dantas, C. A. R., Miyazaki, C. M. and Oliveira Jr, O. N., (2010). Recent advances in electronic tongues, *Analyst*, **135**, pp. 2481-2495.
11. del Valle, M., (2008). Electrochemical micro(bio)sensor arrays, *Microchim. Acta*, **163**, pp. 1-2.
12. del Valle, M., (2011). Bioinspired Sensor Systems, *Sensors*, **11**, pp. 10180-10186.
13. Richards, E., Bessant, C. and Saini, S., (2002). Multivariate data analysis in electroanalytical chemistry, *Electroanalysis*, **14**, pp. 1533.
14. Scott, S. M., James, D. and Ali, Z., (2006). Data analysis for electronic nose systems, *Microchim. Acta*, **156**, pp. 183-207.
15. Ciosek, P. and Wroblewski, W., (2006). The analysis of sensor array data with various pattern recognition techniques, *Sens. Actuators B-Chem.*, **114**, pp. 85-93.
16. Despagne, F. and Massart, D. L., (1998). Neural networks in multivariate calibration, *Analyst*, **123**, pp. 157R-178R.
17. Durán, A., Cortina, M., Velasco, L., Rodríguez, J. A., Alegret, S. and del Valle, M., (2006). Virtual instrument for an automated potentiometric e-tongue employing the SIA technique, *Sensors*, **6**, pp. 19-29.
18. del Valle, M., (2007) in *Comprehensive Analytical Chemistry* (Eds. Alegret, S. and Merkoçi, A.), (Elsevier, Amsterdam), pp. 721-753.
19. Gutiérrez, M., Calvo, D. and del Valle, M., (2007) in *Comprehensive Analytical Chemistry* (Eds. Alegret, S. and Merkoçi, A.), (Elsevier, Amsterdam), pp. e311-e330.
20. Toko, K., (2000). Taste sensor, *Sens. Actuators B-Chem.*, **64**, pp. 205-215.
21. Toko, K. (2000) *Blomimetic Sensor Technology*, (Cambridge University Press, Cambridge), pp. 1.
22. Magalhaes, J. M. C. S. and Machado, A. A. S. C., (2002). Array of potentiometric sensors for the analysis of creatinine in urine samples, *Analyst*, **127**, pp. 1069-1075.
23. Gutiérrez, M., Alegret, S. and del Valle, M., (2008). Bioelectronic tongue for the simultaneous determination of urea, creatinine and alkaline ions in clinical samples, *Biosens. Bioelectron.*, **23**, pp. 795-802.
24. Moser, I., Jobst, G. and Urban, G. A., (2002). Biosensor arrays for simultaneous measurement of glucose, lactate, glutamate, and glutamine, *Biosens. Bioelectron.*, **17**, pp. 297-302.

25. Tønning, E., Sapelinikova, S., Christensen, J., Carlsson, C., Winther-Nielsen, M., Dock, E., Solna, R., Skládal, P., Nørgaard, L., Ruzgas, T. and Ernés, J., (2005). Chemometric exploration of an amperometric biosensor array for fast determination of wastewater quality. *Biosens. Bioelectron.*, **21**, pp. 608-617.
26. Solná, R. and Skládal, P., (2005). Amperometric Flow-Injection Determination of Phenolic Compounds Using a Biosensor with Immobilized Laccase, Peroxidase and Tyrosinase. *Electroanalysis*, **17**, pp. 2137-2146.
27. Torrecilla, J. S., Mena, M. L., Yáñez-Sedeño, P. and García, J., (2007). Quantification of phenolic compounds in olive oil mill wastewater by artificial neural network/laccase biosensor. *J. Agric. Food Chem.*, **55**, pp. 7418-7426.
28. Ghasemi-Varmakhasti, M., Rodríguez-Méndez, M. L., Mohtasebi, S. S., Apetrei, C., Lozano, J., Ahmadi, H., Razavi, S. H. and Antonio de Saja, J., (2012). Monitoring the aging of beers using a bioelectronic tongue. *Food Control*, **25**, pp. 216-224.
29. Elkautit, M., Naranjo-Rodríguez, I., Tamsamani, K. R., Domínguez de la Vega, M. and de Cisneros, J. L. H. H., (2008). Dual Laccase-Tyrosinase Based Sonogel-Carbon Biosensor for Monitoring Polyphenols in Beers. *J. Agric. Food Chem.*, **55**, pp. 8011-8018.
30. Cetó, X., Céspedes, F., Pividori, M. I., Gutiérrez, J. M. and del Valle, M., (2012). Resolution of phenolic antioxidant mixtures employing a voltammetric bio-electronic tongue. *Analyst*, **137**, pp. 349-356.
31. Cortina, M., Del Valle, M. and Marty, J. L., (2008). Electronic tongue using an enzyme inhibition biosensor array for the resolution of pesticide mixtures. *Electroanalysis*, **20**, pp. 54-60.
32. Valdés-Ramírez, G., Gutiérrez, M., del Valle, M., Ramírez-Silva, M. T., Fournier, D. and Marty, J. L., (2009). Automated resolution of dichlorvos and methylparaoxon pesticide mixtures employing a Flow Injection system with an inhibition electronic tongue. *Biosens. Bioelectron.*, **24**, pp. 1103-1108.
33. Alonso, G. A., Domínguez, R. B., Marty, J. L. and Muñoz, R., (2011). An approach to an inhibition electronic tongue to detect on-line organophosphorus insecticides using a computer controlled multi-commuted flow system. *Sensors*, **11**, pp. 3791-3802.
34. Zeravik, J., Lacina, K., Jilek, M., Vlcek, J. and Skládal, P., (2010). Biosensor for determination of carboxylic acids in wines based on the inhibition of sarcosine oxidase. *Microchim. Acta*, **170**, pp. 251-256.
35. Spinks, C. A., (2000). Broad-specificity immunoassay of low molecular weight food contaminants: new paths to Utopia. *Trends in Food Science & Technology*, **11**, pp. 210-217.
36. Mak, A. C., Osterfeld, S. J., Yu, H., Wang, S. X., Davis, R. W., Jejelowo, O. A. and Pourmand, N., (2010). Sensitive giant magnetoresistive-based immunoassay for multiplex mycotoxin detection. *Biosens. Bioelectron.*, **25**, pp. 1635-1639.
37. Wilson, M. S. and Nie, W., (2006). Multiplex Measurement of Seven Tumor Markers Using an Electrochemical Protein Chip. *Anal. Chem.*, **78**, pp. 6476-6483.
38. Calvo, D., Tort, N., Salvador, J. P., Marco, M. P., Centi, F. and Marco, S., (2011). Preliminary study for simultaneous detection and quantification of androgenic anabolic steroids using ELISA and pattern recognition techniques. *Analyst*, **136**, pp. 4045-4052.
39. Roth, K. M., Peyvan, K., Schwarzkopf, K. R. and Ghindilis, A., (2006). Electrochemical detection of short DNA oligomer hybridization using the combiMatrix electraSense microarray reader. *Electroanalysis*, **18**, pp. 1982-1988.
40. Yeung, S.-W., Ming-Hung Lee, T., Cai, H. and Hsing, I.-M., (2006). A DNA biochip for on-the-spot multiplexed pathogen identification. *Nucleic Acids Research*, **34**, pp. e118.
41. Bonanni, A., Calvo, D. and del Valle, M., (2008). Dual-Genic Hybridization Sensor Employing Electrochemical Impedance Spectroscopy. *Electroanalysis*, **20**, pp. 941-948.
42. Xu, D., Han, H., He, W., Liu, Z., Xu, D. and Liu, X., (2006). Electrically Addressed Fabrication of Aptamer-Based Array Electrodes. *Electroanalysis*, **18**, pp. 1815-1820.
43. Taylor, E. H. (1989) in *Chemical Analysis* (Ed. Winefordner, J. D.), (John Wiley & Sons, New York).
44. Burtis, J. D., Ashwood, E.R. (1994) *Tietz Textbook of Clinical Chemistry*. (W.B. Saunders, Philadelphia), pp.
45. Jurkiewicz, M., Alegret, S., Almirall, J., Garcia, M. and Fabregas, E., (1998). Development of a biparametric bioanalyser for creatinine and urea. Validation of the determination of biochemical parameters associated with hemodialysis. *Analyst*, **123**, pp. 1321-1327.
46. Karube, I., Matsuoka, H., Suzuki, S., Watanabe, E. and Toyama, K., (1984). Determination of fish freshness with an enzyme sensor system. *J. Agric. Food Chem.*, **32**, pp. 314-319.
47. Matsumoto, K., Kamikado, H., Matsubara, H. and Osajima, Y., (1988). Simultaneous determination of glucose, fructose, and sucrose in mixtures by amperometric flow-injection analysis with immobilized enzyme reactors. *Anal. Chem.*, **60**, pp. 147-151.
48. Radomska, A., Koncki, R., Pyrzyńska, K. and Glab, S., (2004). Bioanalytical system for control of hemodialysis treatment based on potentiometric biosensors for urea and creatinine. *Anal. Chim. Acta*, **523**, pp. 193-200.
49. Rui, C. S., Sonomoto, K. and Kato, Y., (1992). Amperometric flow-injection biosensor system for the simultaneous determination of urea and creatinine. *Anal. Sci.*, **8**, pp. 845-850.

50. Gutierrez, M., Alegret, S. and del Valle, M., (2008). Bioelectronic tongue for the simultaneous determination of urea, creatinine and alkaline ions in clinical samples. *Biosens. Bioelectron.*, **23**, pp. 795-802.
51. Koncki, R., Radomska, A. and Glab, S., (2000). Bioanalytical flow-injection system for control of hemodialysis adequacy. *Anal. Chim. Acta*, **418**, pp. 213-224.
52. Zhang, L., Liang, Y. Z., Jiang, J. H., Yu, R. Q. and Fang, K. T., (1998). Uniform design applied to nonlinear multivariate calibration by ANN. *Anal. Chim. Acta*, **370**, pp. 65-77.
53. Ebrahimi, D., Chow, E., Gooding, J. J. and Hibbert, D. B., (2008). Multivariate sensing: a chemometrics approach to understanding the merits of electrode arrays versus single electrodes. *Analyst*, **133**, pp. 1090-1096.
54. Cetó, X., Gutiérrez, J. M., Moreno-Barón, L., Alegret, S. and del Valle, M., (2011). Voltammetric Electronic Tongue in the Analysis of Cava Wines. *Electroanalysis*, **23**, pp. 72-78.
55. de Carvalho, R. M., Mello, C. and Kubota, L. T., (2000). Simultaneous determination of phenol isomers in binary mixtures by differential pulse voltammetry using carbon fibre electrode and neural network with pruning as a multivariate calibration tool. *Anal. Chim. Acta*, **420**, pp. 109-121.
56. Gutiérrez-Osuna, R. and Nagle, H. T., (1999). A method for evaluating data-preprocessing techniques for odour classification with an array of gas sensors. *Systems, Man, and Cybernetics, Part B: Cybernetics*. IEEE Transactions on, **29**, pp. 626-632.
57. Moreno-Barón, L., Cartas, R., Merkoçi, A., Alegret, S., del Valle, M., Leija, L., Hernandez, P. R. and Muñoz, R., (2006). Application of the wavelet transform coupled with artificial neural networks for quantification purposes in a voltammetric electronic tongue. *Sens. Actuator B-Chem.*, **113**, pp. 487-499.
58. Cetó, X., Céspedes, F. and del Valle, M., (2012). Bioelectronic Tongue for the quantification of total polyphenol content in wine. *Talanta*, **99**, pp. 544-551.
59. Céspedes, F., Martínez-Fabregas, E. and Alegret, S., (1996). New materials for electrochemical sensing I. Rigid conducting composites. *TrAC Trends Anal. Chem.*, **15**, pp. 296-304.

7. Publicacions fora del
compendi que formen part de
la tesi doctoral com a annex

7. Publicacions fora del compendi que formen part de la tesi doctoral com a annex

- 7.1. *Application of an Electronic Tongue towards the analysis of brandies*
X. Cetó, M. Llobet, J. Marco and M. del Valle
Analytical Methods, **5**(5): 1120-1129
- 7.2. *Comparison of methods for the processing of voltammetric Electronic Tongues data*
X. Cetó, F. Céspedes and M. del Valle
Microchimica Acta, **180**(5-6): 319-330
- 7.3. *Voltammetric Electronic Tongue for the qualitative analysis of beers*
X. Cetó, J.M. Gutiérrez, A. Mimendia, F. Céspedes and M. del Valle
Electroanalysis, **25**(7), 1635-1644
- 7.4. *Cava wine authentication employing a voltammetric Electronic Tongue*
X. Cetó, J. Capdevila, S. Mínguez and M. del Valle
manuscrit
- 7.5. *Voltammetric BioElectronic Tongue for the analysis of phenolic compounds in rosé cava wines*
X. Cetó, J. Capdevila, S. Mínguez and M. del Valle
manuscrit

Article 7

Application of an Electronic Tongue towards the analysis of brandies

X. Cetó, M. Llobet, J. Marco and M. del Valle

Analytical Methods, 5(5): 1120-1129

Analytical Methods

www.rsc.org/methods

Volume 5 | Number 5 | 7 March 2013 | Pages 1073–1356



ISSN 1759-9660

RSC Publishing

PAPER
del Valle *et al.*
Application of an electronic tongue towards the analysis of brandies



1759-9660(2013)5:5;1-0

Application of an electronic tongue towards the analysis of brandiest

Cite this: *Anal. Methods*, 2013, 5, 1120

Xavier Cetó,^a Matias Llobet,^b Joan Marco^b and Manel del Valle^{*a}

This work reports the application of a voltammetric Electronic Tongue (ET) in the analysis of brandies, specifically in their classification according to the scores given by a skilled sensory panel and in the discrimination of different ageing methods. For this purpose, spirits were analyzed with no other pretreatment than their dilution with a saline solution to ensure enough conductivity. Recorded voltammetric signals produced by an array of six modified epoxy-composite sensors were preprocessed employing Fast Fourier Transform in order to reduce the complexity of the input signals while preserving the relevant information. Then, using the obtained coefficients, responses were evaluated using Linear Discriminant Analysis (LDA) as the pattern recognition model used to carry out the classification tasks. In both cases, good prediction ability was attained by the ET (classification rates of 100% and 97%, respectively), therefore permitting the correct classification of the different samples under study. Furthermore, two Artificial Neural Network models were also trained for the semi-quantitative identification of some undesired compound markers of some brandy defects above certain levels (namely butan-2-ol, ethyl acetate, acetaldehyde and butan-1-ol; $r > 0.975$) and the quantification of polyphenol index I_{280} ($r = 0.977$).

Received 18th September 2012

Accepted 14th December 2012

DOI: 10.1039/c2ay26066b

www.rsc.org/methods

1 Introduction

Over the last few years, the use of Electronic Tongues (ETs) is growing as a promising approach to analyze liquid samples.^{1,2} Such analytical systems are formed by an array of sensors where several sensing units, which exhibit different responses to various compounds, are coupled with advanced signal processing methods based on pattern recognition or multivariate response models, which allow for the qualitative or quantitative analysis of different sample parameters.³

In essence, ETs mimic the sense of taste in mammals, where a few receptors can respond to a large variety of substances in a combinatorial approach; this sensing principle is next coupled with a complex data treatment stage, as it is done in the brain, allowing to quantify or to classify perceptions or substances. These biomimetic systems, then in contrast to ever more selective classical methods, assume the combination of low selectivity sensor arrays (or with cross-response features) in order to obtain some added value in the generation of analytical information.

Brandy is a spirit produced from wine distillates finally containing *ca.* 36–45% alcohol by volume (abv).⁴ The production of brandy is a complex process that consists of several major steps including the production of base wine. Among brandies, the most well-known is Cognac; although this refers only to a type of brandy made in that region of France, many producers from different countries try to emulate its production methods. In particular, brandy must be made from specific grape varieties, and it must be twice distilled in copper pot stills and aged for at least two years in French oak barrels from certain regions. Brandy matures in the same way as whisky and wine when aged in barrels, and most brandies are aged considerably longer than the minimum legal requirement to improve their sensorial features. Usually, before blending, the aged brandy is first tasted by a panel to be classified and to assess its quality.

Classification of brandies is usually performed by a skilled sensory panel, given the difficulties to perform this assessment using classical analytical techniques. That is, the main compounds primarily responsible for the style of brandy have not been clearly identified yet, what in turn hinders the use of classical methods which would imply identifying the compounds responsible for those sensations. In contrast, ETs combine all the extracted chemical information from the array of sensors to generate a model like the human brain would do, avoiding focusing on which specific compounds are responsible for those flavours or tastes, but also recognizing a specific fingerprint to identify or quantify them. A correctly trained ET

^aSensors and Biosensors Group, Department of Chemistry, Universitat Autònoma de Barcelona, Edifici Cn, 08193 Bellaterra, Barcelona, Spain. E-mail: manel.delvalle@uab.cat; Fax: +34 93 5812477; Tel: +34 93 5811017

^bMiguel Torres SA, 08720 Vilafranca del Penedès, Barcelona, Spain

† Electronic supplementary information (ESI) available: Tables containing detailed information of the samples under study and examples of the recorded voltammograms for the whole sensor array and both subsets. See DOI: 10.1039/c2ay26066b

would extract knowledge from the expert panel, and apply it in correspondence to the sensor's available information to reproduce operation of the former. Hence, the obtaining of an automated method would allow significant reduction of brandy tasting by the panel, not replacing it, and an increase of the number or frequency of samples that can be analyzed. In this way, ETs may provide a helpful tool that might be used as a screening method either to assess producers' guarantee quality control standards or to detect any adulteration or contamination.

Within this context, ETs have already been successfully applied to the classification of several alcoholic beverages such as wines,⁵ cava wines,⁶ beers,^{7,8} whiskies⁹ or even sakes;¹⁰ besides, ETs have been able to focus in some cases on the detection of inappropriate handling practices,¹¹ the use of alternative ageing methods¹² or the adulteration of samples.⁹ A unique prior attempt to apply ETs towards the qualification of brandies is a conference communication from Legin's group in Saint Petersburg;¹³ also classification of some eau-de-vie samples was attempted by the same group.¹⁴

The main goal of this work was the application of an ET that would allow the recognition and discrimination of the different types of brandies, in a similar way a sensory panel would do, allowing also the detection of adulterations in the production process. To this end, samples were measured using an array of modified graphite-epoxy voltammetric sensors; then, their complex voltammetric responses were preprocessed employing Fast Fourier Transform (FFT) in order to extract and to compress the most significant departure information, next evaluating their responses with a pattern recognition method to discriminate the different samples. In addition, two different Artificial Neural Network (ANN) models were also trained to semi-quantitatively predict the concentration of some undesired compounds which are markers of some defects of brandies when above certain levels, and to quantitatively assess their polyphenol content.

2 Experimental

2.1 Reagents and solutions

All reagents used were of analytical reagent grade and all solutions were prepared using deionised water from a Milli-Q system (Millipore, Billerica, MA, USA). Copper, platinum and silver nanoparticles (<50 nm), cobalt(II) phthalocyanine and polypyrrole were purchased from Sigma-Aldrich (St. Louis, MO, USA). KCl was purchased from Merck KGaA (Darmstadt, Germany).

2.2 Samples under study

A total set of 74 wine spirit samples were analyzed. These samples were divided into two different subsets in order to evaluate the capabilities of the ET in the qualitative analysis of brandies from two different points of view. In the first case, samples considered were selected considering their taste attributes; while in the second, the different methods used in the ageing of brandy were studied. Detailed information of the

samples under study could be found in the ESI, Tables S1 and S2.†

Additionally, in all cases, the alcohol content of samples, as alcohol by volume (abv), was quantified employing routine analysis based on the measurement of sample density, which is inversely proportional to the ethanol content and can be later related to abv through the use of conversion tables.⁴ Concretely, in our case sample density (ergo abv) was measured at 20 °C with an electronic densimeter (ANTON PAAR DMA 4500, Graz, Austria) which already incorporates the correlation tables between density and ethanol content (abv), and thermostats the sample to avoid any error in the determination, given changes in the temperature induce changes in density.

2.2.1 Taste attributes. A total of 41 samples, previously selected and tasted by a sensory panel (formed by oenologist from *Miguel Torres*), were considered. Those samples were classified into three different groups according to their taste attributes;^{4,15} in this manner, most appreciated samples were assigned to category "taste 1", while worst ones were assigned to "taste 3". Hence, in this case the objective was to evaluate if the ET was able to classify the samples in the same manner the sensory panel would do.

Basically, as stated, that classification agrees with usual producers' standards assigning to class 1 the samples that have the best mouthfeeling and aroma (*i.e.* total absence of any defect or unpleasant remark; but also pleasant perceived flavours, aromas and general characteristics of a good wine). The ones with some slight defects are assigned to class 2 (*i.e.* requiring confirmation of their quality by standard methods, such as the determination of higher alcohols by GC to confirm that there is not any marker of brandy defects; besides, their usage should be limited during the blending to avoid demoting of the quality of the final product); while the worst ones are categorized as class 3 (*i.e.* the ones not suitable according to company/sector standards for the production of a good brandy).

2.2.2 Ageing of the brandy. Additionally, another set of 33 samples, all of them from the 2009 vintage (except the ones aged in new barrels), were analyzed. In this case, samples were selected in order to obtain a set of samples that have been aged using different methods, *i.e.* using pieces of wood (chips, sticks, blocks and strips) in the ageing of brandy to resemble its ageing in new barrels while accelerating the process. Thus, the aim was to assess if the ET is capable of distinguishing between the different ageing methods.

To ensure that discrimination was due to the different ageing methods, two precautions were taken into account. Firstly, all samples considered were from the 2009 vintage. However, since the aim of using wood pieces is to accelerate the ageing process, those samples are aged for a short period of time (vintage 2009), while the ones aged in new barrels are aged for longer periods (vintage 2008). Thus, the final product was obtained at the same time. Additionally, polyphenol index I_{280} was also measured to confirm that no trend in its value was found between different methods, hence to ensure that those parameters are not the source of discrimination between samples by the ET.

2.3 Electronic tongue

2.3.1 Sensor array. Based on previous studies in our laboratory,^{6,11,16} an array of 6 voltammetric electrodes were prepared following the conventional methodology by our research group.¹⁷ First, a resin EpoTek H77 (Epoxy Technology, Billerica, MA, USA) and its corresponding hardener compound were mixed in the ratio 20 : 3 (w/w); after that a 15% of graphite (w/w) and a 2% of the modifier (w/w) were added to the previous mixture before hardening, obtaining the composite. Then, it was manually homogenized for 60 min, and afterwards, the paste was allowed to harden for 3 days at 80 °C. Finally, the electrode surface was polished with different sandpapers of decreasing grain size, with a final electrode area of 28 mm².

In this manner, one blank electrode plus 5 different bulk-modified voltammetric sensors were prepared using bare graphite C in the first case, and adding different modifiers such as copper, platinum and silver nanoparticles, cobalt phthalocyanine and polypyrrole to the bulk mixture – one component per electrode.

2.3.2 Apparatus and measurement cell. The amperometric measurement cell was formed by the 6-sensor voltammetric array plus a commercial platinum counter electrode (Model 52-67, Crison Instruments, Barcelona, Spain) and a reference double junction Ag/AgCl electrode (Thermo Orion 900200, Beverly, MA, USA). Cyclic voltammetry measurements were taken using a 6-channel AUTOLAB PGSTAT20 (Ecochemie, The Netherlands), in a multichannel configuration, controlled with GPES Multichannel 4.7 software package. Electrodes were cycled for 3–5 times in saline solution in order to get stable voltammetric responses before performing the measurements with the wine spirit samples.

2.3.3 Voltammetric measurements. Prior to their measurement, samples were firstly diluted in 0.1 M KCl in order to ensure good conductivity of the media. In addition, all samples were diluted to the same abv level (~17.4°) to prevent the high level of alcohol in the original samples damaging the sensors and also to avoid that this factor could be a source of discrimination.⁹

Cyclic voltammetry measurements were carried out at room temperature (25 °C) under quiescent conditions and with no other sample pretreatment than its mere dilution with KCl solution. The potential was swept between –1.0 V and +1.3 V *vs.* Ag/AgCl, with a scan rate of 100 mV s⁻¹ and a step potential of 9 mV. Apart, all experiments were carried out without performing any physical surface regeneration of the working electrodes; only, an electrochemical cleaning stage between each measurement was done in order to prevent the accumulative effect of impurities on the working electrode surfaces. For this, a conditioning potential of +1.4 V was applied for 45 s after each experiment in a cell containing 25 ml of distilled water.⁶

2.4 Data processing

All computations and chemometric processing were carried out by specific routines written by the authors using MATLAB 7.1 (MathWorks, Natick, MA) and its Neural Network Toolbox (v4.0.6). For the qualitative approach Linear Discriminant

Analysis (LDA) was used as the modelling tool, while quantitative models were based on the use of Artificial Neural Networks (ANNs). Sigmaplot 2000 (Systat Software Inc., California, USA) was used for graphic representation of data and results.

3 Results and discussion

3.1 Voltammetric responses

Examples of the different obtained signals produced by the voltammetric ET array for the different samples of both subsets are shown in Fig. 1; additionally, the obtained response for all the sensors could be seen in the ESI, Fig. S1 and S2.† Voltammograms on the left correspond to the samples from the first subset, the one with the qualification of fresh distillate taste, while the ones on the right correspond to the second subset used to evaluate the use of alternative ageing methods.

In the first case, it could be seen how some samples present very similar voltammetric responses in specific regions along the voltammogram; also, those analogous resemblances are repeated along the whole recorded signal, but between different samples. *E.g.* in Fig. 1A it could be seen how the samples A16, A23 and A29 show equivalent anodic response at the region around 1 V; while for the cathodic wave (*ca.* 0 V), A23 shows differentiated response, but still similar signals are obtained for samples A16 and A29, plus sample A27, which presents the same response in this region, but not in the other. Also, differentiated trends are obtained for the different sensors from the ET; *e.g.* in Fig. 1C it could be seen that only samples A1 and A4 show the same response in the anodic region close to –1 V, which could be somehow expected given both samples were classified to taste type 1 by the sensory panel. Thus, a huge source of variability and cross-response features are obtained for the different samples with the proposed sensor array, a highly desirable situation for an ET system.

Similarly, a consistent behaviour was obtained for the samples of the second subset (graphs on the right of Fig. 1). In this case, it is known that some oxidative signals may be due to the total polyphenolic content of the samples,¹⁶ although as it will be shown later, the discrimination is not due to this fact (Section 3.3.2), and as before different trends are shown along the whole voltammogram.

Hence, as can be seen, different voltammetric responses are obtained for each kind of sensor, with different behaviour for each sample, generating very rich data to be used as the departure point. Nevertheless, to fully exploit all the information obtained from each voltammogram, a compression step is required in order to gain advantages in training time, to avoid redundancy in input data and to obtain a model with better generalization ability.¹⁸ Moreover, the main reason is due to the extreme complexity of the considered departure signals that would require instead the use of multiway processing strategies.¹⁹

In order to reduce the large dataset generated for each sample (6 sensors × 512 current values at different polarization potentials), a preprocessing stage employing the Fast Fourier Transform (FFT) was used to compress the original sensor recordings.²⁰ In this manner, signals from each voltammogram

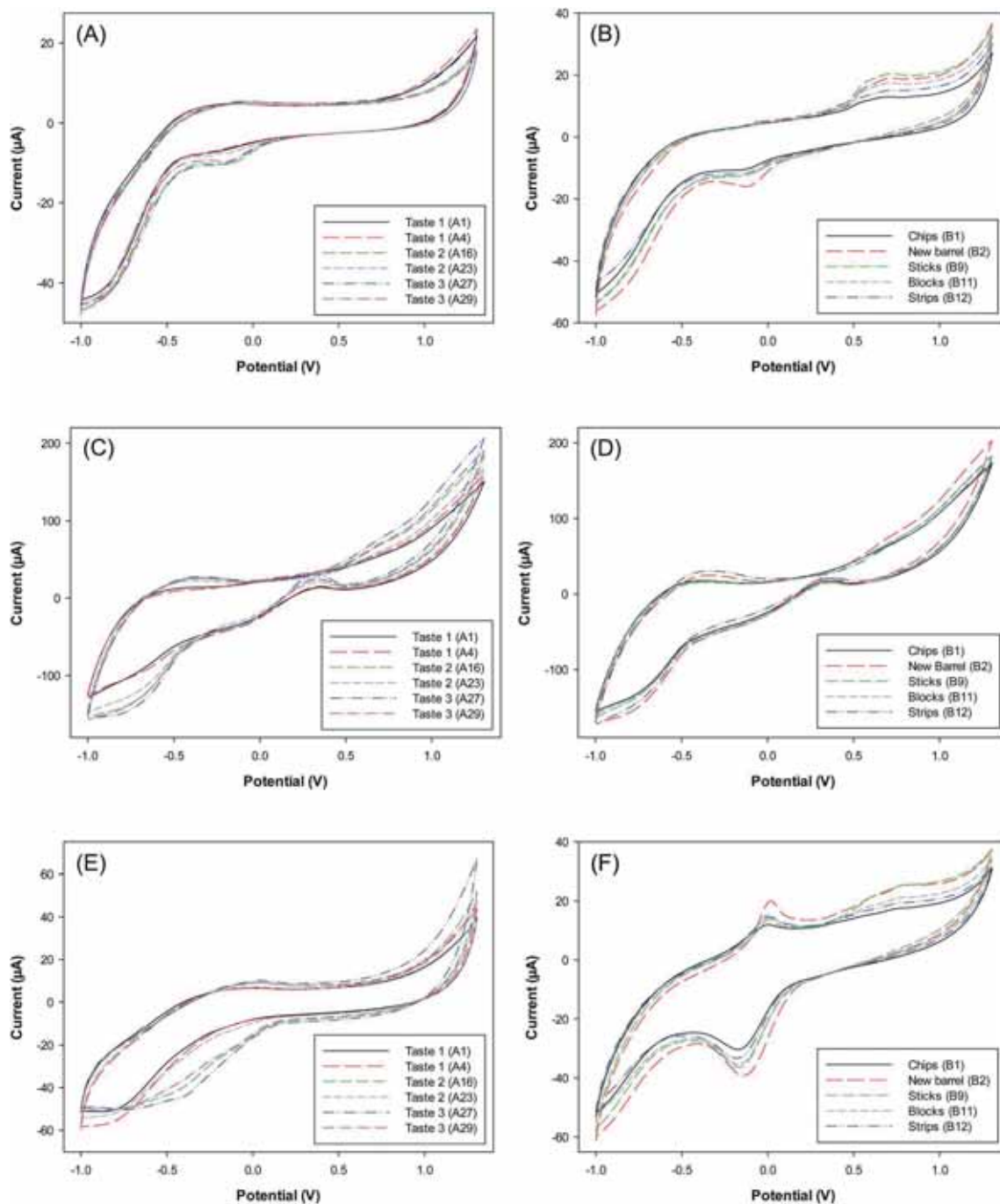


Fig. 1 Example of the different obtained voltammograms with the sensor array: (A and B) graphite-epoxy, (C and D) platinum nanoparticle and (E and F) copper nanoparticle. Graphs on the left correspond to fresh distillates of different quality from the first subset (Section 2.2.1), while the ones on the right correspond to spirits produced by different ageing methods from the second subset (Section 2.2.2). Detailed information of the samples is available in the ESI.†

were reduced down to 16 coefficients without any loss of relevant information, which means a compression ratio of 96.9%.

3.2 LDA classification models

After data compression, obtained coefficients were used as inputs in a LDA classification model (building one independent model for each of the cases). The obtained models will be formed by just $k-1$ linear discriminant functions (the new axis;

k being the number of groups), where these functions correspond to linear combinations of the predictor variables that provide the best discrimination between the groups.^{21,22} Concretely, this discrimination is usually measured employing Wilk's lambda (Λ) statistic, which provides a measure of the classes' separation,²¹ hence being used to evaluate when the optimum performance is reached (best separation between classes).

Additionally, LDA was done using a stepwise inclusion method which allows the variables that have a lower contribution to the classification model to be removed,²¹ this strategy tries to find the best subset to perform the prediction task as well as possible, with as few variables as can be (Table S3†). In this way, modelling accuracy and robustness of the final model are improved, as well as its easier understanding.

