

# Geographical classification of Chilean wines by an electronic nose

Nicolás H Beltrán  
Manuel A Duarte-Mermoud  
Ricardo E Muñoz

Department of Electrical Engineering,  
University of Chile, Santiago, Chile

**Abstract:** This paper discusses the classification of Chilean wines by geographical origin based only on aroma information. The varieties of Cabernet Sauvignon, Merlot, and Carménère analyzed here are produced in four different valleys in the central part of Chile (Colchagua, Maipo, Maule, and Rapel). Aroma information was obtained with a zNose™ (fast gas chromatograph) and the data was analyzed by applying wavelet transform for feature extraction followed by an analysis with support vector machines for classification. Two evaluations of the classification technique were performed; the average percentage of correct classification performed on the validation set was obtained by means of cross-validation against the percentage of correct classification obtained on the test set. This developed technique obtained results on classification rates over 94% in both cases. The geographical origin of a Chilean wine can be resolved rapidly with fast gas chromatography and data processing.

**Keywords:** geographical origin, origin denomination, wine classification, pattern recognition, support vector machines, wavelet analysis, feature extraction

## Introduction

Origin denomination (OD) is generally understood as ascribing specific characteristics of an agriculture product to a geographical region. Those region-specific characteristics mean each product differs from similar products grown in other regions. In wine, the region-specific characteristics are produced by variations in soil and climate. This differentiation is applied to many products such as cattle,<sup>1</sup> potatoes,<sup>2</sup> honey,<sup>3,4</sup> olives oil,<sup>5</sup> onions,<sup>6</sup> etc, but the most well known are related to wine. European countries stand out from other wine-producing countries with many different wine-producing regions assigned different ODs. Prof. Leighton attempted to provide a rigorous definition for OD by defining the concept of “*terroir*” as the relationship between wine and its growing environment, which includes Andreas Smolle’s<sup>7</sup> suggestion to consider soil geology (chemical composition, drain, and structure). France has used this procedure to determine the borders of different wine-growing regions for the “Appellation d’Origine Contrôlée” printed on wine labels.

Knowledge of how soil composition affects wine quality is mainly based on the experience of individual winegrowers. However, Leneuf and Rat<sup>8</sup> formulated a hypothesis to explain how chemical elements such as potassium, manganese, and magnesium influence the growth of grapes and the yield of each vineyard. Nevertheless, this work was not able to conclude definitively on how those chemical elements affect the taste or aroma of wines.

Correspondence: Nicolás H Beltrán  
Department of Electrical Engineering,  
University of Chile, Av. Tupper 2007,  
Casilla 412-3, Santiago, Chile  
Email nicolas.beltran@die.uchile.cl

Climate factors such as temperature, sunlight, and humidity also determine the characteristics of the *terroir* since they affect grape growth. Soil, climate, and each winery's individual wine-making methods are the main elements affecting OD.

Several techniques have been proposed to determine the OD of wine. They range from genetic analysis of grapes to well known techniques used in chemistry or physics studies to determine compounds or elements in a wine sample.

Determination of isotopic ratios by mass spectroscopy and nuclear magnetic resonance (NMR) is probably the most efficient technique available today to detect the geographical origin of a component in a vegetable sample. The measurement of the relationship between a deuterium and hydrogen in ethanol molecules produced by sugar fermentation can be used in some products to determine their geographic origin. This technique works because it is physically impossible to change sugar isotopic ratios to camouflage additives. This was the main reason why the European Union (EU) officially recognized this technique for wine analysis in 1990.

NMR has been used to analyze properties of different materials in polymers or biomedicine. For wine analysis, it is used to compare the deuterium content in a wine sample with another sample from the same region to produce a region-specific pattern and define an OD.<sup>9</sup>

Inductively coupled plasma-mass spectrometry has been used by different researchers to determine the geographic origin of wine based on inorganic component concentrations.<sup>10,11</sup> The isotopic ratio for strontium 87 and 86 ( $^{87}\text{Sr}/^{86}\text{Sr}$ ) for wines from five Portuguese regions and one from France have been determined. Sample pretreatment was carried out to avoid the influence of rubidium (Rb) on the measurement. According to the researchers, the results are promising for the determination of OD. Similar results are reported for wines produced in the Canary Islands.<sup>12</sup> The study was carried out on 153 samples of white, rosé, and red wines from eight different regions on four islands, and differences in the concentrations of 20 elements was observed.

Another technique used for OD was the identification of amino acids by high-performance liquid chromatography (HPLC).<sup>13</sup> Twenty amino acids were analyzed in Spanish wines from the La Mancha OD, which resulted in a greater concentration of arginine and proline.

This short review on the techniques used to determine OD shows that all require sophisticated equipment to be used for a few hours on each sample analysis and longer times for a set of samples to reach a conclusion. On the other hand,

researchers normally use basic pattern recognition techniques such as principal component analysis (PCA) and linear discriminant analysis (LDA).

Another group of techniques extensively used for identifying wine variety is based on electronic noses. Interesting work is being done by Lozano and colleagues<sup>14–18</sup> on Spanish wine discrimination using electronic noses. Lozano and colleagues<sup>14</sup> used two different sampling methods for the electronic nose (purge and trap, and solid phase microextraction) which were analyzed and compared. Multivariate analyses such as PCA and artificial neural networks (ANN) were applied to five different wines and good results were obtained.

García and colleagues<sup>15</sup> used an electronic nose based on metal oxide semiconductor thin-film sensors to characterize and classify four types of red wines of the same grape variety. Data analysis was performed by two pattern recognition techniques: PCA and probabilistic neuronal network (PNN). The results showed that the electronic nose was able to identify wines.

Lozano and colleagues<sup>16</sup> used an electronic nose to recognize and detect wine aging. The same wine was aged in different types of oak barrel (French and American oak). PCA and PNN were used as pattern recognition techniques. Classification rates of 97% and 84% were achieved, respectively.

A portable electronic nose using two different micro-machined resistive-type sensor arrays was developed by Alexandre and colleagues.<sup>17</sup> One used a polysilicon integrated heater and the other used a platinum film. The nose was tested with four different wines from the Madrid region. PCA and PNN were employed as discrimination tools. Correct classification rates of 100% and 88% were reported, respectively.

