https://doi.org/10.1016/j.foodchem.2014.08.057

Electronic Nose and Isotope Ratio Mass Spectrometry in combination with chemometrics for the characterization of the geographical origin of Italian sweet cherries

F. Longobardi[a,]*, G. Casiello[a], A. Ventrella[a], V. Mazzilli[a], A. Nardelli[a], D. Sacco[a], L. Catucci[a,b], A. Agostiano[a,b]

[a] Dipartimento di Chimica, Università di Bari “Aldo Moro”, Via Orabona 4, 70126 Bari, Italy

[b] Consiglio Nazionale delle Ricerche, Istituto per i Processi Chimico-Fisici (IPCF-CNR), sez. di Bari, Via Orabona 4, 70126 Bari, Italy

* Corresponding author. Tel.: +39-080-5442042; fax: +39-080-5443607.
E-mail address: francesco.longobardi@uniba.it

Abstract

In this work sweet cherries coming from two Italian regions, Apulia and Emilia Romagna, were analyzed using Electronic Nose (EN) and Isotope Ratio Mass Spectrometry (IRMS) aiming at distinguishing them according to their geographic origin. The data were elaborated by statistical techniques, examining the EN and IRMS datasets both separately and in combination. Preliminary exploratory overviews were performed and then Linear Discriminant Analyses (LDA) were implemented for the classification aims. Regarding EN, different approaches for variable selection were tested highlighting the most suitable strategies. The LDA classification results were expressed in terms of recognition and prediction abilities and it was found that both EN and IRMS gave interesting classification performances, even if IRMS showed a better cross-validated prediction ability (91.0%); the EN-IRMS combination lead to slightly better results (92.3%). In order to validate the final results, the models were tested employing an external set of samples with very satisfying output.

Keywords: Electronic Nose; Isotope Ratio Mass Spectrometry; Sweet cherry; Geographic origin; Chemometrics

1. Introduction

The fruits of the sweet cherry trees belonging to *Prunus avium L.* species are used for fresh consumption, and for the production of marmalades, juices, jams, jelly fruits, and also alcoholic beverages. Sweet cherries are widely appreciated for their taste and their nutritional qualities, that are a consequence of their chemical profiles. In particular, they show a higher simple sugars content if compared to sour cherries, with glucose and fructose as main responsible for their sweetness; they present a considerable amount of hydrosoluble (C, B) and liposoluble (A, E and K) vitamins, carotenoids (such as β-carotene, lutein and zeaxantine), minerals (calcium, magnesium, phosphorous and potassium), and volatile compounds, such as esters, alcohols, aldehydes, ketones, and terpenoid compounds (Ferretti, Bacchetti, Belleggia, & Neri, 2010; Li, Kang, Hu, Li, & Shen, 2008; Pérez-Sánchez, Gómez-Sánchez, & Morales-Corts, 2010). In addition, natural healthy antioxidant substances like antocyanins and polyphenols are present in significant amounts (Liu et al., 2011).

The main producers of sweet cherries in the world are represented by Turkey, United States, Iran, Italy, France, Spain and Russia (Doymaz & Ismail, 2011; Pérez-Sánchez et al., 2010), and regarding Italian production, it takes place mainly in the regions of Apulia, Campania, Veneto and Emilia Romagna. The varieties principally cultivated in Italy are Bigarreau, Black, Anella, Giorgia and Ferrovia, and some of them are cultivated mainly in specific localities, so showing peculiar traits, that confer them a remarkable quality value, leading local producers to act with the purpose to obtain European marks, such as “protected designation of origin” (PDO), “protected geographical indication” (PGI) and “traditional specialty guaranteed” (TSG), that, in general, help to protect and promote the brand names of Europe’s traditional agricultural produce and foods.

Therefore, it is clear there is economic basis to develop analytical methods able to certify the declared geographical origin of food products in order to protect consumers and honest producers from frauds and unfair competition, respectively; consequently, during the past years, several techniques have been proposed for such purpose on various food matrices with various results (Cajka, Riddellova, Klimankova, Cerna, Pudil, & Hajslova, 2010; Camin, Perini, Bontempo, & Giongo, 2009; Longobardi et al., 2012a; Longobardi et al., 2012b; Longobardi et al., 2013; Torri, Sinelli, & Limbo, 2010).

Among the innovative analytical techniques, the Electronic Nose (EN) has been proven highly useful in studies on food matrices (Benedetti, Buratti, Spinardi, Mannino, & Mignani, 2008; Pacioni, Cerretani, Procida, & Cichelli, 2014; Russo, di Sanzo, Cefaly, Carabetta, Serra, & Fuda, 2013). Briefly, ENs are devices that mimic the sense of smell of mammalians, on the basis of different technologies, to detect volatile analytes in complex matrices (Peris & Escuder-Gilabert, 2009).

