Cava wine authentication employing a voltammetric Electronic Tongue

Xavier Cetó,a Josefina Capdevila,b Anna Puig,b Manel del Vallea*

a Sensors and Biosensors Group, Department of Chemistry, Universitat Autònoma de Barcelona, Edifici Cn, 08193 Bellaterra, Barcelona, Spain
b Estació de Viticultura i Enologia, INCAVI IRTA, 08720 Vilafranca del Penedès, Spain
* e-mail: manel.delvalle@uab.cat

Received: 
Accepted: 

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 samples. Obtained responses were then preprocessed employing FFT and the resulting coefficients were input to a LDA model which allowed the classification of the samples according to its vintage time. Besides, a quantitative model employing ANNs was built for the prediction of the total amount of sugar present in the samples, a parameter also used to classify cava samples.

Keywords: Electronic Tongue, Linear Discriminant Analysis, Artificial Neural Network, cava wine, ageing

DOI: 10.1002/elan

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) [1]. That is the same than the one for French Champagne, but adapted to the local varieties of Catalonía.

Given its DO status, cava is subjected to quality checks carried out by the technical services of the regulatory board; the latter has the attributions to assess that producers comply with the specific technical conditions to produce both the cuvée and the cava, as well as their specific type of labels [2]. These controls include the whole production process, from the grapes to the bottling; i.e. they start in the vineyards assessing grape varieties; continue controlling volumes of production and monitoring of its quality; and last until the 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 cavas will depend on the differing variety combinations, the vintages and also the different ageing periods [3]. 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), Demi-dry (33-50 g·L⁻¹) or Sweet (>50 g·L⁻¹) [2]. 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) [2]. 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 categories, cavas are 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 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) [4]. 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 [5, 6]. These biomimetic systems, opposed to conventional approaches, are directed towards the combination of low selectivity sensors array (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 [7, 8], which have already been successfully applied to the classification of several alcoholic beverages such as wines [9], cava wines [10], brandies [11] or beers [12, 13], between others. Moreover, they have also been used for the detection of inappropriate handling practices or adulteration processes in wine [14, 15]; and even the quantification of several analytical parameters such as phenolic content [16, 17].

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 young cavas 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 [7, 8], which have already been successfully applied to the classification of several alcoholic beverages such as wines [9], cava wines [10], brandies [11] or beers [12, 13], between others. Moreover, they have also been used for the detection of inappropriate handling practices or adulteration processes in wine [14, 15]; and even the quantification of several analytical parameters such as phenolic content [16, 17].

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 young cavas in order to obtain an analytical tool to easily detect frauds. 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 for feature extraction and Linear Discriminant Analysis or Artificial Neural Networks for, respectively, build the qualitative and quantitative prediction models.

2. Materials and Methods

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 [2]: “young” cavas (9 to 15 months), “Reserva” (15 to 30 months) and “Gran Reserva” (more than 30 months). Accordingly, the set was formed by 16 Gran Reserva, 24 Reserva and 25 young cava wines samples.

Those samples were mainly from Penedès region (as is the region where it’s mostly produced), and obtained mainly from Macabeu, Parellada and Xarel-lo grape varieties, although Chardonnay or Subirat grapes may also be used; that is, the five different grape varieties authorised by Consell Regulador del Cava [2]. In this sense, Table S1 summarizes information about the cava wines under study.

Additionally, one Champagne (French sparkling wine) sample was included in the set as outlier. Even its differentiation was not the aim of the present study, this was done given different regulations apply for both Cava and Champagne, hence to check the response of the ET towards it. While at the same time, its usage as control sample along the experiment, also allowed to assess the repeatability of the ET response.

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). Potassium chloride was purchased from Merck KGaA (Darmstadt, Germany). Cobalt (II) phthalocyanine, polyaniline, polypyrrole, copper and platinum nanoparticles (<50 nm), which were used as electrode modifiers, were purchased from Sigma-Aldrich (St. Louis, MO, USA).

