CrossMark

ORIGINAL PAPER

Electronic tongue: a versatile tool for mineral and fruit-flavored waters recognition

Luís G. Dias[1,2] · Zelda Alberto[1,3] · Ana C. A. Veloso[4,5] · António M. Peres[6]

Abstract Natural mineral waters (still), effervescent natural mineral waters (sparkling) and aromatized waters with fruit-flavors (still or sparkling) are an emerging market. In this work, the capability of a potentiometric electronic tongue, comprised with lipid polymeric membranes, to quantitatively estimate routinely quality physicochemical parameters (pH and conductivity) as well as to qualitatively classify water samples according to the type of water was evaluated. The study showed that a linear discriminant model, based on 21 sensors selected by the simulated annealing algorithm, could correctly classify 100 % of the water samples (leave-one out cross-validation). This potential was further demonstrated by applying a repeated K-fold cross-validation (guaranteeing that at least 15 % of independent samples were only used for internal-validation) for which 96 % of correct classifications were attained. The satisfactory recognition performance of the E-tongue could be attributed to the pH, conductivity, sugars and organic acids contents of the studied waters, which turned out in significant differences of sweetness perception indexes and total acid flavor. Moreover, the E-tongue combined with multivariate linear regression models, based on sub-sets of sensors selected by the simulated annealing algorithm, could accurately estimate water's pH (25 sensors: R^2 equal to 0.99 and 0.97 for leave-one-out or repeated K-folds cross-validation) and conductivity (23 sensors: R^2 equal to 0.997 and 0.99 for leave-one-out or repeated K-folds cross-validation). So, the overall satisfactory results achieved, allow envisaging a potential future application of electronic tongue devices for bottled water analysis and classification.

Keywords Natural mineral waters · Fruit-flavored waters · Water quality parameters · Electronic tongue · Discriminant analysis · Multiple linear regression models

✉ António M. Peres
  peres@ipb.pt

1 Escola Superior Agrária, Instituto Politécnico de Bragança, Campus Santa Apolónia, 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 Instituto Superior Politécnico do Kuanza-Sul, Avenida Joaquim Kapango, Sumbe, Angola

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, 5301-855 Bragança, Portugal

Introduction

The quality of tap, spring, natural mineral and effervescent natural mineral waters is greatly influenced by the geographical origin and quality of the raw water as well as by the efficiency of the drinking water production and bottling process [1]. Indeed, some studies report natural variation of certain groups of chemical elements among bottled water brands from different origins, reflecting geological patterns [2, 3]. Natural mineral and effervescent natural mineral waters exploitation and marketing are regulated [4–6] and play an essential role in today's society [7]. These two types of mineral waters are commonly known by consumers as still mineral water (StW) or sparkling mineral water (SpW), including the latter naturally carbonated