The application of the EN in the discrimination of the geographic origin or of the variety of food matrices is well documented, indicating a great potential of this technique in these fields (Cajka et al., 2010; Cynkar, Dambergs, Smith, & Cozzolino, 2010; de las Nieves López de Lerma, Bellincontro, García-Martínez, Mencarelli, & Moreno, 2013).

Another innovative technique for the analysis of food matrices is the Isotopic Ratio Mass Spectrometry (IRMS). Such technique investigates the ratios of the stable isotopes present in a sample, i.e. those isotopes that do not decay through radioactive processes over time.

In the field of authenticity of food, and, in particular, the discrimination of the geographical origin of food matrices, the IRMS has a great potential as demonstrated by the numerous papers (Kelly, Heaton, & Hoogewerff, 2005; Perini & Camin, 2013; Longobardi, Casiello, Sacco, Tedone, & Sacco, 2011; Rossmann, Reniero, Moussa, Schmidt, Versini, & Merle, 1999; Rummel, Hoelzl, Horn, Rossmann, & Schlicht, 2010; Sacco et al., 2009).

However, concerning the use of EN and IRMS on cherry samples, at our knowledge, the examples in the literature are few and mainly deal with the evaluation of fruit ripeness (Benedetti, Spinardi, Mignani, & Buratti, 2010). Therefore, in this work, sweet cherry samples coming from two different Italian regions devoted to the cherry production, Apulia and Emilia Romagna, were analyzed by means of the above mentioned innovative instrumental techniques, i.e. EN and IRMS, with the purpose to discriminate the samples on the basis of their geographic origin.

2. Materials and methods

2.1 Sample collection

A total of 112 cherry samples from two different Italian regions, i.e. 56 Emilian samples and 56 Apulian samples, belonging to three different varieties, i.e. “Bigarreau”, “Giorgia” and “Ferrovia” were collected. Cherries were harvested during the 2010 crop season, between the 3rd decade of May and the 3rd decade of June. Apulian samples came from the areal close to the south-east of Bari whereas Emilian ones from the area between the provinces of Modena and Bologna. The fruits were harvested in a state of consumption maturity, cooled in a few hours, and transported to laboratory assuring the maintenance of the cold chain. Subsequently, the cherries were washed with tap water, carefully wiped with laboratory paper, and, for IRMS analyses, freeze-dried (see below) while for EN analyses, samples were introduced in polyethylene bags, kept frozen and stored at -70°C.

2.2 EN apparatus and analyses

For EN measurements, cherry samples were thawed at +4° C for 2 h, cut in small pieces, and then an aliquot of 2 g was placed into a 10 mL vial hermetically sealed with a cap having a Teflon septum and placed in a thermostatic bath at 40°C for 90 min in order to establish equilibrium between headspace and sample. The cherry headspace was pumped in an Electronic Nose System PEN3 (Airsense Analytics, Schwerin, Germany), equipped with an array of 10 metal oxide semiconductor (MOS) sensors, at a flow rate of 400 mL min^{-1}, and the sensor responses were sampled every 1 s for 120 s. After each sample analysis, the system was purged for 200 s with filtered air prior to the next sample injection to allow re-establishment of the instrument baseline. Each sample was analyzed twice and the average of the results was used for subsequent statistical analysis.

2.3 IRMS apparatus and analyses

For stable isotope ratio analysis, cherry fruits were cut in half, pitted, frozen at -80°C and then freeze-dried for 48h using a Heto Lyolab 3000 freeze dryer (Heto-Holten A/S, Allerød, Denmark). Freeze-dried cherries were powered using a commercial blender and stored in sealed containers under vacuum until analysis.

For $^{13}C/^{12}C$ analysis about 0.3 mg of freeze-dried sample were weighed into tin capsules and directly analyzed, whereas for $^{18}O/^{16}O$ and $^{2}H/^{1}H$ analysis about 1.5 mg of sample were firstly weighed into silver capsules and then stored in a desiccator above P_2O_5 for at least 72 h before analysis.

The analyses were performed using an isotopic ratio mass spectrometer (IRMS, Finnigan Delta V Advantage, Thermo Fisher Scientific, Bremen, Germany) coupled with an elemental analyser (EA, FlashEA 1112 HT, Thermo Fisher Scientific, Bremen, Germany). The EA was equipped with a combustion reactor (held at 1020°C) for $^{13}C/^{12}C$ determination and a pyrolysis reactor (a high-temperature conversion elemental analyser, held at 1450°C) for $^{18}O/^{16}O$ and $^{2}H/^{1}H$ ratios. Samples were introduced into the pyrolysis/combustion column via the autosampler (MAS 200R, Thermo Fisher Scientific, Bremen, Germany), equipped with a suitable cover, where dry conditions were ensured by flushing nitrogen continuously over the samples.