An electronic nose using an array of surface acoustic wave (SAW) sensors was reported by Fernández and colleagues,<sup>18</sup> which used piezoelectric zinc oxide deposited by sputtering over silicon substrates to detect low concentrations of different volatile compounds such as octane, toluene, and methylethylketone. The dual configuration array is composed by seven sensors spray coated with diverse polymer thin films and one reference sensor is left uncoated. PCA, PNN, and partial least squares (PLS) were used to classify samples and predicted gas concentrations reasonably well. It is important to note that the sensor used by Fernández and colleagues<sup>18</sup> is of the same type that Beltrán and colleagues<sup>19</sup> used to successfully perform variety classifications of Chilean wines.

## Motivation

This review is motivated by the need to have an objective technique that quickly delivers reliable results of the variety and/or the OD claimed by the winegrower. This intelligent system certainly assists wine trading and provides importers with a reliable method of checking the received product and a quality control tool for exporters. The system developed by Beltrán and colleagues<sup>19</sup> fulfills those requirements by accomplishing the task of classifying wine varieties in a short time after sample measurement. The results presented in this paper are an extension of this work to identify geographic origin minutes after testing a sample.

In this paper, our data is obtained by a zNose<sup>TM</sup>, which is essentially a fast gas chromatograph. The manufacturer<sup>20</sup> describes the system as consisting of two sections: a one-meter column (capillary tubing) where helium gas flows constantly through and impinges onto the surface of an uncoated SAW detector which is temperature-controlled. The capillary column tubing is made from stainless steel with a polysiloxane coating on the inner surface. A closed-loop microprocessor controls direct column heating, which allows linear ramping of the column temperature at rates from 10 °C/s to 20 °C/s. The second section of the system is used to sample the head space for volatile organic compounds (VOCs) of the wine by pumping the gaseous components through a heated inlet. Linking both sections is a trap (3-cm length of heated resistive metal capillary tubing), which is essentially a preconcentrator.

Obtaining a wine chromatogram with this system follows a sequential process involving both sections. First, VOCs are sampled and preconcentrated on the trap and then switched into the helium section. By rapidly heating the trap to 250 °C, the absorbed VOCs are released and each compound undergoes an absorption interaction as they pass through the capillary column. Linear ramping of the column temperature releases different absorbed compounds according to their characteristic retention times. The compounds exit

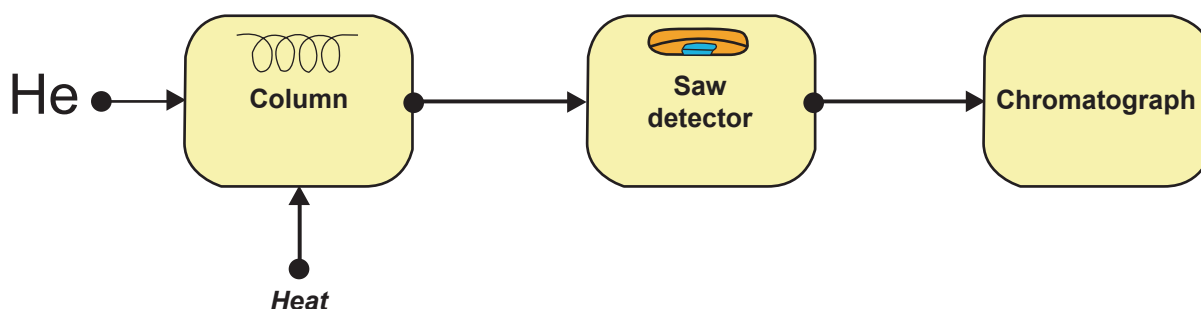
the column onto the surface of the SAW detector. Physically absorbed organic compounds on the crystal surface change the frequency of the resonator by slowing down the mechanical surface waves. The change in frequency is proportional to the absorbed mass on the crystal surface. Once the measurement is completed, reversing the SAW detector desorbs condensed compounds, which allows a new measurement to be carried out. A block diagram of the system is shown in Figure 1.

## Methodology

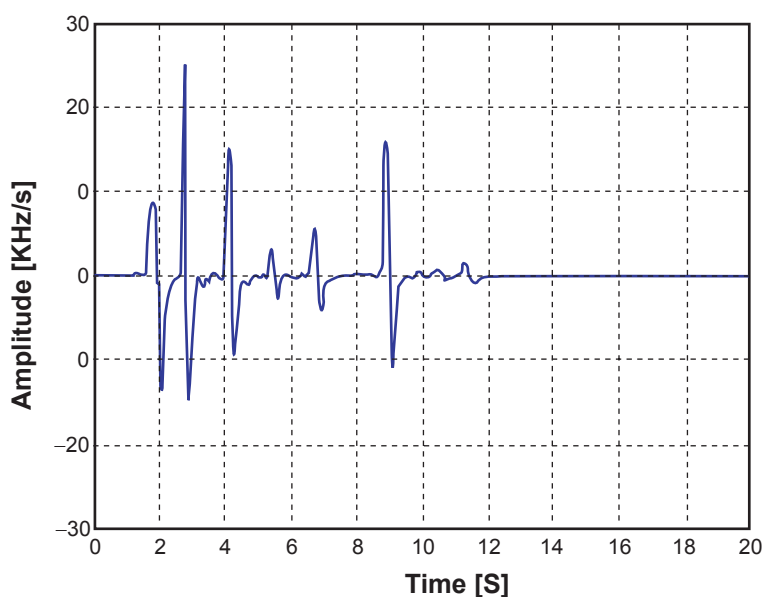
The information provided by the zNose<sup>TM20</sup> is a 20 s time signal, which is taken at a 0.01 s rate (sampling period) and generates a 2,000-point profile. As an example, Figure 2 shows a representative chromatogram from a wine sample (Cabernet Sauvignon) produced in the Colchagua valley. As seen in Figure 2, information content is mainly present in the interval 0–12 s. Thus, the segment from 12 to 20 s was neglected (chopped), generating a 1,200-point profile sampled at a 0.01 s period.

In order to decrease even more data dimension without losing information contained in the chromatograms, organic volatile compound profiles were re-sampled at twice the original sampling period (0.02 s). According to the basic signal theorem, known as the Shannon theorem,<sup>21</sup> there is no loss of information when re-sampling the signal at twice the greatest frequency available (the Nyquist frequency).<sup>22</sup> In this case, after a frequency content analysis of all profiles performed by Fourier analysis, it was determined that the Nyquist frequency was 50 Hz. Thus, the original profile can be re-sampled using a 0.02 s sampling period without losing information contained in the signals. As a consequence of this procedure, the number of profile points to be processed decreases from 1,200 to 600 points in our case.

