

Monovarietal extra-virgin olive oil classification: a fusion of human sensory attributes and an electronic tongue

Luís G. Dias^{1,2} · Nuno Rodrigues³ · Ana C. A. Veloso^{4,5} · José A. Pereira³ · António M. Peres⁶

Received: 27 May 2015 / Revised: 12 August 2015 / Accepted: 22 August 2015 / Published online: 3 September 2015
© Springer-Verlag Berlin Heidelberg 2015

Abstract Olive oil quality grading is traditionally assessed by human sensory evaluation of positive and negative attributes (olfactory, gustatory, and final olfactory–gustatory sensations). However, it is not guaranteed that trained panelist can correctly classify monovarietal extra-virgin olive oils according to olive cultivar. In this work, the potential application of human (sensory panelists) and artificial (electronic tongue) sensory evaluation of olive oils was studied aiming to discriminate eight single-cultivar extra-virgin olive oils. Linear discriminant, partial least square discriminant, and sparse partial least square discriminant analyses were evaluated. The best predictive classification was obtained using linear discriminant analysis with simulated annealing selection algorithm. A low-level data fusion approach (18 electronic tongue signals and nine

sensory attributes) enabled 100 % leave-one-out cross-validation correct classification, improving the discrimination capability of the individual use of sensor profiles or sensory attributes (70 and 57 % leave-one-out correct classifications, respectively). So, human sensory evaluation and electronic tongue analysis may be used as complementary tools allowing successful monovarietal olive oil discrimination.

Keywords Single-cultivar extra-virgin olive oil · Sensory analysis · Potentiometric electronic tongue · Linear multivariate methods · Simulated annealing algorithm

Introduction

There is an increasing need for developing appropriate methodologies for guaranteeing olive oil commercial category (extra virgin, EVOO; virgin, VOO; and lampante olive oils, LOO), as well as for identifying geographical or olive cultivar origin [1–3]. Olive oil organoleptic attributes [1, 4–6] and physicochemical parameters [7, 8] are important for olive oil quality assessment and discrimination, avoiding the indiscriminate use of uncorrected definitions in olive oil labels [1]. These analysis are time-consuming and costly, requiring trained sensory panelists and high-skilled technicians, respectively. Moreover, sensory evaluation is mainly used to classify olive oils as LOO, VOO, or EVOO according to their positive and/or negative organoleptic sensations. In the last years, several analytical approaches, ranging from expensive chromatographic [2, 6, 9–13]-, DNA [14]-, or nondestructive spectroscopy [15, 16]-based methods to fast and low-cost electrochemical devices [3, 13, 17–19], have been reported. These methods showed satisfactory predictive performances regarding olive oil quality assessment and classification, including the successful

✉ António M. Peres
peres@ipb.pt

¹ Escola Superior Agrária, Instituto Politécnico de Bragança, Campus Santa Apolónia, Apartado 1172, 5301-855 Bragança, Portugal
² CQ-VR, Centro de Química – Vila Real, University of Trás-os-Montes e Alto Douro, Apartado 1013, 5001-801 Vila Real, Portugal
³ CIMO - Mountain Research Centre, Escola Superior Agrária, Instituto Politécnico de Bragança, Campus Santa Apolónia, Apartado 1172, 5301-855 Bragança, Portugal
⁴ Instituto Politécnico de Coimbra, ISEC, DEQB, Rua Pedro Nunes, Quinta da Nora, 3030-199 Coimbra, Portugal
⁵ CEB - Centre of Biological Engineering, University of Minho, Campus de Gualtar, 4710-057 Braga, Portugal
⁶ LSRE - Laboratory of Separation and Reaction Engineering - Associate Laboratory LSRE/LCM, Escola Superior Agrária, Instituto Politécnico de Bragança, Campus Santa Apolónia, Apartado 1172, 5301-855 Bragança, Portugal

discrimination of EVOO, VOO, and LOO; the identification of monovarietal EVOO and their geographical origin as well as the identification of potential putative markers for the recognition of olive oil adulteration [2]. Less attention has been given to the possibility of using sensory data for classifying monovarietal EVOO according to its olive cultivar. In a previous work, Dias et al. [3] showed that an electronic tongue (E-tongue) could be employed to discriminate three Portuguese or three Spanish monovarietal EVOOs, from two crop years, according to olive cultivar (100 % and 98 % of correct classifications, respectively). However, when the six single-cultivar EVOOs were evaluated simultaneously, the correct classification percentage dramatically decreased to 43 %. This drawback only overcome when EVOO samples were split by production year, which enabled to reduce data variability within each year group, allowing to increase the correct classification rate of nine single-cultivar EVOOs to 93–95 % [20].

On the other hand, some research works aiming for the improvement of olive oil quality assessment and classification according to olive cultivar or geographical origin reported the successful use of data fusion approaches, which merged data from different instrumental analytical methods, such as headspace mass spectrometry (electronic nose), ultraviolet–visible and near-infrared spectroscopy [21]; near- and mid-infrared spectroscopy [1, 22]; ultraviolet–visible, near- and mid-infrared spectroscopy together with fatty acid composition profile from gas chromatography analysis [23]; electronic nose and tongue [19]; and visible spectroscopic fingerprints and chemical descriptors [24]. Indeed, data fusion enabled to obtain compatible measurements originated from different sources [25]. Data fusion approaches, with different abstraction levels (low, mid, or high level), were effectively applied to differentiate or classify different food matrices or their attributes, namely beers [25, 26], snack foods [27], fruit juices [28], and black teas [29].

In this work and for the first time, the feasibility of a low-level data fusion approach, combining E-tongue potentiometric signals and olfactory and/or gustatory sensations,

was studied aiming to improve the discrimination of high-value monovarietal EVOO according to olive cultivar. Signal profiles were recorded using an E-tongue, with cross-sensitivity nonspecific lipid membranes as previously described [3]. The single-cultivar EVOO organoleptic profiles were evaluated by four trained sensory panelists. Eight Spanish monovarietal EVOOs were used (cvs. Arbosana, Arroniz, Cornicabra, Frantoio, Manzanilla, Redondilla, Royuela, and Zorzal). Four supervised linear multivariate models were tested to establish EVOO classification according to olive cultivar based on the independent variables, potentiometric sensor signals, and/or sensory descriptors. As previously reported [3], hydroethanolic extracts of EVOO, potentially rich in polar compounds that may be correlated with bitterness, astringency, and pungency attributes, were analyzed to overcome the difficulty of carrying out electrochemical assays in viscous non-conductive liquids [30]. In summary, this work aimed to demonstrate the complementary capabilities of human and artificial sensory analysis.