The EA was interfaced with the IRMS through a dilutor (Finnigan Conflo III, Thermo Fisher Scientific, Bremen, Germany) dosing the samples and reference gases. To separate the gases produced (CO_2 during the combustion and CO or H_2 during the pyrolysis) the elemental analyser was supplied with two Porapak QS gas chromatography columns.

The isotopic values were expressed using the conventional δ notation in parts per thousand (‰) vs. V-SMOW (Vienna-Standard Mean Ocean Water) for oxygen and hydrogen, PDB (Pee Dee Belemnite) for carbon, according to the following formula:

$$[(R_{\text{sample}} - R_{\text{standard}}) / R_{\text{standard}}] \times 1000$$

where R represents the ratio between the heavy and light isotopes, in the sample and standard, respectively. Each sample was analysed twice and the isotopic value was reported as mean of the two determinations. The values were calculated against reference gases (i.e. CO_2, CO and H_2) previously calibrated against International Standards supplied by IAEA (International Atomic Energy Agency, Vienna, Austria): USGS 40 for $^{13}C/^{12}C$, IAEA-CH-7 for $^{2}H/^{1}H$, and IAEA-601 for $^{18}O/^{16}O$. To check the accuracy, working in-house standards were analysed in each run. In particular, a commercial casein and benzoic acid (Carlo Erba Reagents, Milan, Italy) was used for $^{13}C/^{12}C$ and $^{18}O/^{16}O$, respectively. For $^{2}H/^{1}H$, supposing the possible exchange of hydrogen with water and/or ambient air moisture, the values were corrected against an Inter-laboratory Comparison Material (ICM) – casein reference material according to the "comparative equilibration technique" (Wassenaar & Hobson, 2003). The precision of measurement, expressed as one standard deviation and obtained measuring a cherry sample 10 times, was ± 3‰ for δD, ± 0.3‰ and ± 0.2‰ for $\delta^{18}O$ and $\delta^{13}C$, respectively.

2.4 Chemometrics

For the statistical analyses, dataset was subdivided into two subsets by exploiting the Kennard and Stone Duplex design (Casale et al., 2012): a modeling set and an external set, containing 78 (39 Emilia Romagna and 39 Apulian samples, i.e. about 70% of total samples), and 34 (17 Emilia Romagna and 17 Apulian samples, i.e. about 30% of total samples) cherry samples, respectively.

Modeling subset were processed by univariate statistics (*t*-test) and the following multivariate techniques: unsupervised (PCA), for an exploratory overview, and supervised discriminant techniques (LDA) in order to build statistical models able to discriminate cherries according to their geographic origin. In the model-building step, each supervised pattern recognition model was evaluated in terms of non-error classification rate, both on the whole modeling set (recognition ability), and on the test set obtained by *k*-fold cross-validation (CV prediction ability) with a k value equal to 5. Finally, the validation and comparison of the models were executed calculating the prediction abilities obtained on the external set. Statistic and chemometric data analyses were performed by using Statistica version 8.0 (StatSoft Italia srl, Padova, Italy) and V-Parvus release 2010 (http://www.parvus.unige.it, Genova, Italy).

3. Results and discussion

3.1 EN results

The whole EN data matrix was constituted by 78 objects (cherry samples) and 1200 variables (120 points for each of the10 sensors). Using all the matrix elements, even if assures bringing all the available information, could require further data treatments to avoid statistical and/or computational problems. Based on these considerations, researchers chose different strategies to sample the sensor signal points in their works, i.e. using one point per sensor or a higher number of points, at selected sampling times (Hai & Wang 2006; Hernández-Gómez, Wang, Hu, & García-Pereira, 2008). In this work, three different ways were tested and then the obtained results were compared and discussed. In particular, the variable employment strategies (VESs) tested herein were:

VES1. For each curve only a single point (sensor response at 110 s) was considered, i.e. almost at the end of the sensor curve, where all the signals could be considered stationary.

VES2. For each curve, variables were selected considering one point every 10 s.

VES3. The whole curve was employed, i.e. no variable choice was applied, so that 120 points were used per each sensor response.

Therefore, the three VESs were at an increasing level of complexity, consequently producing data matrices strongly different in dimensions. Each VES has its own advantages and drawbacks: VES1 leads to a data matrix that can be easily used in the following statistical treatments even if the great part of the curve information is lost, and it is not possible to know a priori if such lost information would be important for the geographic classification aims; VES3, at opposite, takes into account all the information contained in the sensor curve, nothing being left out, and it does not force the operator to decide a particular variable selection strategy, but it generates a huge data matrix containing highly correlated variables, and it requires further treatments for the subsequent applications used herein. Finally, VES2 represents a compromise between VES1 and VES3, since it leads to a data matrix that brings more information than VES1, even not containing the huge amount of highly correlated variables of VES3.