To normalize data amplitude, a scale factor was applied. This factor was related to the standard deviation of the profile.<sup>19</sup> By this procedure, the amplitude of each one of



**Figure 1** Block diagram of the measurement system (zNose<sup>TM</sup>).



**Figure 2** Characteristic profile provided by the zNose™ from a Cabernet Sauvignon wine belonging to Colchagua valley.

the 339 patterns of the data base was scaled according to the scale factor given in equation (1).

$$x_{normalized} = \frac{x_{not\ normalized}}{\sigma_{x_{not\ normalized}}} \quad (1)$$

For each wine sample, 10 chromatograms were obtained from the electronic nose by repeating the experiment 10 times to avoid systematic instrumental errors. The average of these 10 profiles was considered as representative for each sample. The experiments were performed under temperature-controlled conditions of 22 °C.

## Database

To build our database, 339 commercial Chilean wines produced between 1992 and 2005 were tested. Wines of the varieties Cabernet Sauvignon, Merlot, and Carménère produced in Colchagua, Maipo, Maule, and Rapel valleys were characterized by analyzing the headspace of the samples

with a zNose™ FAST GC Analyzer 7100 manufactured by Electronic Sensor Technology (Newbury Park, CA, USA).<sup>20</sup> Sample wine distribution according to geographical origin and the number of wine samples for each variety and region are presented in Table 1.

The total database was divided in two sets. The first contained 80% of the samples chosen randomly and was used for training-validation purposes. The remaining 20% of the samples constituted the test set. Thus, samples used for testing the classification system were not known in advance by the system. Sample distribution of these two sets was as follows:

- Training-validation set: 270 patterns corresponding to 87 from Colchagua valley, 87 Maipo valley, 39 from Maule valley, and 57 from Rapel valley.
- Test set: 69 patterns corresponding to 22 from Colchagua valley, 22 from Maipo valley, 11 from Maule valley, and 14 from Rapel valley.

**Table 1** Wine distribution in database

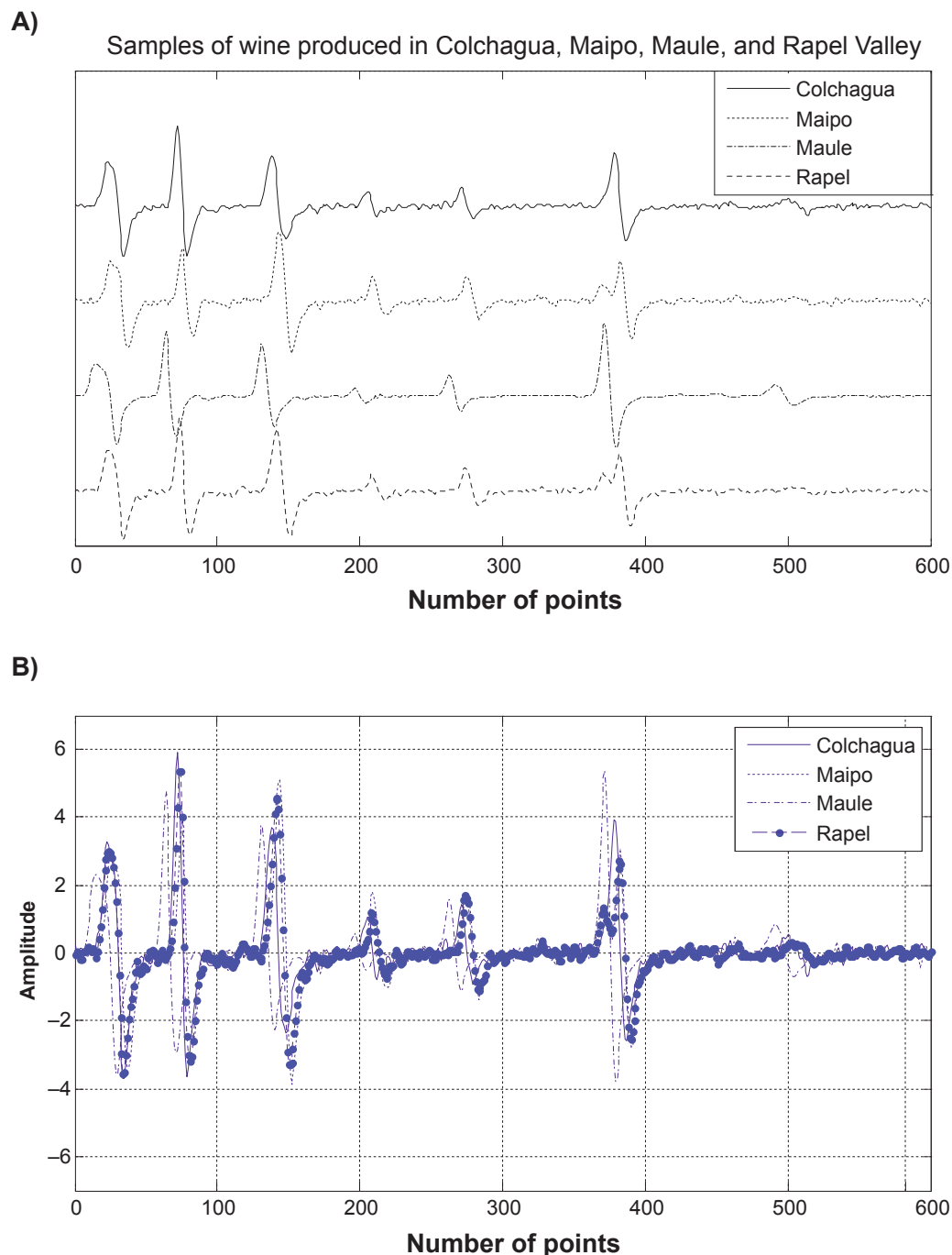
		Varieties			Total
		Cabernet Sauvignon	Merlot	Carménère	
Geographical origin (Valley)	Colchagua	47	33	29	109 (32%)
	Maipo	61	30	18	109 (32%)
	Maule	18	19	13	50 (15%)
	Rapel	30	19	22	71 (21%)
	TOTAL	156	101	82	339

The characteristic profiles of Cabernet Sauvignon wine samples produced in Colchagua, Maipo, Maule, and Rapel valleys are shown in Figure 3 after the chopping and re-sampling process. The profiles are plotted separately to observe possible differences in Figure 3(a).