2.3. Reference methods

To further assess the ET capabilities in the analysis of cava wines, those were also analyzed by classical methods to obtain 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, related to level of dryness and total dry extract 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 Fourier transform infrared spectroscopy (FTIR) with a WineScan FT 120 (FOSS, Barcelona, Spain), which was previously calibrated according to official OIV methods [18, 19].
2.3.2. Total dry extract

Total dry extract or total dry matter includes all matter which is non-volatile under specified physical conditions, which must be such that the matter forming the extract undergoes minimal alteration while the test is being carried out. This extract was calculated indirectly from the specific density and the alcohol by volume (abv) degree of the wine according to regulated methods [18, 19], and expressed in grams per litre.

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 [10, 20]. 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) [21]. In this manner, the first was a carbon transducer, and five of them were modified by adding components as nanoparticles of copper and platinum, conducting polymers in powder like polyaniline and polypyrrole, or cobalt (II) phtalocyanine – one component per sensor. Those modifiers were selected based on previous reported studies with wines, either from other research groups or from our laboratories, 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 [14]; a desired situation in ETs applications.

From an electroanalytical point of view, electrodes modified with phthalocyanines (mainly CoPc and its derivatives) are interesting for being efficient electrocatalysts in the determination of many important inorganic, organic or biological compounds [22]. Similarly, in the last few years, nanoparticles have emerged as interesting electroactive materials in electroanalysis; these are an alternative to bulk metals, with catalytic and electrocatalytic peculiarities, mainly derived from their higher surface/mass ratio [14]. Lastly, conducting polymers represent interesting electrode modifying materials; their advantages are ascribed to their antifouling and electrocatalytic properties, as well to their high sensitivities and low detection limits [23].

Consequently, the type of electroactive compounds interacting with these sensors are essentially reducing sugars, phenolic compounds and other antioxidants presents in the sample [10, 20], as previously observed.

2.4.2. Measurement 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 its GPES Multichannel 4.7 software package.

2.4.3. Measurement cell

Prior to perform cava samples measurements, electrodes were first cycled in saline solution (i.e. 10mM KCl) in order to get stable voltammetric responses, ensuring reproducible signals from the ET array. Afterwards, an aliquot of 25 ml of cava wine was directly used for each measurement, without any sample pretreatment. Once the cava bottles were opened, measures were carried out after a few minutes (ca. 5 min at room temperature) in order to minimize the formation of bubbles onto the electrode surface. That is, no specific removal of oxygen or CO2 was done, but minimization of distortions by air bubbles was attempted.

Then, a complete cyclic voltammogram was recorded 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 any cumulative effect of impurities on the working electrode surfaces [10]. For this, a conditioning potential of +1.5V was applied during 40s in a cell containing 25 ml of distilled water. This step permitted to minimize the electrode fouling, checking that the original signal (baseline) was recovered 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.
Full Paper

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 was generated (either the LDA or ANN model), it could be applied for the prediction (either qualitative or quantitative values) from new samples by simply introducing the readings of the sensors into the model, hence obtaining a powerful analytical tool for rapid screening of cava wine samples.

In all cases, prior to the building of the models, and 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 [17].

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 [24]. Then by plotting the PCs, one can view interrelationships between the different samples, 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 [25]. 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 [26]. 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 acting as the classifier.

The last tool used, ANNs, were designed to mimic the reasoning of the human brain and have been applied to quantitative and qualitative analysis during the last decades [4]. They consist of a number of simple processing units (or neurons) linked by weighted modifiable interconnections [27]. As LDA, and imitating the biological learning, they require a training process where the weights of those connections are adjusted, in this way building a model that will allow to carry out the prediction of the desired parameters (either qualitative or quantitative). ANNs 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, samples were 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 for the different kinds of sensors. Besides, it could be seen how some samples present 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 young cavas present clearly differentiated response for the cathodic wave in the region from -0.5 V to -1.0V; meanwhile, Champagne sample presents a more similar response to young cavas, although still showing slightly differentiated response, but in the region from -0.75V to -1.0V. The later 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 young 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 [20], 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 (Figure 1E). On the other hand, platinum (Figure 1C) and copper nanoparticles (Figure 1F) 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 +1V for the Champagne sample, indicating that its discrimination might be also possible.