On the other hand, given this is a supervised method, the success of classification was evaluated using leave-one-out cross-validation. In this way, each sample is classified by means of the analysis function derived from the other samples (all cases except the case itself). Although results with this strategy sometimes are criticized as overoptimistic, its use is necessary when there is no availability of large sample sets.²³

3.2.1 Taste attributes. The first approach consisted of assessing if the ET was capable of discriminating the samples according to its taste attributes in the same manner a sensory panel would do. In this sense, ETs have a large advantage over classical methods given they are based on a biomimetic approach. That is, classical analytical methods require the identification of the concrete species responsible for those sensations, while the ET combines all the extracted chemical information from the array of sensors to generate a model like the human brain would do, not requiring to focus attention to which specific compounds are responsible for those flavours or tastes, but using a generic signal fingerprint instead. Therefore, samples were first tested by a sensory panel, which classified them into three classes from “1” to “3”, *i.e.* from best to worst taste quality. The LDA model was built employing their scores as the classes to be distinguished, extracting and using the knowledge from the experts.

As can be seen in Fig. 2, clear discrimination for the three qualities of distillates was achieved. Moreover, from the plot it could also be seen that samples are sorted along DF1 according to their taste class, with a complete agreement with the order assigned by the sensory panel.

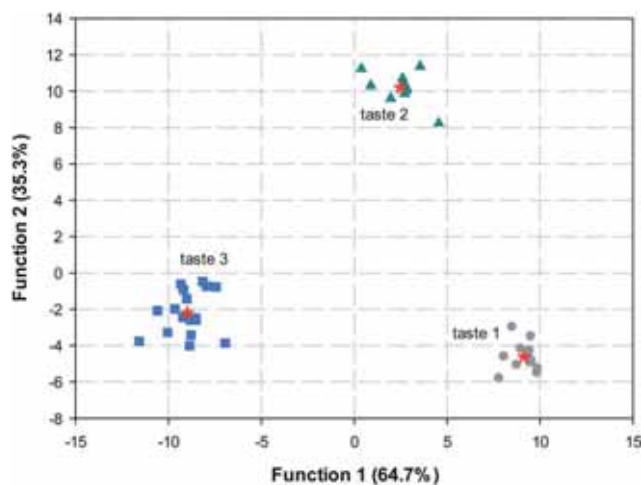


Fig. 2 Score plot of the functions obtained after LDA of the wine spirits, according to their taste category. As can be seen, clear discrimination is obtained for the different spirits: (●) taste 1, (▲) taste 2 and (■) taste 3. Also, the centroid of each class is plotted (★).

Table 1 Confusion matrix built according to the wine spirits' taste category obtained using the LDA model and leave-one-out cross-validation

| Expected | Found | | |
|---------------------------|---------|---------|---------|
| | Taste 1 | Taste 2 | Taste 3 |
| Taste 1 | 14 | 0 | 0 |
| Taste 2 | 0 | 10 | 0 |
| Taste 3 | 0 | 0 | 17 |
| Percentage classification | | | 100% |

Classification results of the LDA leave-one-out cross-validation approach are summarized in Table 1. As expected from the LDA plot, all samples were correctly classified according to their taste class. Hence, the classification rate of the samples reached 100% of accuracy. The efficiency of the classification obtained was also evaluated according to their sensitivity, *i.e.* the percentage of objects of each class identified by the classifier model, and to their specificity, the percentage of objects from different classes correctly rejected by the classifier model. From the statistical definition, both parameters were calculated for each of the classes, averaging them afterwards to obtain the indicator of the global model performance. In this case, the value of sensitivity, averaged for the classes considered, was 100%, and that of specificity was 100%.

3.2.2 Ageing method. As done in the previous case, a new LDA model was built and evaluated using the leave-one-out cross-validation method, now with the second subset of samples. In this case, the target was to see if the ET was able to distinguish the different uses of oak pieces of wood in the ageing of brandy to reproduce its ageing in new barrels while accelerating the process. Thus, in this case samples were distributed in five groups for the model construction: chips, new barrel, sticks, blocks and strips.

As shown in Fig. 3, also in this case, clear discrimination between the five classes of spirits was achieved, where well

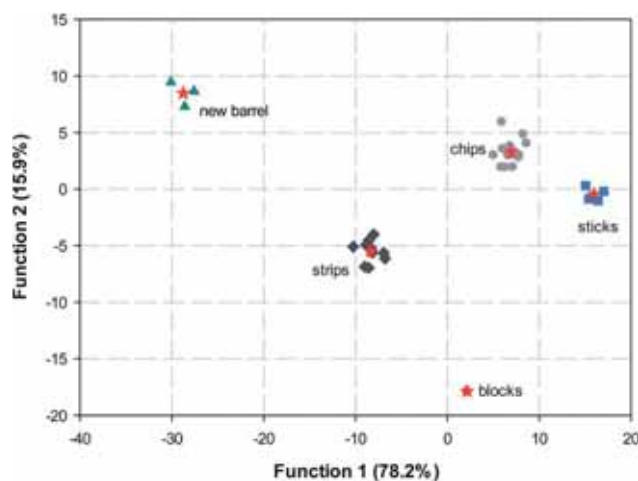


Fig. 3 Score plot of the functions obtained after LDA of the wine spirits, according to the ageing method used. As can be seen, clear discrimination is obtained for the different spirits: (●) wooden chips, (▲) new barrel, (■) wooden sticks, (●) wooden blocks and (◆) wooden strips. Also, the centroid of each class is plotted (★).

Table 2 Confusion matrix built according to the wine spirits' ageing method obtained using the LDA model and leave-one-out cross-validation

| Expected | Found | | | | |
|---------------------------|-------|------------|--------|--------|--------|
| | Chips | New barrel | Sticks | Blocks | Strips |
| Chips | 15 | 0 | 0 | 0 | 0 |
| New barrel | 0 | 3 | 0 | 0 | 0 |
| Sticks | 0 | 0 | 4 | 0 | 0 |
| Blocks ^a | 0 | 0 | 0 | 0 | 1 |
| Strips | 0 | 0 | 0 | 0 | 10 |
| Percentage classification | | | | | 97.0% |

^a Only one sample was available.

established clusters evidence that samples are grouped according to their ageing process. Analyzing more deeply the obtained plot, it could be seen that barrel samples are more distant from the other clusters, while samples using alternative methods are closer to each other. Also the distance of each group from the barrel cluster could give an idea of how different is the proposed method from the original one. In this sense, some authors claim that a key parameter is the size of the pieces of wood.²⁴ For example, the two closer variants being chips and sticks, evidence two procedures only distinguished by the size and shape used.

Again, model accuracy was evaluated employing leave-one-out cross-validation, the results of which are summarized in Table 2. As expected from the LDA plot, nearly all samples were correctly classified, with a classification rate of 97.0% in terms of accuracy, while sensitivity and specificity values were 80.0% and 99.1%. It should be noticed that there was only one sample misclassified, which corresponded to the blocks type. This is due to the fact that only one sample from this class was available for the study; therefore, when describing the model class with the leave-one-out cross-validation method, the absence of other similar samples prevented its correct identification. Thus, when evaluating its class, the pattern recognition method only was able to assign it to the nearest cluster, the strip.

In this way, to improve the performance of the qualitative models, as well as to reduce the risk of misclassifications, model's response can be modified establishing confidence levels around the centroid of each class, so that if one sample is beyond that limit it is rather classified as "unknown". Similarly, further improvements can be done taking into account its Receiver Operating Characteristics (ROC).²⁵ Those are plots that taking into account sensitivity and specificity values illustrate the performance of a binary classifier system as its discrimination threshold is varied.

3.2.3 Vintage identification. Additionally, given the importance of the vintage to the quality of the obtained product (in the case of wine), and given brandies are obtained from its distillation, the effect of vintage was also studied. In fact, the existence of any fingerprint in the brandies related to vintage, that might be detectable by means of the ET, was examined. Hence, employing the samples from the first subset, a new LDA model was built taking into account this parameter. In this case samples were distributed in five groups for the model construction, from 2008 to 2012 vintage.

As shown in Fig. 4, also in this case, clear discrimination between the different vintages of spirits was achieved, where well established clusters evidence that samples are grouped according to their vintage. Analyzing more deeply the obtained plot, it could be seen that samples from 2012 are more distant from the other clusters, as could be expected given that they are less aged, and mainly separated by DF1, while the rest of samples are closer to each other and discriminated also thanks to DF2.

Again, model accuracy was evaluated by employing leave-one-out cross-validation, the results of which are summarized in Table 3. As expected from the LDA plot, nearly all samples were correctly classified, with a classification rate of 95.1% in terms of accuracy, while sensitivity and specificity values were 78.8% and 98.7%. Again, it should be noticed that there were only two samples misclassified, one of those corresponds to a class where only one sample was available for the study which prevented its correct identification.

3.3 ANN quantitative models

As a further ability of electronic tongue sensor systems, a quantitative application in order to predict compounds of relevant significance in brandies was also attempted. In this case, concentrations of the sought compounds were modelled from the set of voltammetric responses, previously compressed with FFT, by means of an ANN model. In this sense, and depending on the qualitative approach, different compounds were analyzed by classical methods for both subsets to extract additional quantitative analytical information that may complement ET qualitative response if their level could be modelled properly.

In the first case, given the importance to keep under certain levels some undesired compounds, mainly alcohols, responsible for taste defects in the fresh wine distillates,²⁶ those were quantified by standard procedures employing Gas

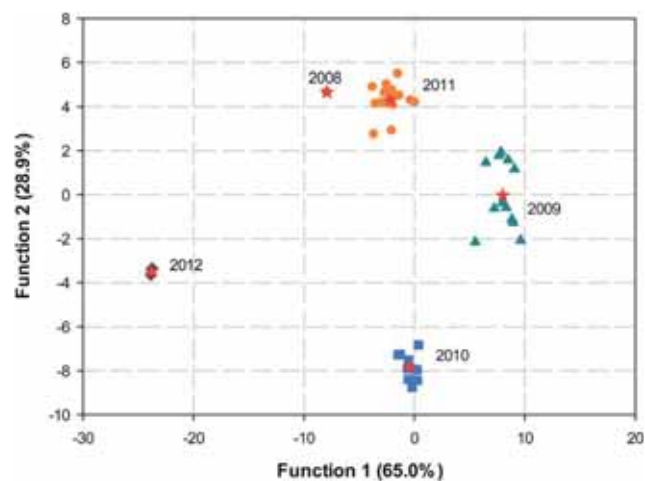


Fig. 4 Score plot of the functions obtained after LDA of the wine spirits, according to their vintage. As can be seen, clear discrimination is obtained for the different spirits: (●) 2008, (▲) 2009, (■) 2010, (○) 2011 and (◆) 2012. Also, the centroid of each class is plotted (★).

Table 3 Confusion matrix built according to wine spirits vintage obtained using the LDA model and leave-one-out cross-validation

| Expected | Found | | | | |
|---------------------------|-------|------|------|------|-------|
| | 2008 | 2009 | 2010 | 2011 | 2012 |
| 2008 ^a | 0 | 0 | 0 | 1 | 0 |
| 2009 | 0 | 12 | 0 | 0 | 0 |
| 2010 | 0 | 0 | 9 | 0 | 0 |
| 2011 | 1 | 0 | 0 | 16 | 0 |
| 2012 | 0 | 0 | 0 | 0 | 2 |
| Percentage classification | | | | | 95.1% |

^a Only one sample was available.

Chromatography (GC). In the second case, given the importance of polyphenols in the ageing of wine, these were quantified employing the polyphenol index I_{280} as the quick reference procedure.²⁷ Then, the objective was to correlate the ET response with specific compounds or with general index concentrations. As before, coefficients obtained after FFT analysis were used as inputs of the ANN model, building two different models for each of the subsets.

The first step, prior to building the ANN model, is the selection of the topology of the network. This is a trial-error process that requires a significant effort of optimization due to the difficulties in predicting the optimum ANN configuration in advance, where several parameters (training algorithms, number of hidden layers, transfer functions, *etc.*) are fine-tuned in order to find the best configuration that optimizes the performance of the neural network model.¹⁸ Thus, multiple ANN architectures and topologies were assayed employing Bayesian regularization algorithms. Comparison graphs of predicted *vs.* expected concentrations for each of the compounds under study were built to check the prediction ability of the ANN.

3.3.1 Markers of some brandy defects. In order to provide more relevant information of the samples that compliments its taste category, the identification of undesired compounds was also attempted in a semi-quantitative approach. Hence, much attention was focused to specific compounds known to induce some defects on brandy samples,⁴ and their quantification was attempted by means of the ET approach.

Generally, alcohols with more than two carbon atoms, known as higher alcohols and mainly produced during fermentation, and also their esters have intense odours that play an important role in wine aromas. At low concentrations, they contribute to their aromatic complexity; while at higher levels, their penetrating odours mask their aromatic finesse.²⁶ Particularly, butan-2-ol, ethyl acetate, acetaldehyde and butan-1-ol were the specific substances under study. Among those, butan-2-ol as such has no sensory impact but is often a marker of bacterial alterations coming from their lees in wines poorly conserved; acetaldehyde is an intermediate compound of the alcoholic fermentation, the concentration of which can be increased due to several factors such as the addition of SO₂; ethyl acetate induces a strange smell (similar to solvent glue) and is generated in sour wine from acetic acid and ethanol.

Table 4 Concentration ranges, expressed in mg L⁻¹, used for the division of samples into the three-level classes used for the semi-quantitative approach

| | Butan-2-ol | Ethyl acetate | Acetaldehyde | Butan-1-ol |
|----|-------------|---------------|--------------|-------------|
| -1 | <1 | <350 | <30 | <3 |
| 0 | 1 ≤ X < 2.5 | 350 ≤ X < 450 | 30 ≤ X < 50 | 3 ≤ X < 6.6 |
| +1 | ≥2.5 | ≥450 | ≥50 | ≥6.6 |

To estimate the presence of these markers, concentration levels of each of the compounds were divided into specific domains to a three-level scale (*i.e.* -1 for normal levels, 0 for slight defects and +1 for clear defects) according to food industry requirements (Table 4),²⁸ the values of which were then modelled employing a fuzzy logic prediction. Although the levels of some specific compounds can be used as markers of some brandy defects, it should be taken into account that they are only markers and do not determine the final taste category. That is, even those samples may have relatively high levels of one of those compounds, this does not imply that the sample taste classification will be demoted given its quality could be improved or demoted due to the presence of other compounds not quantified or identified. Some examples could be found in Table S1;† *e.g.* samples A36 and A37 both have low levels for the four considered compounds, and yet assigned to worst taste quality, while opposite situation could be seen for sample A8.

In this manner, a semi-quantitative ANN fuzzy model for the ordinal quantification of 4 different compounds was trained employing 75% of the data, using the remaining 25% of the data (testing subset) for the evaluation of model's performance, repeating this subdivision 4 times, employing the *k*-fold cross-validation method. That is, data were firstly distributed in four groups randomly; then a model was built using three subsets to train it and the fourth to evaluate its performance, repeating this process four times leaving out a different subset to assess its predictive capabilities each time. Afterwards, once all the models are built data are joined depending on if they have been employed in the training process or in the testing subset.

After some preliminary tests, the final architecture of the ANN model had 96 neurons (6 sensors × 16 coeffs. obtained from the FFT analysis) in the input layer, 6 neurons and *tansig* transfer function in the hidden layer and four neurons and *purelin* transfer function in the output layer, providing sample levels of the four compounds under study.

As can be seen in Fig. 5, where comparison graphs of predicted *vs.* expected levels for each of the compounds were built, good prediction ability was attained by the model. Although larger dispersion was obtained for a few samples, good trend is obtained in general, with fitted regression lines almost indistinguishable from the theoretical ones.

3.3.2 Polyphenol index. After evaluating the total polyphenol content of the spirits using the reference method, it was confirmed that there was no relation between its index value and the type of ageing method employed, and what's more, having nearly the same mean value for all the classes (Fig. 6). Hence, it was confirmed that discrimination of samples by the ET was not due to this parameter.

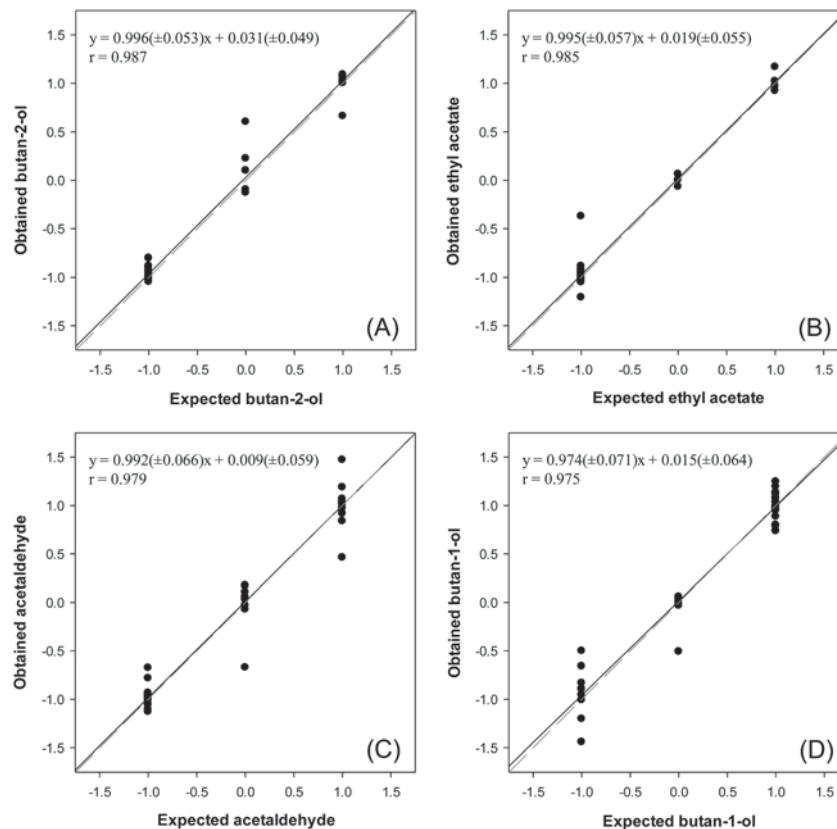


Fig. 5 Modelling ability of the optimized FFT-ANN for the semi-quantitative prediction of four compounds used as markers of some taste defects. Set adjustments of obtained vs. expected values for the testing subset are shown. The dashed line corresponds to theoretical diagonal line.

Nevertheless, capability of the ET system to correctly predict polyphenol index in brandies was evaluated, an approach already tested successfully in the case of wines.¹⁶ For this purpose, a quantitative ANN model was built and optimized. After some preliminary tests, the final architecture of the ANN model had 96 neurons (6 sensors \times 16 coeffs. obtained from the

FFT analysis) in the input layer, 6 neurons and *tansig* transfer function in the hidden layer and one neuron and *purelin* transfer function in the output layer, providing as output the predicted polyphenol index.

Accuracy of the generated model was evaluated employing the repeated random sub-sampling validation method to avoid dependence of predictions from the specific subdivision of data,²⁹ training the model with 75% of the data (25 samples selected at random) and doing its external test with the remaining 25% (8 samples). In this way, train/test data subdivision was repeated 25 times in order to ensure that model's accuracy was satisfactory and that ET performance did not depend on the specific subsets used. Then, once all responses from all the constructed models were obtained, predicted values by each model were grouped depending on if they were used in the training process or in the testing subset. Finally, average values for each sample were calculated, allowing us to calculate model uncertainties and obtain unbiased data.²⁹

Comparison graphs of predicted vs. expected I_{280} index were built to check the prediction ability of the ANN (Fig. 7). As can be observed, the obtained comparison parameters are close to the ideal values, with intercept near to 0 and slope and correlation coefficient around 1, meaning that there are no significant differences between the values predicted by the ANN model and those expected and provided by the reference method.

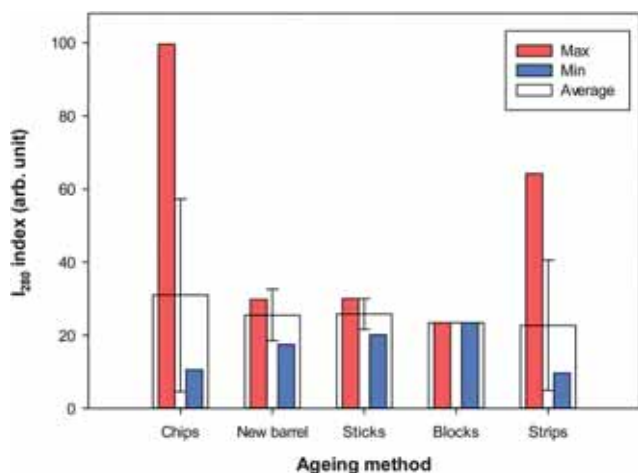


Fig. 6 Polyphenol content of each class as measured by the I_{280} method. Average content and its standard deviation are shown for each class, as well as its maximum and minimum.

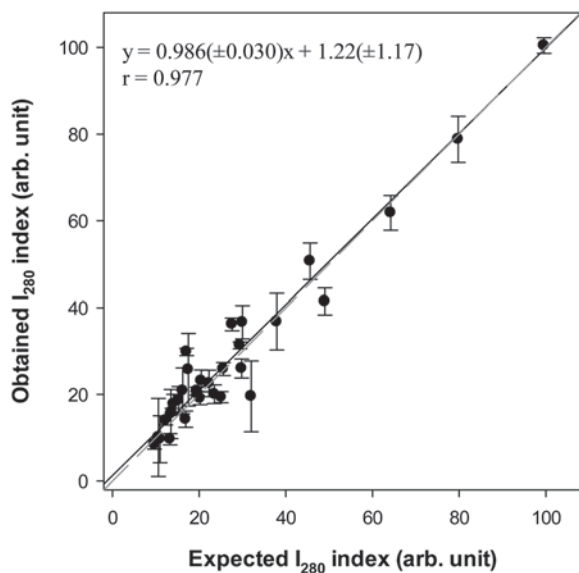


Fig. 7 Modelling ability of the optimized FFT-ANN for the prediction of I_{280} index. Set adjustments of obtained vs. expected values for the testing condition are shown. The dashed line corresponds to theoretical diagonal line. Results provided correspond to the average of the values obtained for each sample after 25 repeated calculations, done with random division of samples for train/test subsets each time. Uncertainties are calculated at the 95% confidence level.

4 Conclusions

Electronic tongues have proved to be a useful analytical tool able to provide information for either the detection of the use of alternative ageing practices such as the use of pieces of wood or to assess a quality index of brandy samples according to skilled sensory panel results. Moreover, it has also allowed the semi-quantitative detection of undesired compounds above certain levels or the quantitative prediction of polyphenol content according to the I_{280} method.

In this sense, ET represents an alternative to traditional methods with advantages over those such as being a fast-response and low-cost method for its characterization, really suitable for screening analysis. Furthermore, given its biomimetic behaviour, it presents a straightforward solution to the absence of the knowledge about which compounds are primarily responsible for the brandy style. What is more, the proposed ET system has been able to extract knowledge from the expertise gathered in the sensory panel and build correspondence with the signal fingerprint provided by the sensor array. This is accomplished overcoming the problems found with conventional analytical techniques such as spectroscopy and chromatography, the application of which to such tasks is really hindered by the fact that responsible compounds may not have been identified yet.

Acknowledgements

Financial support for this work was provided by *Spanish Ministry of Science and Innovation*, MCINN (Madrid) through project CTQ2010-17099 and by program ICREA Academia from

Generalitat de Catalunya. X. Cetó thanks the support of *Dept. d'Innovació, Universitats i Empresa de la Generalitat de Catalunya* for the predoctoral grant.

References

- 1 P. Ciosek and W. Wroblewski, *Analyst*, 2007, **132**, 963–978.
- 2 M. del Valle, *Electroanalysis*, 2010, **22**, 1539–1555.
- 3 Y. Vlasov, A. Legin, A. Rudnitskaya, C. Di Natale and A. D'Amico, *Pure Appl. Chem.*, 2005, **77**, 1965–1983.
- 4 A. J. Buglass (Ed), *Handbook of Alcoholic Beverages: Technical, Analytical and Nutritional Aspects*, Wiley, Chichester, UK, 2011.
- 5 J. Zeravik, A. Hlavacek, K. Lacina and P. Skladal, *Electroanalysis*, 2009, **21**, 2509–2520.
- 6 X. Cetó, J. M. Gutiérrez, L. Moreno-Barón, S. Alegret and M. del Valle, *Electroanalysis*, 2011, **23**, 72–78.
- 7 A. Rudnitskaya, E. Polshin, D. Kirsanov, J. Lammertyn, B. Nicolai, D. Saison, F. R. Delvaux, F. Delvaux and A. Legin, *Anal. Chim. Acta*, 2009, **646**, 111–118.
- 8 P. Ciosek and W. Wróblewski, *Talanta*, 2006, **69**, 1156–1161.
- 9 W. Novakowski, M. Bertotti and T. R. L. C. Paixão, *Microchem. J.*, 2011, **99**, 145–151.
- 10 K. Toko, *Meas. Sci. Technol.*, 1998, **9**, 1919.
- 11 J. M. Gutiérrez, L. Moreno-Barón, M. I. Pividori, S. Alegret and M. del Valle, *Microchim. Acta*, 2010, **169**, 261–268.
- 12 M. Gay, C. Apetrei, I. Nevares, M. del Alamo, J. Zurro, N. Prieto, J. A. De Saja and M. L. Rodríguez-Mendez, *Electrochim. Acta*, 2010, **55**, 6782–6788.
- 13 A. Rudnitskaya, E. Lochner, M. Lambrechts and A. Legin, ISOEN 2007 conference proceedings, St. Petersburg, Russia, 2007.
- 14 A. Legin, A. Rudnitskaya, B. Seleznev and Y. Vlasov, *Anal. Chim. Acta*, 2005, **534**, 129–135.
- 15 R. S. Jackson, *Wine tasting: a professional handbook*, Academic Press, London, 2002.
- 16 X. Cetó, J. M. Gutiérrez, M. Gutiérrez, F. Céspedes, J. Capdevila, S. Mínguez, C. Jiménez-Jorquera and M. del Valle, *Anal. Chim. Acta*, 2012, **732**, 172–179.
- 17 F. Céspedes, E. Martínez-Fàbregas and S. Alegret, *TrAC, Trends Anal. Chem.*, 1996, **15**, 296–304.
- 18 F. Despagne and D. Luc Massart, *Analyst*, 1998, **123**, 157–178R.
- 19 D. Ebrahimi, E. Chow, J. J. Gooding and D. B. Hibbert, *Analyst*, 2008, **133**, 1090–1096.
- 20 X. Cetó, F. Céspedes and M. del Valle, *Talanta*, 2012, **99**, 544–551.
- 21 R. A. Johnson and D. W. Wichain, *Applied multivariate statistical analysis*, Prentice-Hall, Upper Saddle River, N.J., 2007.
- 22 D. J. Hand, *Discrimination and classification*, Wiley, Chichester, UK, 1981.
- 23 D. Kirsanov, O. Mednova, V. Vietoris, P. A. Kilmartin and A. Legin, *Talanta*, 2012, **90**, 109–116.
- 24 M. Del Alamo Sanza and I. Nevares Domínguez, *Anal. Chim. Acta*, 2006, **563**, 255–263.

- 25 O. V. Demler, M. J. Pencina and R. B. D'Agostino, *Stat. Med.*, 2011, **30**, 1410–1418.
- 26 P. Ribéreau-Gayon, Y. Glories, A. Maujean and D. Dubourdiou, *Handbook of Enology, Volume 2, The Chemistry of Wine Stabilization and Treatments*, Wiley-Blackwell, London, 2000.
- 27 A. L. Waterhouse, in R. E. Wrolstad (Ed), *Current Protocols in Food Analytical Chemistry*, Wiley, New York, 2002.
- 28 J. Sourisseau, *Bull. Union Physiciens*, 2002, **96**, 881–892.
- 29 A. M. Molinaro, R. Simon and R. M. Pfeiffer, *Bioinformatics*, 2005, **21**, 3301–3307.

Application of an Electronic Tongue towards the analysis of brandies

Xavier Cetó¹, Matias Llobet², Joan Marco² and Manel del Valle^{1,*}

¹ *Sensors and Biosensors Group, Department of Chemistry, Universitat Autònoma de Barcelona, Edifici Cn, 08193 Bellaterra, Barcelona, SPAIN*

² *Miguel Torres SA, Vilafranca del Penedès, Barcelona, SPAIN*

Supporting information

This supplementary provides further details of the samples under study as well as showing the voltammetric responses obtained for the both subsets of samples and the whole array of sensors.

Additionally, the number of features used in each model (and from which sensors they came from), which were selected by using the stepwise inclusion method, are summarized in Table S3.