In comparing the chromatograms shown in Figure 3(b), where all have been plotted in the same plot, significant

differences were not clearly observed to perform a simple profile analysis. This fact means that classifying different classes from these data becomes an interesting and challenging problem, which clearly requires the use of advanced pattern recognition techniques.

After re-sampling each of the profiles, a wavelet transform is applied for characteristic feature extraction and



**Figure 3** Representative profiles of a Cabernet Sauvignon sample produced in different valleys (after chopping and re-sampling).

followed by classification with support vector machines (SVM) using a multiclass method.

The steps carried out in this work to obtain the classification of Chilean wines are illustrated in the block diagram shown in Figure 4.

## Feature extraction technique

Feature extraction (FE) is a procedure aimed to reduce the dimension of feature vectors. Several FE techniques have been extensively used in artificial intelligence (AI). The most commons are Fourier transform (FT),<sup>23</sup> Fisher transform,<sup>24</sup> PCA,<sup>25</sup> and wavelet transform.<sup>26,27</sup> In this study, we have chosen the FE technique of wavelet transform analysis based on the successful results obtained in varietal classification of Chilean wines.<sup>19,28</sup>

Wavelet transform<sup>26,27</sup> is an interesting alternative to the widely used FT when analyzing nonstationary signals. Wavelet transform uses windows of variable size ie, small windows (detail window) for a fine analysis of high frequency signals and large windows (approximation windows) for a course analysis of low frequency signals.

Formally, a time signal  $f(t)$  of square integral can be expressed in terms of certain functions  $\psi_{j,k}(t)$  and  $\phi_k(t)$  according to the following decomposition:

$$f(t) = \sum_{k=-\infty}^{\infty} c_k \phi_k(t) + \sum_{j=0}^{\infty} \sum_{k=-\infty}^{\infty} d_{j,k} \psi_{j,k}(t), \quad j, k \in \mathbf{Z} \quad (2)$$

where  $\{\phi_k(t)\}$ ,  $\{\psi_{j,k}(t)\}$  are sets of orthonormal functions called *wavelets*. Coefficients  $c_k$  and  $d_{j,k}$  correspond to the approximation and detail coefficients, respectively, of the discrete wavelet transform of  $f(t)$  and they are computed as:

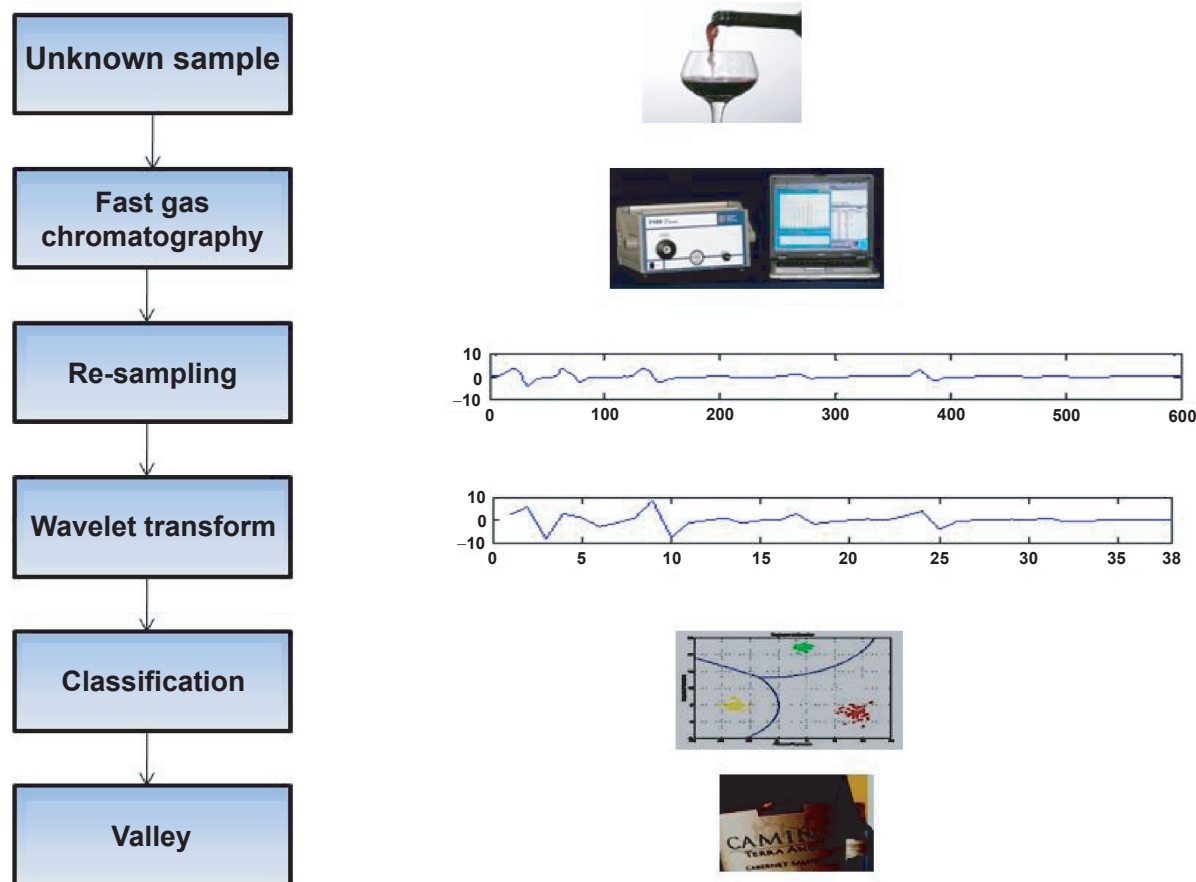
$$d_{j,k} = \int_{-\infty}^{\infty} f^*(t) \psi_{j,k}(t) dt \quad (3)$$

$$c_k = \int_{-\infty}^{\infty} f^*(t) \phi_k(t) dt \quad (4)$$

where  $f^*(t)$  is the conjugate of  $f(t)$ .