3.1.1 Exploratory overview by PCA

In order to get a general overview of the data distributions, the data matrix obtained according to VES1, 2 and 3 were subjected to PCA and the minimum number of PCs explaining more than 95% of cumulative variance, were five, six and seven for VES1, VES2, and VES3, respectively.

By plotting the scores of the samples in the sub-space PC1 vs. PC2, no grouping of objects was observed on the basis of the geographical origin, for all the VES methods (graphs not shown).

Comparing the PC Fisher weights (FW), i.e. a measure of the between-class variance/within-class variance ratio (Harper et al., 1977), it was evidenced that the PCs having better ability to discriminate origins (higher FWs) were not the ones explaining most of the observed variance.

However, it has to be noticed that all FW values were considerably lower than 1, evidencing that no single PC was sufficiently suitable for classification aims, as showed also in Fig. 1a, b and c, where 2D graphs of the PCs, with the highest FWs, showed only partial grouping of objects on the basis of the geographic origin for each VES. This information seemed to indicate that a multivariate approach was advisable. Finally, the VES2 and 3 were found to lead to PC variables with higher FWs than for VES1, indicating that it could be better to use more information from the sensor curves than only one point.

3.1.2 Linear Discriminant Analysis

LDA was applied on all the three VES matrices listed above. For applying discriminant analysis, it is necessary that the number of variables is not too large with respect to the number of objects, mainly due to the overloading of the computations required to calculate the Mahalanobis distances, and also due to the possibility to incur overfitting problems. Therefore, when the variables/objects ratio is too high, a variable reduction must be performed. With regard to the overfit risks, a general rule states that the number of variables should not exceed $(n-g)/3$, where n is the number of objects and g is the number of categories (Berrueta, Alonso-Salces, & Héberger, 2007; Defernez & Kemsley, 1997).

Moreover, it is important to highlight that if variable selection is carried out by means of supervised methods, overoptimistic results could be still obtained, and consequently model results should be accepted only after having performed a cross- and/or external validation of the model.

Considering the data matrices under study, the above mentioned requisite was directly satisfied by VES1 modeling set, whilst variable selection routines were found to be necessary for VES2 and 3 before performing LDA. Two different strategies for selecting the variables were tested, taking into account not to exceed a critical number of 25, i.e. $(n-g)/3$. The first strategy was to choose a set of 25 variables by applying the Parvus SELECT feature (Casale, Casolino, Oliveri, & Forina, 2010):

this technique, is a variable selection that, based on a stepwise decorrelation of the variables, generates a set of decorrelated variables ordered by their classification FWs (according to a response variable, i.e. the geographic origin in this case). SELECT searches, at each step, for the variable with the largest classification weight, that is selected and decorrelated from the other ones. In the second strategy, a forward stepwise LDA was performed onto VES2 and 3; the forward stepwise statistics, with F-to-enter equal to 1.0 and F-to-remove equal to 0.5, selected respectively 22 and 4 variables to be used in the relevant final models. All the recognition and CV prediction abilities of the final obtained models are reported in Table 1 for comments.

As it can be easily observed, VES2 and 3 lead to better classification performances if compared to VES1 LDA results, both considering the recognition and the CV prediction abilities. This could be ascribable to the importance of using more information from the sensor curve rather than the only contained just in a single point, and this seems to be in accordance with the considerations coming from the previously commented PCA analyses.

Moreover, if VES2 and 3 LDA prediction results are compared, it can be noticed that no particular advantage seems to occur by considering all the sensor curve points; rather, considering the stepwise LDA strategy, a slight decrease of the classification performances was obtained going from VES2 to VES3 model; in other words, the results evidenced that LDA should be applied on a data matrix that brings sufficient amount of information of the EN sensor curves, without overloading the variable selection algorithms with an excessive amount of variables to compute.

By computing the Factor Structure Coefficients (FSCs) that express the pooled within-class (groups) correlations of the original EN variables with the discriminant function, it was possible to partly interpret the meaning of the discriminant function getting at the same time information about the most discriminating original EN variables.

In particular, in all the VES methods, a remarkable importance of the sensor 7 variables was highlighted (highest FSCs); more in detail, by considering VES2 and 3, the first part of the sensor 7 curve (comprised in the range 10-20 s) was found to be interestingly important for the discrimination of the geographical origin of samples, together with the final part of the same sensor curve. As indicated by the EN instrument manufacturer, the sensor 7, coded as W1W, is particularly sensitive to terpene molecules, that are important volatile flavor compounds in sweet cherries as reported in literature (Li et al., 2008; Girard & Kopp, 1998). This suggests that terpenes could be useful molecules for the discrimination of the cherries on the basis of their geographic origins, although, due to the lack of specificity of the EN response, it cannot be known with certainty if other classes of compounds have contributed to the good model performances obtained herein.