Materials and methods

Olive oil samples

Eight monovarietal Spanish EVOOs (cv. Arbosana (ARBO), cv. Arroniz (ARR), cv. Cornicabra (COR), cv. Frantoio (FRA), cv. Manzanilla (MAN), cv. Redondilla (RED), cv. Royuela (ROY), and cv. Zorzal (ZOR)), produced at the north of Spain (Valladolid region), were studied. These traditional olive varieties were selected due to their recognized high-quality, peculiar, and differentiated organoleptic characteristics. Moreover, the capability of discriminating these EVOOs has become quite relevant since some of these olive cultivars are usually produced in intensive olive groves (e.g., ARBO and FRA) and so cannot be used to obtain Protected Denomination of Origin olive oils. In total, 22 different single-cultivar EVOOs were obtained directly from olive oil certified producers during 2012 and 2013 (Table 1). For each one of the EVOOs, two

Table 1 Sample details of the Spanish monovarietal EVOO obtained in Valladolid region

Varietal (cvs.)	Number of EVOOs	Number of samples (bottles from different production lots)	Production year(s)
Arbosana (ARBO)	4	8	2012 and 2013
Arroniz (ARR)	2	4	2012
Cornicabra (COR)	4	8	2012 and 2013
Frantoio (FRA)	4	8	2012 and 2013
Manzanilla (MAN)	2	4	2012
Redondilla (RED)	2	4	2012
Royuela (ROY)	2	4	2013
Zorzal (ZOR)	2	4	2013

bottles of different production lots were analyzed, totaling 44 independent olive oil samples. Olive oils were packed and stored in dark at $-20\text{ }^{\circ}\text{C}$ for a 24-h period after their production in olive mills with a two-phase extraction process and kept in those conditions until further analysis [3].

Sensory analysis

Olive oil samples were subjected to organoleptic assessment following the methods and standards adopted by the International Olive Council (IOC or COI) for sensory analysis of olive oils, namely COI/T.20/Doc. No. 15/Rev. 6 [31] and COI/T.30/Doc. No. 17 [32]. Each sample was subjected to the judgment of four trained panel members who classify the samples in a scale according to olfactory sensations, gustatory–retronasal sensations, and final olfactory–gustatory sensations. For olfactory sensations, the following attributes were measured: olive fruitiness (0–7); other fruits (0–3); green (grass/leaves) (0–2); other positive sensations (0–3); and harmony (0–20). Concerning gustatory–retronasal sensations were evaluated the olive fruitiness (0–10); sweet (0–4); bitter (0–3); pungent (0–3); green (grass/leaves) (0–2); other positive sensations (0–3); and harmony (0–20). A final olfactory–gustatory sensation for each sample was also considered, conjugating all the organoleptic sensations, pointing out the complexity (0–10) and persistence (0–10). It should be remarked that harmony increases when the attributes are balanced, and complexity increases with the number and intensity of aromas and flavors. Although COI guidelines recommend the use of 8–12 tasters for each analysis, it was decided to use only four trained panelists (three from Portugal and one from Spain), which are quite familiar with the classification of positive and negative olive oil organoleptic attributes and usually are jury members in olive oil classification in competitions. With this approach, it was intended to guarantee a more consistent sensory classification.

Sample preparation and E-tongue analysis

All samples were electrochemically analyzed in the same day, which turned out in different storage times, namely 1 year or less. The olive oil extraction procedure, using hydroethanolic solutions ($\text{H}_2\text{O}:\text{EtOH}$, 80:20 v/v), was previously reported by Dias et al. [3]. One independent sample was collected and extracted from each bottle of each monovarietal EVOO and analyzed in duplicate (totalizing 22 olive oils, 44 bottles, and 88 analyses). The E-tongue included two print-screen potentiometric devices, containing different cross-sensitivity membranes as chemical sensors [3]: four lipidic additives (octadecylamine, oleyl alcohol, methyltrioctylammonium chloride, and oleic acid from Fluka; corresponding to approximately

3 %), five plasticizers (bis(1-butylpentyl) adipate, dibutyl sebacate, 2-nitrophenyl-octylether, tris(2-ethylhexyl)phosphate, and dioctyl phenylphosphonate from Fluka; representing around 65 %) and high molecular weight polyvinyl chloride (PVC); near 32 %). The type of sensors and polymeric membrane compositions (relative percentage of additive, plasticizer, and PVC) were selected based on a previous work [33] taking into account their satisfactory signal stability over time ($\% \text{RSD} < 5\%$) and repeatability ($0.5\% < \% \text{RSD} < 15\%$) for basic standard taste compounds (e.g., sweet, acid, bitter, salty, and umami). Further details about each sensor are described by Sousa et al. [34]. Also, lipid polymeric membranes were used since they may interact with tastant substances through electrostatic or hydrophobic interactions [35]. Each sensor was identified with a code with a letter S (for sensor) followed by the number of the array (1 or 2) and the number of the membrane (1–20, corresponding to different combinations of plasticizer and additive used).

Data fusion

Three data fusion approaches are usually applied. The main differences rely on the abstraction levels, i.e., the way how data originated from several analytical techniques or different sources, can be merged, and form a consistent concatenated data matrix. In low-level of abstraction data fusion, the data from all sources are simply concatenated before model development, resulting in a data matrix with a dimension of number of samples \times total number of variables from all sources. This approach may improve model classification performance namely multisensor data fusion [19]. In mid-level fusion, a feature extraction is applied to each data source followed by merging the extracted features. For high level of abstraction, data from each source are initially analyzed, and a model is derived, separately, from each data source, being the model classification outputs then merged [26, 28, 36]. In this work, a low-level data fusion was adopted.

Statistical analysis

Different chemometric methods were applied. First, sensory data were subjected to an analysis of variance (two-way ANOVA) to infer about the existence of a significant effect of production year or olive cultivar. If a significant effect was found, the Tukey's post hoc multiple comparison test was further employed.

Four linear multivariate methods were explored to establish monovarietal EVOO classification models using independent predictors among the E-tongue potentiometric signals recorded and/or the sensory attributes assessed by the trained panelists, namely linear discriminant analysis

(LDA) and partial least squares discriminant analysis (PLS-DA), as the most common linear techniques applied in classification studies; sparse partial least squares discriminant analysis (SPLS-DA), as a PLS improved methodology, which allows to obtain a simplified PLS model by applying a feature variable selection; and finally, a LDA approach coupled with meta-heuristic simulated annealing algorithm (LDA-SA), which uses a variable selection step.

The LDA is a supervised technique used as a starting point in classification studies since the model obtained is represented by linear functions of the original variables. So, the importance of the variables in the classification process can be easily perceived [37]. The PLS-DA is a more complex supervised classification technique and should be used if a deeper mathematical treatment is required. This technique is usually applied when p explanatory variables show high multicollinearity and/or are greater than the total number of samples/observations (N). The PLS procedure implies a dimension reduction step by forming new orthogonal variables (linear combinations of the original variables, namely principal component (PCs) functions, which coefficients involve information about the variances and covariances of both the explanatory and response variables. The model obtained is complex since new variables are represented in a new dimensional space [37]. The SPLS-DA is based in the sparsity principle [38], on the basis that not all explanatory variables are needed to the fit model, i.e., meaning that a subset of original variables could be mainly responsible for determining the response. Therefore, SPLS-DA also has the ability to perform feature selection keeping the structure of PLS-DA model. Considering that some variables may enhance noise effects in PLS-DA prediction, the SPLS-DA technique imposes sparsity on the dimension reduction step of PLS by selecting variables and increasing the predictive power of the model [37]. As in PLS technique, SPLS uses a data matrix computed by principal components and not the original information of explanatory and response variables. A major disadvantage of LDA compared with the other two other multivariate techniques is the lack of robustness regarding the existence of collinear variables. To overcome this drawback, LDA was also implemented coupled with a meta-heuristic simulated annealing (SA) variable selection algorithm, enabling the selection of the best subsets of independent predictors among the E-tongue potentiometric signals recorded and/or the sensory sensations assessed by the trained panelists. This algorithm pursues to select the optimal conditions based on the annealing physic process, which mimics the slow controlled cooling process of a heated material in order that the material can reach the most regular possible crystal lattice configuration (free from defects at a minimum energy crystalline state). It is classified as a meta-heuristic algorithm, since it is able to identify a subset of the