Amongst the simplest coefficients there exist a set of orthonormal wavelet basis called Haar wavelets, defined as:<sup>29</sup>

$$\psi_{j,k}(t) = 2^{j/2} \psi(2^j t - k), \quad \text{with } j, k \in \mathbf{Z} \quad (5)$$



**Figure 4** Steps of the method followed in this work to determine geographical origin of a Chilean wine sample.



$$\varphi_k(t) = \phi(t - k) \quad (6)$$

where function  $\psi(t)$  is called mother wavelet and function  $\phi(t)$  is the scale function. They are defined as:

$$\psi(t) = \phi(2t) - \phi(2t - 1) \quad (7)$$

$$\phi(t) = \begin{cases} 1 & \text{for } 0 \leq t \leq 1 \\ 0 & \text{otherwise} \end{cases} \quad (8)$$

where subindex  $j$  defines the decomposition level and the subindex  $k$  is the shift time.

In practice, the effect of wavelet transform is to filter the original signal using a set of low-pass and high-pass filters and is later re-sampled in a process called down-sampling, reducing the number of points of the original signal from  $n$  to  $n/2$  after each filtering processing. The number of times that this procedure is repeated corresponds to the wavelet decomposition level of signal  $S$ . The high-pass filters give the detail coefficients and the low-pass filters the approximation coefficients.

## Pattern recognition technique used

The SVM introduced by Vapnik<sup>30</sup> is a powerful classification and regression method widely used in machine-learning and data-mining.<sup>31</sup> SVM minimizes the empirical risk (defined as the error in the training set) and minimizes the generalization error.<sup>31</sup> The advantage of SVM applied to classification models is that supplies a classifier with a minimum Vapnik–Chervonenkis dimension,<sup>30</sup> which implies a small probability of error in the generalization. Another characteristic is that SVM allows the classification of nonlinearly separate data, since it performs a mapping of the input space onto the characteristic space of a higher dimension, where the data is indeed linearly separable by a hyperplane. This introduces the concept of an optimal hyperplane.

Given a set of training data with their respective outcome  $(x_i, y_i)$ ,  $i = 1, \dots, n$ , where  $x_i \in \mathcal{R}^m$  and  $y_i \in \{1, -1\}^n$ , SVM requires the solution of the following optimization problem:

$$\min_{w, b, \xi} \frac{1}{2} w^T w + C \sum_{i=1}^n \xi_i \quad (9)$$

constrained to  $y_i (w^T \phi(x_i) + b) \geq 1 - \xi_i$

$$\xi_i \geq 0 \quad (10)$$

Training data  $x_i$  are mapped onto a feature space of higher dimension by mean of function  $\phi(\cdot)$ . The solution implies the definition of the kernel function  $K(x_i, x_j) = \phi(x_i)^T \phi(x_j)$  and

$C > 0$  corresponds to the penalization error parameter. Some typical kernel functions commonly used are:

- *Linear*:  $K(x_i, x_j) = x_i^T x_j$
  - *Polynomial*:  $K(x_i, x_j) = (\gamma x_i^T x_j + r)^d$ ,  $\gamma, r > 0$
  - *Radial basis function (RBF)*:  $K(x_i, x_j) = \exp(-1/\sigma^2 \|x_i - x_j\|^2)$ ,  $\sigma > 0$
  - *Sigmoid*:  $K(x_i, x_j) = \tan h(\gamma x_i^T x_j + r)$ ,  $\gamma, r > 0$
- where  $\gamma$ ,  $\sigma$ ,  $r$ , and  $d$  are parameters of kernel functions.

## SVM multiclass

Since we are dealing with a multiclass problem (number of classes higher than two) and SVM can only discriminate between two classes, the strategy proposed by Angulo was applied.<sup>32</sup> Basically a stage is added to the normal application of SVM in order to carry out the classification for a number of classes higher than two.

We briefly describe here some of the techniques used to transform a two-class problem in a multiclass problem (decomposition methods). Furthermore, we present a reconstruction method that allows the fusion of the predictions of several classifiers to end up with a final answer.

The most commonly used decomposition methods to use a two-class classification technique as a multiclass one are *one-versus-rest*, *one-versus-one*, and *error correcting output codes* (ECOC).

The classifier, *one-versus-rest*, separates one class from the remaining  $C - 1$  classes by using the binary classifier. The standard method is to build  $C$  binary classifiers in parallel. When SVM is used, the proposed architecture by Vapnik is applied.<sup>30</sup>

The decomposition method associated to *one-versus-one* consists in building  $C(C - 1)/2$  binary classifiers all in parallel of *one-versus-one* types. Each node is trained by two different types of the  $C$  classes involved in the classification process.<sup>33</sup>

The ECOC technique<sup>34</sup> generates a codification for each class that allows the classification errors to be easily identified. For instance, in the case of four classes we can generate the following codes:

Class	Code
<i>a</i>	1111111
<i>b</i>	0000111
<i>c</i>	0011001
<i>d</i>	0101010

The first classifier predicts 1 if the class is *a* and 0 if it is *b*, *c*, or *d*. The second predicts 1 if the class is *a* or *d* and 0 if it is *b* or *c*, and so on. Instead of building four classifiers, one for each class, seven classifiers are generated.

In the case of  $k$  classes, each code will have  $2^{k-1} - 1$  bits. The first class is built with bits that have value 1. The second class is built with  $2^{k-2}$  zeroes followed by  $2^{k-2} - 1$  ones. The third,  $2^{k-3}$  zeroes, followed by  $2^{k-3}$  ones, and  $2^{k-3}$  zeroes, followed by  $2^{k-3} - 1$  ones, and so on.

Allwein and colleagues suggest an extension of the ECOC method that allows unifying the three decomposition schemes already mentioned.<sup>35</sup> Let us define the decomposition matrix:

$$M \in \{-1, 0, +1\}^{C \times l} \quad (11)$$

For all  $s$  with  $s = 1, \dots, l$  we provide the learning algorithm with labeled information (supervised learning) of the form  $(x_i, M(y_i, s))$  for all the examples  $i$  in the training set, but avoid all the examples for which  $M(y_i, s) = 0$ . The learning algorithm uses these data to generate a hypothesis of  $f_s: \mathcal{X} \rightarrow \mathcal{Y}$ .