* E-mail: manel.delvalle@uab.cat; tel: +34 93 5811017; fax: +34 93 5812379

Table S1. Detailed information for the first subset of samples (section 2.2.1).

| Sample | Vintage | Alcohol by volume | Taste classification | Butan-2-ol | Ethyl acetate | Acetaldehyde | Butan-1-ol |
|--------|---------|----------------------|-------------------------|------------|------------------|--------------|------------|
| A1 | 2008 | 65.4 | 1 | -1 | -1 | -1 | 1 |
| A2 | 2009 | 65.0 | 1 | -1 | -1 | 1 | 0 |
| A3 | 2009 | 77.0 | 1 | -1 | -1 | 0 | 1 |
| A4 | 2010 | 65.0 | 1 | 1 | -1 | 0 | 0 |
| A5 | 2010 | 77.0 | 1 | 0 | 0 | 1 | 1 |
| A6 | 2010 | 65.0 | 1 | 0 | -1 | -1 | 1 |
| A7 | 2010 | 65.4 | 1 | 1 | -1 | 1 | 1 |
| A8 | 2011 | 65.0 | 1 | 1 | 1 | 1 | 1 |
| A9 | 2011 | 65.0 | 1 | -1 | 1 | 1 | 1 |
| A10 | 2011 | 65.2 | 1 | 0 | -1 | 0 | 1 |
| A11 | 2011 | 65.0 | 1 | -1 | 1 | 1 | -1 |
| A12 | 2011 | 77.2 | 1 | -1 | -1 | 1 | 1 |
| A13 | 2011 | 78.0 | 1 | -1 | 0 | 1 | 1 |
| A14 | 2012 | 77.2 | 1 | -1 | 0 | 1 | 1 |
| A15 | 2010 | 77.0 | 2 | 1 | -1 | -1 | 1 |
| A16 | 2010 | 65.0 | 2 | 1 | -1 | 0 | 1 |
| A17 | 2010 | 65.0 | 2 | 0 | -1 | -1 | 1 |
| A18 | 2011 | 65.0 | 2 | -1 | -1 | 0 | 1 |
| A19 | 2011 | 77.0 | 2 | -1 | -1 | -1 | 1 |
| A20 | 2011 | 77.2 | 2 | -1 | -1 | 1 | 1 |
| A21 | 2011 | 77.0 | 2 | -1 | 1 | 1 | 1 |
| A22 | 2011 | 77.0 | 2 | -1 | -1 | -1 | -1 |
| A23 | 2012 | 77.0 | 2 | 1 | -1 | 1 | 1 |
| A24 | 2010 | 70.0 | 2 | -1 | -1 | 1 | 0 |
| A25 | 2009 | 65.0 | 3 | -1 | -1 | -1 | 0 |
| A26 | 2009 | 65.0 | 3 | -1 | -1 | -1 | 0 |
| A27 | 2009 | 65.0 | 3 | -1 | -1 | -1 | 0 |
| A28 | 2010 | 65.0 | 3 | 0 | -1 | 0 | 1 |
| A29 | 2011 | 65.0 | 3 | -1 | -1 | 0 | 1 |
| A30 | 2011 | 77.0 | 3 | 1 | -1 | -1 | 1 |
| A31 | 2011 | 65.0 | 3 | -1 | -1 | 1 | 1 |
| A32 | 2011 | 65.0 | 3 | -1 | -1 | 1 | 1 |
| A33 | 2009 | 70.0 | 3 | -1 | -1 | -1 | -1 |
| A34 | 2009 | 70.0 | 3 | -1 | -1 | -1 | 0 |
| A35 | 2009 | 70.0 | 3 | -1 | -1 | -1 | 0 |
| A36 | 2009 | 70.0 | 3 | -1 | -1 | -1 | -1 |
| A37 | 2009 | 70.0 | 3 | -1 | -1 | -1 | -1 |
| A38 | 2009 | 70.0 | 3 | -1 | -1 | -1 | -1 |
| A39 | 2009 | 70.0 | 3 | -1 | -1 | -1 | 0 |
| A40 | 2011 | 65.0 | 3 | -1 | -1 | -1 | -1 |
| A41 | 2011 | 65.0 | 3 | -1 | -1 | 0 | -1 |

Table S2. Detailed information for the second subset of samples (section 2.2.2)

| Sample | Vintage | Alcohol by volume | Ageing method | I ₂₈₀ index |
|--------|---------|----------------------|------------------|------------------------|
| B1 | 2009 | 69.5 | Chips | 14.1 |
| B2 | 2008 | 69.5 | New barrel | 17.5 |
| B3 | 2009 | 69.5 | Chips | 13.5 |
| B4 | 2009 | 69.5 | Chips | 19.5 |
| B5 | 2009 | 69.5 | Chips | 49 |
| B6 | 2009 | 69.5 | Chips | 25.1 |
| B7 | 2008 | 69.5 | New barrel | 29.4 |
| B8 | 2009 | 69.5 | Chips | 37.9 |
| B9 | 2009 | 69.5 | Strips | 27.6 |
| B10 | 2009 | 69.5 | Strips | 25.6 |
| B11 | 2009 | 69.5 | Blocks | 23.5 |
| B12 | 2009 | 69.5 | Sticks | 16.2 |
| B13 | 2009 | 69.5 | Sticks | 13.7 |
| B14 | 2009 | 69.5 | Sticks | 12.3 |
| B15 | 2009 | 69.5 | Sticks | 13.3 |
| B16 | 2009 | 69.5 | Sticks | 22.3 |
| B17 | 2009 | 69.5 | Sticks | 45.6 |
| B18 | 2009 | 69.5 | Sticks | 64.2 |
| B19 | 2008 | 69.5 | New barrel | 29.8 |
| B20 | 2009 | 69.5 | Chips | 15.3 |
| B21 | 2009 | 69.5 | Chips | 17.1 |
| B22 | 2009 | 69.5 | Sticks | 9.7 |
| B23 | 2009 | 69.5 | Sticks | 16.9 |
| B24 | 2009 | 69.5 | Sticks | 13.5 |
| B25 | 2009 | 69.5 | Chips | 10.6 |
| B26 | 2009 | 69.5 | Chips | 19.6 |
| B27 | 2009 | 69.5 | Chips | 11 |
| B28 | 2009 | 69.5 | Chips | 20.5 |
| B29 | 2009 | 69.5 | Chips | 32 |
| B30 | 2009 | 69.5 | Chips | 79.8 |
| B31 | 2009 | 69.5 | Chips | 99.6 |
| B32 | 2009 | 69.5 | Strips | 20.2 |
| B33 | 2009 | 69.5 | Strips | 30 |

Table S3. Number of features (FFT coefficients) employed for building the different LDA models.

| <i>Sensor</i> | <i>Taste attributes</i> | <i>Ageing method</i> | <i>Vintage identification</i> |
|----------------------------------|-------------------------|----------------------|-------------------------------|
| Graphite-epoxy | 6 | 4 | 5 |
| Cobalt (II) phthalocyanine | 2 | 1 | 3 |
| Platinum nanoparticle | 4 | 4 | 4 |
| Silver nanoparticle | 9 | 5 | 5 |
| Polypyrrole | 3 | 4 | 3 |
| Copper nanoparticle | 6 | 3 | 5 |
| <i>Total coefficients</i> | 30 | 21 | 25 |

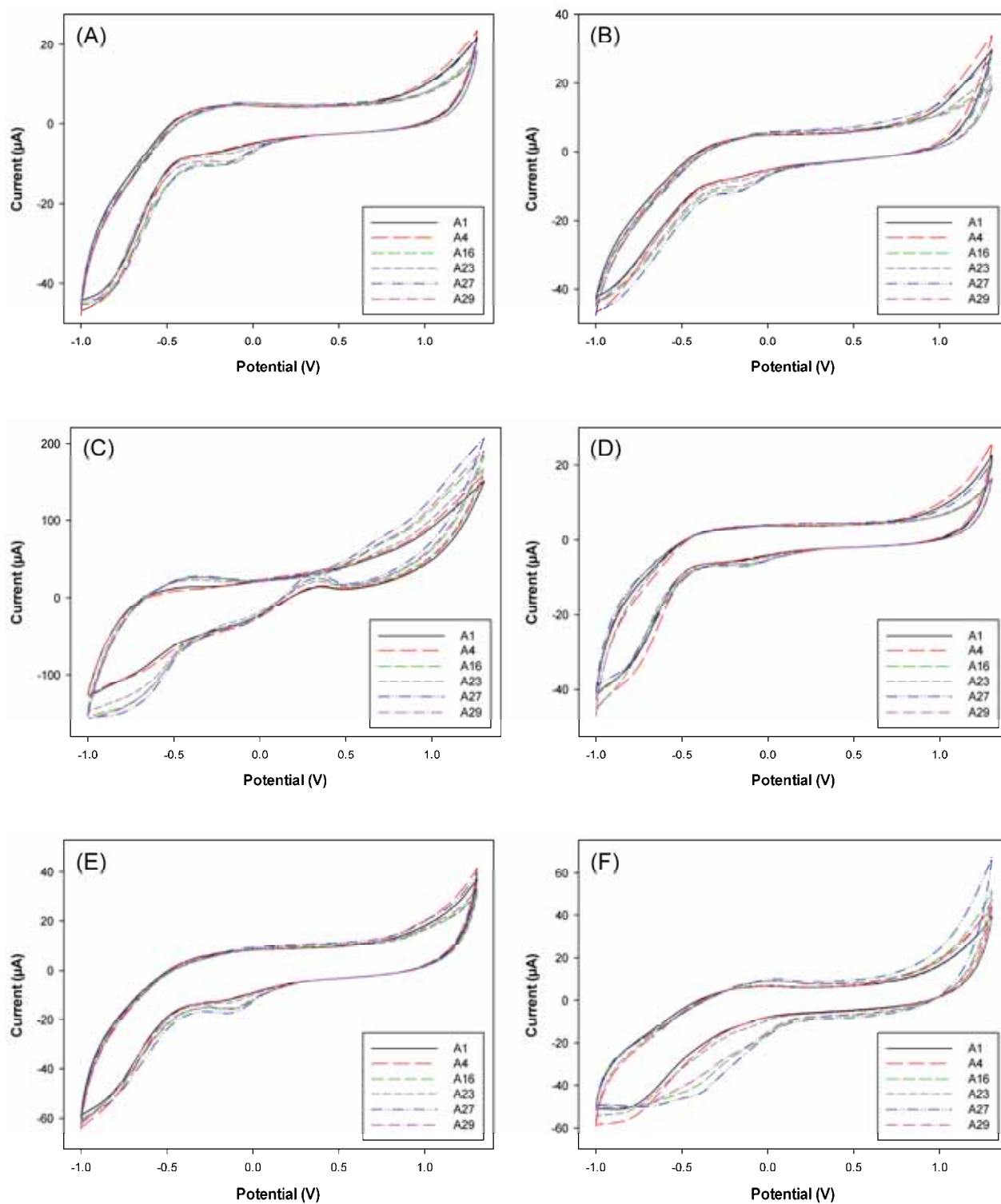


Figure S1. Example of the different obtained voltammograms for the first data set (section 2.2.1) obtained with the sensor array: (A) graphite-epoxy, (B) phthalocyanine, (C) platinum nanoparticle, (D) silver nanoparticle, (E) polypyrrole and (F) copper nanoparticle.

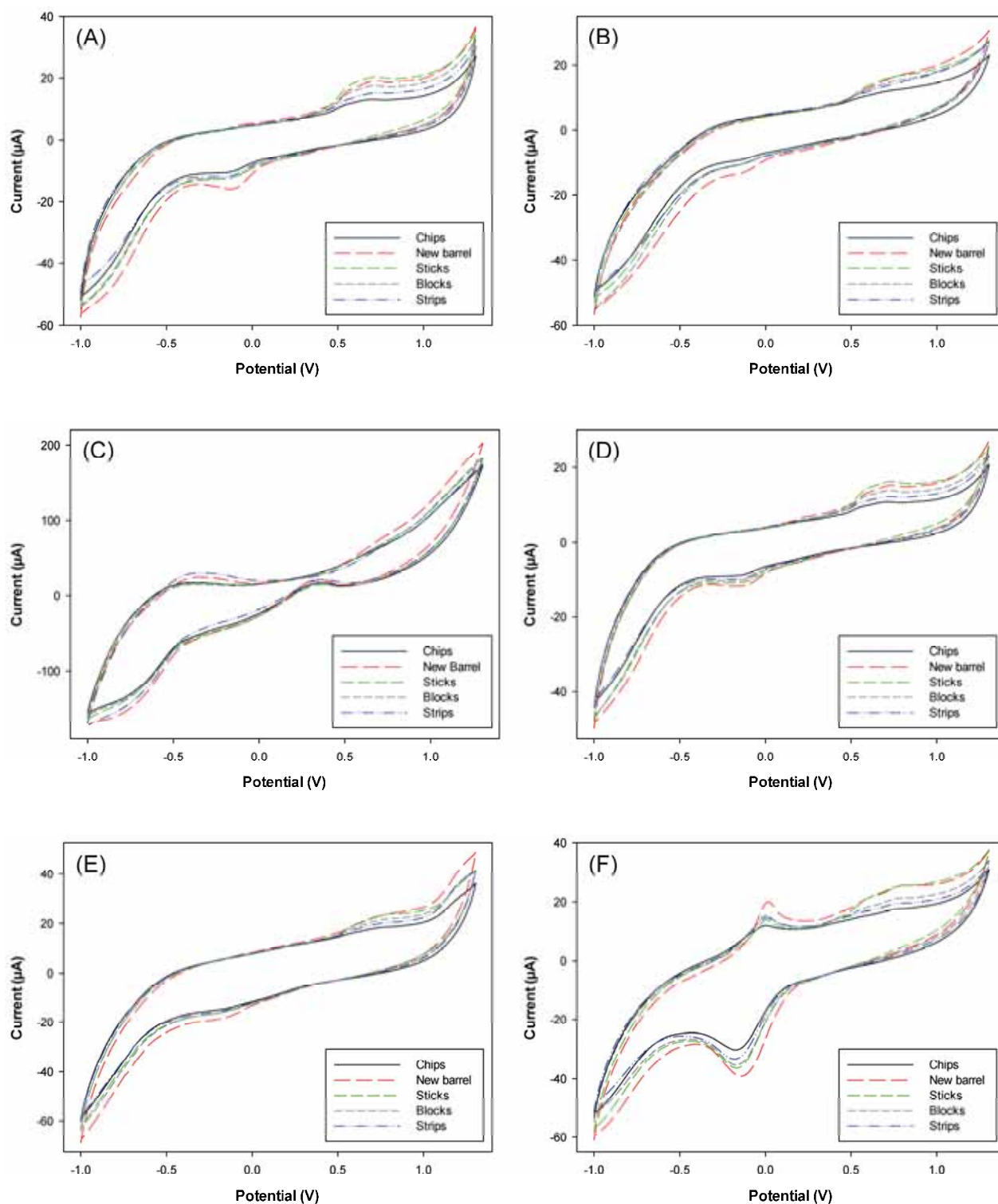


Figure S2. Example of the different obtained voltammograms for the second data set (section 2.2.2) obtained with the sensor array: (A) graphite-epoxy, (B) phthalocyanine, (C) platinum nanoparticle, (D) silver nanoparticle, (E) polypyrrole and (F) copper nanoparticle.

Article 8

Comparison of methods for the processing of voltammetric Electronic Tongues data

X. Cetó, F. Céspedes and M. del Valle

Microchimica Acta, **180**(5-6): 319-330

Comparison of methods for the processing of voltammetric electronic tongues data

Xavier Cetó · Francisco Céspedes · Manel del Valle

Received: 27 September 2012 / Accepted: 28 December 2012 / Published online: 22 January 2013
© Springer-Verlag Wien 2013

Abstract We are making a numerical comparison of various preprocessing strategies for dealing with data from voltammetric electronic tongues in order to reduce the high dimensionality of the response matrices. Different modelling tools are presented and briefly described. We then compare combinations of four preprocessing strategies (principal component analysis, fast Fourier transform, discrete wavelet transform, voltammogram-windowed slicing integral) with four modelling alternatives (principal component regression, partial least squares regression, multi-way partial least squares regression, artificial neural networks) by employing data from a voltammetric bioelectronic tongue, an array formed by enzyme-modified biosensors and applied to the discrimination and quantification of phenolic compounds.

Keywords Electronic tongue · Artificial neural network · Partial least squares regression · Multivariate data analysis · Data preprocessing · Voltammetric sensors

Introduction

In the recent years, there has been a demand of new analytical methods with the high sensitivity and selectivity, and fast response needed to meet new challenges in environmental monitoring, food safety and public health. In this fashion, industry is increasingly focused on fast-response and low-cost methods, as those based on chemical sensors, that may

Electronic supplementary material The online version of this article (doi:10.1007/s00604-012-0938-7) contains supplementary material, which is available to authorized users.

X. Cetó · F. Céspedes · M. del Valle (✉)
Sensors and Biosensors Group, Department of Chemistry,
Universitat Autònoma de Barcelona, Edifici Cn,
08193 Bellaterra, Barcelona, Spain
e-mail: manel.delvalle@uab.es

be used for screening or detecting any adulteration or contamination of the products, either during or after its production, or to assess they guarantee quality control standards [1, 2].

In this sense, classical research lines in the field of chemical sensors have focused on the development of ever more selective devices towards a particular species, and sensitive to lower concentrations at the same time. Unfortunately, there are few optimally operating chemical sensors that function without any interference or matrix effect in the required conditions.

The difficulty to obtain sensors with appropriate selectivity and sensitivity for a given analyte has led to the appearance of new working strategies. Accordingly, from the late 1990s a new concept in the field of sensors has appeared to tackle these problems: Electronic Tongues (ETs) [3]. These analytical systems are inspired by the sensory ability of taste in mammals, where a few receptors can respond to a large variety of substances. This principle is coupled with complex data treatment as applied in the brain, which allows to quantify or to classify a large amount of substances. These biomimetic systems, opposed to conventional approaches, are directed towards the combination of low selectivity sensors array response (or cross response features) in order to obtain some added value in the generation of analytical information.

According to the agreed IUPAC definition [4], an Electronic Tongue is “a multisensor system, which consists of a number of low-selective sensors and uses advanced mathematical procedures for signal processing based on Pattern Recognition (PARC) and/or Multivariate data analysis [artificial neural networks (ANNs), principal component analysis (PCA), etc.]”. Hence, these systems consist of the coupling of an array of sensors capable of giving a wide and complete response of the analyzed species, plus a chemometric processing tool able to interpret and extract meaningful data from the complex readings, relating them with their analytical meaning [5]. Moreover, the data processing stage may offset any matrix or interference

effect from the sample itself or allow the resolution of the interferences, drifts or non-linearity obtained with the sensors [6]. Thus, with this methodology, it is even possible to achieve a parallel determination of a large number of different species, while interference effects are solved by counterbalancing their presence [7].

Until now, ETs have provided alternatives to many classical methods, especially in the food analysis field [2, 5, 8]. Besides the wide variety of samples which they are applied to, there are two key parameters that determine the information obtained; i.e. the modelling tool and the type of sensors used.

On that account, data analysis and pattern recognition, are a fundamental part of any sensor array system [9]. This computer modelling stage is usually realized in two steps. First, the data is preprocessed in order to make it independent from units, remove redundant information and to enhance signal-to-noise ratio. Following this, the model describing the relation between readings and outputs is then created. Among the various numeric procedures that can be implemented, Principal Component Analysis (PCA), Partial Least Squares (PLS) and Artificial Neural Networks (ANNs) are the most widely used for ETs applications [2]. Concretely, PCA is the most common one being used either as a qualitative visualization tool or as a preprocessing step; whereas most advanced qualitative modelling may be achieved with the use of PLS-Discriminant Analysis (PLS-DA), Linear Discriminant Analysis (LDA) or Support Vector Machine (SVM). While in the case of quantitative applications, easier models can be built either using Multiple Linear Regression (MLR) or Principal Components Regression (PCR) [10], although better results can be achieved by using more powerful methods such as PLS or ANNs. Lastly, the new trends in data analysis are mainly related to the use of trilinear approaches such as PARAFAC or Tucker models for qualitative analysis, and Multi-Way PLS (*n*PLS) for quantitative models [11].

On the other hand, among the different families of chemosensors that may form the ET, potentiometric, voltammetric or impedimetric sensors are the main variants used [5, 12]. Furthermore, it is even possible to combine the responses obtained from different sensors families, a strategy known as data fusion [13]. Nevertheless, in this work we will focus on voltammetric Electronic Tongues, one of the most complex cases, given the high dimensionality of the considered signals which hinders its treatment, usually requiring the use of a preprocessing stage as stated below.

Therefore, by using ETs it is possible to provide multicomponent analytical information in real time with a direct, relatively simple measuring setup and low cost. In the same way, it is feasible to overcome the difficulties of analyzing the sensors' raw data by building calibration computer models. Some previous studies dealing with the preprocessing of data from voltammetric ETs have been reported, although they are

mainly focused on classification tasks rather than quantification [14–17]. On that account, in this work we will focus on the different preprocessing strategies that can be used to reduce its dimensionality (i.e. compression of the original signal) and the different multivariate calibration methods (i.e. modelling sensors response) found in the literature. Finally, in order to illustrate the differences between different preprocessing options (e.g. PCA, FFT or DWT) and modelling methods (i.e. PCR, PLS, *n*PLS and ANN), a whole comparison of the performance of the proposed strategies was applied to the discrimination and quantification of phenolic compounds mixtures with data collected from a voltammetric biosensors array.

Experimental

Data preprocessing - compression

A known problem when voltammetric sensors are used is the large dimensionality of the generated data which hinders their treatment. That is, when a complete voltammogram is recorded for each sensor from the array. This is perhaps the main reason why this approach has been rarely used in the literature; especially if ANNs are to be used, in which case departure information needs to be preprocessed. Although signal preprocessing is not always strictly necessary (e.g. the whole data set may be employed in the case of PLS or PCR), it has been demonstrated that even in these cases, its use improves model predictive behaviour [18, 19]. Hence, this stage is necessary given the sensor array contains complex voltammetric responses encompassing global information about the compounds analyzed through the entire voltammetric curve. Despite differentiated response in terms of specificity being found with conventional amperometry treatment (currents measured at a selected potential), significant result improvements can be obtained by using the whole curve instead of particular peaks [20].

The straightforward solution is the use of multi-way processing methods (samples \times sensors \times polarization potential) like *n*PLS, but the intricacy of the approach is also critical [21, 22]. While still another option is the unfolding of original data into a 2D array [23], and the use of standard linear or non-linear chemometric calibration tools such as Principal Component Regression (PCR), Partial Least Squares regression (PLS) or Artificial Neural Network (ANN) models, this approach is normally avoided, given the distortions it introduces in the data.

In this sense, an attractive solution when dealing with a set of voltammograms is to employ a preprocessing stage for data reduction prior to modelling (Fig. 1). The main objective of this step is to reduce the complexity of the input signal preserving the relevant information and making it

compatible with ANNs or other numeric modellers, which facilitates an advantageous reduction in training time, to avoid redundancy in input data and to obtain a model with better generalization ability [24]. This compression stage may be achieved by the use of methods such as feature selection [25], Principal Component Analysis (PCA) [26], “kernels” [27], Discrete Wavelet Transform (DWT) [28] or even Fast Fourier Transform (FFT) [29].

Feature selection

The simplest solution in these cases is the mere selection of the data points that most contribute to the analyte signals and their inclusion into the modelling stage. Even though using the whole voltammetric curve always conveys useful information, there are advantages in identifying and removing the less-relevant inputs even if a large ANN or any other model is well trained: better modelling accuracy and higher robustness than the initial model; knowledge acquisition for the identification of the important inputs; larger ratio of examples to model connection weights (thus a better generalization capacity); ease of model comprehension [24]. Thus, better models are obtained with a reduced input set. For these reasons, different strategies for this feature selection, that attempt to capture the most important data points, are presented.

- *Knowledge-based*: in some cases through accumulated experience of the analyst, and specially when working with biosensors or some modified sensors which present very specific peaks, the potential in which voltammetric response appears has been previously identified. Thus, the main relevant information along the voltammetric curve is focused in these points, and by including them to the modelling stage, reduction of the data dimension can be achieved. Despite the favourable results obtained in this way, this procedure is not the

best choice given that it requires prior knowledge of the system, selection of the inputs is very subjective, it can not be ensured that all the information contained on the raw data is included and the more relevant points cannot always be easily identified.

- *Statistical parameters/Stepwise inclusion methods*: in the same sense, selection of the relevant data points that must be included in the modelling stage can be achieved by selecting them according to some statistical parameters such as: correspondence analysis (evaluation of correlation between the data points and removal of the correlated ones), factor analysis extraction [30], genetic algorithms [31], etc.
- *Causal Index analysis (CI)*: another possibility for feature extraction is the usage of ANNs as feature selection tools [32]. The use of this method for variable selection is the location of an optimal set of inputs that can successfully classify or predict the desired outputs. This is a feed-forward network that combines a back-propagation algorithm which is used to train the network according to a learning rule [33]. Therefore, to carry out the dimensionality reduction and feature selection, the following steps are followed: (1) First of all, the ANN architecture should be selected prior to perform modelling steps. For this, the number of neurons in the hidden layer and the transfer functions between the layers are selected in advance to represent the intrinsic dimensionality of the dataset. This represents a limit on the number of hidden neurons in the initial ANN. Based on the experience and literature, the suggested number of neurons in the hidden layer should be already very close to the optimal size before the network reduction procedures are applied, usually five neurons is a good number. Thus, most of the effort is directed at reducing the number of network inputs since the number of outputs is fixed by the requirements of the model. (2) Afterwards, it comes the determination of

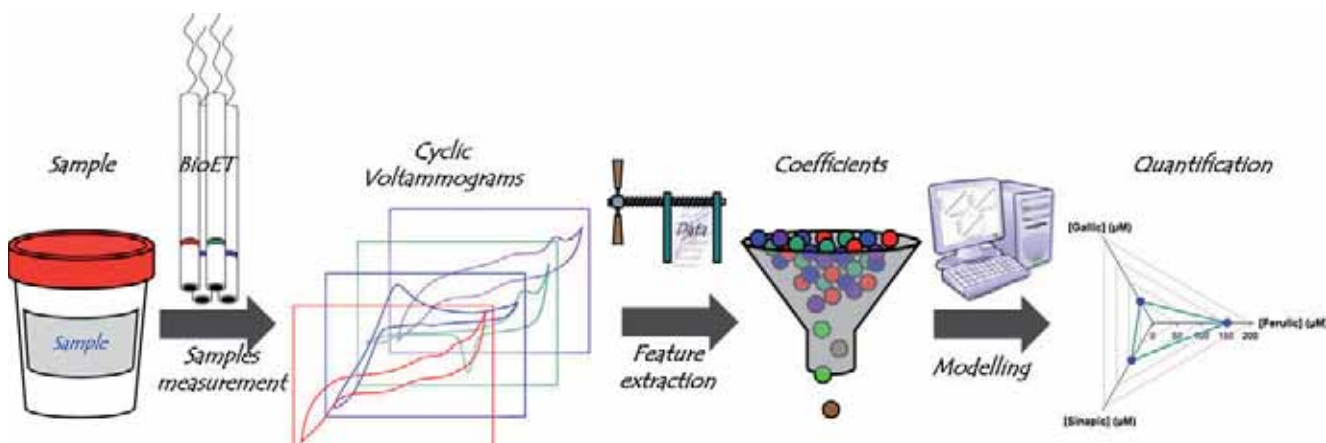


Fig. 1 Processing scheme for building the quantification model

the relevance of network inputs. Once the ANN model is trained, the analysis of its connection weights can easily identify the important inputs. In this manner, the relative contributions of the inputs to the variance of the output layer in the trained network are calculated [24]. Inputs that make relatively small contributions to the variance can be discarded, viz. pruned. These small contribution values indicate that either the input does not change significantly or that the training of the ANN has allocated small connection weights to scale the input. (3) Repeating the training process of the ANN model with the reduced input set and selecting the most relevant inputs can proceed until a near-optimal, small, set of inputs is identified [24, 32]. Consequently, better models are built with the reduced input set, and this process is repeated until the model accuracy is degraded. A known drawback is that some of the resulting minimal sets are not unique, which will depend on the ANN architecture used in the training.

Voltammogram windowed slicing integral

This method considers the area under the voltammetric response curve, which can also be interpreted as the scaled average of the data points [27]. Thus, in this case the voltammogram is divided into k sections and the area under each section is used as input data point to the modelling stage.

The idea behind this compression technique is to capture some information about the voltammetric waveform characteristics. The number of sections (k) into which voltammogram is divided is arbitrary and it depends on the response profile of the data; by increasing k , the original information in the voltammetric data is best reproduced, but at expenses of the compression ratio. When the k value is close to the number of points in the original signal, the obtained coefficients are just a scaled value of the original signal; that is, the original value multiplied by a constant factor. It is for this reason that a compromise between the compression ratio and signal reproduction must be taken into account.

Principal Component Analysis (PCA)

A viable option for data reduction is the use of PCA as feature extraction tool. This is a useful procedure able to reduce the dimensionality of large data sets, such as those from voltammetric sensor arrays; and it might also be used to find signals in noisy data.

PCA allows for the projection of the information carried by the original variables onto a smaller number of underlying (“latent”) variables called principal components (PCs) with

new coordinates called scores, obtained after data transformation. The first PC covers as much of the variance in the data as possible; the second principal component is orthogonal to the first and covers as much of the remaining variation as possible, and so on. Then by plotting the PCs, one can view interrelationships between different variables, and thus detect and interpret sample patterns, groupings, similarities or differences [34].

As in the previous cases, there is a compromise in the number of extracted features (PCs in this case) used as inputs into the modelling stage. In this case, the accumulated variance could be used as a measure of the reconstruction degree, obtaining an indicator of the number of PCs that must be taken into account.

Fast Fourier Transform (FFT)

FFT is an efficient algorithm to compute the Discrete Fourier Transform (DFT) much faster; the reason might be explained by the fact that computing it directly from the definition is often too slow to be practical, thus with the use of FFT, the difference in speed can be substantial, especially for long data sets (in practice, the computation time can be reduced by several orders of magnitude) [35]. DFT is an important tool in digital signal processing which decomposes a sequence of values into components of different frequencies.

The processing of the signals using Fourier transform could be considered as the decomposition of the signal using a sine/cosine function pair at different frequencies, calculating a coefficient for each one taking into account its contribution to the original signal. When DFT analysis is applied for signal decomposition, an equal number of coefficients and points from the original signal are obtained, where every contribution represents the information of frequency bands contained in a specific segment of the analysed signals. As in the case of PCA, the most relevant information is kept by the first coefficients, ergo good signal reconstruction could be achieved just preserving the first coefficients and discarding the rest. In this treatment, the Nyquist theorem must be recalled, limiting the highest frequency that carries significant information to the discretization interval [36]. Moreover, an additional advantage of this approach is the facility to remove the noise present in signals thorough the removal of the high frequency contributions, normally related to it [37]. Finally, the number of coefficients to be kept can be optimized through the inverse Fourier transform, which can be applied to evaluate the degree of signal reconstruction.

Discrete Wavelet Transform (DWT)

The Wavelet transform (WT) is a high performance signal processing technique developed from the Fourier Transform during the late 1980s [38]; instead of projecting the signal

down to a cosine or sinus wave, the WT projects the signal down to a wavelet function, with the key advantage over FFT being its temporal resolution: it captures both frequency and location information (location in time). In the last decade, numerous applications of the WT have been proposed for chemical analysis, including denoising and smoothing, data compression, baseline correction, resolution of overlapping signals, electroanalytical chemistry and chromatography [39].

WT is used for signal decomposition onto a set of basis functions. These functions are obtained by dilations and translations of a unique function called mother wavelet [40], *Daubechies* wavelets being the most commonly used set of discrete wavelet transforms [37]. Thus, WT decomposes the signal in labelled and located contributions for one scale and one position parameter. This is implemented using Mallat's pyramidal algorithm [38], which operates over a single discrete signal of length M by decomposing it into orthogonal subspaces of length ca. $M/2$ in each step [28]. In this way, increasing the decomposition level (which means repeating the decomposition process) k times, the signal compression ratio can be increased. Nevertheless, the compression ratio is enhanced at the expense of signal reconstruction, given signal reconstruction accuracy is degraded for each step.

Also, as in the case of Fourier transform, DWT presents the additional advantage of its ability to easily decrease the noise present in signals. Furthermore, WT can be used for data compression if only the most significant decomposition coefficients are kept for further treatment. WT has shown high efficiency in this situation, since it offers the advantage of performing data compression, feature extraction and noise reduction at the same time [28]. As in the previous case, this transformation can be undone to evaluate the signal reconstruction degree, which permits to optimize the number of inputs.

Data processing – modelling

After applying the desired preprocessing method, the next step is the modelling of the target variables using the obtained coefficients as inputs into the modelling stage. This modelling stage may be achieved through the use of linear methods such as MLR [18], PCR [41], PLS [21] or nPLS [21] or with non-linear ones such as ANNs amongst others [7]. This sorting takes into account the form in which regression analysis is done; that is, in the case of non-linear models, data is modelled by a function which is a nonlinear combination of the model parameters and depends on one or more independent variables, fitted by a method of successive approximations. While in the case of linear models, data is fitted using linear regression, usually by means of least squares approach. In the next sections, basic principles and discussion of the idea behind each modelling method are introduced.

Principal Component Regression (PCR)

PCR is a method for relating the variance in a target variable (y ; usually the expected concentrations) to the variance of several predictors (x ; response matrix of the ET array), for explanatory or predictive purposes. PCR is a two-step procedure which first decomposes the data response matrix with the use of PCA, and then fits a Multiple Linear Regression (MLR) model, using the PCs scores instead of the original response variables as predictors [34].

MLR is a classical regression method that combines a set of several predictors or response variables, in linear combinations, which correlate as closely as possible to a corresponding single response [10]. Thus, even the results for several predictors can be obtained in one run, these are based on individually built separate models. The same limitation is obtained in the case of PLS-1, where only 1 response variable could be modelled at once. In this sense, PCR could be considered as an alternative to PLS-1 [10].

This method performs particularly well when the various response variables express common information, i.e. when there is a large amount of correlation, or even collinearity [42]. Since the scores are orthogonal, the MLR solution is stable and therefore the PCR model does not suffer from collinearity effects. For its optimization, only the number of PCs used for the MLR regression must be taken into account, thus selecting those that lead to the lower RMSE achieved by the model for predicted values.

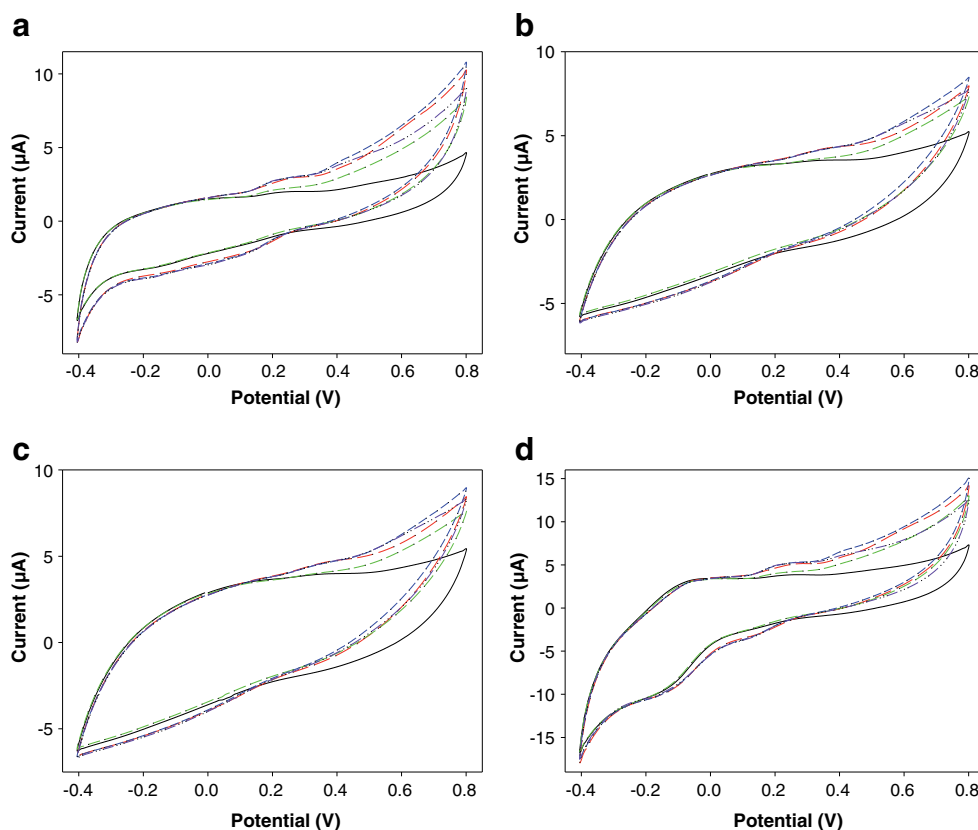
Partial Least Squares regression (PLS-2)

PLS is a very similar linear transformation procedure to PCR; it models both the ET response (x) and target (y) matrices simultaneously to find the latent (or hidden) variables in x that will best predict the latent variables in y . These PLS components are similar to PC's, but will be referred to as latent variables or factors [43]. Indeed, instead of finding hyperplanes of maximum variance between the response and independent variables, it finds a linear regression model by projecting the predicted variables and the observable variables to a new space. Because both the x and y data are projected to new spaces, the PLS family of methods are known as bilinear factor models.