original independent variables that correspond to a global optimum for a given approximation criterion, selected within a large search space of other possible subsets of variables [39]. This means that for optimizing a system by selecting the best subset of k variables considering this type of thermodynamic behavior, in each iteration, values of two solutions (the current subset of k variables and the new subset to be tested, also with k variables) are compared according to a criterion that measures the quality of those subsets of variables (for instance, accuracy of a classification model). The new solution is randomly selected in the neighborhood of the current solution and tested according to the simulated annealing rules, becoming the new solution if the criterion has best values than the original. The algorithm continues the search for new solutions until it reaches the maximum number of interactions established at the beginning of procedure.

To evaluate the classification models' stability and quality, a leave-one-out cross-validation (LOO-CV) procedure was applied. Although this methodology may be considered an overoptimistic procedure, the use of a more robust cross-validation approach (e.g., based on an external validation set) was not possible due to the large number of groups considered (eight monovarietal EVOOs) and the reduction in the number of independent olive oil samples for each group (between 2 and 4). Indeed, when the number of samples is small, LOO-CV procedure has proven to be an adequate procedure to test model prediction capability and to minimize the risk of overoptimistic predictive model performance [3, 25, 40].

So, as a first step, the discriminative capability of the olfactory–gustatory sensations was tested, being the quality of the established models (LDA, PLS-DA, SPLS-DA, and LDA-SA) were evaluated based on the LOO-CV procedure. Secondly, the same procedure was applied to the potentiometric signal profiles, and as before, their classification capabilities were also evaluated using the LOO-CV methodology. To the best model, the spatial distribution of the single-cultivar EVOO original groups was tentatively correlated with the taste sensory attributes evaluated by the trained panelists, by means of the linear Pearson correlation coefficient (R -Pearson), for inferring about the practical potential of the E-tongue in mimicking the human taste sensory perception of the monovarietal EVOO analyzed.

Finally, a low-level of abstraction data fusion approach was applied, combining sensor signal profiles and sensory descriptors of the olive oils, to evaluate synergies in the information contained in both data sets. Data set size of each source is of similar magnitude (44×40 and 44×14 ; for E-tongue and sensory trained panel analysis, respectively) minimizing the possibility of the largest data set collapsed the other [28, 41, 42]. To normalize the weight of each variable for each of the final linear multivariate

classification models, variable scaling and centering procedures were applied. The main goal of this approach was to improve Spanish monovarietal EVOO discrimination according to olive cultivar.

All statistical analyses were performed using the caret [37, 43], Subselect [39, 44], pls [37, 45], spls [37, 46], and MASS [47] packages of the open-source statistical program R (version 2.15.1).

Results and discussion

Sensory data evaluation

Each monovarietal EVOO sample was evaluated by four trained sensory panelists and classified according to 14 organoleptic descriptors including five olfactory sensations, seven gustatory–retronasal sensations, and two final olfactory–gustatory sensations, according to the IOC directives [31, 32]. The two-way ANOVA showed that some olive oil sensory descriptors were significantly affected by the production year (olfactory sensations: olive fruitiness, green, and harmony; gustatory–retronasal sensations: bitter, pungent, green, and harmony; final olfactory–gustatory sensations: persistence; p values <0.0154) and olive oil cultivar (olfactory sensations: olive fruitiness and harmony; gustatory–retronasal sensations: sweet, pungent, other positive sensations, and harmony; final olfactory–gustatory sensations: persistence; p values <0.0357), but the interaction effects were not statistically significant (p value >0.1050) at a 5 % significance level. The differences found between production years may be related to the different edaphoclimatic conditions. However, in this work, olive oils were used, regardless of their production year, trying to evaluate the potentiality of the data fusion approach. The significant effect of olive cultivar in some olive oil sensory sensations may be attributed to their different physicochemical composition, which affects the organoleptic characteristics such as pungency, bitterness, and astringency. In Fig. 1, a comparison of mean sensory data from the quantitative descriptive analysis (QDA), for the eight single-cultivar EVOO evaluated, is shown. The radar plot, for each sensory descriptor, also shows which olive cultivar means are significantly different at a 5 % level, according to the post hoc Tukey's multiple comparison test. Globally, it can be stated that EVOO from cvs. Redondilla and Cornicabra are those that show more differences within olfactory sensations. Based on the gustatory–retronasal sensations, it was found that EVOOs from cvs. Manzanilla, Redondilla, and Ruyuela were those that showed the greater differences.

Finally, the capability of the sensory descriptors to discriminate the different single-cultivar EVOO was assessed by means of the supervised linear techniques: LDA,

PLS-DA, SPLA-DA, and LDA-SA. Table 2 shows the accuracy obtained for all these multivariate techniques. The LDA and PLS-DA results were obtained considering all the independent variables in the model. For PLS-DA and SPLS-DA approaches, the number of PCs in the model was selected from a search process with a maximum number of 14 possible PCs [equal to the number of PCs defined by principal component analysis (PCA)]. Also, the SPLS-DA performed feature variable selection within the PCs allowing to increase the predictive power of the model. The LDA-SA uses SA algorithm to identify a subset of the original independent variables. To establish all the multivariate classification models, the global optimum criterion used was the accuracy (i.e., the level of correct classifications using the LOO-CV procedure also called sensitivity).

The best result concerning single-cultivar EVOO classification based on olive oil sensory attributes (greater accuracy percentage) was obtained from the LDA-SA technique (57 % of correct classifications for LOO-CV, according to olive cultivar). The worst performances were achieved by LDA and PLS-DA methods (39 % of correct classifications). Since among the organoleptic attributes no multicollinearity issues were found (R^2 -Pearson <0.48), the poor results of LDA and PLS-DA techniques may be due to the presence of variables that increase noise effects affecting their prediction capabilities, being this problem slightly reduced, but not completely eliminated, with the variable selection procedure of SPLS-DA technique.