3.2 IRMS results

As showed in Fig. 1d, representing the cherry samples in the space defined by the three IRMS variables (i.e. $\delta^{13}C$, $\delta^{18}O$, and δD), a partial grouping of cherry samples coming from the two different geographical origins was observed. In order to quantitatively find out which of the three original IRMS variables were more discriminating on the basis of cherry geographic origins, a univariate t-test was carried out highlighting statistically significant differences ($p < 0.05$) only in the mean values of $\delta^{18}O$ and δD. In particular, as reported in Table 2, $\delta^{18}O$ and δD showed mean values that increased from north (Emilia) to south (Apulia), i.e. a $\delta^{18}O$ mean value of 33.2 ‰ for Emilia and of 35.4 ‰ for Apulia, and a δD mean value of -38.5 ‰ and of -30.7 ‰ for Emilia and Apulia, respectively. Therefore, all results reported above confirm that the $\delta^{18}O$ and δD are good parameters to differentiate geographic origins of foodstuffs. Indeed, stable carbon isotope ratios of plants, are primarily related to the photosynthetic pathway used by a plant even if $\delta^{13}C$ in foodstuffs exhibits some geographical dependence linked to water stress and humidity during cultivation although these differences are often very small in comparison to other isotopes (Hurley, West, & Ehleringer, 2010; Longobardi et al., 2011). On the country, hydrogen and oxygen stable isotopes of plant materials are strongly linked to the climatic conditions (relative humidity, temperature, amount of precipitation) and geographical characteristics (distance from the sea or other evaporation source, altitude, latitude) of the area where the plants grow (Bontempo, Camin, Larcher, Nicolini, Perini, & Rossmann, 2008; Hermann & Voerkelius, 2008; Iacumin, Bernini, & Boschetti, 2009). In particular, the $\delta^{18}O$ of the plant products reflects the isotopic composition of groundwater and average precipitation in the region (mainly related to geographic coordinates i.e. latitude, distance from the sea and altitude) and the extent of evapotranspiration (mainly influenced by humidity and temperature) (Rossmann et al., 1999). Similarly, the hydrogen present in plant material originates from the water taken up by the roots (Ziegler, Osmind, Stichler, & Trimborn, 1976) and the subsequent evapotranspiration process of water through the leaf stomata enriches the remaining water in deuterium.

Subsequently, to assess discrimination efficiency for cherry origin, LDA was performed by using all the isotopic ratio values obtaining recognition and CV prediction abilities of 94.9% and 94.1%, respectively (Table 1). These excellent performance slightly decreased when LDA model was built considering only $\delta^{18}O$ and δD, in particular the recognition and CV prediction abilities resulted to be 92.3 and 91.2%, respectively. Therefore, unlike what has been shown by univariate analysis, $\delta^{13}C$, used in multivariate combination with other isotopic indicators, results to be a useful parameter for tracing cultivation areas of cherry samples, demonstrating the powerful of a multiple stable isotopes composition analysis in geographic discrimination aims of food (Kelly et al., 2005; Zhao, Zhang, Chen, Chen, Yang, & Ye, 2014).

3.3 External validation of the classification models

Among the LDA models here obtained, the most promising ones in terms of classification performances were subjected to an external validation procedure to verify their real reliabilities (Table1).

In particular, regarding the EN technique, the VES2 and VES3 recognition and CV results were found to be remarkable, while in the case of IRMS the attention was focused on the approach considering all the isotopic variables. As results, with regard to EN, it was found that the VES2 strategy produced the same prediction abilities (82.4%) independently on the variable selection method (SELECT or forward stepwise statistics) adopted, indicating a considerable stability of both models. At contrary, the external validation of the VES3 based LDA models evidenced that the prediction ability was remarkably dependent on the variable selection routine, since while the forward stepwise LDA gave a prediction ability comparable to the ones obtained in VES2 approach, the SELECT based LDA lead to a considerably worse prediction performance (76.5%).
These results showed that, the higher the number of variables and the correlation among them (as in VES3 with respect to VES2 matrix), the more the LDA depends on the variable selection technique adopted. Moreover, clearly these findings highlight how can be important to apply an external validation procedure to the classification models, especially when obtained after supervised variable selection routines on huge data matrices as the one obtained herein.
Regarding the IRMS, the particularly good classification and CV performances, were confirmed by the external prediction ability, equal to 94.1%, evidencing the reliability of the IRMS LDA model, that was found to classify incorrectly only one external sample per each class.
Finally, in order to find if the combination of the two instrumental techniques could produce even more interesting results, here the EN and IRMS variables were considered together, and among all the possibilities tested, the one that gave the best results was to consider the EN variables coming from VES2 design together with $\delta^{13}C$, $\delta^{18}O$, and δD variables coming from IRMS, and to apply a forward stepwise LDA. The forward stepwise statistics selected 9 variables: $\delta^{13}C$, $\delta^{18}O$, δD, S5p79, S7p9, S7p19, S7p99, S7p109, S9p69 that allowed obtaining recognition and CV prediction abilities equal to 96.2 and 92.3%, respectively. The obtained model was then subjected to the external validation procedure, showing a prediction ability of 94.1%. Therefore, considering this slight

improvement of the model performances, it can be asserted that the EN-IRMS synergistic approach is not a necessary step to obtain reliable and acceptable results, obtainable by the single techniques here tested.