The number appended at the end of the acronym identifies the number of modes of the responses, which in the case of PLS it is restricted up to 2 [23]. That is, PLS1 where one block of data consists of a single variable or PLS2, being both blocks multidimensional arrays. Accordingly, unlike PCR or PLS-1, PLS-2 analysis could be carried out with one or more y variables, meaning that multiple y responses can be modelled at once during regression stage.

PLS simultaneously decomposes x and y into a set of latent variables that maximizes the covariance between the

Fig. 2 Raw transient response of the 4-voltammetric sensors array: **a** graphite–epoxy sensor, **b** tyrosinase biosensor, **c** laccase biosensor and **d** copper nanoparticle modified sensor. Signals obtained for different mixtures of the three phenols are shown



independent and dependent data. In this case, convergence of the system to a minimum residual error is often achieved in fewer factors than using PCR. As before, only the number of latent variables must be optimized in order to get the lowest RMSE.

Multi-way Partial Least Squares Regression (nPLS)

This method is an extension of standard PLS regression methods and can be used to regress on data structures of a varying number of dimensions (multiway data or tensor) [11, 22]. As previously stated, this method has the advantage that it can relate the variations in one of several response variables (y) arranged in a 2D table to the variations of several predictors arranged in a 3D table (primary and secondary x variables), with explanatory or predictive purposes. Thus, it overcomes the disadvantages caused by the unfolding process, that is, the

discontinuities created from one recording to the next, which causes the loss of the original relationship and creates limitations for first-order calibration [44].

As an extension to two-way PLS regression, n-PLS regression builds a calibration model incorporating a relationship between the sets of predictors and responses based on the multiway structure of the arrays (tensor). The main difference between n-PLS compared to two-way PLS regression is the use of loading matrices of higher dimensions for decomposing the multiway array (tensor). The goal of the methodology is to simultaneously fit multilinear models formed by score and loading vectors from both predictors and response arrays, along with a regression model relating the two decomposition models. According to the underlying theory of PLS regression, the model is fitted such that the score vectors have maximum covariance with the unexplained part of the dependent variable [23].

Table 1 Number of features used as input in the modelling stage and compression ratio for the different preprocessing techniques

| | Number of features | Number of inputs | Compression ratio | Explained variance | R ² | fc |
|----------|--------------------|------------------|-------------------|--------------------|----------------|-------|
| CI-ANN | – | 25 | 97.7 % | – | – | – |
| Integral | 11 | 44 | 95.9 % | – | – | – |
| PCA | 4 | 16 | 98.5 % | 99.9 % | – | – |
| FFT | 16 | 64 | 94.0 % | – | 0.999 | 0.975 |
| DWT | 23 | 92 | 91.4 % | – | 0.997 | 0.969 |

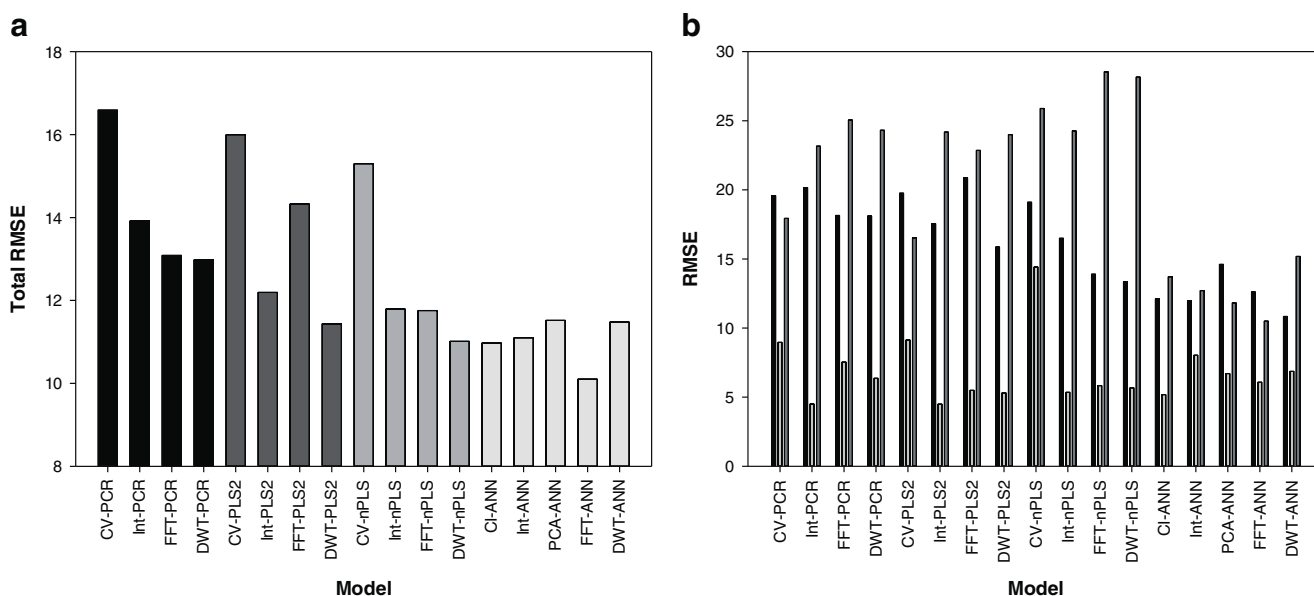


Fig. 3 Obtained RMSE values (expressed in μM) for the testing subset and the different models constructed. **a** Total RMSE and **b** individual RMSE for each of the considered phenols: (■) ferulic, (▣) gallic and (●) sinapic acids

In addition to this, models obtained by applying first-order methods to data whose multiway structure is ignored can be less robust and hard to interpret, on the opposite to models based on multiway methods, which also yields better predictions and possibilities for exploring the data [22, 45].

Artificial Neural Networks (ANNs)

ANNs were originally designed to mimic the function of the human brain and applied to quantitative and qualitative analysis during the last decades [46]. They consist of a number of simple processing units (or neurons) linked by weighted modifiable interconnections [47].

Compared with previous presented methods, ANNs are more flexible modelling methodologies, since both linear and non-linear functions can be used (or combined) in the processing units, thus they are specially suited to being used with non-linear sensor responses. This also allows for more complex relationships between a high-dimensional descriptor space and the given retention data, and may lead to better predictive power of the resulting ANN model compared with other linear methods.

However, the major disadvantage of ANNs is also directly related to the complex model infrastructure, which requires some previous knowledge for the selection of the topology of the network and significant effort must be reckoned on the optimization due to the difficulties in predicting the optimum ANN configuration in advance [48]. This is a trial-error process where several parameters (training algorithms, number of hidden layers, transfer functions, etc.) are fine-tuned in order to find the best configuration that optimizes the performance of the neural network model. Moreover, as previously

discussed, before building the ANN model a preprocessing step for reducing the input dimensionality is usually required.

Software

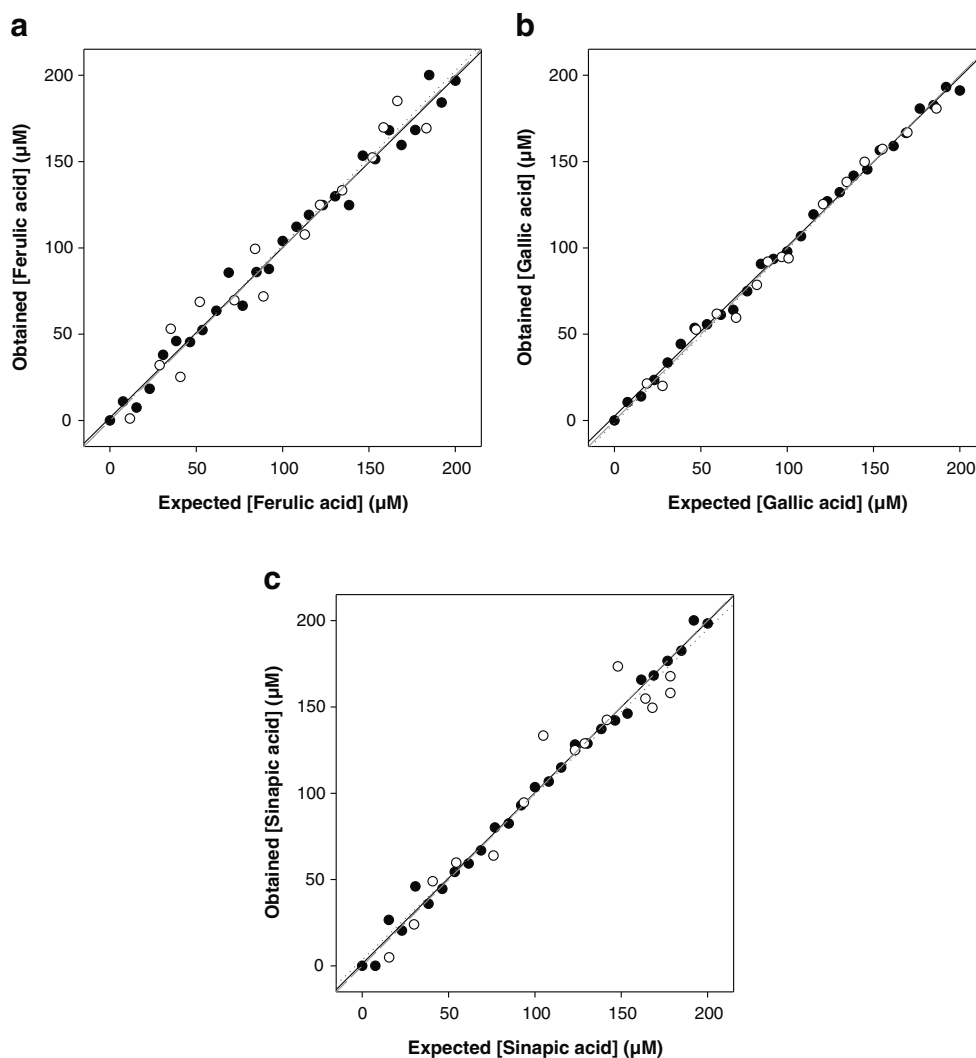
Chemometric processing was done by specific routines in MATLAB 7.10 (MathWorks, Natick, MA) written by the authors, using Neural Network Toolbox (v6.0.4), Wavelet Toolbox (v4.5) and Statistics Toolbox (v7.3). Sigmaplot 2000 (Systat Software Inc, California, USA) was used for graphic representations of data and results.

Results and discussion

In order to illustrate the differences between the various preprocessing and modelling methods, their performance was evaluated in the discrimination and quantification of polyphenol mixtures with data collected from a voltammetric BioElectronic Tongue employing an array of enzymatic biosensors [25] (details on sensors preparation can be found in Supporting Information).

The study case was the resolution of three phenols mixtures formed by ferulic, gallic and sinapic acids, three antioxidant compounds usually found in beers [49]. To this end, a total set of 42 samples were manually prepared with a concentration range for the three species from 0 to 200 μM for each polyphenol. The set of samples was divided into two data subsets: a training subset formed by 27 samples (64 %), which were distributed in a cubic design and used to establish the response model [25]; plus 15 additional samples (36 %) for the testing subset, distributed randomly along the experimental domain,

Fig. 4 Modelling ability of the optimized CI-ANN model. Comparison graph of expected vs. obtained concentrations for **a** ferulic, **b** gallic and **c** sinapic acids, both for training (●, *solid line*) and testing subsets (○, *dotted line*). Dashed line corresponds to theoretical diagonal line



and used to evaluate the models predictive ability. In this way, the set of samples was measured employing the 4-sensor array, recording a cyclic voltammogram for each sensor. Examples of the different obtained signals for some of the phenolic mixtures are shown in Fig. 2. As can be seen, differentiated signals were observed for the different sensors, with complex and overlapping voltammetric responses that contain global information about the compounds analyzed along all the voltammogram. However the amount of information and the dimensionality of the signals (4 sensors x 268 current values x 42 samples) generate great difficulties in data treatment. Therefore, to fully exploit all the information obtained from each voltammogram, a compression step is required.

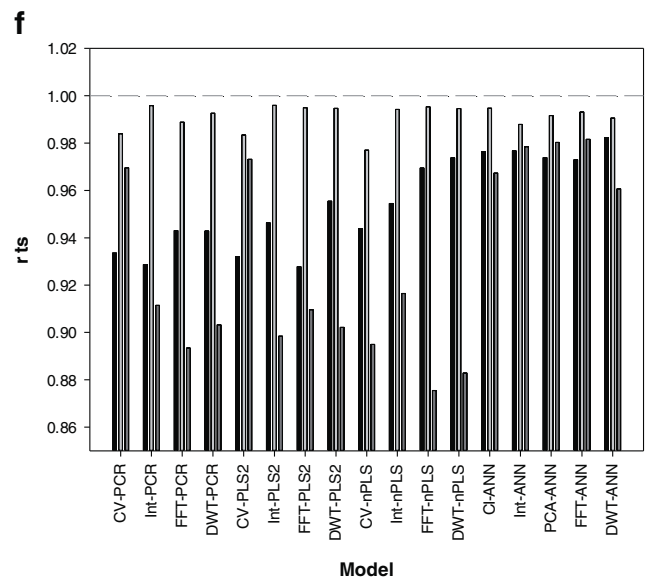
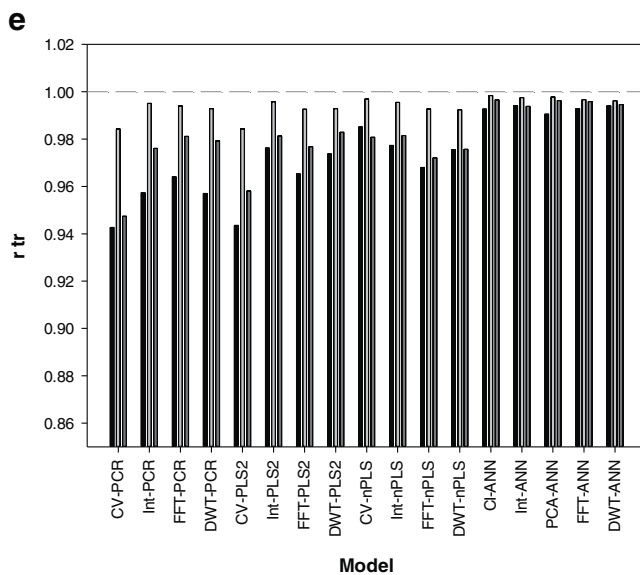
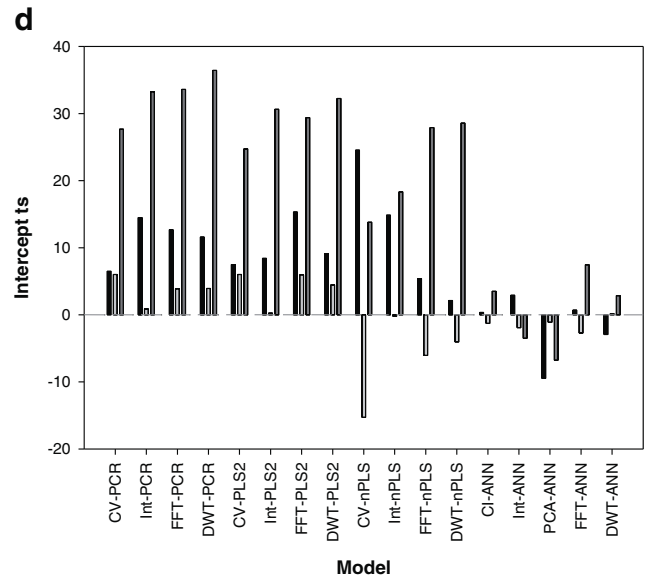
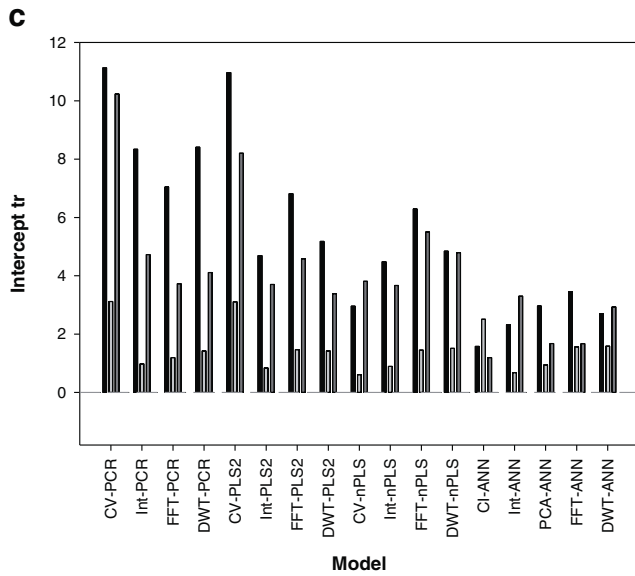
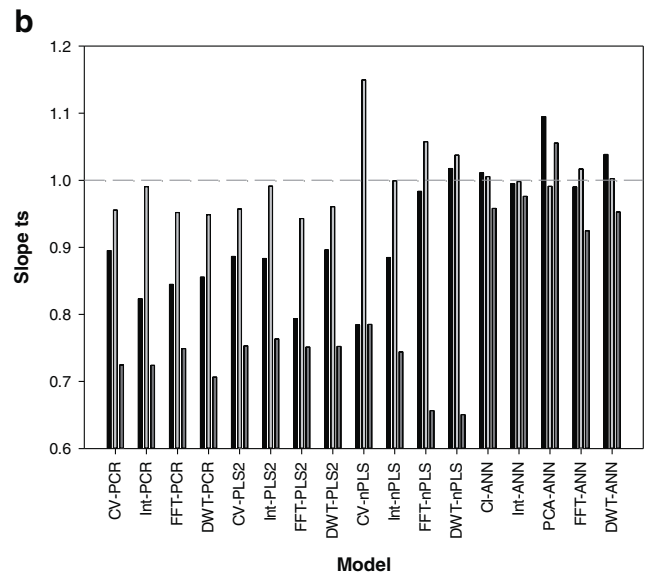
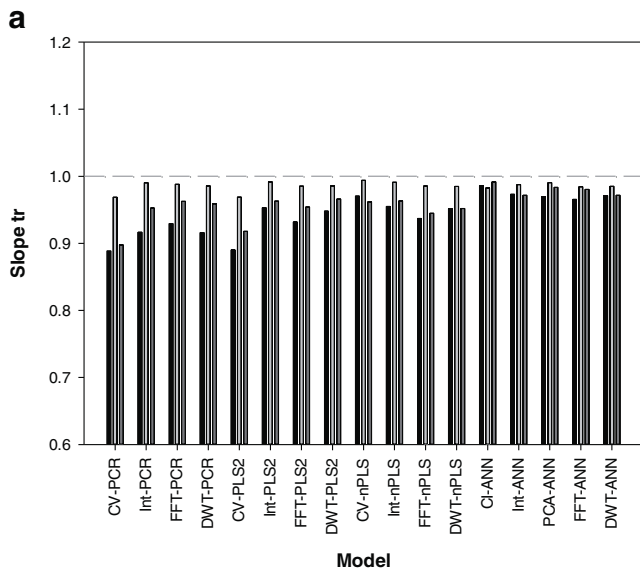
Data preprocessing - compression

In order to quantify the reconstruction degree in the cases where the signals could be reconstructed from the obtained coefficients, the coefficient of determination (R^2) and a

comparison factor named f_c , that considers the area under both signals when superimposed, where used [28]. The f_c is defined as the ratio of the area intersected by both curves to the total area under both curves. From the set theory, with A and B as the areas under each curve, f_c can be expressed as $f_c = (A \cap B) / (A \cup B)$. This factor ranges from 0 to 1 depending on signals similarity; it values 0 when two signals have nothing in common and increases its value proportionally with similarity. In this sense, f_c computes similarity in a way related to a correlation coefficient R, but being more sensitive to small differences.

Table 1 summarizes the results obtained after applying the different preprocessing methods. In the different cases, parameters of the algorithms were selected to ensure that practically all the information contained in the original data

Fig. 5 Fitted regression lines of the comparison between obtained vs. expected results attained with the different models for the three considered phenols (■ ferulic, ■ gallic and ■ sinapic acids), both for training and testing subsets respectively. **a, b** Slopes, **c, d** Intercepts (μM) and **e, f** correlation coefficients



is preserved, as demonstrated in the case of PCA, FFT and DWT. Moreover, it should be noted that in all cases compression ratio was above 90 %, such a huge value will allow a reduction in training time and model complexity, as well as obtaining models with better generalization ability.

Beyond the compression ratio provided by each method, some of the advantages and disadvantages of each method should be taken into account and thus, depending on the case select the most appropriated one. For example, the use of CI is quite a time-consuming procedure, nevertheless it is a useful way to identify the significant features on the response profile; namely, which sensors provide more significant information and which potentials are associated with the electrochemical processes taking place. In the case of windowed slicing integral, the number of sections is arbitrary, thus in our case a value of $k=11$ was selected, despite good results were obtained even with less coefficients. This value was preferred given it was more comparable to the other methods, where a similar number of coefficients was used. In this sense and depending on the case, this parameter can be changed in order to improve the detail representation of the original signal or to reduce the number of inputs. The use of PCA is another simple way to reduce the high-dimensionality of the input signal, despite this method not being recommended in cases where there is some drift on the measurements given it is very sensitive to it, thereby providing coefficients more related to the drift than to samples information [50]. Also, its usage is not recommended in all the cases, given it focuses on the maximum variance between samples, hence giving less importance to small differences which in some cases could be significant. That is, if in our set we have a sample that is very different from the rest, that variance would not allow us to see the smaller differences between the rest of the samples (in this case, can't see the trees for the forest). Finally, even though it is not so clear in this case, it must be acknowledged that in complex and highly overlapped responses, some methods like FFT or DWT are more efficient [51]. The reason is due to the fact that the obtained coefficients are a representation of the entire voltammetric signal, which is also denoised during the feature extraction process.

On the other hand, it must be considered that the good results (large explained variance with only a few PCs) obtained with the PCA analysis are mainly due to the type of sensors used and the study case. That is, in this application the voltammetric ET was formed by an array which incorporated some specific biosensors selective towards phenolic compounds, and in addition synthetic samples were used. Thus, sensors' response is clearly related to desired compounds concentration and could be more easily identified at specific regions along the voltammogram. In this situation, some research groups [52] claim that the use of only one half of the voltammetric wave is enough given that following a reversible

electrochemical process, the same information could be obtained either from the cathodic or the anodic waves. In this manner, the described preprocessing methods could be further simplified by just focusing on one part of the voltammetric response, either the cathodic or anodic component, or even a more restricted region.

Data processing - modelling

Applying the different preprocessing strategies to the raw data obtained from the ET sensors array, the different coefficients were obtained, which were used as inputs in the modelling stage. In the case of linear methods (PCR, PLS and nPLS), models were also built using the whole voltammetric curve as input; this was not done in the case of the ANN given this method is more sensitive to the number of inputs, consequently obtaining less robust models [48]. In this case, CI analysis was applied, which is based on the use of the ANN as a feature extraction tool rather than as a quantification tool. Also, given the use of PCA is an implicit step of PCR when building the models, it was not used in this case; likewise, for the similarity that would be between PCR and PLS models if it was used, in the case of PLS, it was also discarded in the latter.

After building all the models using the different obtained coefficients as inputs and training them with the training subset, the models' accuracies were evaluated using the samples of the external test subset; the results obtained were also used to carry out the selection of the optimal model in each case (i.e. ANN architecture, number of PCs or latent variables). Once the modelling was completed, its Root Mean Square Error (RMSE) value was calculated and an overall comparison was performed, as shown in Fig. 3 and Table S2. From this, it can be seen that larger errors are obtained when using the whole raw voltammetric data, with similar results for the rest of the cases. Also, it could be seen that, in general, best results are obtained when using DWT as a preprocessing method for the different modelling tools. Moreover, it should be noted that the good results that were obtained employing the *windowed slicing integral*, which represents the implementation of a new and simple preprocessing method. As stated above, it has the disadvantage that the selection of coefficients obtained is arbitrary, thus requiring some previous knowledge or trials to ensure a good departure point. Despite this, it has been proved to be a good way to reduce signal dimensionality, its simplicity being its main advantage.

In addition to RMSE values, comparison graphs of predicted vs. expected concentration for each phenolic compound were built to check the prediction ability of the obtained models, both for train and test subsets. Then slope values and correlation coefficients of the comparison were taken into account in order to help in selecting the best configuration. Therefore, the optimum model would have a small RMSE,

meaning there are no differences between the values predicted by the model and the real ones. Also the comparison graphs should display a correlation value and slope close to 1 and an intercept close to 0, discarding any model error.

One example of these graphs is shown in Fig. 4. It may be seen that a satisfactory trend is obtained in all the cases, with regression lines almost indistinguishable from the theoretical ones. Also, as usual in multivariate calibration models, lower dispersion and uncertainties are obtained for the training subsets.

In this manner, comparison graphs were built for all cases; the obtained regression parameters are summarized in Fig. 5, Table S1 and Table S2. As said, it could be seen that better trend is obtained in all the cases for the training subset, where slopes and correlation coefficients are close to 1 and intercepts closer to 0, with small differences between the various preprocessing and modelling methods. However, despite the low and similar RMSE values obtained between the different modelling methods for the testing subset, higher differences are attained with the comparison graphs. Firstly, it could be seen that sinapic acid is the one presenting the worst models' results in all the cases, implying that its signal is masked by the other compounds.

Leaving that aside, in all cases it could be seen that correlation is still good ($r > 0.875$ in the worst case), but the main difference can be found in the slopes and intercepts (values down to ca. 0.65 and up to 36.5 μM respectively), where a clear decline is obtained particularly with linear models. In this sense, ANNs are the ones presenting the best results with slopes higher than 0.92 and intercepts under 9.5 μM in the worst situation (PCA-ANN).

Hence, for the selection of the best model, both RMSE and comparison graph results must be considered. In this case, the best results were obtained employing FFT-ANN. While in the case of linear models, best results were obtained employing DWT preprocessing of the signal. As stated, these improvements are due to both methods providing a good compression of the data, preserving relevant information and denoising the original signal.

Finally, the good results obtained using *windowed slicing integral method* should be noted, which despite its simplicity proved to be a satisfactory variant. Also, the good performance obtained using CI analysis; although this method has the drawback that the reduced sets are not unique and the arrival to an optimal set depends on the architecture of the ANN used. Moreover, the amount of time required for the feature extraction process is much larger. In spite of this, it allows for the identification of the most relevant sensors and inputs from the voltammograms that contribute to the discrimination and the quantification of the compounds under scrutiny. And consequently allowing the selection of which sensors must be kept for further experiments or the modification of the voltammetric parameters.

Conclusions

In this article, an overview of different preprocessing and processing strategies that can be used when dealing with voltammetric Electronic Tongue data has been presented. In this line, obtained responses were first preprocessed in order to reduce the huge dimensionality of the obtained matrix responses, which allowed for the acquisition of advantages in the modelling stage such as the reduction in training time, avoiding redundancy in input data and obtaining a model with better generalization ability. Afterwards, obtained coefficients were used as inputs in the modelling stage in order to predict the concentration of the desired compounds. Finally, in order to assess the predictive capabilities, RMSE values were calculated and comparison graphs were built, which allowed the selection of the best model.

In order to illustrate the differences between the various methods, the resolution of phenolic antioxidant mixtures was selected as the study case. Numerical results merely present some general trends, given the optimal method could change depending on the study case. However, we want to emphasize that the objective of this article has not been to determine the best technique or set of techniques, but to give an overview of them and to provide the information that would help in the selection of the most suitable one in each case.

Acknowledgments Financial support for this work was provided by *Spanish Ministry of Science and Innovation*, MCINN (Madrid) through projects CTQ2010-17099 and by program ICREA Academia. X. Cetó thanks the support of *Dept. d'Innovació, Universitat i Empresa de la Generalitat de Catalunya* for the predoctoral grant.