For *one-versus-rest*,  $M$  is a  $C \times C$  matrix with all diagonal elements +1 and the reminder set to -1. For instance in the case of four classes,  $C = 4$ , we have:

$$M_{1-v-r} = \begin{pmatrix} +1 & -1 & -1 & -1 \\ -1 & +1 & -1 & -1 \\ -1 & -1 & +1 & -1 \\ -1 & -1 & -1 & +1 \end{pmatrix} \quad (12)$$

In the case of *one-versus-one*,  $M$  is a  $C \times \binom{C}{2}$  matrix where each column corresponds to a different pair of classes. Let us consider two different classes  $r_1, r_2$ : this column of  $M$  has +1 on row  $r_1$ , -1 on row  $r_2$ , and zero on the rest of the rows. For example, in the cases of four classes,  $C = 4$ , we get:

$$M_{1-v-1} = \begin{pmatrix} +1 & +1 & +1 & 0 & 0 & 0 \\ -1 & 0 & 0 & +1 & +1 & 0 \\ 0 & -1 & 0 & -1 & 0 & +1 \\ 0 & 0 & -1 & 0 & -1 & -1 \end{pmatrix} \quad (13)$$

A reconstruction method is proposed by Allwein and colleagues, which allows a final decision based on the binary classifiers.<sup>35</sup> For notation purposes  $M(r)$  denotes the row  $r$  of  $M$  and  $f(x)$  is the prediction vector for one example  $x$  with:

$$f(x) = (f_1(x), \dots, f_l(x)) \quad (14)$$

Given predictions  $f_s S$  for a pattern of the test set, let us say  $x$ , which of the  $C$  labels should be assigned to the pattern?

There exist numerous reconstruction methods to combine the  $f_s S$ , but here we will focus on only two which are rather simple to apply. The basic idea in both methods is to assign the label  $r$  whose row  $M(r)$  is the closest to the

predictions  $f(x)$ . In other words, to assign the label  $r$  that minimizes  $d(M(r), f(x))$  for measure  $d$ . This formulation requires to measure the distance between both vectors.

One way to measure the distance is to count the number of places in  $s$  in which the sign of the prediction  $f_s$  differs from the matrix input  $M(r, s)$ . Formally this means that the distance measure is:

$$d_H(M(r), f(x)) = \sum_{s=1}^l \left( \frac{1 - \text{sign}(M(r, s)f_s(x))}{2} \right) \quad (15)$$

where  $\text{sign}(z)$  is +1 if  $z > 0$ , -1 if  $z < 0$ , and 0 if  $z = 0$ . This is the computation of the Hamming distance between rows  $M(r)$  and the signs of the  $f_s$ . For a given pattern  $x$  and a matrix  $M$ , the label to be assigned is:

$$\hat{y} = \arg \min_r d_H(M(r), f(x)) \quad (16)$$

This reconstruction method is called Hamming decoding.<sup>35</sup> One disadvantage of this method is that it completely ignores the magnitude of the predictions that in some cases could be considered as a measure of the prediction's confidence.

There exists a second method that makes use of this information as well as the loss function " $L$ ". The idea is to choose the label  $r$  which is more coherent with predictions  $f_s(x)$  in the sense that if a pattern  $x$  was labeled as  $r$ , the total loss over a pattern  $(x, r)$  should be minimized by choosing  $r \in Y$ . Formally this means that our distance measure is the total loss over  $(x, r)$ :

$$d_L(M(r), f(x)) = \sum_{s=1}^l L(M(r, s)f_s(x)) \quad (17)$$

Similarly to Hamming decoding, the label assigned  $\hat{y} \in \{1, \dots, k\}$  is

$$\hat{y} = \arg \min_r d_L(M(r), f(x)) \quad (18)$$

This reconstruction method is called loss-based decoding.<sup>35</sup> Some of the loss functions that can be considered are:

$$L(z) = (1 - z)_+$$

$$L(z) = e^{-z}$$

$$L(z) = (1 - z)^2$$

$$L(z) = \log(1 + e^{-2z})$$

In summary, to implement a multiclass classifier based on SVM, a *kernel* function and its parameters have to be defined, with coefficient  $C$  for (9) and the loss function  $L$  (17) to be used in the reconstruction method.

In this paper, the loss-based decoding reconstruction method was used since it takes into account the information on the prediction magnitude delivered by the binary classifier (SVM).



## Results

The results obtained in this study by the methodology described above were compiled into a database.

As mentioned in the Methodology section, we used the Haar wavelet as a FE technique in this study.<sup>27</sup> As a pattern recognition technique, we used SVM<sup>30</sup> with a radial basis function kernel. The SVM technique is then characterized by parameters  $C$  (penalization error parameter) and  $\sigma$ . Since it is being used in a multiclass problem, loss function  $L$  also has to be defined. We used  $L(z) = (1 - z)$  in our study due to its simplicity and good results obtained in preliminary simulations. The best values for  $C$  and  $\sigma$  were determined by means of cross-validation using the training-validation set.

The obtained results are summarized in Figures 5 and 6. Best classification results for the four valleys under study using the complete database (ie, not separated by the grape varieties of Cabernet Sauvignon, Merlot, and Carménère) are shown in Figure 5 for different wavelet decomposition levels (0, 1, ..., 5). For each decomposition level, the best values of  $C$  and  $\sigma$  are different in general. It is interesting to note that the best averages of correct classification are over 94% in the training-validation set (for the case where no wavelet extraction is used) and over 95% for the test set (for a wavelet decomposition level of 5).

In Table 2, the confusion matrix for the best case of the test set (level 5) is shown. The results are shown in terms of the number of samples as well as a percentage with respect to the total number of samples in the test set. Percentages on the diagonal represent the percentages of patterns correctly classified. Off-diagonal terms indicate the patterns are wrongly classified.

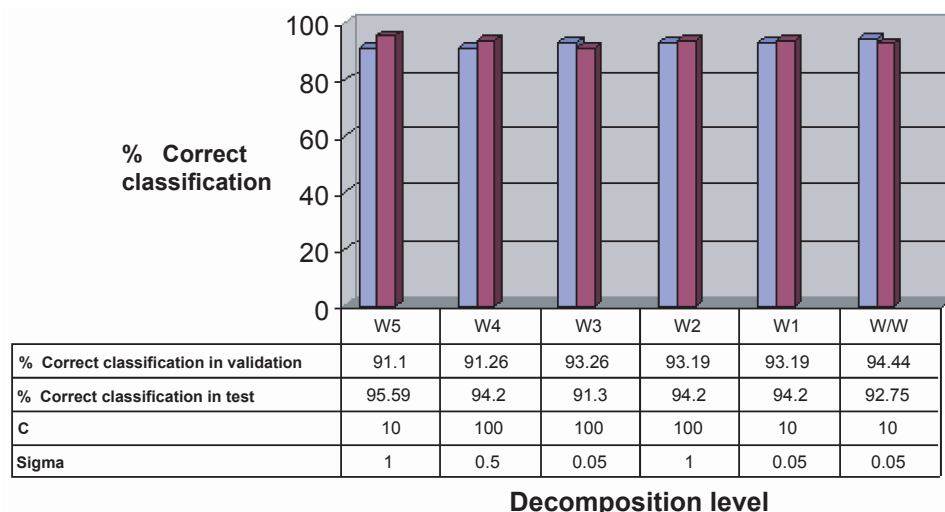
Table 2 shows that all samples from Colchagua and Rapel valleys are correctly classified. The greatest confusion occurs in the Maule valley where 18.18% of the samples are classified as Maipo provenance. This is due in part since the training-validation set has a larger number of samples from Colchagua and Maipo valleys.