4. Conclusions

In this paper, EN and/or IRMS data were used, in combination with LDA, to discriminate Italian sweet cherries coming from two different geographic origins.

Regarding EN analyses, the results demonstrated that the selection of the variables to be considered in the LDA building was decisive to obtain good and stable model performances; in particular, the best prediction abilities ranged from 85.9% to 89.7% and from 82.4% to 85.3% for an internal (CV) and an external validation, respectively. Better results were obtained with IRMS especially by using all the isotopic ratios gaining a CV prediction ability of 91.0% and an external prediction equal to 94.1%. No significant improvement was obtained combining isotopic and electronic nose data.

In conclusion, it can be asserted that both techniques represent valid tools for geographic discrimination of Italian cherries but some considerations should be taken into account. On one hand, the IRMS is more accurate and allows obtaining stable databases overtime although it requires more expensive equipment and skilled operators. On the other hand, the EN even giving less accurate prediction results depending on the poor sensor selectivity, and even requiring some approaches for compensating sensors drift, shows important advantages in terms of portability, price, and ease of use; therefore, it can be easily adopted for industrial routine controls.

Acknowledgements

This work has been carried out within “Intervento Reti di Laboratori Pubblici di Ricerca cofinanziato dall’Accordo di Programma Quadro in materia di Ricerca Scientifica – II Atto Integrativo – PO Puglia FESR 2007-2013, Asse I, Linea 1.2 – PO Puglia FSE 2007–2013 Asse IV” (“Apulian Food Fingerprint” project).

The authors gratefully thank the “Direzione Qualità di Coop Italia” (via del Lavoro 6/8, 40033, Casalecchio di Reno, Bologna, Italy), Ortofrutticoli La Pernice S.r.l. (via Casamassima, 70010 Turi, Bari, Italy), Giuliano S.r.l. (S.P 65 km 4.8, 70010, Turi, Bari, Italy), Valenzano Vittorio (via Conversano, Km. 2, 70018, Rutigliano, Bari, Italy), and Professor Francesco D’Aprile, for supplying the cherry samples.

Special thanks are for the retired Professor Antonio Sacco of the Department of Chemistry (University of Bari “Aldo Moro”) for his fundamental help in this research work.

Figure captions

Figure 1. PCA scatter plots for cherry samples data obtained by EN with VES1 (a), VES2 (b) and VES3 (c); for each panel the two PCs with the highest FWs are reported as axes. Three dimensional scatter plots (d) for cherry samples data obtained by IRMS, considering the isotopic variables $\delta^{13}C$, $\delta^{18}O$, and δD. Geographical origins: Emilia Romagna (□), Apulia (+).

References

Alonso-Salces, R. M., Héberger, K., Holland, M. V., Moreno-Rojas, J. M., Mariani, C., Bellan, G., Reniero, F., & Guillou, C. (2010). Multivariate analysis of NMR fingerprint of the unsaponifiable fraction of virgin olive oils for authentication purposes. *Food Chemistry, 118,* 956–965.

Benedetti, S., Buratti, S., Spinardi, A., Mannino, S., & Mignani, I. (2008). Electronic nose as a non-destructive tool to characterise peach cultivars and to monitor their ripening stage during shelf-life. *Postharvest Biology and Technology, 47,* 181–188.

Benedetti, S., Spinardi, A., Mignani, I., & Buratti, S. (2010). Non-destructive evaluation of sweet cherry (prunus avium L.) ripeness using an electronic nose. *Italian Journal of Food Science, 22,* 298-304.

Berrueta, L. A., Alonso-Salces, R. M., & Héberger, K. (2007). Supervised pattern recognition in food analysis. *Journal of Chromatography A, 1158,* 196–214.

Bontempo, L., Camin, F., Larcher, R., Nicolini, G., Perini, M., & Rossmann, A. (2008). Discrimination of Tyrrhenian and Adriatic Italian olive oils using H, O, and C stable isotope ratios. *Rapid Communications in Mass Spectrometry, 23,* 1043–1048.