References

1. Scampicchio M, Ballabio D, Arecchi A, Cosio SM, Mannino S (2008) Amperometric electronic tongue for food analysis. *Microchim Acta* 163:11–21
2. Ciosek P, Wróblewski W (2007) Sensor arrays for liquid sensing - electronic tongue systems. *Analyst* 132:963–978
3. Vlasov Y, Legin A (1998) Non-selective chemical sensors in analytical chemistry: from “electronic nose” to “electronic tongue”. *Fresenius J Anal Chem* 361:255–260
4. Vlasov Y, Legin A, Rudnitskaya A, Di Natale C, D'Amico A (2005) Nonspecific sensor arrays (“electronic tongue”) for chemical analysis of liquids (IUPAC Technical Report). *Pure Appl Chem* 77:1965–1983
5. del Valle M (2010) Electronic Tongues Employing Electrochemical Sensors. *Electroanal* 22:1539–1555
6. Hossein-Babaei F, Ghafarinia V (2010) Compensation for the drift-like terms caused by environmental fluctuations in the responses of chemoresistive gas sensors. *Sensor Actuat B-Chem* 143:641–648
7. Mimendia A, Gutiérrez JM, Opalski LJ, Ciosek P, Wróblewski W, del Valle M (2010) SIA system employing the transient response from a potentiometric sensor array—Correction of a saline matrix effect. *Talanta* 82:931–938

8. Zeravik J, Hlavacek A, Lacina K, Skladal P (2009) State of the Art in the Field of Electronic and Bioelectronic Tongues - Towards the Analysis of Wines. *Electroanal* 21:2509–2520
9. Riul A Jr, Dantas CAR, Miyazaki CM, Oliveira ON Jr (2010) Recent advances in electronic tongues. *Analyst* 135:2481–2495
10. Martens H, Naes T (1991) *Multivariate calibration*. Wiley, Chichester
11. Bro R (2006) Review on multiway analysis in chemistry - 2000-2005. *Crit Rev Anal Chem* 36:279–293
12. del Valle M (2008) Electrochemical micro(bio)sensor arrays. *Microchim Acta* 163:1–2
13. Gutiérrez JM, Haddi Z, Amari A, Bouchikhi B, Mimendia A, Cetó X, del Valle M (2013) Hybrid Electronic Tongue based on Multisensor Data Fusion for Discrimination of Beers. *Sensors Actuators B Chem* 177:989–996
14. Holmin S, Spångeus P, Krantz-Rülcker C, Winquist F (2001) Compression of electronic tongue data based on voltammetry — a comparative study. *Sensor Actuat B-Chem* 76:455–464
15. Palit M, Tudu B, Bhattacharyya N, Dutta A, Dutta PK, Jana A, Bandyopadhyay R, Chatterjee A (2010) Comparison of multivariate preprocessing techniques as applied to electronic tongue based pattern classification for black tea. *Anal Chim Acta* 675:8–15
16. Scott SM, James D, Ali Z (2006) Data analysis for electronic nose systems. *Microchim Acta* 156:183–207
17. Richards E, Bessant C, Saini S (2002) *Multivariate Data Analysis in Electroanalytical Chemistry*. *Electroanal* 14:1533–1542
18. Palacios-Santander JM, Cubillana-Aguilera LM, Cocchi M, Ulrici A, Naranjo-Rodríguez I, Seeber R, Hidalgo-Hidalgo de Cisneros JL (2008) Multicomponent analysis in the wavelet domain of highly overlapped electrochemical signals: Resolution of quaternary mixtures of chlorophenols using a peg-modified Sonogel–Carbon electrode. *Chemometr Intell Lab* 91:110–120
19. Cetó X, Gutiérrez JM, Gutiérrez M, Céspedes F, Capdevila J, Mínguez S, Jiménez-Jorquera C, del Valle M (2012) Determination of total polyphenol index in wines employing a voltammetric electronic tongue. *Anal Chim Acta* 732:172–179
20. Alsberg BK, Woodward AM, Winson MK, Rowland JJ, Kell DB (1998) Variable selection in wavelet regression models. *Anal Chim Acta* 368:29–44
21. Ebrahimi D, Chow E, Gooding JJ, Hibbert DB (2008) Multi-analyte sensing: a chemometrics approach to understanding the merits of electrode arrays versus single electrodes. *Analyst* 133:1090–1096
22. Bro R (1998) *Multi-way analysis in the food industry. Models, algorithms and applications*. University of Amsterdam
23. Cartas R, Mimendia A, Legin A, del Valle M (2011) Multiway Processing of Data Generated with a Potentiometric Electronic Tongue in a SIA System. *Electroanal* 23:953–961
24. Boger Z (2003) Selection of quasi-optimal inputs in chemometrics modeling by artificial neural network analysis. *Anal Chim Acta* 490:31–40
25. Cetó X, Céspedes F, Pividori MI, Gutiérrez JM, del Valle M (2012) Resolution of phenolic antioxidant mixtures employing a voltammetric bio-electronic tongue. *Analyst* 137:349–356
26. de Carvalho RM, Mello C, Kubota LT (2000) Simultaneous determination of phenol isomers in binary mixtures by differential pulse voltammetry using carbon fibre electrode and neural network with pruning as a multivariate calibration tool. *Anal Chim Acta* 420:109–121
27. Gutiérrez-Osuna R, Nagle HT (1999) A method for evaluating data-preprocessing techniques for odour classification with an array of gas sensors. *IEEE T Syst Man Cy B* 29:626–632
28. Moreno-Barón L, Cartas R, Merkoçi A, Alegret S, del Valle M, Leija L, Hernandez PR, Muñoz R (2006) Application of the wavelet transform coupled with artificial neural networks for quantification purposes in a voltammetric electronic tongue. *Sensor Actuat B-Chem* 113:487–499
29. Cetó X, Céspedes F, del Valle M (2012) BioElectronic Tongue for the quantification of total polyphenol content in wine. *Talanta* 99:544–551
30. Harman HH (1976) *Modern Factor Analysis*, 3rd edn. University of Chicago Press, Chicago
31. Vesin J-M, Grüter R (1999) Model selection using a simplex reproduction genetic algorithm. *Signal Process* 78:321–327
32. Xiaobo Z, Jiewen Z, Povey MJW, Holmes M, Hanpin M (2010) Variables selection methods in near-infrared spectroscopy. *Anal Chim Acta* 667:14–32
33. Smits JRM, Melssen WJ, Buydens LMC, Kateman G (1994) Using artificial neural networks for solving chemical problems: Part I. Multilayer feed-forward networks. *Chemometr Intell Lab* 22:165–189
34. Jolliffe IT (2002) *Principal Component Analysis*. Springer, New York
35. Rader C, Brenner N (1976) A new principle for fast Fourier transformation. *IEEE T Acoust Speech* 24:264–266
36. Griffiths PR, De Haseth JA (1986) *Fourier transform infrared spectrometry*, vol v. 83, 2nd edn. Wiley, Hoboken
37. Tchistiakov V, Ruckebusch C, Duponchel L, Huvenne J-P, Legrand P (2000) Neural network modelling for very small spectral data sets: reduction of the spectra and hierarchical approach. *Chemometr Intell Lab* 54:93–106
38. Mallat SG (1989) A theory for multiresolution signal decomposition - the wavelet representation. *IEEE Trans Pattern Anal Mach Intell* 11:674–693
39. del Valle M, Muñoz R, Gutierrez JM (2012) *Wavelets: Classification, Theory and Applications*. Nova Science, New York
40. Grossmann A, Morlet J (1984) Decomposition of hardy functions into square integrable wavelets of constant shape. *SIAM J Math Anal* 15:723–736
41. Lvova L, Kim SS, Legin A, Vlasov Y, Yang JS, Cha GS, Nam H (2002) All-solid-state electronic tongue and its application for beverage analysis. *Anal Chim Acta* 468:303–314
42. Esbensen KH (2002) *Multivariate Data Analysis - in practice*. CAMO Software, Oslo
43. Abdi H (2010) Partial least squares regression and projection on latent structure regression (PLS Regression). *WIREs Comput Stat* 2:97–106
44. Faber NM, Ferre J, Boque R, Kalivas JH (2002) Second-order bilinear calibration: the effects of vectorising the data matrices of the calibration set. *Chemometr Intell Lab* 63:107–116
45. Smilde AK, Bro R, Geladi P (2004) *Multi-way analysis: applications in the chemical sciences*. John Wiley & Sons Ltd., Chichester
46. Zupan J, Gasteiger J (1991) Neural networks: A new method for solving chemical problems or just a passing phase? *Anal Chim Acta* 248:1–30
47. Despagne F, Luc Massart D (1998) Neural networks in multivariate calibration. *Analyst* 123:157R–178R
48. Hush DR, Horne BG (1993) Progress in supervised neural networks. *IEEE Signal Process Mag* 10:8–39
49. del Valle M, Cetó X, Gutiérrez M (2013) Chapter 6 BioElectronic Tongues: When the Sensor Array Incorporates Biosensors. In: Lvova L, Kirsanov D, Di Natale C, Legin A (eds) *Multisensor systems for chemical analysis: materials and sensors*. Pan Stanford Publishing, Salem MA
50. Tønning E, Sapelnikova S, Christensen J, Carlsson C, Winther-Nielsen M, Dock E, Solna R, Skladal P, Nørgaard L, Ruzgas T, Emnéus J (2005) Chemometric exploration of an amperometric biosensor array for fast determination of wastewater quality. *Biosens Bioelec* 21:608–617
51. Moreno-Barón L, Cartas R, Merkoçi A, Alegret S, Gutiérrez JM, Leija L, Hernández PR, Muñoz R, del Valle M (2005) Data compression for a voltammetric electronic tongue modelled with artificial neural networks. *Anal Lett* 38:2189–2206
52. Apetrei C, Apetrei IM, Saja JAD, Rodriguez-Mendez ML (2011) Carbon Paste Electrodes Made from Different Carbonaceous Materials: Application in the Study of Antioxidants. *Sensors* 11:1328–1344

Comparison of Methods for the Processing of Voltammetric Electronic Tongues Data

Xavier Cetó, Francisco Céspedes and Manel del Valle *

Sensors and Biosensors Group, Department of Chemistry, Universitat Autònoma de Barcelona, Edifici Cn, 08193 Bellaterra, Barcelona, SPAIN

Supporting Information

This supplementary information gives detailed information of the results obtained for each of the calculated models as well as details on sensors preparation and usage.

BioET sensor array

Based on previous studies with polyphenols in our laboratory [25], an array of four working electrodes was prepared, following the conventional methodology in our laboratories, using bare graphite C and adding different modifiers such as tyrosinase, laccase and copper nanoparticles to the bulk mixture – one component per electrode plus a blank electrode without any modifier.

The electrode fabrication begins with the preparation of the composite paste. For this, the resin EpoTek H77 (Epoxy Technology, Billerica, MA, USA) and its corresponding hardener compound were mixed in the ratio 20:3 (w/w). Afterwards 15% (w/w) of graphite (50 mm BDH Laboratory Supplies) and 2% (w/w) of the modifier (either the enzyme or the catalyst) were added to the previous mixture before hardening. Then, it was manually homogenized for 60 min, and afterwards, the paste was allowed to harden for 7 days at 40 °C. Finally, polishing electrodes surface with different sandpapers of decreasing grain size, with a final electrode area of 28 mm².

The measurement cell was formed by the 4-sensor voltammetric array plus a reference double junction Ag/AgCl electrode (Thermo Orion 900200, Beverly, MA, USA) and a commercial platinum counter electrode (Model 52-67, Crison Instruments, Barcelona, Spain). Cyclic voltammetry measurements were carried out at room temperature (25 °C) under quiescent condition using a 6-channel AUTOLAB

* E-mail: manel.delvalle@uab.cat; tel: +34 93 5811017; fax: +34 93 5812379

PGSTAT20 (Ecochemie, Netherlands) in a multichannel electrode configuration, controlled by GPES Multichannel 4.7 software package.

To avoid possible fouling of electrodes surface, an electrochemical cleaning stage was applied between each measurement [25]. Moreover, a control sample was periodically measured during the measurements to ensure there was not any drift or loss of sensitivity; i.e. checking that same voltammetric response was obtained during the whole experiment.

Evaluation of models behaviour

In order to evaluate models' behaviour, two parameters are suggested and taken into account: RMSE for predicted values and comparison graphs between obtained vs. expected concentrations for the compounds under study.

Residual values were calculated between expected and found concentration values, for each sample (i) and for each of the three analytes (j) considered according to equation 1:

$$RMSE = \sqrt{\frac{\sum_{ij} (c_{ij} - \hat{c}_{ij})^2}{3n-1}} \quad (1)$$

In addition to RMSE values, comparison graphs of predicted vs. expected concentration for each polyphenol were built to check the prediction ability of the obtained models, both for train and test subsets. Then slope values and correlation coefficients of the comparison were taken into account in order to help in deciding the best configuration. Therefore, the optimum model would have a small RMSE, meaning there are no differences between the values predicted by the model and the real ones. Also the comparison graphs should display a correlation value and slope close to 1 and intercept close to 0, discarding any model error.

Table S1. RMSE and fitted regression lines of the comparison between obtained vs. expected results for the training subset. Three best and worst values for each parameter are highlighted in green and red, respectively.

| Model | Preprocessing | Total RMSE | Gallic acid | | | | Ferulic acid | | | | Sinapic acid | | | |
|-------|---------------|---------------|-------------|----------|----------|-------|--------------|----------|----------|------|--------------|----------|----------|-------|
| | | | <i>m</i> | <i>b</i> | <i>r</i> | RMSE | <i>m</i> | <i>b</i> | <i>r</i> | RMSE | <i>m</i> | <i>b</i> | <i>r</i> | RMSE |
| PCR | CV | 16.86 | 0.889 | 11.13 | 0.943 | 19.99 | 0.969 | 3.12 | 0.984 | 9.51 | 0.898 | 10.23 | 0.947 | 19.04 |
| | Int | 11.16 | 0.917 | 8.34 | 0.957 | 17.30 | 0.990 | 0.98 | 0.995 | 6.76 | 0.953 | 4.73 | 0.976 | 12.40 |
| | FFT | 11.75 | 0.929 | 7.05 | 0.964 | 15.91 | 0.988 | 1.19 | 0.994 | 5.77 | 0.963 | 3.73 | 0.981 | 11.30 |
| | DWT | 12.60 | 0.916 | 8.41 | 0.957 | 17.37 | 0.986 | 1.42 | 0.993 | 6.13 | 0.959 | 4.11 | 0.979 | 11.71 |
| PLS | CV | 16.00 | 0.890 | 10.97 | 0.944 | 19.85 | 0.969 | 3.10 | 0.984 | 8.96 | 0.918 | 8.20 | 0.958 | 17.15 |
| | Int | 10.20 | 0.953 | 4.70 | 0.976 | 12.98 | 0.992 | 0.84 | 0.996 | 5.11 | 0.963 | 3.70 | 0.981 | 10.84 |
| | FFT | 12.09 | 0.932 | 6.82 | 0.965 | 15.64 | 0.985 | 1.47 | 0.993 | 6.59 | 0.954 | 4.58 | 0.977 | 12.25 |
| | DWT | 10.66 | 0.948 | 5.18 | 0.974 | 13.64 | 0.986 | 1.42 | 0.993 | 6.53 | 0.966 | 3.39 | 0.983 | 10.59 |
| nPLS | CV | 9.24 | 0.970 | 2.96 | 0.985 | 10.30 | 0.994 | 0.61 | 0.997 | 4.57 | 0.962 | 3.81 | 0.981 | 11.37 |
| | Int | 10.10 | 0.955 | 4.49 | 0.977 | 12.69 | 0.991 | 0.89 | 0.996 | 5.36 | 0.963 | 3.67 | 0.981 | 10.79 |
| | FFT | 12.34 | 0.937 | 6.29 | 0.968 | 15.03 | 0.985 | 1.45 | 0.993 | 6.61 | 0.945 | 5.50 | 0.972 | 13.69 |
| | DWT | 11.16 | 0.952 | 4.85 | 0.975 | 13.19 | 0.985 | 1.51 | 0.992 | 6.76 | 0.952 | 4.80 | 0.976 | 12.40 |
| ANN | CI | 7.35 | 0.986 | 1.58 | 0.993 | 9.72 | 0.983 | 2.51 | 0.998 | 4.81 | 0.992 | 1.20 | 0.997 | 6.66 |
| | Int | 8.00 | 0.973 | 2.32 | 0.994 | 8.82 | 0.988 | 0.68 | 0.998 | 5.75 | 0.972 | 3.30 | 0.994 | 9.02 |
| | PCA | 8.17 | 0.970 | 2.96 | 0.990 | 11.09 | 0.990 | 0.94 | 0.998 | 5.34 | 0.984 | 1.68 | 0.996 | 6.96 |
| | FFT | 8.01 | 0.965 | 3.46 | 0.993 | 9.70 | 0.984 | 1.56 | 0.997 | 6.61 | 0.981 | 1.67 | 0.996 | 7.38 |
| | DWT | 8.16 | 0.971 | 2.70 | 0.994 | 8.85 | 0.985 | 1.59 | 0.996 | 7.05 | 0.972 | 2.93 | 0.995 | 8.46 |

PCR: Principal Component Regression; PLS: Partial Least Squares regression; nPLS: Multi-way Partial Least Squares regression; ANN: Artificial Neural Network

CV: Cyclic Voltammogram raw data; Int: Voltammogram windowed slicing integral; PCA: Principal Component Analysis; FFT: Fast Fourier Transform; DWT: Discrete Wavelet Transform; CI: Causal Index selection

Table S2. RMSE and fitted regression lines of the comparison between obtained vs. expected results for the test subset. Three best and worst values for each parameter are highlighted in green and red, respectively.

| Model | Preprocessing | Total RMSE | Gallic acid | | | | Ferulic acid | | | | Sinapic acid | | | |
|-------|---------------|---------------|-------------|----------|----------|-------|--------------|----------|----------|-------|--------------|----------|----------|-------|
| | | | <i>m</i> | <i>b</i> | <i>r</i> | RMSE | <i>m</i> | <i>b</i> | <i>r</i> | RMSE | <i>m</i> | <i>b</i> | <i>r</i> | RMSE |
| PCR | CV | 16.59 | 0.895 | 6.47 | 0.934 | 19.58 | 0.956 | 6.00 | 0.984 | 8.96 | 0.725 | 27.68 | 0.970 | 17.94 |
| | Int | 13.92 | 0.823 | 14.49 | 0.929 | 20.16 | 0.991 | 0.87 | 0.996 | 4.50 | 0.724 | 33.25 | 0.911 | 23.16 |
| | FFT | 13.09 | 0.845 | 12.68 | 0.943 | 18.15 | 0.952 | 3.87 | 0.989 | 7.53 | 0.749 | 33.61 | 0.893 | 25.05 |
| | DWT | 12.98 | 0.855 | 11.60 | 0.943 | 18.11 | 0.949 | 3.91 | 0.993 | 6.36 | 0.706 | 36.44 | 0.903 | 24.32 |
| PLS | CV | 16.00 | 0.887 | 7.49 | 0.932 | 19.76 | 0.957 | 6.00 | 0.983 | 9.13 | 0.753 | 24.74 | 0.973 | 16.53 |
| | Int | 12.19 | 0.883 | 8.43 | 0.946 | 17.56 | 0.991 | 0.25 | 0.996 | 4.49 | 0.763 | 30.64 | 0.898 | 24.18 |
| | FFT | 14.33 | 0.794 | 15.31 | 0.928 | 20.87 | 0.943 | 5.95 | 0.995 | 5.49 | 0.751 | 29.39 | 0.910 | 22.86 |
| | DWT | 11.43 | 0.896 | 9.11 | 0.955 | 15.87 | 0.961 | 4.45 | 0.995 | 5.31 | 0.752 | 32.26 | 0.902 | 24.00 |
| nPLS | CV | 15.30 | 0.784 | 24.55 | 0.944 | 19.10 | 1.150 | -15.28 | 0.977 | 14.42 | 0.785 | 13.82 | 0.895 | 25.89 |
| | Int | 11.79 | 0.885 | 14.87 | 0.954 | 16.50 | 0.999 | -0.21 | 0.994 | 5.35 | 0.744 | 18.32 | 0.916 | 24.26 |
| | FFT | 11.75 | 0.984 | 5.39 | 0.969 | 13.90 | 1.058 | -6.05 | 0.995 | 5.83 | 0.657 | 27.91 | 0.875 | 28.54 |
| | DWT | 11.01 | 1.018 | 2.12 | 0.974 | 13.34 | 1.037 | -4.04 | 0.995 | 5.67 | 0.651 | 28.57 | 0.883 | 28.16 |
| ANN | CI | 10.97 | 1.011 | 0.34 | 0.976 | 12.10 | 1.006 | -1.28 | 0.995 | 5.18 | 0.958 | 3.49 | 0.967 | 13.71 |
| | Int | 11.10 | 0.995 | 2.90 | 0.977 | 11.97 | 0.998 | -1.93 | 0.988 | 8.03 | 0.976 | -3.50 | 0.978 | 12.71 |
| | PCA | 11.52 | 1.095 | -9.46 | 0.974 | 14.61 | 0.991 | -1.11 | 0.992 | 6.70 | 1.055 | -6.77 | 0.980 | 11.82 |
| | FFT | 10.10 | 0.990 | 0.69 | 0.973 | 12.61 | 1.017 | -2.71 | 0.993 | 6.08 | 0.925 | 7.46 | 0.982 | 10.50 |
| | DWT | 11.48 | 1.039 | -2.93 | 0.982 | 10.84 | 1.003 | 0.15 | 0.991 | 6.86 | 0.953 | 2.84 | 0.961 | 15.19 |

PCR: Principal Component Regression; PLS: Partial Least Squares regression; nPLS: Multi-way Partial Least Squares regression; ANN: Artificial Neural Network

CV: Cyclic Voltammogram raw data; Int: Voltammogram windowed slicing integral; PCA: Principal Component Analysis; FFT: Fast Fourier Transform; DWT: Discrete Wavelet Transform; CI: Causal Index selection

Article 9

Voltammetric Electronic Tongue for qualitative analysis of beers

X. Cetó, J. M. Gutiérrez, A. Mimendía, F. Céspedes and M. del Valle

Electroanalysis **25**(7): 1635-1644

Voltammetric Electronic Tongue for qualitative analysis of beers

Xavier Cetó¹, Juan Manuel Gutiérrez², Aitor Mimendia^{1,3}, Francisco Céspedes¹ and Manel del Valle^{1,*}

¹ Sensors and Biosensors Group, Department of Chemistry, Universitat Autònoma de Barcelona, Edifici Cn, 08193 Bellaterra, Barcelona, SPAIN

² Bioelectronics Section, Department of Electrical Engineering, CINVESTAV, 07360 Mexico D.F., Mexico

³ IRTA Aquatic Ecosystems, Carretera de Poble Nou, Km 5.5, 43540 St. Carles de la Ràpita, Catalonia, Spain

Abstract

This paper deals with the application of a voltammetric Electronic Tongue towards beers classification. For this purpose, samples were analyzed using cyclic voltammetry without performing any sample pretreatment, albeit its dilution with distilled water. The voltammetric signals were first preprocessed employing Fast Fourier Transform. Then, using the obtained coefficients, responses were evaluated using three different clustering techniques: Principal Component Analysis (PCA), Partial Least Squares Discriminant Analysis (PLS-DA) and Linear Discriminant Analysis (LDA). In this case, the ET has demonstrated a good capability to correctly discriminate and classify the different beer samples according to its type (Lager, Stout and IPA) and manufacture process (commercial and craft).

Keywords: Electronic Tongue; Linear Discriminant Analysis; voltammetric sensors; Fourier feature extraction; beer

1. Introduction

Novel methodologies based on the concept of Electronic Tongues (ETs) represent a promising chemical system for the discrimination and identification of liquid samples, being able to distinguish particularities such as flavour or to provide their characterization. In this context, ETs have demonstrated to be suitable as screening methods to achieve new challenges in quality control or food and environmental analysis [1]. The fundamentals of ETs imply the use of an array of non-specific, poorly selective, chemical sensors, with cross-sensitivity towards different compounds in a solution, plus an appropriate chemometric tool for data processing [2].

In this sense, ETs start to be significant in foodstuff analysis, mainly devoted to the identification of different types/varieties of samples or the quantification of components present in those, such as milk, fruit juices, coffee, wine, etc. [1, 3-5]. Among the different sensors that can be used for developing ETs, potentiometric, voltammetric and impedimetric ones are the most commonly used [6]. Specifically, in this paper we will focus on the use of a voltammetric ET for qualitative analysis of beers.

Beer is one of the most popular beverages around the world, mainly distinguished by characteristics such as foam, colour and flavour that are granted by the fermentation process of grains (malts). In this sense, there is an effort to differentiate and categorize beers by various factors such as colour, flavour, strength, production method, fermentation method, etc.

Basically, beer is categorized into one of the two different styles: Ales (i.e. beer brewed from malted barley

using a warm-fermentation with a strain of brewers' yeast) and Lagers (i.e. beers fermented and conditioned at low temperatures). In this sense, beers such as Indian Pale Ale (IPA) and Stout are varieties of Ales while Pilsner and American styles are categorized as Lagers [7]. On the other hand, as a consequence of the large number of producers advocating for the usage of traditional methods during beer elaboration, this factor has also become an important aspect for beer categorization due to the added value it provides, both from the sensory and economic points of view.

Nevertheless, there is a lack of analytical methods able to objectively perform such classification, requiring usually its taste by a sensory panel in order to achieve its correct classification. Thus, an important task during brewing is to characterize beer flavour as well as its colour, strength and bitterness, which levels can be used to assess beers style.

In this direction, there are some works in the literature that attempt to use ETs for beer analysis and recognition; thus representing the obtaining of a useful tool for quality control analysis. ETs try to mimic the sense of taste in mammals, focusing in the extraction of a characteristic fingerprint rather than the identification of specific compounds, although also allowing its quantification if properly trained. The goal of these contributions is focused mainly on the discrimination of samples of different varieties/origins or the monitoring of beer aging and fermentation [8, 9]; there are also some interesting works focused on the prediction of certain chemical [10, 11] or taste attributes in comparison to a sensory panel [12].

Although the progress in the manufacture of electrochemical sensors is a well established area, there are crucial innovative aspects in the implementation of ETs associated with the analysis of analytical information

* E-mail: manel.delvalle@uab.cat; tel: +34 93 5811017; fax: +34 93 5812379

obtained from them, which are pending to be developed. This is especially true in the case of ETs employing voltammetric sensors. In this case, the followed approach leads to a challenging situation: the high dimensionality of departure data, formed by the set of voltammograms. The final processing is cumbersome because of the multiway nature of the data (sensors \times polarization voltage \times samples) as this structure is hardly compatible with the most common pattern recognition or data processing strategies. At this point, trilinear approaches might be employed, such as PARAFAC or Tucker models [13]; however, they also have their own intricacy. Another strategy that could be used is the data reduction (compression) of the original data set, together with the unfolding of the different channels (sensors) in a unique row [14]. Some of the approaches in this case are the use of kernels [15], the discrete Wavelet transform (DWT) or the fast Fourier transform (FFT) [14]. For this aim, first the set of voltammograms corresponding to each sample are preprocessed, being reduced to a number of coefficients; and next, the pattern recognition algorithms are applied. Alternatively, for data pretreatment, feature selection can be performed which would be likely to enhance the distinguishing ability [16].

In this sense, most common pattern recognition methods include the use of Principal Component Analysis (PCA) coupled with a modelling tool such as an Artificial Neural Network (ANN) model [3], Partial Least Squares Discriminant Analysis (PLS-DA) [17] and Linear Discriminant Analysis (LDA) [18]. Nevertheless, there are other supervised methods from machine learning such as Support Vector Machines (SVMs) or ANNs [19], which can lead to a superior performance of the built models thanks to the use of non-linear functions.

The aim of the present work is the use of a voltammetric ET based on differently modified graphite-epoxy electrodes plus one enzymatic biosensor coupled with data processing techniques to identify either different beer style (Lager, Stout and IPA) or commercial and craft manufacturing process. Interpretation models were achieved using the whole cyclic voltammetric signals obtained from the electrodes as main departure information. Nevertheless, the complexity of the recordings requires a preprocessing stage capable of reducing the information without loss of significant information. For this purpose, a preprocessing stage for the voltammetric data was considered using FFT. Finally, the discrimination stage was performed using PCA, PLS-DA and LDA.

2. Experimental

2.1 Reagents and solutions

All reagents used were of analytical grade and all solutions were prepared using deionised water from a Milli-Q system (Millipore, Billerica, MA, USA). Glucose oxidase from *Aspergillus niger* (EC 1.1.3.4, 174.4 U \cdot mg⁻¹), D(+)-glucose monohydrate, polypyrrole, copper and platinum nanoparticles (<50 nm), and Cobalt (II) phtalocyanine were purchased from Sigma-Aldrich (St. Louis, MO, USA). KCl was purchased from Merck KGaA (Darmstadt, Germany).

2.2 Beer samples

A total set of 25 samples of different brands and varieties were provided by a craft brewery (Art Cervecers Company, Canovelles, Barcelona, Spain) or purchased at a local supermarket, preparing only one sample from each of the beers considered and not replicating them. This was preferred to avoid that usage of replicate measurements of the same beer could result in an overfitting of the model, i.e. that training of the model could result in a “memorization” of the observed signals if the same sample is used both in train and test.

Samples were selected according to two features: the manufacture method (commercial or craft beers) and the kind of beer (Lager, Stout and IPA). Thus, the set was formed by 15 commercial beers plus 10 craft ones; and also, taking into account the beer type, the set was formed by 17 lager, 5 Stout and 3 IPA beers.

2.3 Voltammetric sensor array

Based on previous experience in our laboratory, an array of 6 voltammetric electrodes were prepared following the conventional methodology [20]. A resin EpoTek H77 (Epoxy Technology, Billerica, MA, USA) and its corresponding hardener compound were mixed in the ratio 20:3 (w/w); after that a 15% of graphite (w/w) and a 2% of the modifier (w/w) were added to the previous mixture before hardening, obtaining the composite. Then, it was manually homogenized for 60 min, and afterwards, the paste was allowed to harden for 3 days at 80 °C (7 days at 40 °C in the case of GOX biosensor). Finally, the electrode surface was polished with different sandpapers of decreasing grain size. The final electrode area was 28 mm².

In this manner, one blank electrode plus 5 different bulk-modified voltammetric sensors were prepared using bare graphite C and adding different modifiers such as cobalt phtalocyanine, platinum and copper nanoparticles, polypyrrole and GOX enzyme to the bulk mixture – one component per electrode. Those modifiers were selected according to previous studies in our laboratory with wines and also following results from other research groups [21–25], in order to obtain a variety of electrodes with significant cross-selectivity and complementary electroactive properties, that allow the obtaining of rich information to enhance modelling capabilities.

From an electroanalytical point of view, conducting polymers represent interesting modifying electrode materials; their advantages are ascribed to their antifouling and electrocatalytic properties, as well to their high sensitivities and low detection limits [24]. Also, in the last few years, nanoparticles have emerged as interesting electroactive material in electroanalysis; these are alternative to bulk metals, with catalytic and electrocatalytic peculiarities; mainly derived from their higher surface/mass ratio [21]. Finally, electrodes modified with phtalocyanines (mainly CoPc and its derivatives) are interesting for being efficient electrocatalysts in the determination of many important inorganic, organic or biological compounds [25].

On the other hand, even normally beer is not associated with sweet taste or a sugar containing product, in the case of purely craft beers, those are bottled with the addition of some sugar and yeast so that a second fermentation takes

place in the bottle, generating in this manner alcohol and CO₂, responsible for its effervescence [26]. Therefore, the detection with GOX biosensor of the sugar that has not fermented may help indicating a craft-made beer.

2.4 Measuring procedure

The amperometric measurement cell was formed by the 6-sensor voltammetric array plus a commercial platinum counter electrode (Model 52–67, Crison Instruments, Barcelona, Spain) and a reference double junction Ag/AgCl electrode (Thermo Orion 900200, Beverly, MA, USA). Cyclic Voltammetry measurements were taken using a 6-channel AUTOLAB PGSTAT20 (Ecochemie, Netherlands), in a multichannel configuration, using GPES Multichannel 4.7 software package.

Electroanalytical experiments were carried out at room temperature (25 °C) under quiescent conditions and without any sample pretreatment, just diluting the sample in distilled water. In this way, each beer sample was opened prior starting the measurement, thus to avoid the loss of any volatile compound or oxidation of beer components. Then, it was diluted by mixing it with distilled water in the ratio 15:35, respectively [12]. This dilution was done to minimize the apparition of bubbles onto the electrode surface while the measurement was done, and at the same time, to reduce any possible matrix effect. A dilution step was preferred rather than degassing of the beer to simplify and reduce preparative steps prior to perform samples measurement on one side, and to avoid the loss of any volatile compound or oxidation of beer components such as antioxidants. Afterwards, sample measurement was done immediately after its dilution.

Electrodes were cycled for 3-5 times in saline solution in order to get stable voltammetric responses before performing the measurements of beer samples. For samples measurement, potential was sweep between -1.0 V and 1.2 V vs Ag/AgCl, with a scan rate of 100 mV·s⁻¹ and a step potential of 9 mV. All experiments were carried out without performing any physical surface regeneration of the working electrodes. In order to prevent the accumulative effect of impurities on the working electrode surfaces, an electrochemical cleaning stage was done between each measurement applying a conditioning potential of +1.4 V during 45s after each experiment in a cell containing 50 ml of distilled water [22].

2.5 Data processing

Chemometric processing was done by specific routines written by the authors in MATLAB 7.10 (MathWorks, Natick, MA), based on already preprogrammed standard functions using Neural Network Toolbox (v.4.0.6) and Statistics Toolbox (v7.3). Sigmaplot 2000 (Systat Software Inc, California, USA) was used for graphic representations of data and results.

Firstly, obtained responses were preprocessed employing Fast Fourier Transform (FFT). Then obtained coefficients were used as input into the three different methods: Principal Component Analysis (PCA), Partial Least Squares Discriminant Analysis (PLS-DA) and Linear Discriminant Analysis (LDA). Finally, in the case of PLS-DA and LDA, given that those are supervised methods, classification success was evaluated using a

‘leave one out’ cross-validation scheme and also *repeated random sub-sampling validation* to further validate proposed approach.

The main difference between those models is that instead of generating new axis based on the directions of maximum variance of response matrix, as done in PCA, LDA generates the new axis based on the maximum discrimination between sample groups (Figure 1). Another difference is the number of axis generated. That is, in the case of PCA as many “axis” (Principal Components) as variables are generated; while in the case of LDA the model will be formed by just $k-1$ linear discriminant functions (the new axis; being k the number of groups), where these functions correspond to linear combinations of the predictor variables that provide the best discrimination between the groups. Finally, PLS-DA is somehow related to PCA, being the main difference that instead of finding hyperplanes of minimum variance between the response and independent variables, it finds a linear regression model by projecting the predicted variables and the observable variables to a new space.

Additionally, in the case of LDA analysis, it was done using a stepwise inclusion method [27]. In this manner, and based on a statistical criteria, variables were included using Wilk’s lambda (λ) statistic [27, 28], which is used in this case as a measure of the class centroids separation. Thus, variables that will be introduced in the equation were selected basing on their contribution to decrease λ value. Thus, in each step the variable that minimizes global λ is introduced until optimum performance is reached (best separation between classes). Wilk’s lambda is defined as the proportion of the group inertias which is not explained by the response variable (that identifies the classes) in the classical scheme of variance decomposition [27, 28]. It is therefore the ratio of the intra-class inertia and the total inertia. Therefore being its value between 0 and 1. If only a small fraction of the total inertia is not explained by the existence of groups, then these groups are well separated, and their means are significantly different. Hence, a small (close to 0) value of λ means that the groups are well separated; while, a large (close to 1) value of λ means that the groups are poorly separated.