The SA algorithm used with the LDA technique, allowed selecting a set of seven sensory attributes (two olfactory sensations: olive fruitiness and other fruits; four gustatory–retronasal sensations: sweet, bitter, and pungency; and one final olfactory–gustatory sensation: complexity) from the 14 available organoleptic sensations, resulting in a linear discriminant model with seven functions (100 % of explained variance). The model obtained enabled a correct classification percentage of only 84 and 57 % for the original data group (Fig. 2, for the first three discriminant functions) and for the LOO-CV procedure. So, it is clear that, although the above-mentioned sensory descriptors may contribute to EVOO classification according to olive cultivar, they cannot be used for discrimination issues, and alone, these attributes cannot be applied to guarantee correct monovarietal EVOO labeling. Indeed, except for EVOO from cvs. Redondilla and Zorzal, all other single-cultivar EVOOs present misclassification problems. Nevertheless, it should be noticed that from the sensory descriptors assessed by the panelists, gustatory–retronasal sensations were those that most contributed for olive cultivar recognition. In conclusion, the known capability of human sensory evaluation to assess organoleptic characteristics of virgin olive oil used for quality classification seems to be not sufficient to discriminate high-quality

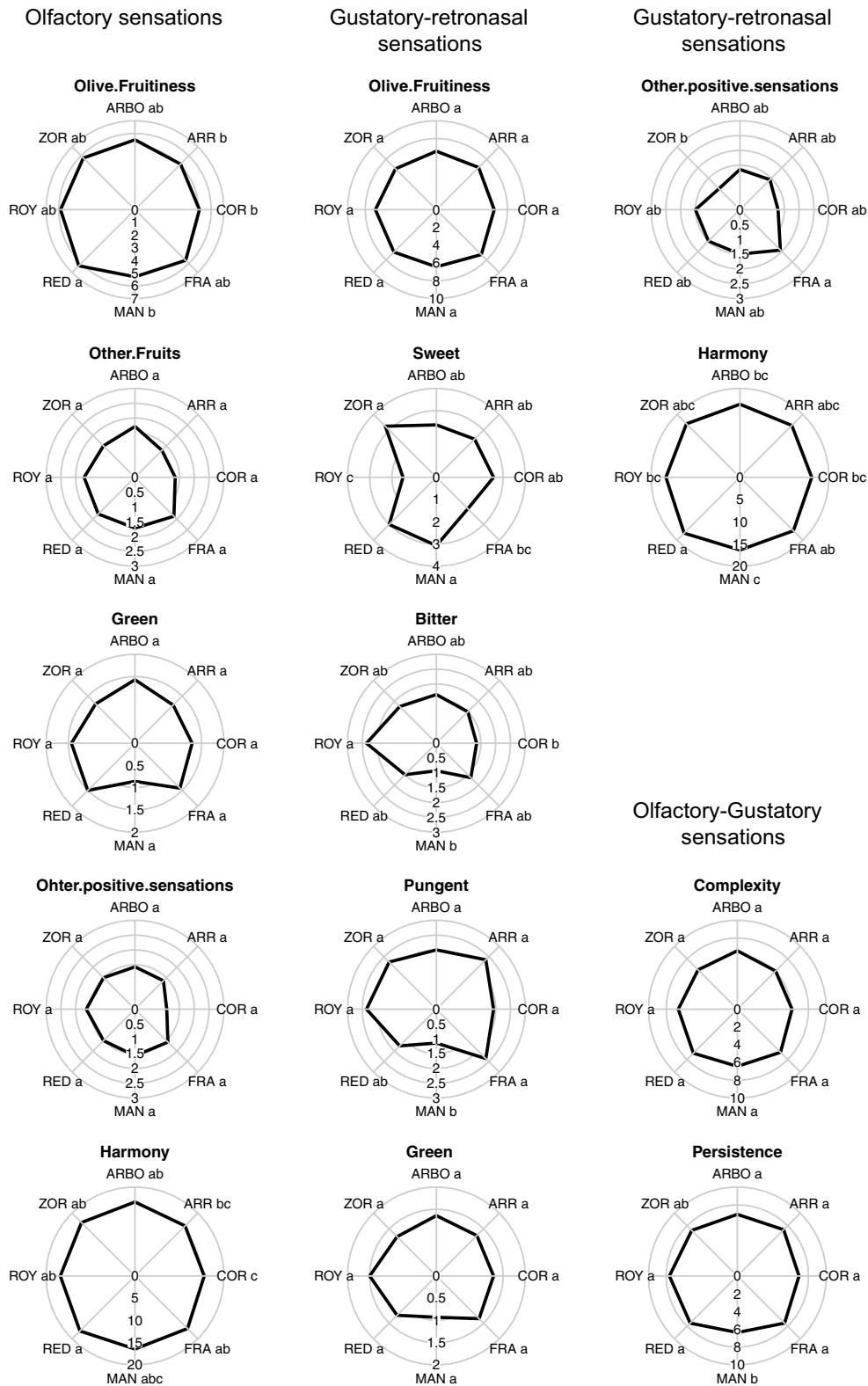


Fig. 1 Radar plot of the olive oil sensory descriptors assessed by the trained panelists for each monovarietal EVOO (ARB: cv. Arbosana; ARR: cv. Arroniz; COR: cv. Cornicabra; FRA: cv. Frantoio; MAN: cv. Manzanilla; RED: cv. Redondilla; ROY: cv. Royuela; and ZOR:

cv. Zorzal). Different letters for the same sensory descriptor point out significant statistical differences identified by Tukey’s post hoc multiple comparison test ($p < 0.05$)

Table 2 Discriminative capability of the sensory descriptors (olfactory–gustatory sensations), potentiometric signal profiles, and low-level of abstraction data fusion approach was applied, combining sensor signal profiles and sensory descriptors of the olive oils, toward single-cultivar EVOO differentiation, based on the accuracy values in the LOO-CV procedure

Data matrix	Accuracy for LOO-CV (%) ^a			
	LDA	PLS-DA	SPLS-DA	LDA-SA
Sensory attributes	39	39	45	57
No of selected sensory attributes	14	14	7	7
No of principal components	–	8	4	–
E-tongue sensor signal profiles	23	27	27	70
No of selected E-tongue signals	40	40	29	26
No of principal components	–	24	1	–
Low-level data fusion	7	41	48	100
No of selected sensory attributes	14	14	4	9
No of selected E-tongue signals	40	40	8	18
No of principal components	–	10	5	–

^a Accuracy = % of correct classifications

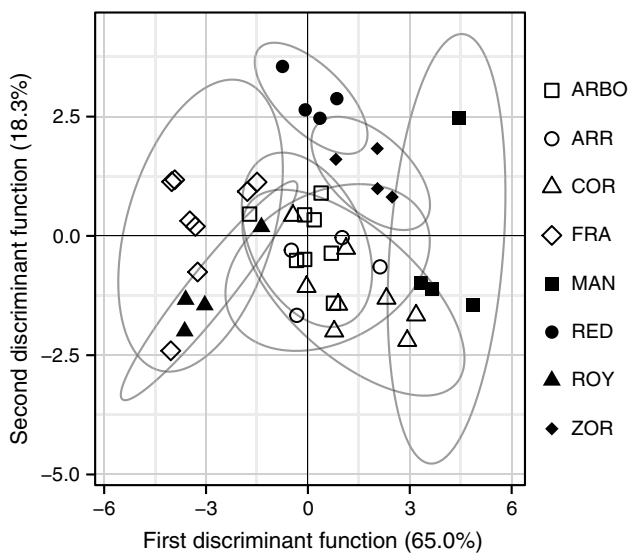


Fig. 2 Spanish monovarietal EVOO classification performance for the sensory attributes data: first and second LDA functions based on the best subset of sensory descriptors (seven olfactory, gustatory–retro-nasal, and final olfactory–gustatory sensations), assessed by trained panelists, selected using the SA meta-heuristic algorithm (ARB: cv. Arbosana; ARR: cv. Arroniz; COR: cv. Cornicabra; FRA: cv. Frantoio; MAN: cv. Manzanilla; RED: cv. Redondilla; ROY: cv. Royuela; and ZOR: cv. Zorzal)

monovarietal EVOO according to olive cultivar. Nevertheless, its potential exists and could be complemented by an artificial electronic sensory device like the electronic tongue.