Cajka, T., Riddellova, K., Klimankova, E., Cerna, M., Pudil, F., & Hajslova, J. (2010). Traceability of olive oil based on volatiles pattern and multivariate analysis. *Food Chemistry, 121,* 282–289.

Camin, F., Perini, M., Bontempo, L., & Giongo, L. (2009). Multi-element (H, C, N, O) stable isotope characterization of blueberries. *Acta Horticulturae, 810,* 697-703.

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 Chemistry, 118,* 163–170.

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. *Analytica Chimica Acta, 712,* 56– 63.

Cynkar, W., Dambergs, R., Smith, P., & Cozzolino, D. (2010). Classification of Tempranillo wines according to geographic origin: Combination of mass spectrometry based electronic nose and chemometrics. *Analytica Chimica Acta, 660,* 227–231.

Defernez, M., & Kemsley, E. K. (1997). The use and misuse of chemometrics for treating classification problems. *Trends in Anaytical Chemistry, 16,* 216–221.

de las Nieves López de Lerma, M., Bellincontro, A., García-Martínez, T., Mencarelli, F., & Moreno, J. J. (2013). Feasibility of an electronic nose to differentiate commercial Spanish wines elaborated from the same grape variety. *Food Research International, 51,* 790–796.

Doymaz, I., & Ismail, O. (2011). Drying characteristics of sweet cherry. *Food and Bioproducts Processing, 89,* 31–38.

Ferretti, G., Bacchetti, T., Belleggia, A., & Neri, D. (2010). Cherry Antioxidants: From Farm to Table. *Molecules, 15*, 6993-7005.

Girard, B., & Kopp, T. G. (1998). Physicochemical Characteristics of Selected Sweet Cherry Cultivars. *J. Agric. Food Chemistry, 46,* 471−476.

Z. Hai, J. Wang (2006). Electronic nose and data analysis for detection of maize oil adulteration in sesame oil. *Sensors and Actuators B, 119,* 449–455.

Harper, A. M., Duewer, D. L., Kowalski, B. R., & Fashing J. L. (1977). ARTHUR and Experimental Data Analysis: The Heuristic Use of a Polyalgorithm. In: B. R. Kowalski (Ed.), *Chemometrics: Theory and Application* (pp. 14-52). ACS Symposium Series 52. Washington: American Chemical Society.

Hermann, A., & Voerkelius, S. (2008). Meteorological impact on oxygen isotope ratios of German wines. *American Journal of Enology and Viticulture, 59,* 194–199.

Hernández-Gómez, A., Wang, J., Hu, G., & García-Pereira, A. (2008). Monitoring storage shelf life of tomato using electronic nose technique. *Journal of Food Engineering, 85*, 625-631.

Hurley, J. M., West, J. B., & Ehleringer, J. R. (2010). Stable isotope models to predict geographic origin and cultivation conditions of marijuana. *Science & Justice, 50,* 86–93.

Iacumin, P., Bernini, L., & Boschetti, T. (2009). Climatic factors influencing the isotope composition of Italian olive oils and geographic characterisation. *Rapid Communications in Mass Spectrometry, 23,* 448–454.

Kelly, S., Heaton, K., & Hoogewerff, J. (2005). Tracing the geographical origin of food: The application of multi-element and multi-isotope analysis. *Trends in Food Science and Technology, 16,* 555–567.

Li, X. L., Kang, L., Hu, J. J., Li, X. F., & Shen X. (2008). Aroma volatile compound analysis of SPME headspace and extract samples from crabapple (Malus sp.) fruit using GC–MS. *Agricultural Science in China, 7,* 1451–1457.

Liu, Y., Liu, X., Zhong, F., Tian, R., Zhang, K., Zhang, X., & Li, T. (2011). Comparative Study of Phenolic Compounds and Antioxidant Activity in Different Species of Cherries. *Journal of Food Science*, *76*, C633-C638.

Longobardi, F., Casiello, G., Sacco, D., Tedone, L., Sacco, A. (2011). Characterisation of the geographical origin of Italian potatoes, based on stable isotope and volatile compound analyses. *Food Chemistry*, *124*, 1708–1713.

Longobardi, F., Ventrella, A., Casiello, G., Sacco, D., Catucci, L., Agostiano, A., & Kontominas, M. G. (2012a). Instrumental and multivariate statistical analyses for the characterization of the geographical origin of Apulian virgin olive oils. *Food Chemistry*, *133,* 579–584.

Longobardi, F., Ventrella, A., Casiello, G., Sacco, D., Tasioula-Margari, M., Kiritsakis, A. K., & Kontominas, M. G. (2012b). Characterisation of the geographical origin of Western Greek virgin olive oils based on instrumental and multivariate statistical analysis. *Food Chemistry, 133*, 169–175.