Additionally, to ensure repeatability of the sensor array response, Champagne sample was used as control sample and its measurement was replicated four times along the
Fig. 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, (D) polyaniline, (E) polypyrrole and (F) copper nanoparticle modified sensors.

whole experiment; obtained voltammograms are shown in Figure S1. As expected, small differences attained between different replicates corroborate the fact that different voltammetric responses are obtained depending on the type of cava wine, and that those are not due to sensors’ drift.

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 x sensors x potentials) while preserving the relevant information. In addition to data reduction, compression is intended for extracting significant features from the departure information, while
Fig. 2. Score plot of the first three components obtained after PCA analysis of the cava wine samples: (●) Gran Reserva, (▲) Reserva, (■) Young and (♦) French Champagne.

Fig. 3. Score plot of the functions obtained after LDA analysis of the cava wine samples according to the ageing time: (●) Gran Reserva, (▲) Reserva, (■) Young and (♦) French Champagne; also the centroid of each class is plotted (★).

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 and Young 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 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 3B) and one without them (Figure 3A).

LDA model was trained and validated using k-fold cross-validation method with k=6 (data was split in 6 subsets, using k-1 subsets to build the model and the remaining one as testing subset; this process is repeated k times leaving each time out one of the k subsets, and finally results from the different folds are averaged). Afterwards, confusion matrix was built (Table 1), allowing calculating 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, given there were...
Table 1. Confusion matrix built according cava wine vintage obtained using LDA model and k fold cross validation method with k=6.

<table>
<thead>
<tr>
<th></th>
<th>GR\textsuperscript{a}</th>
<th>R\textsuperscript{a}</th>
<th>Young\textsuperscript{a}</th>
</tr>
</thead>
<tbody>
<tr>
<td>GR</td>
<td>16</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>R</td>
<td>0</td>
<td>24</td>
<td>0</td>
</tr>
<tr>
<td>Young</td>
<td>0</td>
<td>0</td>
<td>25</td>
</tr>
</tbody>
</table>

\textsuperscript{a} Expected. \textsuperscript{b} Found.

3.3. Estimation of cava indexes

As a further ability of electronic tongue systems, a quantitative application in order to predict compounds of relevant significance in cava wines was also attempted. Concretely, quantification of total sugars and total dry extract was undertaken. As stated, cava can be mainly classified from three different points of view, viz. its “colour” (white or rosé), its level of dryness or 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 its ageing (time of aging on yeast lees), as this would give a double functionality. 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 \cite{gran_reserva_label}.

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.

The first step in building the ANN model is selecting the topology of the neural network used. Given the difficulties to predict the optimum configuration in advance, this consists in a trial-and-error process where several parameters (training algorithm, number of hidden layers, number of neurons, transfer functions, etc.) are fine-tuned in order to find the best configuration that optimizes the performance of the neural network model \cite{network_configuration}.

After a systematic study optimizing those parameters, 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 \textit{tansig} transfer function in the hidden layer and two neurons and \textit{tansig} transfer function in the output layer, viz. the total amount of sugar and the total dry extract.

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); from a total of 65 samples, as there were two whose reference analysis could not be performed. 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 it can be observed, a satisfactory trend was 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 ANN model and those expected and provided by the reference method.

4. Conclusions

In this work, it has been demonstrated that the combination of voltammetric measurements with chemometric tools is an analytically promising approach for the characterization of cava wines according to two main categorization parameters, i.e. its level of dryness and ageing time (time of aging on yeast lees).

Thanks to the use of qualitative chemometric tools it has been possible to develop a direct and simple procedure for authentication application or detection of frauds in cava wine elaboration. Concretely, LDA allowed the building of qualitative models able to classify cava wine samples according to its vintage, while ANN allowed the quantification of its dryness level. Additionally, given the biomimetic character of the developed approach, it presents a straightforward solution to the absence of the knowledge about which compounds are primarily responsible for the class that hinders the development of analytical procedures for cava categorization according to its ageing time.

Finally, future efforts with this approach may involve its further validation (e.g. assessing the capabilities of the ET to assess not only the class but also the ageing time), or the miniaturization of the system.

In this fashion, ETs should be considered as an attractive alternative to other classical methods or to sensory panels, especially suitable for screening purposes, with huge advantages as might be its simplicity and low cost.

Acknowledgements

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, Universitats i Empresa de la Generalitat de Catalunya for the predoctoral grant.

References