E-tongue signals analysis

On the whole, 88 assays were carried out (22 samples of eight EVOO × 2 bottles × 1 extraction of each bottle × 2 analysis), each providing 40 potentiometric signals, varying from -0.03 V to +0.25 V (20 different sensor membranes used in duplicate: S1:1–S1:20 and S2:1–S2:20, as previously discussed [3, 40]). Since the voltage signals, recorded during the analysis of the hydroethanolic extracts of the monovarietal EVOO studied, are similar for all sensors and the respective range of variation is relatively narrow, as shown in Fig. 3 for two EVOOs (e.g., ARR and MAN), there was no need of data scaling. In fact, for the duplicate analysis of the hydroethanolic extract of each of the 44 olive oil bottles, the %RSD values of the signals recorded with each one of 40 lipid membranes varied from 0.25 to 1.8 %; for the analysis of the two bottles collected for each monovarietal EVOO, the %RSD ranged from 5.6 to 8.7 %; and finally, for the analysis of all bottles collected, the %RSD varied from 4.6 to 6.9 %. These results showed satisfactory accuracy of the E-tongue analysis.

The results from the application of the four linear multivariate techniques using the E-tongue potentiometric signal profiles to discriminate the different single-cultivar EVOOs are presented in Table 2. As in the previous section, the comparison between the four techniques was made based on the accuracies for LOO-CV procedure. In this case, the optimum number of PCs for PLS-DA and SPLS-DA models was selected within 38 possible PCs (number of PCs defined by PCA). Again, the best performance was obtained from the LDA-SA technique (70 % of correct classifications for LOO-CV) compared with the three other linear multivariate techniques (LDA, PLS-DA, and SPLS-DA: ≤27 % of correct classifications). Indeed, the SA variable selection procedure allowed choosing the most relevant signals for establishing the best predictive LDA model, minimizing the risk of using high collinear variables. So, the resulting LDA-SA classification model had three significant discriminant functions, explaining 98.86 % of the original data variability. The model was based on 26 non-collinear sensor signals (S1:2, S1:4, S1:5, S1:7, S1:9–S1:13, S1:15; S1:17, S1:18, S2:1–S2:3, S2:5–S2:7, S2:9–S2:11, S2:13, S2:14; S2:17–S2:19), allowing 100 and 70 % of correct classifications for original data (Fig. 4) and LOO-CV procedure, respectively. The SA

Fig. 3 E-tongue potentiometric signal profiles of the hydroethanolic extracts of two monovarietal EVOO (ARR: cv. Arroniz; MAN: cv. Manzanilla)

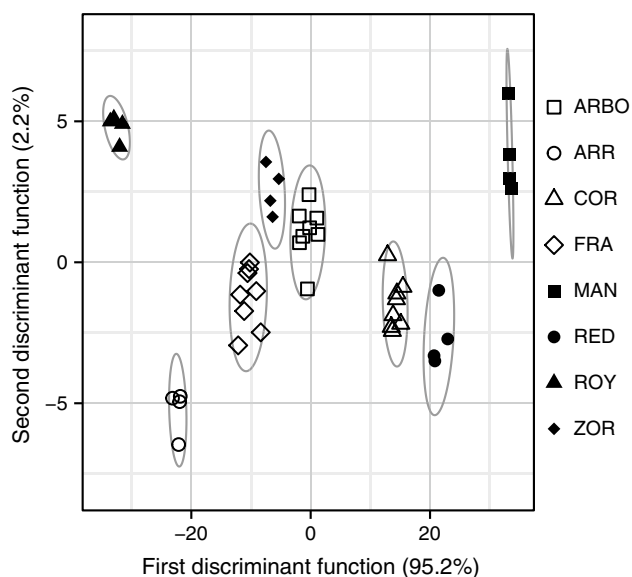
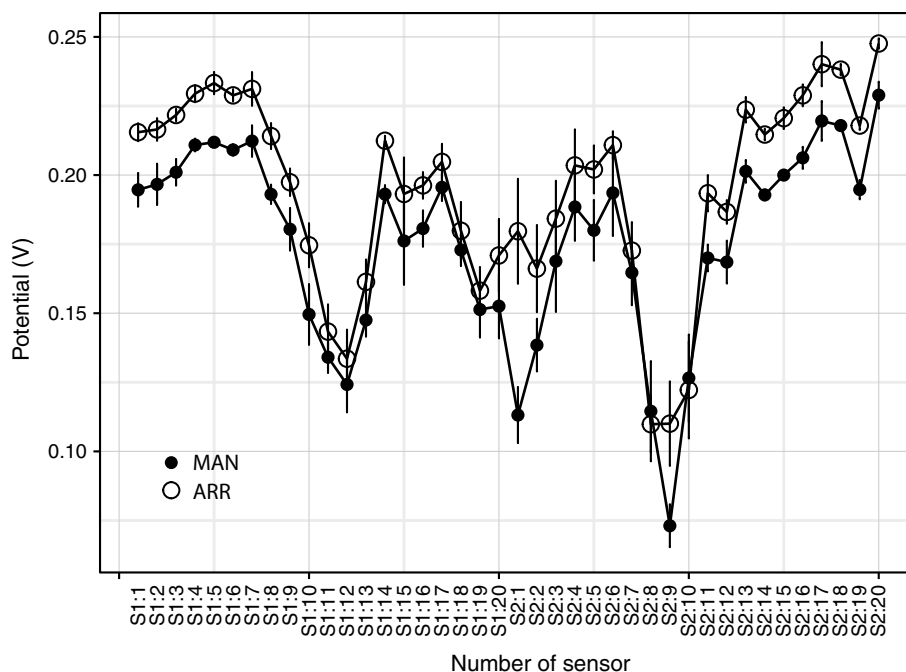


Fig. 4 Spanish monovarietal EVOO classification performance for the sensor signals data: first and second LDA functions based on the best subset of E-tongue potentiometric signals (26 sensors) selected using the SA meta-heuristic algorithm (ARB: cv. Arbosana; ARR: cv. Arroniz; COR: cv. Cornicabra; FRA: cv. Frantoio; MAN: cv. Manzanilla; RED: cv. Redondilla; ROY: cv. Royuela; and ZOR: cv. Zorzal)

algorithm selected seven repeated sensors, with the same membrane composition. The inclusion of repeated sensors could be justified by their slight different physical properties due to the possible inhomogeneous membrane surface, which could be attributed to the drop-by-drop membrane preparation technique [40]. On the other hand, the inclusion

of repeated sensors in a multivariate model may increase its qualitative and/or quantitative performance [48].