Longobardi, F., Ventrella, A., Bianco, A., Catucci, L., Cafagna, I., Gallo, V., Mastrorilli P., & Agostiano, A. (2013). Non-targeted 1H NMR fingerprinting and multivariate statistical analyses for the characterisation of the geographical origin of Italian sweet cherries. *Food Chemistry*, *141,* 3028–3033.

Pacioni, G., Cerretani, L., Procida, G., & Cichelli, A. (2014). Composition of commercial truffle flavored oils with GC–MS analysis and discrimination with an electronic nose. *Food Chemistry, 146,* 30–35.

Pérez-Sánchez, R., Gómez-Sánchez, M. Á., & Morales-Corts, M. R. (2010). Description and quality evaluation of sweet cherries cultured in Spain. *Journal of Food Quality*, *33,* 490–506.

Perini, M., & Camin, F. (2013). δ^{18}O of Ethanol in Wine and Spirits for Authentication Purposes. *Journal of Food Science, 78,* C839-C844.

Peris, M., & Escuder-Gilabert, L. (2009). A 21st century technique for food control: electronic noses. *Analytica Chimica Acta*, *638,* 1-15.

Rossmann, A., Reniero, F., Moussa, I., Schmidt, H.-L., Versini, G., & Merle, M. H. (1999). Stable oxygen isotope content of water of EU data-bank wines from Italy, France and Germany. *Zeitschrift fur Lebensmittel-Untersuchung und-Forschung A–Food Research and Technology, 208*, 400–407.

Rummel, S., Hoelzl, S., Horn, P., Rossmann, A., & Schlicht, C. (2010). The combination of stable isotope abundance ratios of H, C, N and S with ^{87}Sr/^{86}Sr for geographical origin assignment of orange juices. *Food Chemistry, 118,* 890-900.

Russo, M., di Sanzo, R., Cefaly, V., Carabetta, S., Serra, D., & Fuda, S. (2013). Non-destructive flavour evaluation of red onion (Allium cepa L.) Ecotypes: An electronic-nose-based approach. *Food Chemistry, 141,* 896–899.

Sacco, D., Brescia, M. A., Sgaramella, A., Casiello, G., Buccolieri, A., Ogrinc, N., et al. (2009). Discrimination between Southern Italy and foreign milk samples using spectroscopic and analytical data. *Food Chemistry*, *114*, 1559–1563.

Torri, L., Sinelli, N., & Limbo, S. (2010). Shelf life evaluation of fresh-cut pineapple by using an electronic nose. *Postharvest Biology and Technology*, *56*, 239-245.

Wassenaar, L. I., & Hobson K. A. (2003). Comparative equilibration and online technique for determination of nonexchangeable hydrogen of keratins for use in animal migration studies. *Isotopes in Environmental and Health Studies*, *39*, 211-217.

Zhao, Y., Zhang, B., Chen, G., Chen, A., Yang, S., & Ye, Z. (2014). Recent developments in application of stable isotope analysis on agro-product authenticity and traceability. *Food Chemistry, 145,* 300-305.

Ziegler, H., Osmind, C. B., Stichler, W., Trimborn, P. (1976). Hydrogen isotope discrimination in higher plants: Correlations with photosynthetic pathway and environment. *Planta, 128,* 85-92.

Table(s)

Table 1. LDA recognition and prediction abilities for the models classifying Italian sweet cherries according to their geographical origin (Apulia and Emilia Romagna), on the basis of the EN, IRMS and the relevant combined data. Regarding EN, the performances obtained considering different variable employment strategies are reported, i.e. VES1, VES2, and VES3, using 1, 10, and 120 points per sensor curve, respectively.

	LDA performances (%)						
Dataset	**EN VES1**	**EN VES2**		**EN VES3**		**IRMS**	**EN VES2+ IRMS**
Variable selection procedure	no variable reduction	SELECT routine	Stepwise statistics	SELECT routine	Stepwise statistics	no variable reduction	Stepwise statistics
Recognition (modeling set)	85.9	97.4	97.4	98.7	87.2	94.9	96.2
CV prediction (*k*=5)	80.8	85.9	89.7	89.7	85.9	91.0	92.3
External prediction (external set)	82.4	82.4	82.4	76.5	85.3	94.1	94.1

Table(s)

Table 2. Means and standard deviations (SD) of isotopic ratios obtained for the Italian cherries. Results for geographical origin of t-test are reported: groups of one row with different letters are statistically different ($p < 0.05$).

	Italian region			
	Emilia Romagna		Apulia	
Isotopic ratio	**mean**	**SD**	**mean**	**SD**
$\delta^{13}C$	-26.5^{a}	0.9	-26.4^{a}	0.8
$\delta^{18}O$	33.2^{a}	1.2	35.4^{b}	0.9
δD	-38.5^{a}	4.8	-30.7^{b}	5.5

Figure(s)