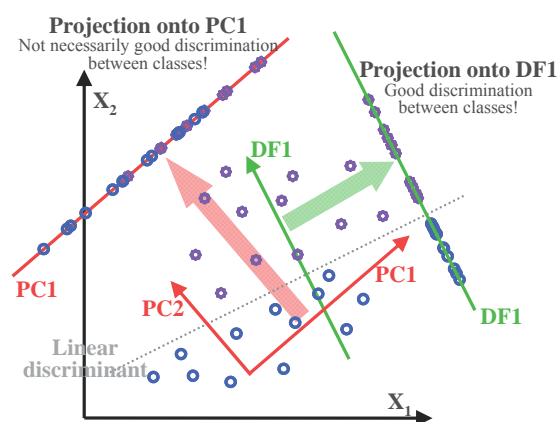


Fig 1 Schematic comparison between PCA and LDA methods. Briefly, PCA finds the direction of maximum variance between the variables, while LDA seeks for the direction that provide maximum separation between the classes. Also note that in the case of PCA as many new PCs as variables in the original data are obtained, whereas in the case of LDA only as many #groups-1 DFs are obtained.

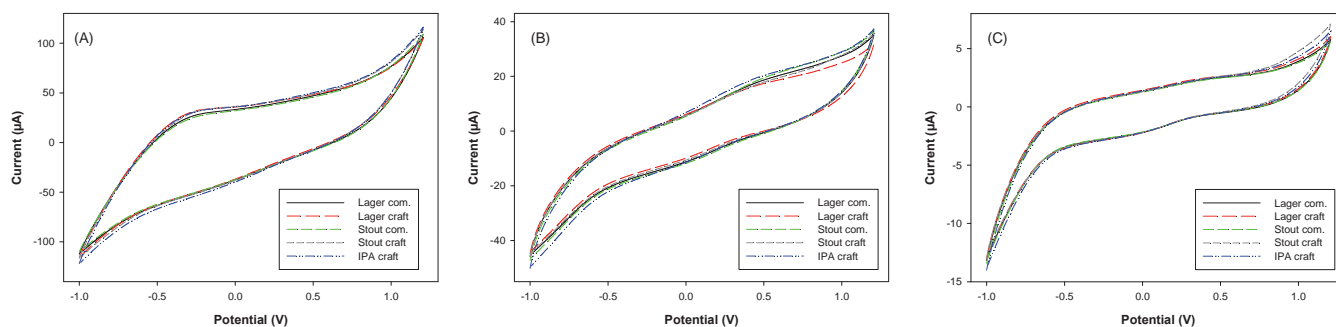


Fig. 2 Example of the different obtained voltammograms with the modified sensor array: (A) platinum nanoparticle sensor, (B) polypyrrole sensor and (C) GOX biosensor. Average responses for beer samples are shown based on its type (Lager, Stout and IPA) and manufacture process (commercial and craft).

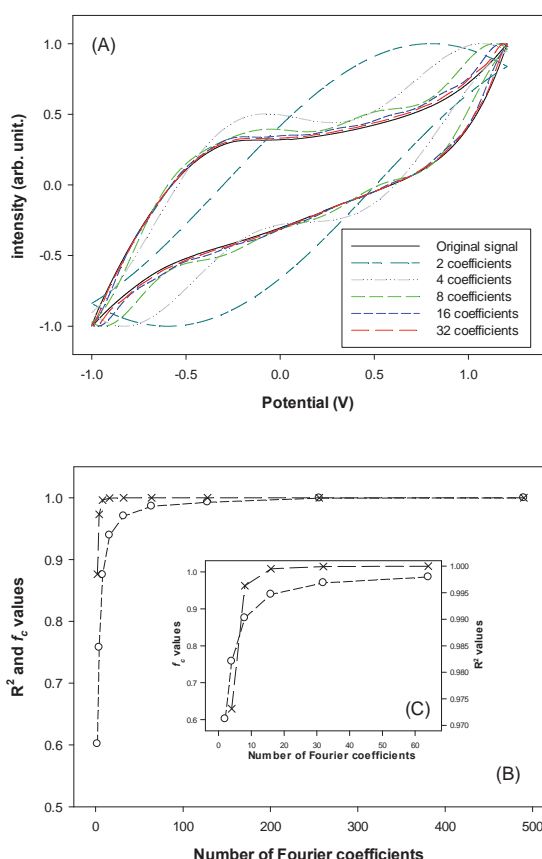


Fig. 3 (A) Reconstruction of the voltammetric signal according to the number of Fourier coefficients employed, platinum nanoparticle sensor shown as example. (B) Representation of the coefficient of determination (R^2 , x) and f_c (o) against the number of Fourier coefficients used from comparing the raw voltammograms with the reconstructed signals. (C) Enlargement of the previous plot.

3. Results and Discussion

3.1 Voltammetric responses

Examples of the responses of the array towards the different samples are shown in Figure 2. As can be seen, the shape and currents of the CVs depend on the nature of the modifier employed, obtaining differentiated response profile for each type of sensor. Despite some similarities are seen along the CV curve between different samples, a higher discrimination can be seen at higher potentials. For example, the GOX biosensor (Fig 2C), which is known to respond to glucose concentration at around 0.9V [29],

seems to discriminate the Lager and the IPA samples from the Stout ones; also, obtaining higher signals for craft beers, as can be expected due to the addition of sugar during craft beers bottling (Section 2.3). Similarly, the platinum nanoparticle sensor (Fig 2A) seems to discriminate the manufactured beers from the craft ones, where a clearly differentiated response can be noticed in the anodic wave around the region of $-0.25V$. Also, polypyrrole sensor (Fig 2B) shows distinguished response in the anodic wave starting at 0V, or at the cathodic one starting also at around $+0.25V$.

This situation, with marked mix-response and differentiated signals obtained from the different electrodes, is highly desirable for studies with ET systems. This very rich data set generated is the ideal departure point, especially if colinearity between sensors is absent. At this point, given the complexity of the departure data, a preprocessing step was used in order to reduce the high dimensionality requirements of the processing.

3.2 FFT signal preprocessing

Given the large dimensionality of the generated data (samples \times sensors \times polarization potential), which hinders their chemometric treatment, and to fully exploit all the information obtained from each voltammogram, a preprocessing stage for data reduction was performed. The main objective of this step was to reduce the complexity of the input signal while preserving the relevant information. Normally, this approach also permits to gain advantages in modelling time, to avoid redundancy in input data and to obtain a model with better generalization ability [30]; i.e. to improve model performance. In our case this compression stage was achieved by the use of Fast Fourier Transform (FFT) [31]. Hence, each voltammogram was substituted by a number of Fourier coefficients (selected from the minimum allowing a good reconstruction of the original data), accomplishing in this way the data reduction.

In order to quantify the reconstruction degree, the coefficient of determination (R^2) and a comparison factor named f_c that considers the area under both signals when superimposed were used [32]. Both factor range from 0 to 1 depending on signals similarity; it values 0 when two signals have nothing in common and increases its value as similarity does. In this sense, f_c computes similarity in a way related to a correlation coefficient r , but being more sensitive to small differences.

Despite reconstruction degree is critical in order to

Table 1. Number of features (FFT coefficients) selected using the stepwise inclusion method employed for building the classification models.

| Sensor | Beer type | Manufacture procedure | Beer type & manufacture procedure |
|---------------------------|-----------|-----------------------|-----------------------------------|
| Graphite-epoxy | 3 | 1 | 4 |
| CoPc | 2 | 1 | 3 |
| Pt nanoparticle | 1 | 1 | 1 |
| Cu nanoparticle | 1 | 2 | 4 |
| Polypyrrole | 4 | 4 | 1 |
| GOX biosensor | 4 | 4 | 2 |
| Total coefficients | 15 | 13 | 15 |

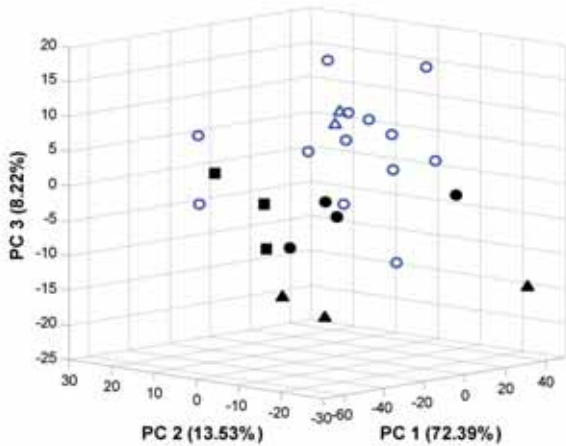


Fig 4. Score plot of the first three components obtained after PCA analysis of the beer samples. As can be seen, no clear discrimination is obtained for the different beer classes: (●) Lager, (▲) Stout and (■) IPA, or the manufacture process: (blue, hollow) commercial or (black, solid) craft beers.

ensure extracted coefficients represent correctly all the information contained in the raw data, it must also be taken into account that increasing the coefficients used also hinders the data treatment. Thus, a compromise is needed; in our case, the number of coefficients were selected by checking the correlation of the reconstructed vs. original signal representation, as the first time that f_c surpassed 0.97 (a large value that ensures the best reproduction/compression ratio of the original signal) [31]. This was achieved with the first 32 Fourier coefficients (Figure 3), which allowed a compression of the original information (490 current values) up to 93.5% without any loss of significant information.

Although the number of coefficients used could still seem quite large given 32 coefficients were used for each sensor, it was preferred to ensure the best reconstruction degree, given LDA analysis was done using a stepwise inclusion method which allows to remove the variables that have a lower contribution to the classification model. Also, in the case of PCA this does not represent any problem, given it is a powerful method that allows to summarize almost all variance contained in the departure information onto a fewer number of directions (the PCs) with new coordinates called scores, obtained after data transformation.

3.3 Classification of beer samples

Discrimination of beer samples was attempted using three different methods for grouping samples: PCA,

PLS-DA and LDA. In the first case, no prior considerations were required given it is an unsupervised pattern recognition method. Thus clustering was evaluated after model construction from the obtained scores values for each sample. In the second case, given both LDA and PLS-DA, unlike PCA, are supervised methods, different models for each case were built. In a first step, discrimination between different beer kinds (Lager, Stout and IPA) was attempted; while in the second approach the objective was to discriminate between the commercial beers from the craft-made ones.

As said, LDA analysis was done using a stepwise inclusion method [27]. This method is very useful in order to select and remove the variables that do not contribute at all to the prediction success. Thus, having a list of independent variables, some of which may be useful predictors, but some of which are almost certainly useless, the aim is to find the best subset to complete the prediction task as well as possible, with as few variables as possible.

This precaution was repeated for each one of the built models (Table 1), allowing an improvement of model performance and reduction of its complexity as well as an easier understanding. Additionally, this step proved the importance and differences of the different sensors forming the array, given it also can be used to evaluate how significantly each sensor contributes to the final classification model; i.e. if none of the features coming from one sensor are enough significant, all coefficients will be kept out by the model, thus not interviewing at all that sensor in the classification success.

3.3.1 Principal Component Analysis

Taking the obtained coefficients from the FFT analysis as input variables, PCA analysis was done. This was first selected given PCA is a powerful linear unsupervised pattern recognition method that reduces the dimensionality of a multivariate problem and helps to visualize the different categories, if directly seen. Thus, samples are not grouped taking into account prior expected similarities, but based only on their response profile.

In this sense, PCA was employed as a preliminary recognition model in order to check the trend of the measures and if there was a relation between them and maybe the kind of beer or the manufacture process. The obtained PCA plot is shown in Figure 4; with the three first PCs, the accumulated variance was ca. 94.1%. Despite this large value, not enough clear discrimination between the different beers was achieved, although some initial patterns could be seen. That is, commercial beers seem to group on the upper right part of the plot, while craft ones appear grouped in the middle; also, the case of IPA beers, which are close to each other. Therefore, to confirm these initial patterns seen by the use of PCA, the next step was the use of LDA and PLS-DA as the pattern recognition methods given an expected superior performance [33].

3.3.2 Beer type discrimination

As before, employing the obtained coefficients from the FFT analysis, LDA analysis was performed. Given this is a supervised method, classification success was evaluated using leave-one-out cross validation. In this manner, each sample is classified by means of the analysis functions derived from the other samples (all cases except

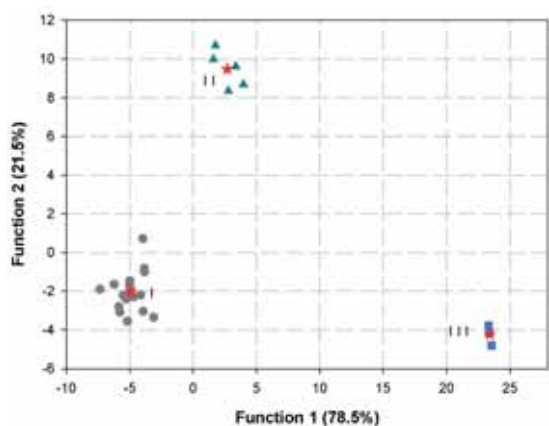


Fig 5. Score plot of the functions obtained after LDA analysis of the beer samples, according to its kind. As can be seen, a clear discrimination is obtained for the different beer classes: (●) Lager, (▲) Stout, (■) IPA. Also, the centroid of each class is plotted (★).

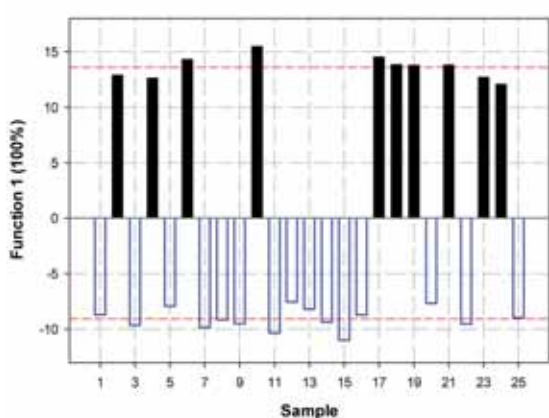


Fig 6. Score plot of the function obtained after LDA analysis of the beer samples against beer samples, according to manufacture process. In this case, from a single discriminant coordinate given only 2 groups were considered. As can be seen, a clear discrimination is obtained depending on fabrication method: (□) commercial and (■) craft beers. Also, the centroid of each class is plotted in the graph (---).

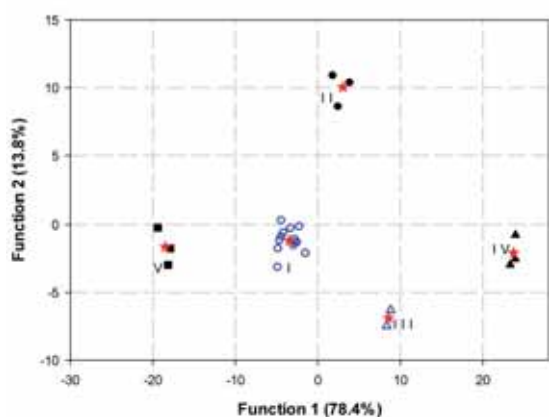


Fig 7. Score plot of the functions obtained after LDA analysis of the beer samples, according to its kind and manufacture process. As can be seen, also in this case clear discrimination is obtained for the different groups: (○) Commercial Lager, (●) Craft Lager, (△) Commercial Stout, (▲) Craft Stout and (■) Craft IPA. Also, the centroid of each class is plotted (★).

the case itself). This process was repeated 25 times (as many as samples) leaving out one different sample each time, the one that must be classified, which acts as model

validation sample. Thus, with this approach all samples are used once as validation.

Although results obtained with the leave-one-out strategy sometimes are criticized as overoptimistic, its use is necessary when there is no availability of large sample sets [34]. In this respect, to demonstrate, the efficacy of the proposed method, model classification success was further validated utilizing the *repeated random sub-sampling validation* technique which, in turn, enabled the estimation of the associated standard errors of the performance indicators and ensured the obtainance of unbiased data [35]. This method has the advantage that avoids dependence of predictions from the specific subdivision of data. In this way, train/test data subdivision was repeated randomly 15 times leaving 6 samples out each time, generating a new model each time and evaluating its response for the different test validation subsets, obtaining a more realistic estimation of model performance.

As can be seen in Figure 5, a clear discrimination between the three varieties of beer was achieved. Patterns in the figure evidence that samples are grouped according to the kind of beer, where these well established clusters clearly separate the main classes of samples corresponding to: Lager, Stout and IPA.

Classification results of LDA leave-one-out cross-validation and *repeated random sub-sampling validation* approach are summarized in Table 2. As expected from the LDA plot, all samples were correctly classified according to its beer type. Hence, the classification rate of the three kinds of beer reached 100% of accuracy (classification rate). The efficiency of the classification obtained was also evaluated according to its sensitivity, i.e. the percentage of objects of each class identified by the classifier model, and to its specificity, the percentage of objects from different classes correctly rejected by the classifier model. The value of sensitivity, averaged for the classes considered was, 100 %, and that of specificity was 100 %. It should be noticed that also in the case of *repeated random sub-sampling validation*, where 6 samples are used each time for the evaluation of model performance, all samples were correctly classified in all the built models; thus, demonstrating the robustness and high discriminative power of the proposed approach.

3.3.3 Manufacture procedure discrimination

As stated, beyond the discrimination of the three kinds of beer, the other objective was to evaluate ET capability to discriminate between manufactured and craft beers. Like the previous case, a new LDA model was built and evaluated using leave-one-out cross validation method. Given in this case only two classes were being considered (i.e. craft and manufactured beers), the LDA model was just formed by one discriminant function. Hence, to better visualize the results, obtained scores were plotted against beer sample number.

As shown in Figure 6, also in this case, clear discrimination between the two types of manufactured beers was achieved. Patterns in the figure evidence that samples are grouped according to the manufacture process. Moreover, as already did when discriminating the kinds of beer, confusion matrix was built employing leave-one-out cross-validation and *repeated random sub-sampling validation* approach, which results are found in Table 2. As

Table 2. Summary of ET capabilities for the classification of the different beers according to expected classes employing either PLS-DA or LDA.

| | <i>PLS-DA & Leave-one-out</i> | | | <i>LDA & Leave-one-out</i> | | | <i>LDA & Split sample resampling</i> | | |
|-----------------------------------|-----------------------------------|-------------|-------------|--------------------------------|--------|--------|--|-------------|-------------|
| | %Class | Sensitivity | Specificity | %Class | %Class | %Class | %Class | Sensitivity | Specificity |
| Beer type | 100% | 100% | 100% | 100% | 100% | 100% | 100±0% | 100±0% | 100±0% |
| Manufacture procedure | 100% | 100% | 100% | 100% | 100% | 100% | 100±0% | 100±0% | 100±0% |
| beer type & manufacture procedure | 88.0% | 81.7% | 95.8% | 100% | 100% | 100% | 93.3±8.5% | 95.0±7.3% | 98.5±1.9% |

* Expressed uncertainties correspond to the standard deviation calculated from the average of the different resampling models

expected from the LDA plot, all samples were correctly classified, with a classification rate of 100% in terms of accuracy. Also, values of 100% were obtained for the method sensitivity and specificity.

3.3.4 Simultaneous discrimination of beer type and manufacture procedure

Finally, given the good discrimination obtained with the previous LDA models, a new model taking into account both parameters (beer type and manufacture process) was built. Thus, in this case samples were distributed in five groups for the model construction: (I) commercial Lager, (II) craft Lager, (III) commercial Stout, (IV) craft Stout and (V) craft IPA beers.

As expected from the previous results, all samples were correctly grouped as can be seen in Figure 7. Moreover, these results are consistent with the previous ones, given groups I and II, which correspond to manufactured and craft Lager beers respectively, are close one to each other; and mainly separated by the second discriminant function. Similar behaviour is obtained for Stout beers (clusters III and IV). Thus, if these clusters are joined (cluster I with II and III with IV, considering them as one class instead of two), LDA model could be summarized as the discrimination between different types of beer (Figure 5).

Again, model accuracy was evaluated employing leave-one-out cross-validation and *repeated random sub-sampling validation* approach, which results are summarized on Table 2. As expected from the LDA plot, all samples were correctly classified, with a classification rate of 100% in terms of accuracy, sensitivity and specificity. However, in this case, a slight degradation of the model performance is seen when evaluating it by using *repeated random sub-sampling validation* approach; a fact that confirms that sometimes the use of leave-one-out approach can reach to an overoptimistic evaluation of model performance.

3.3.5 Partial Least Squares Discriminant Analysis

To further validate the results obtained employing LDA, PLS-DA was also used as a comparison pattern recognition method. As previously done in the case of LDA, three different situations were evaluated; viz. beer type, manufacture procedure and the combination of both. As LDA, PLS-DA is also a supervised method, hence classification success was again evaluated using leave-one-out cross validation; which results are summarized in Table 2.

As can be seen, high performance is still obtained for the discrimination of beer type and manufacture process,

thus demonstrating the capabilities of the ET to correctly distinguish those classes. However, not such a good model is obtained for the simultaneous discrimination of both, which could be somehow expected given the higher complexity of the case and the smaller ratio of samples for each class available for the model to extract the characteristic fingerprint of each class. Nevertheless, this does not represent a huge drawback since correct classification of samples could be achieved separately from the individual models, just by using the obtained coefficients as input in two different models without requiring any extra measurement step of the sample.

Lastly, it should be taken into account that this application is presented as an illustrative study case. That is, with the same experimental setup, the approach proposed herein may be alternatively applied for the classification of other beers into different classes, with a proper set of samples and retraining of the system. Therefore, representing the obtaining of an automated tool able to classify beers in the same manner a sensory panel would do; really suitable for screening purposes and quality control during its elaboration or transportation.

4. Conclusions

A novel method based on a multisensory system to evaluate the kind of beer and its manufacture process has been established. The voltammetric Electronic Tongue (ET) system reported is based on an array of bulk-modified sensors using on one hand different electrochemical modifiers plus a GOX biosensor on the other hand. Samples were measured without any pretreatment, just diluting them in distilled water to reduce matrix and bubbling effects on measurements. The obtained voltammetric responses of the array were preprocessed employing Fast Fourier Transform (FFT) which provided a good reconstruction and compression degree of the data, while preserving relevant information from the raw responses. Lastly, discrimination of beer samples was achieved by the use of Linear Discriminant Analysis (LDA) and Partial Least Squares regression Discriminant Analysis (PLS-DA).

In summary, the power of ETs in food and beverages field as classification tools has been demonstrated again, providing a fast-response and low-cost method for blind samples characterization, really suitable as a screening method for quality control or samples authentication. In addition, it represents a response to the demand of new analytical methods with high sensitivity, good selectivity and fast response needed to meet new challenges in environmental monitoring, food safety, and public health.

Finally, the use of an array of non-specific sensors, based on bulk-modification of graphite-epoxy sensors has been proved as an easy way to obtain sensors with differentiated and cross-selectivity response towards desired samples; while the combination of FFT and LDA provide new strategies for voltammetric data analysis, an approach rarely used.

Acknowledgments

Financial support for this work was provided by Spanish Ministry of Science and Innovation, MCINN (Madrid) through project CTQ2010-17099, by the Mexican National Council of Science and Technology CONACYT (Mexico) through the project 154243 and by the Catalanian program ICREA Academia. X. Cetó thanks the support of Dept. d'Innovació, Universitat i Empresa de la Generalitat de Catalunya for the predoctoral grant. The authors wish to thank Francisco Sánchez and Art Cervecers company – Cervesa Artesanal (Canovelles, Barcelona Spain) for providing samples for the development of this work.

References

- [1] L. Escuder-Gilabert, M. Peris, *Anal. Chim. Acta* **2010**, *665*, 15.
- [2] M. Holmberg, M. Eriksson, C. Krantz-Rülcker, T. Artursson, F. Winquist, A. Lloyd-Spetz, I. Lundström, *Sensor Actuat. B-Chem.* **2004**, *101*, 213.
- [3] P. Ciosek, W. Wroblewski, *Analyst* **2007**, *132*, 963.
- [4] A. Riul Jr, C.A.R. Dantas, C.M. Miyazaki, O.N. Oliveira Jr, *Analyst* **2010**, *135*, 2481.
- [5] A. Rudnitskaya, A. Legin, *CAB Reviews: Perspectives in Agriculture, Veterinary Science, Nutrition and Natural Resources* **2010**, *5*, 1.
- [6] M. del Valle, *Electroanal.* **2010**, *22*, 1539.
- [7] G. Fix, *Principles of brewing science: a study of serious brewing issues*, Brewers Publications, Boulder, CO/US, 1999.
- [8] A. Kutyła-Olesiuk, M. Zaborowski, P. Prokaryn, P. Ciosek, *Bioelectrochemistry* **2012**, *87*, 104.
- [9] M. Ghasemi-Varnamkhashi, M.L. Rodríguez-Méndez, S.S. Mohtasebi, C. Apetrei, J. Lozano, H. Ahmadi, S.H. Razavi, J. Antonio de Saja, *Food Control* **2012**, *25*, 216.
- [10] E. Polshin, A. Rudnitskaya, D. Kirsanov, A. Legin, D. Saison, F. Delvaux, F.R. Delvaux, B.M. Nicolai, J. Lammertyn, *Talanta* **2010**, *81*, 88.
- [11] Á.A. Arrieta, M.L. Rodríguez-Méndez, J.A. de Saja, C.A. Blanco, D. Nimubona, *Food Chem.* **2010**, *123*, 642.
- [12] A. Rudnitskaya, E. Polshin, D. Kirsanov, J. Lammertyn, B. Nicolai, D. Saison, F.R. Delvaux, F. Delvaux, A. Legin, *Anal. Chim. Acta* **2009**, *646*, 111.
- [13] R. Bro, *Crit. Rev. Anal. Chem.* **2006**, *36*, 279.
- [14] X. Cetó, F. Céspedes, M. del Valle, *Microchim. Acta*, doi: 10.1007/s00604-012-0938-7.
- [15] M. Gay, C. Apetrei, I. Nevares, M. del Alamo, J. Zurro, N. Prieto, J.A. De Saja, M.L. Rodríguez-Méndez, *Electrochim. Acta* **2010**, *55*, 6782.
- [16] O. Oliveira, Jr., F. Pavinatto, C.L. Constantino, F. Paulovich, M. Oliveira, *Biointerphases* **2012**, *7*, 1.
- [17] V.E. Vinzi, *Handbook of Partial Least Squares: Concepts, Methods and Applications*, Springer, Berlin, DE, 2010.
- [18] P. Mitteroecker, F. Bookstein, *Evol. Biol.* **2011**, *38*, 100.
- [19] D. Meyer, F. Leisch, K. Hornik, *Neurocomputing* **2003**, *55*, 169.
- [20] F. Céspedes, E. Martínez-Fàbregas, S. Alegret, *TrAC-Trend. Anal. Chem.* **1996**, *15*, 296.
- [21] J.M. Gutiérrez, L. Moreno-Barón, M.I. Pividori, S. Alegret, M. del Valle, *Microchim. Acta* **2010**, *169*, 261.
- [22] X. Cetó, J.M. Gutiérrez, L. Moreno-Barón, S. Alegret, M. del Valle, *Electroanal.* **2011**, *23*, 72.
- [23] X. Cetó, J.M. Gutiérrez, M. Gutiérrez, F. Céspedes, J. Capdevila, S. Mínguez, C. Jiménez-Jorquera, M. del Valle, *Anal. Chim. Acta* **2012**, *732*, 172.
- [24] V. Martina, K. Ionescu, L. Pigani, F. Terzi, A. Ulrici, C. Zanardi, R. Seeber, *Anal. Bioanal. Chem.* **2007**, *387*, 2101.
- [25] M.L. Rodríguez-Méndez, V. Parra, C. Apetrei, S. Villanueva, M. Gay, N. Prieto, J. Martínez, J.A. de Saja, *Microchim Acta* **2008**, *163*, 23.
- [26] M. McKay, A.J. Buglass, C.G. Lee, Beer, in: A.J. Buglass (Ed.) *Handbook of Alcoholic Beverages*, John Wiley & Sons, Ltd, 2010, pp. 132.
- [27] R.A. Johnson, D.W. Wichin, *Applied multivariate statistical analysis*, Pearson Education, Harlow, GB, 2007.
- [28] D.J. Hand, *Discrimination and classification*, John Wiley, Chichester, UK, 1981.
- [29] B. Pérez, M. Pumera, M. del Valle, A. Merkoçi, S. Alegret, *J. Nanosci. Nanotechnol.* **2005**, *5*, 1694.
- [30] F. Despagne, D. Luc Massart, *Analyst* **1998**, *123*, 157R.
- [31] X. Cetó, F. Céspedes, M. del Valle, *Talanta* **2012**, *99*, 544.
- [32] L. Moreno-Barón, R. Cartas, A. Merkoçi, S. Alegret, M. del Valle, L. Leija, P.R. Hernandez, R. Muñoz, *Sensor Actuat. B-Chem.* **2006**, *113*, 487.
- [33] Q. Ouyang, J. Zhao, Q. Chen, H. Lin, *Food Chem.* *138*, 1320.
- [34] D. Kirsanov, O. Mednova, V. Vietoris, P.A. Kilmartin, A. Legin, *Talanta* **2012**, *90*, 109.
- [35] A.M. Molinaro, R. Simon, R.M. Pfeiffer, *Bioinformatics* **2005**, *21*, 3301.

Article 10

Cava wine authentication employing a voltammetric Electronic Tongue

X. Cetó, J. Capdevila, S. Mínguez and M. del Valle

manuscript

Cava wine authentication employing a voltammetric Electronic Tongue

Xavier Cetó¹, Josefina Capdevila¹, Santiago Mínguez¹ and Manel del Valle^{1,*}

¹ Sensors and Biosensors Group, Department of Chemistry, Universitat Autònoma de Barcelona, Edifici Cn, 08193 Bellaterra, Barcelona, SPAIN

² Estació de Viticultura i Enologia, INCAVI, Vilafranca del Penedes, Spain

Abstract

The application of an Electronic Tongue for the classification of cava samples based on the different ageing time is reported. As such, voltammetric responses were obtained from an array of six bulk-modified graphite-epoxy electrodes, which exhibited marked mix-responses towards the different types of samples analyzed. The obtained responses were then preprocessed employing Fast Fourier Transform (FFT) and the resulting coefficients were input to a Linear Discriminant Analysis model which allowed the discrimination of the different samples classes according to its vintage time. Beyond, the qualitative model, a quantitative model employing Artificial Neural Networks (ANNs) was built for the prediction of the total amount of sugar present in the samples, a parameter also used to classify cava samples. Therefore, allowing to correctly categorize cava wine samples according to legal requirements, besides detecting any handling and labelling error or fraud.

Keywords: Electronic Tongue; Linear Discriminant Analysis; Artificial Neural Network; voltammetric sensors; cava wine; ageing

1. Introduction

Cava is a sparkling wine under the denomination of origin status (DO) produced mostly in Catalonia, specially in *Penedès* region, following the traditional method (*méthode champenoise*) (Buxaderas & López-Tamames, 2003). That is the same than the one for French Champagne, but adapted to the local varieties of Catalonia. While in the past, it was sometimes referred as “Spanish champagne”, this is no longer permitted; and legally, nowadays, it is defined by law as a “quality sparkling wine produced in a designated region” (VECPDR). Unlike the French name which comes from the region where it is produced, the name “cava” comes from the place where the bottles are fermented; a dark and damp cave.

Given its DO status, cava is subjected to quality checks carried out by the technical services of the regulatory board to assess that producers comply with the specific technical conditions to obtain the right to produce both the *cuvée* and the *cava*, as well as their specific types labels (Ministerio de Agricultura, 20 November 1991). These controls include the whole production process, from the grapes to the bottling; i.e. they start in the vineyards assessing grapes varieties; continue controlling volumes of production and monitoring of its quality; and last until bottles stoppering and labelling in the cellar (cava must be marked with the month and year of the operation) so as to guarantee the minimum ageing periods required by the Regulations are fulfilled.