To geometrically illustrate the obtained results, a “cluster” plot is shown in Figure 6. This plot corresponds to the Fisher transform where each pattern vector of dimension 600 has been projected onto a three-dimensional space (number of classes minus one) using the Fisher criterion. From this plot it is seen that patterns belonging to different classes (valleys) are not separated, nevertheless the classification process is able to separate the patterns into four classes with a 94% success rate.

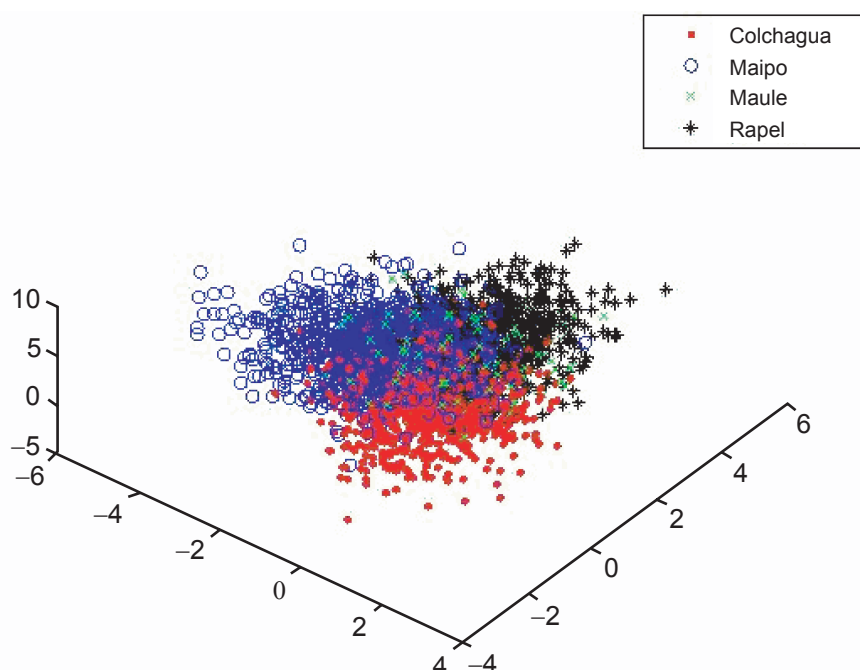
## Discussion and conclusions

From the results obtained in this paper, it can be observed that although Chile is a country with new soils from a geological point of view (the concept of *terroir* has not been yet defined by the regulation authority in Chile), organic volatile compounds taken by a zNose<sup>TM</sup> give enough information to determine the geographical origin of the product with advanced classification techniques.

Another interesting point arises from analyzing confusion matrixes computed during the classification procedure. This analysis revealed that wines labeled as produced in Rapel valley presented distinct characteristics from those wines labeled as produced in Colchagua valley, although Colchagua valley is geographically close to Rapel valley. Chilean authorities consider Colchagua valley as a subset of



**Figure 5** Best classification results for different decomposition levels of wavelet transform.



**Figure 6** Projection of patterns onto the Fisher space.

Rapel valley for OD purposes. After an exhaustive analysis of the data, it was observed that those wines labeled as Rapel valley were produced with grapes that are grown in the north and towards the mountain end of the valley, whereas those labeled as Colchagua valley were produced with grapes grown in vineyards located towards the coastal end of the valley (about 80 km away from the mountain end of the Rapel valley). This is an interesting observation, which reveals that our method is sensitive enough to differentiate geographical location within a same region.

Finally, it should be pointed out that the classification system presented here for Chilean wines was designed using a database containing samples of Chilean wines from four valleys of the central part of the country. However, the system is general enough to classify wines from any geographical region of the world provided that reliable samples from that region are collected and processed. The only constraint for getting reliable results is to build a sufficiently populated database.

**Table 2** Confusion matrix for best results on the test set (Level 5,  $C=1$  and  $\sigma=1$ )

Valley	Colchagua	Maipo	Maule	Rapel
Colchagua	22 (100%)	1 (4.54%)	0 (0%)	0 (0%)
Maipo	0 (0%)	21 (95.45%)	2 (18.18%)	0 (0%)
Maule	0 (0%)	0 (0%)	9 (81.81%)	0 (0%)
Rapel	0 (0%)	0 (0%)	0 (0%)	13 (100%)

## Acknowledgments

The results presented in this work were supported by CORFO-INNOVA under grant 05-CTE02-03 and CONICYT-Chile, under grant FONDEF D011-1016, “Chilean Red Wine Classification by means of Intelligent Instrumentation”. The authors report no conflicts of interest in this work.