It should be noticed that 65 % of the E-tongue sensors (17 of the 26) used in the classification model are the same as reported by Dias et al. [3] for successfully discriminating three Spanish monovarietal EVOOs (cvs. Arbosana, Hojiblanca, and Picual), which were also produced in the same geographical region as those analyzed in the present work. Although the predictive LDA-SA model sensitivity was not very satisfactory (only EVOO samples from cvs. Arbosana, Arroniz, and Royuela were not misclassified), a clear increase was obtained compared to that of the LDA-SA model established using sensory descriptors evaluated by the trained panelists. Also, as already pointed out [3], the capability of the potentiometric E-tongue to correctly classify single-cultivar EVOO according to olive cultivar may be attributed to the nature and level of the polar compounds present in the EVOO extracts analyzed, which could be related to sensory olive oil attributes such as bitterness, astringency, and pungency. Indeed, this evidence was further demonstrated in the present work, since the spatial distribution of the EVOO group centroids (mean value of the discriminant scores for a given category) of the first linear discriminant function (cvs. centroids order: Royuela < Arroniz < Frantoio < Zorzal < Arbosana < Cornicabra < Redondilla < Manzanilla), which explained more than 95 % of the potentiometric E-tongue signals variability, could be statistically correlated with gustatory–retronasal descriptors assessed by the trained panelists: bitter (R -Pearson = -0.83 ; p value = 0.0110), pungent (R -Pearson = -0.91 ; p value = 0.0017), and green

(R -Pearson = -0.80 ; p value = 0.0180) sensations. These results strengthen the idea that the potentiometric E-tongue can mimic olive oil gustatory–retronasal human sensations, which was shown for the first time using sensory sensations data. Then again, this behavior was expected since this type of lipid polymeric membranes could recognize and assess standard solutions of basic tastant substances such as acid, bitter, salty, sweet, and umami [33].

Low-level data fusion of sensory attributes and E-tongue signals

The results previously obtained using sensory attributes or E-tongue signals individually showed that both methodologies could contribute to monovarietal EVOO discrimination by olive cultivar although with unsatisfactory performances. So, the simultaneous use of both human and artificial evaluations was further explored trying to demonstrate their complementary performances. For that, a low-level fusion approach was implemented, merging the 40 potentiometric signals recorded during the electrochemical analysis of the monovarietal EVOO hydroethanolic extracts, with the 14 sensory attributes evaluated by the trained panelists. A single matrix containing 44 samples by 54 variables was obtained.

The accuracy of the low-level fusion approach of the sensory attributes and E-tongue signals to discriminate the different single-cultivar EVOO obtained by applying each of the four linear multivariate techniques is presented in Table 2. In this study, the number of PCs established for PLS-DA and SPLS-DA techniques was set based on a search procedure within 42 possible PCs (number of PCs defined by PCA). Once again, the results confirmed that the LDA-SA technique established the most powerful predictive model (100 % of correct classification for LOO-CV), followed by SPLS-DA, although with a quite lower classification predictive capability (only 48 % of correct classified samples), showing nevertheless that the variable selection procedures are advantageous, allowing to obtain simpler models with better predictive capabilities. It should be remarked that the low-level data fusion was only an effective classification procedure when combined with LDA-SA technique. Indeed, simple human sensory evaluation coupled with SPLS-DA or LDA-SA, or E-tongue data coupled with LDA-SA showed similar or higher classification performances (ranging from 45 to 70 % of correct classifications as shown in Table 2).

Regarding the best linear multivariate technique, a LDA-SA model with two discriminant functions, explaining 100 % of the original data variability (99.91 and 0.03 %, respectively), was established based on signal profiles of 18 potentiometric E-tongue sensors (S1:1, S1:3, S1:6–S1:8, S1:12, S1:13, S1:19, S1:20, S2:9, S2:10, S2:12–S2:15,

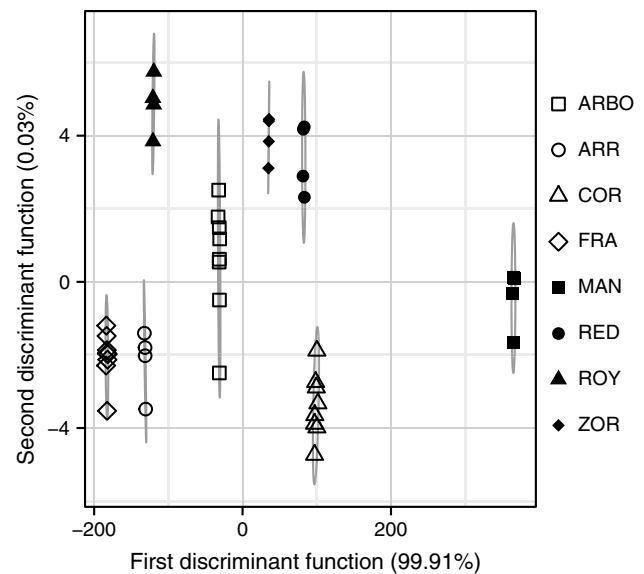


Fig. 5 Spanish monovarietal EVOO classification performance for the E-tongue data using a low-level data fusion approach: first and second LDA functions based on the best subset E-tongue (18 sensors) and sensory descriptors (nine olfactory, gustatory–retronasal, and final olfactory–gustatory sensations), selected using the SA meta-heuristic algorithm (ARB: cv. Arbosana; ARR: cv. Arroniz; COR: cv. Cornicabra; FRA: cv. Frantoio; MAN: cv. Manzanilla; RED: cv. Redondilla; ROY: cv. Royuela; and ZOR: cv. Zorzal)

S2:17, S2:18, and S2:20) and nine sensory attributes (olfactory sensations: other fruits and other positive sensations; gustatory–retronasal sensations: olive fruitiness, bitter, pungent, green, other positive sensations, and harmony; final olfactory–gustatory sensations: complexity), selected by the SA algorithm. The SA variable selection algorithm enabled the inclusion of only three repeated sensors, with the same membrane composition, in the discrimination model, which could improve the global model of the multivariate statistical performance [48]. Although the second discriminant function LDA-SA obtained model was not significant in the classification of groups of different varieties of olive oil, it was used to establish a two-dimensional discriminant graphic in order to better visualize the groups separation.

The proposed model allowed a 100 % correct classification of the original data (Fig. 5), as well as 100 % predictive correct classification for the LOO-CV procedure, as previously mentioned. This result shows that the E-tongue device fused with sensory attributes could be used to correctly classify monovarietal EVOO according to correct olive cultivar and so could be used as practical tool to guarantee the correct labeling of single-cultivar EVOO for cvs. Arbosana, Arroniz, Cornicabra, Frantoio, Manzanilla, Redondilla, or Royuela.

So, this work showed the potential use of electrochemical sensors to evaluate olive oil quality. Indeed, electronic

noses (E-noses) and electronic tongues (E-tongues), individually or combined, have been proposed in the last decade for olive oil characterization using different multivariate statistical techniques. For instance, an E-nose has been successfully used for the discrimination of different Mediterranean single-cultivar EVOO [49]. More recently, our research team has shown that an E-tongue could be applied to successfully discriminate Portuguese (100 % of correct classification) or Spanish (97.5–100 % of correct classifications) monovarietal EVOO according to the olive cultivar, depending on the number of olive cultivars evaluated (3 or 11) [3, 20].