The characteristics of cava will depend on the differing variety combinations, vintages and also the different ageing periods (Stevenson, 2005). Mainly, it can be categorized depending on three factors. The first one deals with the grape varieties used or its “colour”, viz. it may be white or rosé. Secondly, it can also be classified according to its level of dryness; these ratings refer to the amount of sugar added during its production. In this case, cava can be categorized as *Brut nature* (<3 g·L⁻¹, no added sugar), *Extra Brut* (0-6 g·L⁻¹), *Brut* (<15 g·L⁻¹), *Extra Dry* (12-20 g·L⁻¹), *Dry* (17-35 g·L⁻¹), *Semi-dry* (33-50 g·L⁻¹) or *Sweet* (>50 g·L⁻¹) (Ministerio de Agricultura, 20 November 1991). Lastly, cava can also be distinguished according to the ageing time on yeast lees in bottle (*crianza*); namely, *cava* or *jove* (young; aged for 9 months or more in the same cellars), *Reserva* (>15 months) and *Gran Reserva* (>30 months) (Ministerio de Agricultura, 20 November 1991). Additionally, the indication “*Gran Reserva*” may only be used for *Brut Nature*, *Extra Brut* and *Brut*. In all cases, the label must necessarily show this indication and the year of the harvest. In this respect, Cava is the only sparkling wine which can bear this distinction.

Therefore, to qualify such distinctions, cava is subjected to special quality controls including, sensory and analytical examinations, a process during which regular inspections, stock statements and verification of movements are made. Nevertheless, despite all those controls, there is a demand of new analytical low-cost methods, to be used specially for screening purposes, with high sensitivity, good selectivity and fast response needed

* E-mail: manel.delvalle@uab.cat; tel: +34 93 5811017; fax: +34 93 5812379

to assess they guarantee quality control standards or to detect any fraud, either during or after its elaboration.

Particularly, this is critical in the case of the ageing, were unfortunately there is a lack of methods able to classify such samples given the difficulties to perform this assessment using classical analytical techniques as there is not any specific compound primarily responsible for their class; hence, being usually performed by a skilled sensory panel.

On that account, from the late 1990s a new concept in the field of sensors has appeared to tackle these problems: Electronic Tongues (ETs) (del Valle, 2010). These analytical systems are inspired in the taste sense of mammals, where a few receptors can respond to a large variety of substances. This principle is coupled with complex data treatment applied in the brain, which allows to quantify or to classify a large amount of substances (Esteban, Ariño, & Díaz-Cruz, 2006; Ni & Kokot, 2008). These biomimetic systems, opposed to conventional approaches, are directed towards the combination of low selectivity sensors array responses (or with cross response features) in order to obtain some added value in the generation of analytical information.

Within the context, ETs are arising as a promising approach to analyze liquid samples (P. Ciosek & Wroblewski, 2007; Riul Jr, Dantas, Miyazaki, & Oliveira Jr, 2010), which have already been successfully applied to the classification of several alcoholic beverages such as wines (Zeravik, Hlavacek, Lacina, & Skladal, 2009), cava wines (Cetó, Gutiérrez, Moreno-Barón, Alegret, & del Valle, 2011) or beers (Patrycja Ciosek & Wróblewski, 2006; Rudnitskaya et al., 2009) between others. Moreover, they have also been used for the detection of inappropriate handling practices (Gutiérrez, Moreno-Barón, Pividori, Alegret, & del Valle, 2010) or adulteration processes in wine (Parra, Arrieta, Fernández-Escudero, Rodríguez-Méndez, & De Saja, 2006); and even the quantification of several analytical parameters such as polyphenol content (Cetó, Céspedes, & del Valle, 2012; Gay et al., 2010).

The main goal of the present work is the application of an ET towards the classification of cava samples according to its ageing time; i.e. to distinguish *Gran Reserva*, *Reserva* and *jove* cavas in order to obtain an analytical tool able to detect any fraudulent activity. The method proposed herein couples an array of bulk modified voltammetric sensors which allowed the extraction of the samples characteristic fingerprint, with chemometric tools such as Fast Fourier Transform (FFT) for feature extraction and Linear Discriminant Analysis (LDA) or Artificial Neural Networks (ANNs) for, respectively, build the qualitative and quantitative prediction models.

2. Experimental

2.1 Samples under study

A total set of 65 cava wine samples were analyzed. Those samples were selected according to its type, taking into account vintage time according to the regulations of *Consell Regulador del Cava* (Ministerio de Agricultura, 20 November 1991): “*Jove*” cavas (9 to 15 months), “*Reserva*” (15 to 30 months) and “*Gran Reserva*” (*more*

than 30 months). Thus, the set was formed by 16 *Gran Reserva*, 24 *Reserva* and 25 *jove* cava wines samples. Additionally, one Champagne (French) sample was included in the set; which was replicated four times, considering each replicate as an independent sample.

2.2 Reagents and solutions

All reagents used were analytical grade and all solutions were prepared using deionised water from a Milli-Q system (Millipore, Billerica, MA, USA). Cobalt (II) phthalocyanine, polyaniline, polypyrrole, copper and platinum nanoparticles (<50 nm) were purchased from Sigma-Aldrich (St. Louis, MO, USA).

2.3 Reference methods

To further asses ET capabilities in the analysis of cava wine samples, those were also analyzed by classical methods to extract additional quantitative analytical information that may complement ET qualitative response if their level could be modelled properly. Particularly, two parameters were considered: the total amount of sugar which is related to samples level of dryness and dry weight which is related to all the compounds added or present in the wine.

2.3.1 Total sugars

The analysis of total sugar content was made by infrared spectrometry with Fourier transformation (FTIR) with a WineScan FT 120 (FOSS, Barcelona, Spain), which was previously calibrated according to official OIV methods (Office International de la Vigne et du Vin (OIV), 2005).

2.3.2 Dry weight

Dry weight was calculated indirectly from the specific density and the alcohol by volume (abv) degree of the wine according to regulated methods (Commission Regulation (EEC), 17 September 1990).

2.4 Electrochemical measurements

2.4.1 Electrode array

An array of 6 graphite-epoxy voltammetric sensors made with different modifiers added to the bulk mixture was selected according with previous studies in our laboratory with cava wines and wines (Cetó, Gutiérrez, et al., 2012; Cetó et al., 2011).

Standard graphite-epoxy composites were prepared using 50 µm particle size graphite powder (BDH laboratory Supplies, UK) and Epotek H77 resin and hardener (both from Epoxy Technology, USA) (Céspedes, Martínez-Fàbregas, & Alegret, 1996). In this manner, five of them were modified by adding components as nanoparticles of copper and platinum, or conducting polymer in powder like polyaniline and polypyrrole – one component per sensor.

2.4.2 Measurements cell

The measurement cell was formed by the 6-sensor voltammetric array and a reference double junction

Ag/AgCl electrode (Thermo Orion 900200, Beverly, MA, USA) plus a commercial platinum counter electrode (Model 52-67, Crison Instruments, Barcelona, Spain). Voltammetric measurements were taken using a 6-channel AUTOLAB PGSTAT20 (Ecochemie, Netherlands), in a multichannel configuration, using GPES Multichannel 4.7 software package.

2.4.3 Measuring procedure

Prior to perform cava samples measurements, electrodes were first cycled in saline solution in order to get stable voltammetric responses, ensuring reproducible responses from the ET array. Afterwards, a sample volume of 25 ml of cava wine was directly used for each measurement, without any sample pretreatment. Once cava bottles were opened, measures were carried out after a few minutes (ca. 5 min at room temperature) in order to avoid the formation of bubbles into the electrode surface. That is, no specific removal of oxygen and CO₂ was done, but minimization of distortions by air bubbles was attempted.

Then a complete voltammogram was registered for each sample by cycling the potential between -1.0V and +1.3V vs. Ag/AgCl with a step potential of 9 mV and a scan rate of 100 mV·s⁻¹. Besides, all experiment measurements were carried out without performing any physical surface regeneration of the working electrodes. In this sense, an electrochemical cleaning stage was carried out between each measurement to prevent the accumulative effect of impurities on the working electrode surfaces (Cetó et al., 2011). For this, a conditioning potential of +1.5 V was applied during 40 s in a cell containing 25 ml of distilled water. This step ensures that electrode fouling is minimized, recovering the original signal (baseline) after each measurement.

2.5 Data processing

Chemometric processing of the data was done in MATLAB 7.1 (MathWorks, Natick, MA) using specific routines written by the authors, and also Neural Network Toolboxes (v.4.0.6). Sigmaplot (Systat Software Inc, California, USA) was used for graphic representations of data and results.

On the one hand, Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) were both used for the qualitative analysis of the results; while on the other hand, quantitative analysis was achieved by means of Artificial Neural Networks (ANNs). Afterwards, once the corresponding model is generated (either the LDA or ANN model), it can be applied to the prediction of the outputs (either qualitative or quantitative values) from new samples by simply introducing the readings of the sensors for those to the model, hence obtaining a powerful analytical tool for rapid analysis of cava wine samples.

In all cases, prior to build the models, in order to reduce the large data set generated for each sample, a preprocessing stage employing the Fast Fourier Transform (FFT) was used to compress the original sensors information (Cetó, Céspedes, et al., 2012).

PCA allows the projection of the information carried by the original variables onto a smaller number of underlying (“latent”) variables called principal components

(PCs) with new coordinates called scores, obtained after data transformation (Jolliffe, 2002). Then by plotting the PCs, one can view interrelationships between different variables, and detect and interpret sample patterns, groupings, similarities or differences.

LDA is closely related to PCA as they both look for linear combinations of variables which best explain the data (Coomans, Massart, & Kaufman, 1979). However, the main difference lies in the fact that LDA explicitly attempts to model the difference between the classes of data; while PCA does not take into account any difference between classes, providing only a visualization of the variability of the data, not implying any clustering of it, although formation of groups of samples can be a possible result. That is, the use of a supervised method instead of an unsupervised one in the machine learning task (Ghosh & Konar, 2013). For this reason, PCA is normally used just as a visualization tool that permits to check if the samples group together in classes, and cannot be considered as a properly pattern recognition method; whereas to be used as a classifier it must be coupled with a modelling tool such as ANNs. On that account, another option is the use of an actual classifier such as LDA; which provides both a visualization of samples clustering and builds the qualitative model that acts as the classifier.

ANNs were designed to mimic the function of the human brain and have been applied to quantitative and qualitative analysis during the last decades (del Valle, 2010). They consist of a number of simple processing units (or neurons) linked by weighted modifiable interconnections (Despaigne & Luc Massart, 1998). As LDA, and imitating the biological learning, they require a training process where the weights of those connections are adjusted, and build a model that will allow to carry out the prediction of the desired parameters (either qualitative or quantitative). Their main advantages include a high modelling performance, being specially suited to non-linear sensor responses, and the fact of being very much related to human pattern recognition.

3. Results and Discussion

3.1 Voltammetric responses and feature extraction

Under the conditions and procedure described in section 2.4, the set of samples was measured employing the ET array, obtaining a whole cyclic voltammogram for each of the sensors. An extract of the results corresponding to the obtained voltammograms for the different types of cava analyzed is shown in Figure 1. As can be seen, complex and highly overlapped signals are obtained along the whole voltammogram, with differentiated signals obtained for the different kinds of sensors. Besides, it could be seen how some samples present very similar voltammetric responses in specific regions along the voltammogram; also, that analogous resemblances are repeated along the whole recorded signal, but between different samples.

E.g. in Figure 1A it could be seen how *Gran Reserva*, *Reserva* and *jove* cavas present clearly differentiated response for the cathodic wave in the region from -0.5 V to -1.0V; while *Champagne* sample presents a more similar response to *jove* cavas, although still showing

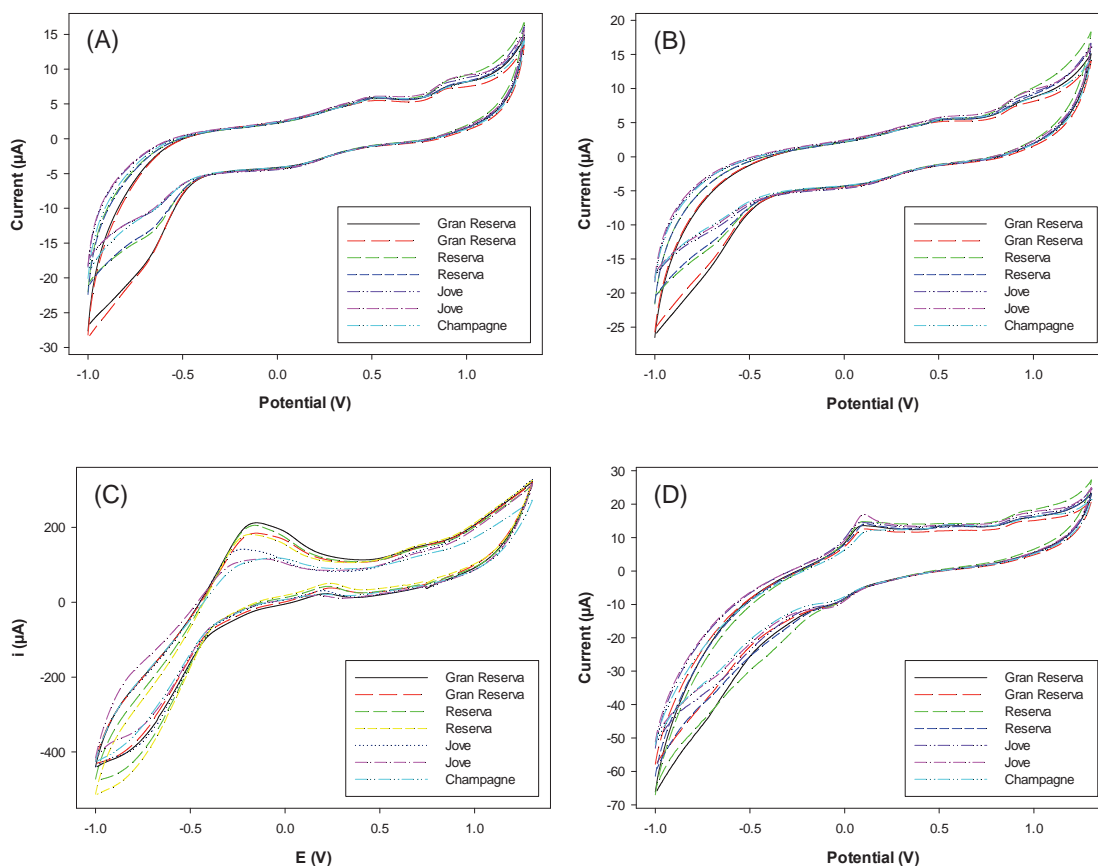


Figure 1. Voltammetric responses obtained with the different sensors forming the ET array for certain arbitrary cava wine samples are shown as example. Signals provided correspond to: (A) graphite-epoxy sensor, (B) cobalt (II) phthalocyanine, (C) platinum nanoparticle and (D) copper nanoparticle modified sensors.

slightly differentiated response, but in the region from -0.75V to -1.0V . The latter behaviour could be somehow expected given both types of sparkling wine are produced following the same procedure, but employing different grape varieties; thus, expecting more resemblances with *jove* cavas as, albeit not existing such classification for *Champagne*, they were aged for the same period time. Whereas differentiated responses shown in the anodic wave could be attributed to the polyphenolic content of samples (Cetó, Gutiérrez, et al., 2012), although no such resemblances between samples for the same class are observed. Similar trends were observed for other sensors such as cobalt (II) phthalocyanine (Figure 1B) or polypyrrole. On the other hand, platinum (Figure 1C) and copper nanoparticles (Figure 1D) sensors bring completely new information to the system, while still showing some previous discussed resemblances; in this aspect, it should be noticed the differentiated response observed for platinum sensor at the anodic region close to $+1\text{V}$ for champagne sample, indicating that its discrimination might be also possible.

This situation, where the different sensors display marked distinct features for the different samples is an ideal departure point for developing an ET approach. Nevertheless, to fully exploit all the information obtained from each voltammogram, a preprocessing compression step was used to reduce the high dimensionality of the recorded signals (samples \times sensors \times potentials) while preserving the relevant information. In addition to data reduction,

compression is intended for extracting significant features from the departure information, while gaining many advantages in the modelling stage (Cetó, Céspedes, & Valle, 2013). As stated, in our case, this variable reduction was achieved by means of FFT, which allowed a compression of the original data up to 96.9% without any loss of relevant information.

3.2 Classification of cava samples

To confirm the initial trends seen in the voltammetric responses and to corroborate its analytical content, the corresponding compressed voltammograms were processed applying PCA analysis. Thus, a preliminary recognition model was performed in order to display the different kinds of cava, according to measurements done. The PCA plot with the different clusters is shown in Figure 2; where with the three first PCs, the explained accumulated variance was ca. 97.8%. The patterns in the figure evidence that cava samples are grouped based on ageing time, with well established clusters clearly separating the main classes of samples corresponding to: *Gran Reserva*, *Reserva*, *jove* cavas and *Champagne*.

To estimate the success rate in sample identification, a LDA model was built to act as an actual classifier given PCA only provides a visualization of the different grouping regions. In this case, obtained Fourier coefficients were used as inputs to the model while the target of the model was to distinguish between the different types of cava and

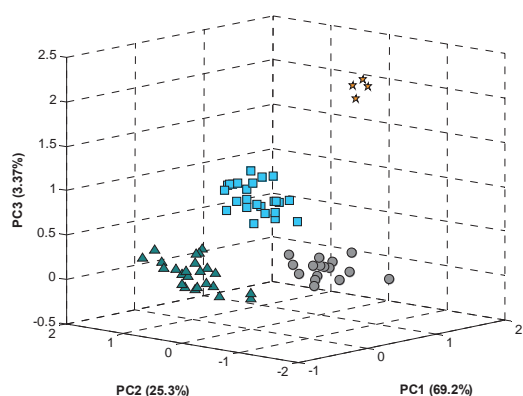


Figure 2. Score plot of the first three components obtained after PCA analysis of the cava wine samples: (●) *Gran Reserva*, (▲) *Reserva*, (■) *Joves* and (★) *Champagne*.

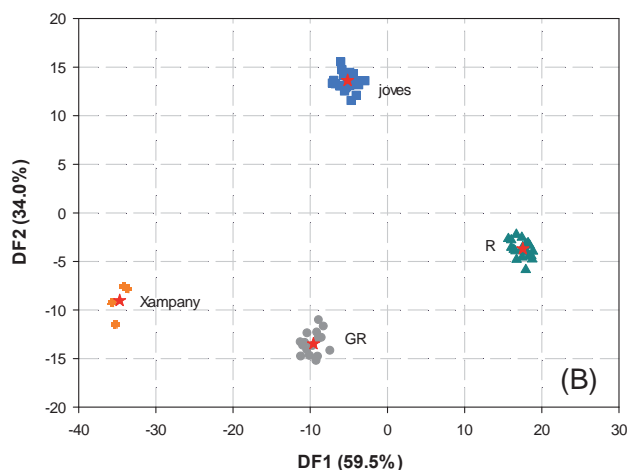
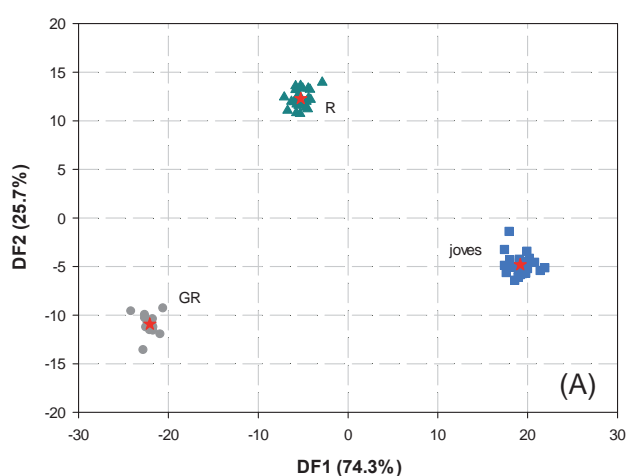


Figure 3. Score plot of the functions obtained after LDA analysis of the cava wine samples according to the ageing time: (●) *Gran Reserva*, (▲) *Reserva*, (■) *Joves* and (★) *Champagne*. Also, the centroid of each class is plotted (★).

Table 1. Confusion matrix built according cava wine vintage obtained using LDA model and k-fold cross-validation method with k=6.

| | | Predit | | |
|---------|--------------|--------|----|-------|
| | | GR | R | Joves |
| Esperat | Gran Reserva | 16 | 0 | 0 |
| | Reserva | 0 | 24 | 0 |
| | Joves | 0 | 0 | 25 |

Champagne. As expected from previous PCA graph, clear discrimination for the four classes considered was achieved as can be seen in Figure 3. Additionally, in this case two different models were built; one including *Champagne* samples (Figure 3A) and one without them (Figure 3B).

LDA model was trained and validated using k-fold cross-validation method with k=6. Afterwards, confusion matrix was built (Table 1), allowing to calculate also the performance of the model by means of three different indicators: classification rate, sensitivity and specificity. On the one hand, the first one corresponds to the ratio between the number of samples correctly classified and the total number of samples. On the other hand, sensitivity is calculated as the percentage of objects of each class identified by the classifier model, and specificity as the percentage of objects from different classes correctly rejected by the classifier model; averaging those for the classes. In this case, reaching 100% for the three statistical indicators.

3.3 Estimation of cava indexes

As a further ability of electronic tongue sensors system, a quantitative application in order to predict compounds of relevant significance in cava wines was also attempted. Concretely, quantification of total sugars and dry weight was attempted. As stated, cava can be mainly classified from three different points of view, viz. its “colour” (white or rosé), level of dryness and vintage. Thus, it would be very useful, from the analytical point of view, to be able to quantify the total amount of sugar present in the cava at the same time than the ageing (time of aging on yeast lees) is identified as this allows to kill two birds with one stone. That is, on one hand it allows to classify also the samples according to its level of dryness; while on the other hand it might also allow to confirm that the amount of sugar added is below the amount required by the DO to issue the *Gran Reserva* label (Ministerio de Agricultura, 20 November 1991).

As before, concentrations of the sought compounds were modelled from the set of voltammetric responses, previously compressed with FFT, but using ANNs as the modelling tool instead of LDA; therefore, building a quantitative model rather than a qualitative one. After some preliminary tests, the final architecture of the ANN model had 96 neurons (6 sensors x 16 coeffs. obtained from the FFT analysis) in the input layer, 4 neurons and *tansig* transfer function in the hidden layer and two neurons and *tansig* transfer function in the output layer, viz. the total amount of sugar and the dry weight.

Accuracy of the generated model was evaluated employing 76% of the data (48 samples) for training the model and tested with the remaining 24% (15 samples). Afterwards, comparison graphs of predicted vs. expected concentrations were built, both for train and test subsets, to check the prediction ability of the ANN (Figure 4). As can be observed, a satisfactory trend is obtained for both indexes, with regression lines almost indistinguishable from the theoretical ones. Additionally, the obtained comparison parameters were calculated and close to the ideal values, with intercepts near to 0 and slopes and correlation coefficients around 1, meaning that there are no significant differences between the values predicted by the

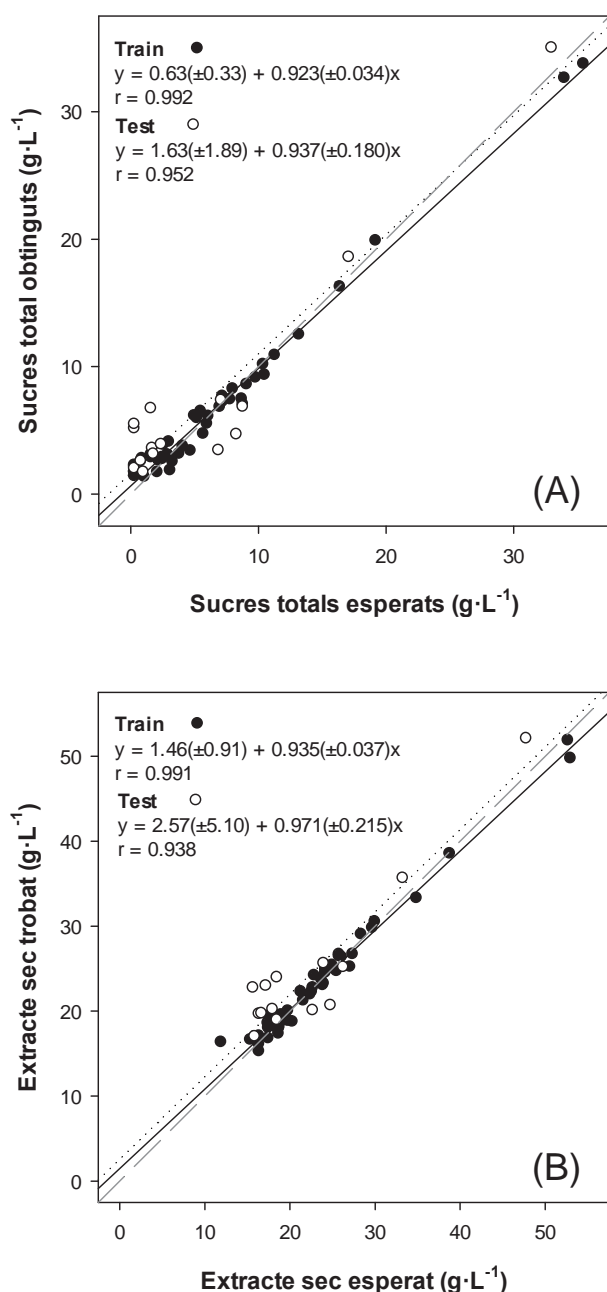


Figure 4. Modelling ability of the optimized ANN. Sets adjustments of obtained vs. expected concentrations for (A) total sugars and (B) dry weight content, both for training (●, solid line) and testing subsets (○, dotted line). Dashed line corresponds to theoretical diagonal line.

ANN model and those expected and provided by the reference method.

Therefore, it has been demonstrated that the combination of voltammetric measurements with chemometric tools is an analytically powerful approach for the characterization of cava samples, either from the qualitative or quantitative points of view. In this fashion, the approach presented herein could be considered as an attractive alternative to other classical methods or the use of sensory panels, being really suitable for screening purposes

4. Conclusions

With this and the previous results, proposed ET has demonstrated its powerful applications in wine making

industry in the qualitative analysis of cava wines according to two main categorization parameters; i.e. level of dryness and ageing time (time of aging on yeast lees). Moreover, in this case we have confirmed the trend previously observed, were ET seemed capable to beyond predicting class derived from the amount of sugar, it might also be able to quantitatively predict it.

On the other hand, the use of chemometric tools have proved to be an easy procedure to enhance the performance of electroanalysis providing a powerful tool for screening or detection of frauds in food analysis. Concretely, LDA allowed building qualitative models able to classify cava wine samples according to its vintage while ANNs allowed quantifying its level of dryness; two parameters that determine the type of cava according to DO requirements.

Additionally, given its biomimetic behaviour, it presents a straightforward solution to the absence of the knowledge about which compounds are primarily responsible for the style that hinders the development of analytical procedures for cava categorization according to its ageing time. Therefore, representing the obtaining of an alternative approach able to perform such categorization, really suitable for screening purposes, which performance characteristics may satisfy food industry requirements of precision, rapidity, sensitivity, simplicity and low cost required to be considered as a useful analytical tool.

Acknowledgments

Financial support for this work was provided by Spanish Ministry of Science and Innovation, MCINN (Madrid) trough project CTQ2010-17099 and by program ICREA Academia from Generalitat de Catalunya. X. Cetó thanks the support of Dept. d'Innovació, Universitat i Empresa de la Generalitat de Catalunya for the predoctoral grant.

References

- Buxaderas, S., & López-Tamames, E. (2003). WINES | Production of Sparkling Wines. In C. Editor-in-Chief: Benjamin (Ed.), *Encyclopedia of Food Sciences and Nutrition (Second Edition)* (pp. 6203-6209). Oxford: Academic Press.
- Céspedes, F., Martínez-Fàbregas, E., & Alegret, S. (1996). New materials for electrochemical sensing I. Rigid conducting composites. *TrAC Trends in Analytical Chemistry*, 15 (7), 296-304.
- Cetó, X., Céspedes, F., & del Valle, M. (2012). BioElectronic Tongue for the quantification of total polyphenol content in wine. *Talanta*, 99 (0), 544-551.
- Cetó, X., Céspedes, F., & Valle, M. (2013). Comparison of methods for the processing of voltammetric electronic tongues data. *Microchimica Acta*, doi: 10.1007/s00604-012-0938-7.
- Cetó, X., Gutiérrez, J. M., Gutiérrez, M., Céspedes, F., Capdevila, J., Mínguez, S., Jiménez-Jorquera, C., & del Valle, M. (2012). Determination of total polyphenol index in wines employing a voltammetric electronic tongue. *Analytica Chimica Acta*, 732 (0), 172-179.

- Cetó, X., Gutiérrez, J. M., Moreno-Barón, L., Alegret, S., & del Valle, M. (2011). Voltammetric Electronic Tongue in the Analysis of Cava Wines. *Electroanalysis*, 23 (1), 72-78.
- Ciosek, P., & Wroblewski, W. (2007). Sensor arrays for liquid sensing - electronic tongue systems. *Analyst*, 132 (10), 963-978.
- Ciosek, P., & Wróblewski, W. (2006). The recognition of beer with flow-through sensor array based on miniaturized solid-state electrodes. *Talanta*, 69 (5), 1156-1161.
- Commission Regulation (EEC). (17 September 1990). Community methods for the analysis of wines In *Official Journal L 272* (Vol. No 2676/90, pp. 0001 - 0192): Publication Office of the European Union.
- Coomans, D., Massart, D. L., & Kaufman, L. (1979). Optimization by statistical linear discriminant analysis in analytical chemistry. *Analytica Chimica Acta*, 112 (2), 97-122.
- del Valle, M. (2010). Electronic Tongues Employing Electrochemical Sensors. *Electroanalysis*, 22 (14), 1539-1555.
- Despagne, F., & Luc Massart, D. (1998). Neural networks in multivariate calibration. *Analyst*, 123 (11), 157R-178R.
- Esteban, M., Ariño, C., & Díaz-Cruz, J. M. (2006). Chemometrics for the analysis of voltammetric data. *TrAC Trends in Analytical Chemistry*, 25 (1), 86-92.
- Gay, M., Apetrei, C., Nevares, I., del Alamo, M., Zurro, J., Prieto, N., De Saja, J. A., & Rodríguez-Méndez, M. L. (2010). Application of an electronic tongue to study the effect of the use of pieces of wood and micro-oxygenation in the aging of red wine. *Electrochimica Acta*, 55 (22), 6782-6788.
- Ghosh, S., & Konar, A. (2013). An Overview of Computational Intelligence Algorithms. In *Call Admission Control in Mobile Cellular Networks* (Vol. 437, pp. 63-94): Springer Berlin Heidelberg.
- Gutiérrez, J. M., Moreno-Barón, L., Pividori, M. I., Alegret, S., & del Valle, M. (2010). A voltammetric electronic tongue made of modified epoxy-graphite electrodes for the qualitative analysis of wine. *Microchimica Acta*, 169 (3-4), 261-268.
- Jolliffe, I. T. (2002). *Principal Component Analysis*. New York, US: Springer New York.
- Ministerio de Agricultura, Pesca y Alimentación. (20 November 1991). Orden de 14 de noviembre de 1991 por la que se aprueba el Reglamento de la Denominación «Cava» y de su Consejo Regulador. In *BOE-A-1991-28079* (pp. 37587-37593). Madrid: Secretaria General Técnica.
- Ni, Y., & Kokot, S. (2008). Does chemometrics enhance the performance of electroanalysis? *Analytica Chimica Acta*, 626 (2), 130-146.
- Office International de la Vigne et du Vin (OIV). (2005). *Recueil des méthodes internationales d'analyse des vins et des moûts* (2005 ed. Vol. 1). Paris: OIV.
- Parra, V., Arrieta, Á. A., Fernández-Escudero, J.-A., Rodríguez-Méndez, M. L., & De Saja, J. A. (2006). Electronic tongue based on chemically modified electrodes and voltammetry for the detection of adulterations in wines. *Sensors and Actuators B: Chemical*, 118 (1-2), 448-453.
- Riul Jr, A., Dantas, C. A. R., Miyazaki, C. M., & Oliveira Jr, O. N. (2010). Recent advances in electronic tongues. *Analyst*, 135 (10), 2481-2495.
- Rudnitskaya, A., Polshin, E., Kirsanov, D., Lammertyn, J., Nicolai, B., Saison, D., Delvaux, F. R., Delvaux, F., & Legin, A. (2009). Instrumental measurement of beer taste attributes using an electronic tongue. *Analytica Chimica Acta*, 646 (1-2), 111-118.
- Stevenson, T. (2005). *The Sotheby's wine encyclopedia*. London: DK.
- Zeravik, J., Hlavacek, A., Lacina, K., & Skladal, P. (2009). State of the Art in the Field of Electronic and Bioelectronic Tongues - Towards the Analysis of Wines. *Electroanalysis*, 21 (23), 2509-2520.