## References

1. Eschnauer H, Neeb R. Microelement analysis in wine and grapes. In: Linskens HF, Jackson JF, editors. *Wine Analysis*. Berlin: Springer-Verlag; 1988.
2. Eschnauer H, Jakob L, Melerer H, Neeb R. Use and limitations of ICP-OES in wine analysis. *Mikrochimica Acta*. 1989;3:291–298.
3. Estévez-Pérez D. Caracterización y tipificación de los vinos de las Islas Canarias con denominación de Origen por su contenido en iones metálicos. Doctoral thesis. Tenerife: Universidad de La Laguna; 2002.
4. Fournier JB. Contribution à L'étude de L'évolution des teneurs en éléments traces dans les produits de la vigne durant le cycle reproductif et la vinification. Nantes: Thèse de L'Université de Nantes, Chimie Biologie-Sciences Agroalimentaires; 1998.
5. Gonzales-Larraina M, Gonzales A, Medina B. Les ions métalliques dans la différenciation des vins rouges, des trois régions d'appellation d'origine Rioja. *Connaissance Vigne Vin*. 1987;21(2):126–140.
6. Greenough JD, Longerich HP, Jackson SE. Element fingerprint of Okanagan Valley wines using ICP-MS. Relationships between wine composition, vineyard and wine colour. *Aust J Grape Wine Res*. 1997; 3:75–83.
7. Smolle A. Terroir: A study of the relationship between the vine and its environment. Berkeley, CA: Class notes of MCB 14 Enology Course; University of California, Berkeley; 1993.
8. Kallithraka S, Arvanitoyannis IS, Kefalas P, Eip-Zajouli A, Soufleros E, Psarra E. Instrumental and sensory analysis of Greek wines; implementation of principal component analysis (PCA) for classification according to geographical origin. *Food Chem*. 2001;73:501–514.

9. Kwan WO, Kowalski BR, Skogerboe RK. Pattern recognition analysis of elemental data. Wines of *Vitis vinifera* cv. Pinot Noir from France and the United States. *J Agric Food Chem*. 1979;27(6):1321–1326.
10. Latorre MJ, Herrero C, Medina B. Utilisation de quelques éléments minéraux dans la différenciation des vins de Galice. *Journal International Sciences Vigne Vin*. 1992;26(3):185–193.
11. Latorre MJ, Garcia-Jares C, Medina B, Herrero C. Pattern recognition analysis applied to classification of wines from Galicia (Northwestern Spain) with certified brand of origin. *J Agric Food Chem*. 1994;42:1451–1455.
12. Maarse H, Slump AC, Schaefer J. Classification of wines according to type and region based on their composition. *Z Lebensm Unters Forsch*. 1987;184:198–203.
13. Martin GJ, Mazure M, Jouitteau C, Martin YL, Aguile L, Allain P. Characterization of the geographic origin of Bordeaux wines by a combined use of isotopic and trace element measurements. *Am J Enol Vitic*. 1999;50(4):409–417.
14. Lozano J, Santos JP, Horrillo MC. Enrichment sampling methods for wine discrimination with gas sensors. *J Food Compos Anal*. 2008;21(8):716–723.
15. García M, Aleixandre M, Gutiérrez J, Horrillo MC. Electronic nose for wine discrimination. *Sens Actuators B Chem*. 2006;113(2):911–916.
16. Lozano J, Arroyo T, Santos JP, Cabellos JM, Horrillo MC. Electronic nose for wine ageing detection. *Sens Actuators B Chem*. 2008;133(1):180–186.
17. Aleixandre M, Lozano J, Gutiérrez J, Sayago I, Fernández MJ, Horrillo MC. Portable e-nose to classify different kinds of wine. *Sens Actuators B Chem*. 2008;131(1):71–76.
18. Fernández MJ, Fontecha JL, Sayago I, et al. Discrimination of volatile compounds through an electronic nose based on ZnO SAW sensors. *Sens Actuators B Chem*. 2007;127(1):277–283.
19. Beltrán NH, Duarte-Mermoud MA, Soto VA, Salah SA, Bustos MA. Chilean wines classification using volatile organic compounds data obtained with a fast GC analyzer. *IEEE Transactions on Instrumentation and Measurement*. 2008;57(11):2421–2436.
20. Electronic Sensor Technology. *7100 Fast GC analyzer: Operation manual*. Newbury Park, CA: Electronic Sensor Technology; 1999.
21. Shannon C. Communication in the presence of noise. *Proc Institute of Radio Engineers*. 1949;37(1):10–21.
22. Middleton RH, Goodwin GC. *Digital control and estimation. A unified approach*. Upper Saddle River, NJ: Prentice Hall International; 1990.
23. Fleiss JL. *Statistical methods for rates and proportions*. 2nd Ed. New York, NY: John Wiley; 1981.
24. Fukunaga K. *Introduction to statistical pattern recognition*. 2nd Ed. New York, NY: Academic Press; 1990.
25. Jolliffe LT. *Principal component analysis: A beginner's guide*. New York, NY: Springer-Verlag; 1990.
26. Theodoridis S, Koutroumbas K. *Pattern recognition*. New York, NY: Academic Press; 1999.
27. Antonini M, Barlaud M, Mathieu P, Daubechies I. Image coding using wavelet transform. *IEEE Trans Image Process*. 1992;1(2):205–220.
28. Beltrán NH, Duarte-Mermoud MA, Bustos MA, et al. Feature extraction and classification of Chilean wines. *J Food Eng*. 2006;75(1):1–10.
29. Hastie T, Tibshirani R, Friedman J. *The Elements of Statistical Learning. Data Mining, Interference and Prediction*. New York, NY: Springer; 2001.
30. Vapnik V. *Statistical Learning Theory*. New York, NY: John Wiley and Sons Inc; 1998.
31. Cristianini N, Shawe-Taylor J. *An introduction to support vector machines and other kernel-based learning methods*. Cambridge, MA: Cambridge University Press; 2000.
32. Angulo C. Aprendizaje con máquinas de núcleo en entornos de multclasificación. Tesis Doctoral, Doctor en Ciencias, especialidad Matemáticas. Departament d'Enginyeria de Sistemes, Automàtica i Informàtica Industrial, Universitat Politècnica de Catalunya; 2001.
33. Hastie T, Tibshirani R. Classification by pairwise coupling. *Ann Stat*. 1998;26(2):451–471.
34. Dietterich T, Bakiri G. Solving multiclass learning problems via error-correcting output codes. *Journal of Artificial Intelligence Research*. 1995;2:263–286.
35. Allwein E, Shapire R, Singer Y. Reducing multiclass to binary: A unifying approach for margin classifiers. *Proceedings of the Seventeenth International Conference on Machine Learning*. 2000:9–16.

## International Journal of Wine Research

### Publish your work in this journal

The International Journal of Wine Research is an international, peer-reviewed open-access, online journal focusing on all scientific aspects of wine, including: vine growing; wine elaboration; human interaction with wine; and health aspects of wine. The journal provides an open access platform for the reporting

Submit your manuscript here: <http://www.dovepress.com/international-journal-of-wine-research-journal>

of evidence based studies on these topics. The manuscript management system is completely online and includes a very quick and fair peer-review system, which is all easy to use. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from some of our published authors.