Regarding monovarietal olive oil classification according to olive cultivar, other approaches have been described in the literature, reporting in some cases slightly less accurate correct classification performances (ranging from 70 to 100 %) depending on the analytical technique used (e.g., NMR, NIR, and MIR spectroscopy; PTR-MS; HS-SPME-GC/MS; SNP-based PCR–RFLP capillary electrophoresis; SNP-based CAPS assays), the number of olive cultivars under study (3, 5, or 10 cultivars) and chemometric tool used (e.g., LDA PLS-DA and canonical discriminant analysis) [13, 14, 22, 50–52].

Conclusions

Previously it was proven that a potentiometric E-tongue could be used to classify monovarietal EVOO. However, the successful discrimination performance was obtained when few single-cultivar EVOOs were evaluated. In this work, it was demonstrated that the proposed approach could be extended to a greater number single-cultivar olive oils when a low-level data fusion approach was adopted, merging E-tongue and human sensory analysis. The proposed strategy enables a full predictive LOO-CV discrimination of olive oil samples from eight monovarietal EVOOs (100 % correct classifications), produced in two different years. Moreover, for the first time, it was shown that the E-tongue could mimic human gustatory–retronasal sensations of EVOO. The successful classification was achieved using a LDA-SA model, based on the most informative potentiometric signals and sensory descriptors, chosen using the SA variable selection algorithm. This methodology proved to be more powerful than the two most common linear supervised multivariate techniques, LDA and PLS-DA, and even SPLS-DA, a PLS improved methodology that allows feature variable selection. So, it could be concluded that, from a practical point of view, E-tongue and sensory analysis performed by trained panelists are complementary techniques, contributing to improve the classification rate of monovarietal EVOO according to olive cultivar. Also, it is shown that the traditional use of olive oil panel sensory

analysis could be extended to the classification of monovarietal EVOO according to olive cultivar if a complementary approach was implemented by fusing human (sensory panel) and artificial (electronic tongue) sensory assessment. So, the use of this type of tool may indeed allow guaranteeing single-cultivar EVOO correct labeling, minimizing the risk of frauds of this high-value product.

Acknowledgments This work was co-financed by FCT/MEC and FEDER under Program PT2020 (Project UID/EQU/50020/2013); by Fundação para a Ciência e Tecnologia under the strategic funding of UID/BIO/04469/2013 unit; and by Project POCTEP through Project RED/AGROTEC—Experimentation network and transfer for development of agricultural and agro industrial sectors between Spain and Portugal.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Compliance with Ethics Requirements This article does not contain any studies with human or animal subjects.

References

1. Sinelli N, Cerretani L, Di Egidio V, Bendini A, Casiraghi E (2010) Application of near (NIR) infrared and mid (MIR) infrared spectroscopy as a rapid tool to classify extra virgin olive oil on the basis of fruity attribute intensity. *Food Res Int* 43:369–375
2. Garcia R, Martins N, Cabrita MJ (2013) Putative markers of adulteration of extra virgin olive oil with refined olive oil: prospects and limitations. *Food Res Int* 54:2039–2044
3. Dias LG, Fernandes A, Veloso ACA, Machado AASC, Pereira JA, Peres AM (2014) Single-cultivar extra virgin olive oil classification using a potentiometric electronic tongue. *Food Chem* 160:321–329
4. Lerma-García MJ, Simó-Alfonso EF, Bendini A, Cerretani L (2009) Metal oxide semiconductor sensors for monitoring of oxidative status evolution and sensory analysis of virgin olive oils with different phenolic content. *Food Chem* 117:608–614
5. Rotondi A, Beghè D, Fabbri A, Ganino T (2011) Olive oil traceability by means of chemical and sensory analyses: a comparison with SSR biomolecular profiles. *Food Chem* 129:1825–1831
6. Lauri I, Pagano B, Malmendal A, Sacchi R, Novellino E, Randazzo A (2013) Application of “magnetic tongue” to the sensory evaluation of extra virgin olive oil. *Food Chem* 140:692–699
7. Carrasco-Pancorbo A, Gómez-Caravaca AM, Cerretani L, Bendini A, Segura-Carretero A, Fernández-Gutiérrez A (2006) Rapid quantification of the phenolic fraction of Spanish virgin olive oils by capillary electrophoresis with UV detection. *J Agric Food Chem* 54:7984–7991
8. Matos LC, Cunha SC, Amaral JS, Pereira JA, Andrade PB, Seabra RM, Oliveira BPP (2007) Chemometric characterization of three varietal olive oils (cvs. Cobrançosa, Madural and Verdeal Transmontana) extracted from olives with different maturation indices. *Food Chem* 102:406–414
9. Bakhouché A, Lozano-Sánchez J, Beltrán-Debón R, Joven J, Segura-Carretero A, Fernández-Gutiérrez A (2013) Phenolic characterization and geographical classification of commercial Arbequina extra-virgin olive oils produced in southern Catalonia. *Food Res Int* 50:401–408