Article 11

Voltammetric BioElectronic Tongue for the analysis of phenolic compounds in
rosé cava wines

X. Cetó, J. Capdevila, S. Mínguez and M. del Valle
manuscrit

Voltammetric BioElectronic Tongue for the analysis of phenolic compounds in rosé cava wines

Xavier Cetó¹, Josefina Capdevila¹, Santiago Mínguez¹ and Manel del Valle^{1,*}

¹ Sensors and Biosensors Group, Department of Chemistry, Universitat Autònoma de Barcelona, Edifici Cn, 08193 Bellaterra, Barcelona, SPAIN

² Estació de Viticultura i Enologia, INCAVI, Vilafranca del Penedes, Spain

Abstract

A BioElectronic Tongue (BioET) based on a sensor array comprising 4 voltammetric sensors plus pattern recognition and multivariate calibration data processing tools was applied towards the analysis of cava rosé wines. A total of 20 different samples were analyzed using cyclic voltammetry without any sample pretreatment. Obtained responses were preprocessed employing windowed slicing integral method in order to compress and extract significant features from the recorded data. Extracted coefficients were then evaluated by means of Principal Component Analysis to visualize some initial patterns, while quantification of different phenolics indexes was achieved by an Artificial Neural Network (ANN) model. In this manner, three different classical indexes related to total phenolic content (i.e. I_{280} , I_{320} and Folin-Ciocalteu index) plus two indexes related to more specific families of those (i.e. total tannins and anthocyanins content) were correlated with sensors responses.

Keywords: Electronic Tongue; voltammetric biosensors; cava wine; phenolic compounds; tannin; anthocyanin

1. Introduction

Phenolic compounds in wine include a large group of chemical compounds that affect its taste, colour and mouthfeel [1]. These compounds include phenolic acids, stilbenoids, flavonols, dihydroflavonols, anthocyanins, flavanol monomers (catechins) and flavanol polymers (proanthocyanidins); which can be broadly separated into two main categories: flavonoids and non-flavonoids. Flavonoids include the anthocyanins and tannins which contribute to the colour and mouthfeel of the wine [2]; while non-flavonoids include the stilbenoids such as resveratrol, and phenolic acids such as benzoic, caffeic and cinnamic acids.

Beyond its contribution to the wine sensorial features (e.g. to its colour, body and astringency), most of these compounds are powerful antioxidants with great health benefits derived from their action as free radical scavengers and inhibitors of lipoprotein oxidation [3]. In this sense, there are evidences that wine antioxidant properties are due to phenolic components, which total content is directly correlated with their antioxidant capacity [4], and that wine without these components loses these properties.

A part from total phenolic content, tannins and anthocyanins are two classes of flavonoids with huge importance, especially in the case of rosé and red wines, as they have a clear influence in its colour and mouthfeel, particularly in its astringency [5]. On the one hand, tannins can affect the colour, ageing ability and texture of the wine. While tannins cannot be smelled or tasted, they can be

perceived during wine tasting by the tactile drying sensation and sense of bitterness that they can leave in the mouth [5]. On the other hand, anthocyanins are odourless and nearly flavourless compounds, contributing to wine taste as a moderately astringent sensation; besides the concentration of those dictates the colour of the wine [5].

As a result, given its importance, several methods to quantify phenolic compounds (either total content or individual identification) are reported in the literature [6]; over those, most common ones include chromatography or spectrophotometry. However, these procedures usually require additional preparative steps, the use of heavy, dedicated laboratory instruments and are not suitable for on-site analysis.

On that account, biosensors are arising as an alternative to traditional laboratory techniques given their low cost and their ease of use to carry out on-field analyses. In the case of phenolic compounds, previous attempts are based on the development of amperometric biosensors based on the immobilization of basically three different enzymes: laccase, tyrosinase and/or peroxidase [7]. Although the applicability of biosensors to the analysis of antioxidant compounds is promising and represents an attractive alternative for their detection, further work is required to avoid and/or take into account the interference problem [8].

To overcome such difficulties, chemometric tools such as Principal Component Analysis or Artificial Neural Networks (ANNs) can be used [9, 10]. This coupling consists in the integration of an array of sensors (with marked mix-response towards the desired species) and a

* E-mail: manel.delvalle@uab.cat; tel: +34 93 5811017; fax: +34 93 5812379

chemometric processing tool (able to interpret and extract meaningful data from the complex readings); an approach known as BioElectronic Tongue [11].

The present work reports the application of a voltammetric BioET towards the analysis of phenolic compounds in cava rosé wines, for the prediction of both global phenolic content and total classes content. As such, it combines the responses from an array of voltammetric biosensors, plus an advanced response model employing a specifically trained Artificial Neural Network, with pretreatment of data employing *windowed slicing integral*.

2. Experimental

2.1 Reagents and solutions

All reagents used were analytical grade and all solutions were prepared using deionised water from a Milli-Q system (Millipore, Billerica, MA, USA). Tyrosinase from mushroom (EC 1.14.18.1, 4276 U·mg⁻¹), laccase from *Trametes versicolor* (EC 1.10.3.2, 21 U·mg⁻¹), copper nanoparticles (< 50 nm) were purchased from Sigma-Aldrich (St. Louis, MO, USA). Folin-Ciocalteu's reagent and sodium carbonate were purchased from Panreac Química (Barcelona, Spain).

2.2 Samples under study

A total set of 20 rosé cava wine samples were analyzed. Those samples were selected so as to obtain a set of samples with sufficiently differentiated total phenolic content and grape varieties (e.g. *Pinot noir*; *Grenache*; *Mourvèdre*, also known as *Monastrell*; *Trepát* or *Xarel-lo*). Moreover, although all the samples considered were from Catalonia region, different wine regions inside this were also considered; i.e. different denomination of origin (D.O.). In this context, Table 1 summarizes detailed information about the brand of the samples under study as well as analytical information of those.

2.3 Standard methods

For comparison purposes, different phenolics indexes (related to different classes of them) were analyzed by classical methods for both subsets to extract additional quantitative analytical information that may complement ET qualitative response if their level could be modelled properly. On the one hand, three indexes related to its total content were evaluated, i.e. Folin-Ciocalteu, I_{280} and I_{320} [12]; while on the other hand, total tannins and anthocyanins (two classes of phenolic compounds) were also quantified by standard procedures.

2.3.1 Folin-Ciocalteu index

Firstly, the Folin-Ciocalteu (FC or Folin) index of the cava wines was determined spectrophotometrically [13]. FC method is a colorimetric assay measuring the amount of phenol needed to inhibit the oxidation of the Folin-Ciocalteu reagent (a mixture of phosphomolybdate and phosphotungstate, which are reduced to the respective oxides).

The FC test was carried out according to the following procedure: 1 mL of sample (cava wines were previously diluted 1:100 or 1:50), 6 mL of deionized water, 0.5 mL of Folin-Ciocalteu reagent and 2 mL of a 20% sodium carbonate solution were added in this order to a 10 mL beaker and diluted to volume with deionized water. The resulting solution was stirred and allowed to react for half an hour at room temperature in darkness. The absorbance was then read at 760 nm using a spectrophotometer PerkinElmer Lambda 20 UV/VIS (MA, USA). Total phenolic content, expressed in gallic acid equivalents, was evaluated from the absorbance value by interpolation into solutions, multiplying the resulting value by 10 and by the proper dilution factor. Different dilution factors were applied given when carrying out spectrophotometric measurements absorbance value should be around 0.3 [13].

Table 1. Detailed information of the cava wine samples under study.

| Sample | Folin index (mg·L ⁻¹) | I_{280} (arb. unit) | I_{320} (arb. unit) | Tannins (mg·mL ⁻¹) | Anthocyanins (mg·mL ⁻¹) | Producer |
|--------|--------------------------------------|--------------------------|--------------------------|-----------------------------------|--|-----------------------|
| S1 | 166 | 8.50 | 5.41 | 100 | 9 | Llopart |
| S2 | 356 | 18.55 | 7.87 | 393 | 26 | Colet |
| S3 | 170 | 7.84 | 4.21 | 189 | 11 | Berberana |
| S4 | 154 | 8.51 | 5.36 | 80 | 11 | Freixenet |
| S5 | 192 | 11.12 | 5.08 | 119 | 14 | Dibon |
| S6 | 242 | 12.04 | 5.98 | 281 | 22 | Saint Clair |
| S7 | 246 | 10.99 | 6.30 | 216 | 17 | Canals & Munné |
| S8 | 208 | 11.19 | 5.40 | 244 | 23 | Castell Peralada |
| S9 | 147 | 8.84 | 4.39 | 85 | 9 | Don Román |
| S10 | 242 | 11.79 | 6.51 | 229 | 17 | Mont-Ferrant |
| S11 | 185 | 8.96 | 5.10 | 116 | 5 | Codorniu |
| S12 | 128 | 7.41 | 3.81 | 50 | 8 | Canals Nadal |
| S13 | 162 | 8.89 | 4.59 | 101 | 10 | Castell de la Comanda |
| S14 | 139 | 8.49 | 3.96 | 150 | 7 | Oriol Rossell |
| S15 | 208 | 10.85 | 6.37 | 123 | 4 | Titiana |
| S16 | 204 | 9.25 | 4.27 | 173 | 5 | Vallformosa |
| S17 | 147 | 6.90 | 4.23 | 110 | 0 | Raventós i Blanc |
| S18 | 246 | 11.88 | 6.06 | 173 | 9 | Fuchs de Vidal |
| S19 | - | 8.15 | 4.44 | - | - | Cavas Hill |
| S20 | - | - | - | - | - | INCAVI |

2.3.2 I_{280} and I_{320} indexes

Additionally, for the determination of the total content of phenolic compounds in cava wines the indexes I_{280} and I_{320} were also considered. While FC is the recommended reference method, the use of those two indexes has arisen as an alternative to the first given its shorter time of analysis and simplicity. Although between them, the most common one is the I_{280} index, I_{320} index is also considered to be important as sometimes there could be a shift on the maximum of absorbance, especially in white wines [14]. Additionally, the latter is related to hydroxycinnamate compounds, a group of phenolic compounds [12].

For its determination, cava wine was diluted with water (1:100 or 1:50) and the absorbance was measured directly at 280 nm and 320 nm. The values of I_{280} and I_{320} for each sample were given as the absorbance multiplied by the proper dilution rate.

2.3.3 Tannins and anthocyanins

Lastly, tannins and anthocyanins, which are two main classes of phenolic compounds belonging to flavonoids group, were also analyzed given its clear influence in the colour and mouthfeel of the wine [5].

For the detection of tannins, the analytical method applied was the acid butanol assay [14]. This method is based on the acid-catalysed oxidative cleavage of the C-C interflavanic bond of proanthocyanindins in butanol-HCL. On the other hand, total anthocyanins were determined by the colour variation in function of pH [15].

2.4 Electrochemical measurements

2.4.1 Voltammetric sensor array

The voltammetric array was formed by four graphite-epoxy voltammetric sensors prepared using bare graphite C and adding different modifiers such as tyrosinase, laccase and copper nanoparticles to the bulk mixture – one component per electrode plus a blank electrode without any modifier.

Electrodes were prepared following the conventional methodology previously described [16]. Electrode fabrication begins with the preparation of the composite paste. For this, resin EpoTek H77 (Epoxy Technology, Billerica, MA, USA) and its corresponding hardener compound were mixed in the ratio 20:3 (w/w); afterwards a 15% (w/w) of graphite (50 μm BDH Laboratory Supplies) and a 2% (w/w) of the modifier (either the enzyme or the catalyst) were added to the previous mixture before hardening. Then, it was manually homogenized for 60 min, and afterwards, the paste was allowed to harden for seven days at 40 °C. Finally, electrode surface was polished with different sandpapers of decreasing grain size, with a final electrode area of 28 mm².

2.4.2 Measuring procedure

The amperometric measurement cell was formed by the 4-sensor voltammetric array and a reference double junction Ag/AgCl electrode (Thermo Orion 900200, Beverly, MA, USA) plus a commercial platinum counter electrode (Model 52-67, Crison Instruments, Barcelona, Spain). Cyclic Voltammetry measurements were taken

using a 6-channel AUTOLAB PGSTAT20 (Ecochemie, Netherlands), in a multichannel configuration, using GPES Multichannel 4.7 software package. For this, potential was cycled between -0.4 V and +0.8 V vs. Ag/AgCl, with a scan rate of 100 mV·s⁻¹ and a step potential of 9 mV.

Electroanalytical experiments were carried out at room temperature (25 °C) under quiescent conditions, and without any pretreatment or dilution of the sample. Prior to perform cava samples measurements, electrodes were first cycled in saline solution in order to get stable voltammetric responses, ensuring reproducible responses from the BioET array. Apart, all experiments were carried out without performing any physical surface regeneration of the working electrodes. In order to prevent the accumulative effect of impurities on the working electrode surfaces, an electrochemical cleaning stage was done between each measurement applying a conditioning potential of +1.0 V during 40 s after each experiment, in a cell containing 25 mL of distilled water [17].

2.5 Data processing

Chemometric processing of the data was done in MATLAB 7.1 (MathWorks, Natick, MA) using specific routines written by the authors, and also Neural Network Toolbox (v.4.0.6). Sigmaplot 2000 (Systat Software Inc, California, USA) was used for graphic representations of data and results.

The whole cyclic voltammograms obtained from each sensor from the proposed sensor array were included in the data processing stage. In order to reduce the multidimensional data matrix generated in each measurement, a preprocessing stage employing the *windowed slicing integral* method was used [18]. In this way, the corresponding compressed voltammograms were processed employing either Principal Component Analysis (PCA) or Artificial Neural Networks (ANNs) models. The first allowed to visualize some initial patterns, while the second one allowed the quantification of different analytical parameters related to phenolic content of cava samples.

Principal Component Analysis (PCA) allows the projection of the information carried by the original variables onto a smaller number of underlying (“latent”) variables called principal components (PCs) with new coordinates called scores, obtained after data transformation. Then by plotting the PCs, one can view interrelationships between different variables, and detect and interpret sample patterns, groupings, similarities or differences [19]. Moreover, PCA is a useful method to reduce the dimensionality of large data sets, such as those from voltammetric sensor arrays.

Artificial Neural Networks (ANNs) consist of a number of simple processing units (or neurons) linked by weighted modifiable interconnections [20], originally designed to mimic the function of the human brain and applied to quantitative and qualitative analysis during the last decades [11]. Imitating the biological learning, they require a training process where the weights of those connections are adjusted, building a model that will allow to carry out the prediction of the desired parameters (either qualitative or quantitative). Then, once the corresponding model is generated, it can be further applied to the

prediction of the outputs for new samples by simply introducing the readings of the sensors of those to the model, hence obtaining a powerful analytical tool for rapid analysis of cava wine samples.

In order to find the appropriate ANN model, significant effort is needed to optimize the configuration details that determine its operation. Normally, this is a trial-and-error process, where several parameters (training algorithms, number of hidden layers, number of neurons in the different layers, transfer functions, etc.) are fine-tuned in order to find the best configuration that optimizes the performance of the model [21].

3. Results and Discussion

3.1 Voltammetric responses

Under the conditions described in Section 2.4, a total of 20 samples were analyzed, registering a complete voltammogram for each of the samples. As can be seen in Figure 1, where examples of the different responses obtained for each kind of sensor are shown, currents monotonously increase as FC index (and other phenolic indexes) increases, with some differentiated behaviour for each sensor. As a general trend, oxidation of phenolic compounds onto electrode surface could be seen in all the cases, while also some reductive currents close to the region of 0 V are obtained for both biosensors and copper modified sensors due to its catalytic effect.

For developing an ET, the first necessary condition is to have analytical signals responding to the phenomena to which the objective is aimed, with variability among them and the different sensors forming the sensor array. Thus, proposed BioET seems to provide very useful departure point, generating very rich data. However, the extreme complexity of the generated signals (the set of voltammograms) hinders the processing step; especially if ANNs are to be used. As already commented, all these data is pretreated using a compression step, required to gain advantages in training time, to avoid redundancy in input data and to obtain a model with better generalization ability. In our case this was accomplished by the use of the *windowed slicing integral* [18].

3.2 Qualitative approach

As deduced from the voltammograms in Figure 1, it could be stated that BioET array seems to clearly respond to phenolic content of samples. Nevertheless, to detect any other similarities or capabilities to distinguish some extra features of cava samples, the corresponding compressed signals were processed employing PCA analysis (Figure 2). Despite PCA cannot be considered as a properly pattern recognition method, as it only provides a visual representation of the relationships between samples and variables, it is a very useful tool due to it provides insights into how measured variables cause some samples to be similar to, or how they differ from each other.

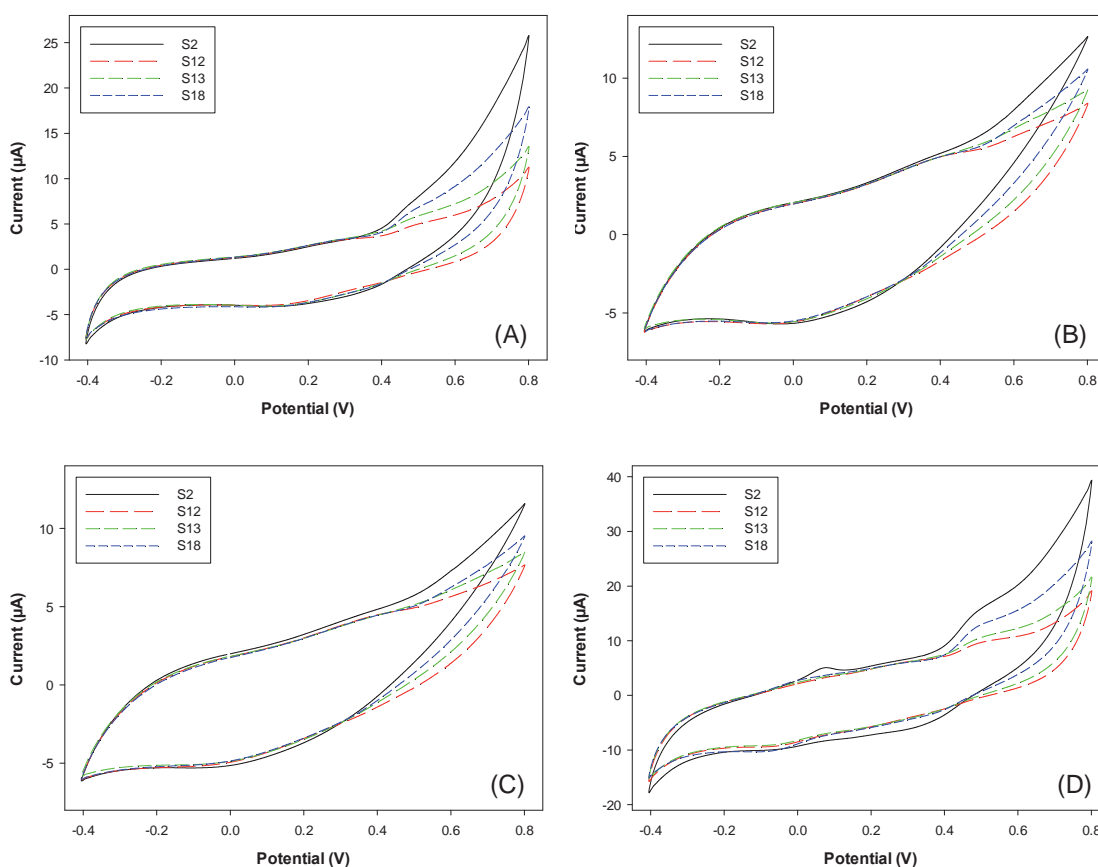
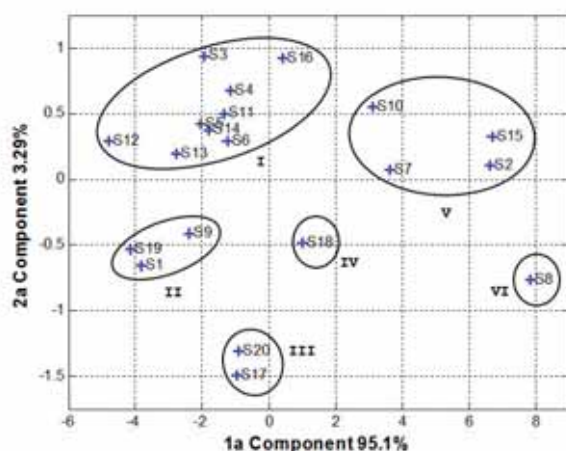
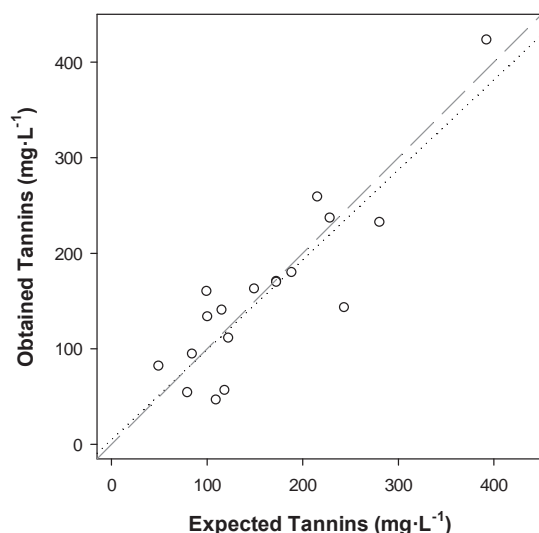


Figure 1. Example of the different voltammograms obtained with the different sensors forming the BioET array and for certain arbitrary cava wine samples are shown. Signals provided correspond to: (A) graphite-epoxy sensor, (B) tyrosinase biosensor, (C) laccase biosensor and (D) copper nanoparticle modified sensor.

Table 2. Results of the fitted regression lines for the comparison between obtained vs. expected values, both for the training and testing subsets of samples and the five considered phenolic compounds indexes (intervals calculated at the 95% confidence level).

| Training subset | | | | | |
|--|-------------|-----------|-------------|------|-------------|
| Phenolic indexes | Slope | Intercept | Correlation | RMSE | Total NRMSE |
| Folin index ($\text{mg}\cdot\text{L}^{-1}$) | 1.018±0.147 | -1.4±29.9 | 0.965 | 15.1 | |
| I ₂₈₀ index (arb. unit) | 0.955±0.109 | 0.45±1.14 | 0.977 | 0.55 | |
| I ₃₂₀ index (arb. unit) | 0.896±0.236 | 0.61±1.27 | 0.895 | 0.49 | 0.077 |
| Tannins ($\text{mg}\cdot\text{L}^{-1}$) | 0.987±0.060 | 2.1±11.0 | 0.993 | 9.50 | |
| Anthocyanins ($\text{mg}\cdot\text{L}^{-1}$) | 0.869±0.167 | 1.2±2.2 | 0.940 | 2.36 | |
| Testing subset | | | | | |
| Phenolic indexes | Slope | Intercept | Correlation | RMSE | Total NRMSE |
| Folin index ($\text{mg}\cdot\text{L}^{-1}$) | 0.977±0.257 | 8.9±52.4 | 0.896 | 26.6 | |
| I ₂₈₀ index (arb. unit) | 0.973±0.308 | 0.51±3.22 | 0.858 | 1.52 | |
| I ₃₂₀ index (arb. unit) | 0.872±0.358 | 0.82±1.92 | 0.791 | 0.74 | 0.151 |
| Tannins ($\text{mg}\cdot\text{L}^{-1}$) | 0.942±0.261 | 4.8±47.7 | 0.886 | 41.4 | |
| Anthocyanins ($\text{mg}\cdot\text{L}^{-1}$) | 0.739±0.340 | 1.9±4.6 | 0.755 | 4.89 | |


Figure 2. Score plot of the first two components obtained after PCA analysis of the cava wine samples.

Figure 3. Example of the modelling ability of the optimized ANN showing the set adjustments of obtained vs. expected tannins content for testing subset (\circ , dotted line). Dashed line corresponds to theoretical diagonal line.

Firstly, it should be noticed that with only the first two PCs, the accumulated explained variance was ca. 98.4%; a large value which means that nearly all the variance contained in the original information is now represented by only these two new coordinates. Secondly, as can be seen by simply

visually analyzing the plot, some clusters are obtained after this transformation, thus indicating some similarities between those samples.

Again, as could be expected from the voltammograms, it could be seen how samples seem to group depending on their phenolic content; e.g. cluster II seems to group samples with very low phenolic content, while cluster V seems to involve samples with the highest content. Moreover, S18 correspond to a cava from *Brut Nature* class, while almost all the others belong to *Brut* class; thus BioET seems to be able also to distinguish this fact [17]. Moreover, a part from phenolic content BioET seems to be capable to distinguish other features such as the different ageing or region. For example, S20 belongs to a low aged cava, thus it might indicate that also S17 is a young one. Another particularity of S8 is its region, which belongs to *l'Empordà* while other samples are mainly from *Penedès* region.

After constructing this preliminary visualization model which permitted recognizing some initial patterns and similarities between samples, while confirming that BioET was also responding to phenolic content of the samples, the next step is the construction of an ANN model that allows the quantitative prediction of phenolic content of rosé cava wine samples.

3.3 Quantitative approach

After preprocessing the recorded voltammograms with *windowed slicing integral*, the obtained coefficients fed an ANN model in order to predict the total phenolic indexes in wines. After a systematic evaluation of topologies, the final architecture of the ANN model had 44 neurons (4sensors x 11 coeffs. obtained from the preprocessing stage) in the input layer, 10 neurons and *purelin* transfer function in the hidden layer and 5 neurons and *purelin* transfer function in the output layer, providing the five phenol indexes considered.

To evaluate the BioET response, leave-one-out method was used given the reduced data set. In this manner, each sample is quantified by means of the model derived from the other samples (all cases except the case itself). This process is repeated k times (as many as samples) leaving out one different sample each time, the one to be quantified, which acts as model validation sample. Thus, with this approach all samples are used once as validation.

Finally, all data is grouped depending if it was intervening in the training process or used in the external test subset, building the response model.

Comparison graphs of predicted vs. expected concentration for the five indexes were built, both for train and test subsets, to check the prediction ability of the obtained ANN model. As an example of those, Figure 3 shows the obtained plot for the prediction of tannins content of rosé cava wine samples. It may be seen that a satisfactory trend is obtained, with regression lines almost indistinguishable from the theoretical ones. Also, as usual in ANN models, lower dispersion is obtained for the training subset.

Similarly, same plots were built for the other four indexes and regression lines were fitted, which regression parameters are summarized in Table 2. As expected from the comparison graphs, a good linear trend is attained for all the cases, but with better correlation coefficients in the training subsets due to the lower dispersion. Despite this, the results obtained for both subsets are close to the ideal values, with intercepts close to 0 and slopes and correlation coefficients close to 1.

Therefore, the presented approach herein represents the obtaining of an alternative analytical tool that allowed the simultaneous determination of five different phenolic compounds index in a simply, rapid and inexpensive way. Furthermore, with the same experimental setup, the proposed approach may be alternatively applied for the quantification of other specific compounds.

4. Conclusions

In summary, a voltammetric BioElectronic Tongue has been applied in cava rosé wine analysis in order to create a tool capable of quantifying total phenolic content as well as concrete classes total content. Concretely, proposed BioET was formed by an array of four biosensors modified with enzymes such as tyrosinase and laccase on one side and copper nanoparticles on the other so as to obtain a set of electrodes responding to the phenomena to which they are aimed and with some variability and cross-response features among them. Additionally, the use of chemometric tools such as ANNs allowed the quantification of five different phenolic compounds indexes widely used in wine sector; viz. Folin-Ciocalteu, I₂₈₀, I₃₂₀, total tannins and total anthocyanins.

In this sense, such strategy has demonstrated to be a powerful and much promising approach, with huge applications in wine making industry, as it allows reducing considerably analysis time, avoids any sample pretreatment or the use of additional reagents, and what's more, allows the simultaneous determination of several indexes at the same time. Moreover, its performance characteristics may satisfy food industry requirements of precision, rapidity, sensitivity, simplicity and low cost required to be considered as a useful analytical tool.

Acknowledgments

Financial support for this work was provided by Spanish Ministry of Science and Innovation, MCINN

(Madrid) through project CTQ2010-17099 and by program ICREA Academia from *Generalitat de Catalunya*. X. Cetó thanks the support of *Dept. d'Innovació, Universitats i Empresa de la Generalitat de Catalunya* for the predoctoral grant.

References

- [1] P. Ribéreau-Gayon, Y. Glories, A. Maujean, D. Dubourdieu, Phenolic Compounds, in: *Handbook of Enology*, John Wiley & Sons, Ltd, Chichester (UK), 2006, pp. 141.
- [2] J.A. Kennedy, M.A. Matthews, A.L. Waterhouse, *Am. J. Enol. Viticult.* **2002**, *53*, 268.
- [3] C. Sánchez-Moreno, J.A. Larrauri, F. Saura-Calixto, *Food Res. Int.* **1999**, *32*, 407.
- [4] D.M.A. Gil, M.J.F. Rebelo, *Eur. Food Res. Technol.* **2010**, *231*, 303.
- [5] R.S. Jackson, *Wine tasting: a professional handbook*, Academic Press, London, 2002.
- [6] D. De Beer, J.F. Harbertson, P.A. Kilmartin, V. Roginsky, T. Barsukova, D.O. Adams, A.L. Waterhouse, *Am. J. Enol. Viticult.* **2004**, *55*, 389.
- [7] R. Solná, P. Skládal, *Electroanal.* **2005**, *17*, 2137.
- [8] V.L. Singleton, R. Orthofer, R.M. Lamuela-Raventós, [14] Analysis of total phenols and other oxidation substrates and antioxidants by means of folin-ciocalteu reagent, in: P. Lester (Ed.) *Methods in Enzymology*, Academic Press, 1999, pp. 152.
- [9] M. Esteban, C. Ariño, J.M. Díaz-Cruz, *TrAC-Trend. Anal. Chem.* **2006**, *25*, 86.
- [10] V. Pravdová, M. Pravda, G.G. Guilbault, *Anal. Lett.* **2002**, *35*, 2389.
- [11] M. del Valle, *Electroanal.* **2010**, *22*, 1539.
- [12] E. Elejalde Caravaca, *Formula* **1999**, *5*, 39.
- [13] Commission Regulation (EEC), in: *Official Journal L 272*, Publication Office of the European Union, 17 September 1990, pp. 0001
- [14] M. Gessner, D. Steiner, Acid Butanol Assay for Proanthocyanidins (Condensed Tannins), in: M. Graça, F. Bärlocher, M. Gessner (Eds.) *Methods to Study Litter Decomposition*, Springer Netherlands, 2005, pp. 107.
- [15] Y. Glories, *Connaiss. Vigne Vin* **1984**, *18*, 253.
- [16] X. Cetó, F. Céspedes, M.I. Pividori, J.M. Gutiérrez, M. del Valle, *Analyst* **2012**, *137*, 349.
- [17] X. Cetó, J.M. Gutiérrez, L. Moreno-Barón, S. Alegret, M. del Valle, *Electroanal.* **2011**, *23*, 72.
- [18] X. Cetó, F. Céspedes, M. del Valle, *Electroanal.* **2013**, *25*, 68.
- [19] I.T. Jolliffe, *Principal Component Analysis*, Springer New York, New York, US, 2002.
- [20] F. Despagne, D. Luc Massart, *Analyst* **1998**, *123*, 157R.
- [21] A. Mimendia, A. Legin, A. Merkoçi, M. del Valle, *Sensor Actuat. B-Chem.* **2010**, *146*, 420.

La constant demanda de nous mètodes analítics capaços d'assolir els nous reptes en camps tals com el monitoratge ambiental, la seguretat alimentària i la salut pública, sumada a la dificultat d'obtenir sensors químics envers qualsevol analític, que operin de forma òptima en les condicions requerides, ha donat lloc a l'aparició de noves estratègies per tal de donar resposta a aquests problemes.

És en aquest context, que ha tingut lloc l'aparició de sistemes biomimètics tals com les llengües electròniques. Bàsicament, aquests sistemes analítics consisteixen en la imitació del sentit del gust dels mamífers, on uns pocs receptors són capaços de respondre a una gran varietat de substàncies, un principi que juntament amb una etapa complexa de tractament de la informació, tal com la que té lloc en el cervell, permet la quantificació o classificació d'un gran nombre de substàncies.

UAB

Universitat Autònoma de Barcelona



**GRUP
DE SENSORS I BIOSENSORS**