10. Karabagias I, Michos Ch, Badeka A, Kontakos S, Stratis I, Kontominas MG (2013) Classification of Western Greek virgin olive oils according to geographical origin based on chromatographic, spectroscopic, conventional and chemometric analyses. *Food Res Int* 54:1950–1958
11. Longobardi F, Ventrella A, Napoli C, Humpfer E, Schütz B, Schäfer H, Kontominas MG, Sacco A (2012) Classification of olive oils according to geographical origin by using ^1H NMR fingerprinting combined with multivariate analysis. *Food Chem* 130:177–183
12. Romero C, Brenes M (2012) Analysis of total contents of hydroxytyrosol and tyrosol in olive oils. *J Agric Food Chem* 60:9017–9022
13. Ruiz-Samblás C, Tres A, Koot A, van Ruth SM, González-Casado A, Cuadros-Rodríguez L (2012) Proton transfer reaction-mass spectrometry volatile organic compound fingerprint for monovarietal extra virgin olive oil identification. *Food Chem* 134:589–596
14. Bazakos C, Dulger AO, Uncu AT, Spaniolas S, Spano T, Kalaitzis P (2012) A SNP-based PCR–RFLP capillary electrophoresis analysis for the identification of the varietal origin of olive oils. *Food Chem* 134:2411–2418
15. Dais P, Hatzakis E (2013) Quality assessment and authentication of virgin olive oil by NMR spectroscopy: a critical review. *Anal Chim Acta* 765:1–27
16. Nunes CA (2013) Vibrational spectroscopy and chemometrics to assess authenticity, adulteration and intrinsic quality parameters of edible oils and fats. *Food Res Int* 60:255–261
17. Apetrei IM, Apetrei C (2013) Voltammetric e-tongue for the quantification of total polyphenol content in olive oils. *Food Res Int* 54:2075–2082
18. Escuderos ME, Sánchez S, Jiménez A (2011) Quartz Crystal Microbalance (QCM) sensor arrays selection for olive oil sensory evaluation. *Food Chem* 124:857–862
19. Haddi Z, Alami H, El Bari N, Tounsi M, Barhoumi H, Maaref A, Jaffrezic-Renault N, Bouchikhi B (2013) Electronic nose and tongue combination for improved classification of Moroccan virgin olive oil profiles. *Food Res Int* 54:1488–1498
20. Peres AM, Veloso ACA, Pereira JA, Dias LG (2014) Electrochemical multi-sensors device coupled with heuristic or meta-heuristic selection algorithms for single-cultivar olive oil classification. *Proced Eng* 87:192–195
21. Casale M, Casolino C, Oliveri P, Forina M (2010) The potential of coupling information using three analytical techniques for identifying the geographical origin of Liguria extra virgin olive oil. *Food Chem* 118:163–170
22. Casale M, Sinelli N, Oliveri P, Di Egidio V, Lanteri S (2010) Chemometrical strategies for feature selection and data compression applied to NIR and MIR spectra of extra virgin olive oils for cultivar identification. *Talanta* 80:1832–1837
23. Casale M, Oliveri P, Casolino C, Sinelli N, Zunin P, Armanino C, Forina M, Lanteri S (2012) Characterisation of PDO olive oil Chianti Classico by non-selective (UV–visible, NIR and MIR spectroscopy) and selective (fatty acid composition) analytical techniques. *Anal Chim Acta* 712:56–63
24. Pizarro C, Rodríguez-Tecedor S, Pérez-del-Notario N, Esteban-Díez I, González-Sáiz JM (2013) Classification of Spanish extra virgin olive oils by data fusion of visible spectroscopic fingerprints and chemical descriptors. *Food Chem* 138:915–922
25. 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. *Sens Actuators B* 177:989–996
26. Vera L, Aceña L, Guasch J, Boqué R, Mestres M, Busto O (2011) Characterization and classification of the aroma of beer samples by means of an MS e-nose and chemometric tools. *Anal Bioanal Chem* 399:2073–2081
27. Bruwer M-J, MacGregor JF, Bourg WM Jr (2007) Fusion of sensory and mechanical testing data to define measurements of snack food texture. *Food Qual Prefer* 18:890–900
28. Haddi Z, Mabrouk S, Bougrini M, Tahri K, Sghaier K, Barhoumi H, El Bari N, Maaref A, Jaffrezic-Renault N, Bouchikhi B (2014) E-nose and e-tongue combination for improved recognition of fruit juice samples. *Food Chem* 150:246–253
29. Banerjee R, Modak A, Mondal S, Tudu B, Bandyopadhyay R, Bhattacharyya N (2013) Fusion of electronic nose and tongue response using fuzzy based approach for black tea classification. *Procedia Technol* 10:615–622
30. Apetrei C, Apetrei IM, Villanueva S, de Saja JA, Gutierrez-Rosales F, Rodriguez-Mendez ML (2010) Combination of an e-nose, an e-tongue and an e-eye for the characterisation of olive oils with different degree of bitterness. *Anal Chim Acta* 663:91–97
31. International Olive Council (2013) Sensory analysis of olive oil—method for the organoleptic assessment of virgin olive oil. COI/T.20/Doc. No. 15/Rev. 6 November 2013. <http://www.internationaloliveoil.org/>
32. International Olive Council (2014) IOC Mario Solinas quality award—rules of the international competition for extra virgin olive oils. T.30/Doc. No. 17 June 2014. <http://www.internationaloliveoil.org/>
33. Dias LG, Peres AM, Veloso ACA, Reis FS, Vilas Boas M, Machado AASC (2009) An electronic tongue taste evaluation: identification goat milk adulterations with bovine milk. *Sens Actuators B* 136:209–217
34. Sousa MEBC, Dias LG, Veloso ACA, Estevinho L, Peres AM, Machado AASC (2014) Practical procedure for discriminating monofloral honeys with a broad pollen profile variability using an electronic tongue. *Talanta* 128:284–292
35. Kobayashi Y, Habara M, Ikezazki H, Chen R, Naito Y, Toko K (2010) Advanced taste sensors based on artificial lipids with global selectivity to basic taste qualities and high correlation to sensory scores. *Sensors* 10:3411–3443
36. Liu Y, Brown SD (2004) Wavelet multiscale regression from the perspective of data fusion: new conceptual approaches. *Anal Bioanal Chem* 380:445–452
37. Kuhn M, Johnson K (2013) Applied predictive modeling, features. Springer, 17 May 2013
38. Chun H, Keleş S (2010) Sparse partial least squares regression for simultaneous dimension reduction and variable selection. *J R Stat Soc B* 72(Part 1):3–25
39. Cadima J, Cerdeira JO, Minhoto M (2004) Computational aspects of algorithms for variable selection in the context of principal components. *Comput Stat Data Anal* 47:225–236
40. Dias LG, Sequeira C, Veloso ACA, Sousa MEBC, Peres AM (2014) Evaluation of healthy and sensory indexes of sweetened beverages using an electronic tongue. *Anal Chim Acta* 848:32–42
41. Söderström C, Rudnitskaya A, Legin A, Krantz-Rülcker C (2005) Differentiation of four *Aspergillus* species and one *Zygosaccharomyces* with two electronic tongues based on different measurement techniques. *J Biotechnol* 119:300–308
42. Rudnitskaya A, Kirsanov D, Legin A, Beullens K, Lammertyn J, Nicolaï BM, Irudayaraj J (2006) Analysis of apples varieties—comparison of electronic tongue with different analytical techniques. *Sens Actuators B* 116:23–28
43. Kuhn M (Contributions from Wing J, Weston S, Williams A, Keefer C, Engelhardt A, Cooper T, Mayer Z, and the R Core Team) (2014) caret: classification and regression training. R package version 6.0-24. <http://CRAN.R-project.org/package=caret>
44. Cerdeira JO, Silva PD, Cadima J, Minhoto M (2012) subselect: selecting variable subsets. R package version 0.12-2. <http://CRAN.R-project.org/package=subselect>

45. Mevik B-H, Wehrens R, Liland KH (2011) pls: partial least squares and principal component regression. R package version 2.3-0. <http://CRAN.R-project.org/package=pls>
46. Chung D, Chun H, Keles S (2012) spls: sparse partial least squares (SPLS) regression and classification. R package version 2.1-2. <http://CRAN.R-project.org/package=spls>
47. Venables WN, Ripley BD (2002) Modern applied statistics with S (Statistics and Computing), 4th edn. Springer, New York. ISBN 978-0-387-21706-2
48. Correia DPA, Magalhães JMCS, Machado AASC (2005) Array of potentiometric sensors for simultaneous analysis of urea and potassium. *Talanta* 67:773–782
49. Cimato A, Dello Monaco D, Distanto C, Epifani M, Siciliano P, Taurino AM, Zuppa M, Sani G (2006) Analysis of single-cultivar extra virgin olive oils by means of electronic nose and HS-SPME/GC/MS methods. *Sens Actuators B* 114:674–680
50. Agiomyrgianaki A, Petrakis PV, Dais P (2012) Influence of harvest year, cultivar and geographic origin on Greek extra virgin olive oils composition: a study by NMR spectroscopy and biometric analysis. *Food Chem* 135:2561–2568
51. Pouliarekou E, Badeka A, Tasioula-Margari M, Kontakos S, Longobardi F, Kontominas MG (2011) Characterization and classification of Western Greek olive oils according to cultivar and geographical origin based on volatile compounds. *J Chromatogr A* 1218:7534–7542
52. Uncu AT, Frary A, Doganlar S (2015) Cultivar origin and admixture detection in Turkish olive oils by SNP-based CAPS assays. *J Agric Food Chem* 63:2284–2295