natural mineral water, natural mineral water fortified with gas from the spring and carbonated natural mineral water. Still or sparkling waters with incorporated flavors are considered soft-drinks and their commercialization is also regulated [8, 9]. So, monitoring the quality of these waters, during their production line as well as of the final bottled product is of utmost importance. However, the characterization of liquid substances is often inexact due to non-homogeneity of the species in the liquid and to fluctuations rose due to the slight variations of the surface or ground raw water [10]. In 2007, Ciosek and Wróblewski [11] reviewed several applications of potentiometric and voltammetric electronic tongue (E-tongue) sensor arrays for liquid sensing, showing that E-tongue devices have been successfully used for water analysis in food, environmental and industrial fields (e.g., mineral waters recognition, assessment of ion levels and total hardness in natural waters, analysis of pollutants (metal ions, organic compounds or microorganisms) in natural waters, monitoring of spoilage of seawater and freshwater fish, analysis of paper-mill wastewaters, determination of anionic and nonionic surfactants in natural waters). Indeed in the last decade, the use of potentiometric or voltammetric E-tongues has been increasingly reported since they have been proved to be useful in assessment of complex liquid media. Regarding drinking water analysis, E-tongues are mainly applied for: (i) monitoring water quality and detection of specific chemical compounds, possible contaminants, microorganisms, determination of sewage water concentrations in drinking water [1, 7, 12–18]; or, (ii) water classification/authentication according to commercial brands or type of water [10, 19–28]. Water safety monitoring technologies based on E-tongues have been recently discussed by Vagin and Winquist [29], which demonstrated the potential of E-tongues for high throughput monitoring of edible or potable products. For the latter type of application, potentiometric E-tongue devices with ion-selective membranes, ion sensitive field effect transistors or polymer membranes with active ionic functional sites are usually described and successfully applied allowing the correct recognition of the water brand or type, namely tap water, spring water or natural mineral and/or effervescent natural mineral waters [19–22, 27, 28, 30, 31]. A potentiometric E-tongue system with multi-electrodes assembly was also capable to recognize different classes of drinking water namely mineral, sparkling and tap water [19]. A voltammetric E-tongue with two working electrodes (Ag and Pt) was also reported for mineral water classification and authentication [10]. A novel E-tongue based on Fourier transform impedance spectroscopy was also used for classifying bottled mineral waters according to brand [24]. More recently, Sipos et al. [32] used a commercial potentiometric electronic tongue for predicting of sensory

attributes of flavored mineral waters. Oliveira et al. [33] used flow-cell electronic tongues with nanocomposite films for discriminating potable waters samples (purified, treated and natural) and nonpotable water samples contaminated with metals or pesticides. Recently, Nery et al. [28] proposed a low cost potentiometric paper-based E-tongue capable of discriminating forged water samples, enabling distinguishing tap and lake water from mineral water samples (commercially available or directly obtained from spring).

In all these works, different multivariate techniques are applied to extract the valuable information contained in the E-tongue signal profiles: hierarchical clustering analysis [21], principal component analysis (PCA) [10, 16, 19, 21, 22, 24, 26–28, 31–33], partial least squares (PLS) [10, 17, 32], soft independent modeling class analysis [21], multi-layer feed-forward (MLFF) networks [23], fuzzy-logic or fuzzy-ARTMAP classifiers [16, 20, 23, 26], support vector machines [24, 25], discriminant factorial analysis [22], linear discriminant analysis (LDA) [23, 31, 32], M-robust regression [17] and K-nearest neighbor [28].

In this work, a potentiometric E-tongue with cross-sensitivity lipid membranes, previously applied for qualitative and quantitative analysis of carbonated soft-drinks [34–36], was used to discriminate different types of commercial bottled waters. Commercial still mineral waters (StW), sparkling mineral waters (SpW), fruit-flavored still waters (FStW) and fruit-flavored sparkling waters (FSpW) were evaluated aiming, and for the first time, to establish a predictive discrimination model for assessing the type of drinking water. Fruit-flavored waters are being introduced in the market in the last years trying to meet consumers' expectations and as a marketing response towards the increasing consumption of other traditional soft-drinks. The E-tongue signal patterns were treated using LDA combined with the simulated annealing (SA) algorithm, which is a meta-heuristic variable selection technique aiming to improve the discrimination capability of the E-tongue device. Finally, the capability of the E-tongue to quantify classic water quality parameters (pH and conductivity), for which routine analysis are legally required [37], by means of multivariate linear regression (MLR) models, based on sub-sets of sensors selected using the SA algorithm, was also assessed.

Materials and methods

Samples

In this work, 34 commercial bottled drinking waters, from 20 different brands (Portuguese, French and Italian), were purchased at local supermarkets (Bragança, Portugal) and

stored at room temperature until analysis. The samples included 6 still mineral waters (StW), 8 flavored still mineral waters (FStW), 12 sparkling waters (SpW) and 8 flavored sparkling waters (FSpW). More details regarding the type of mineral water (according to European or Portuguese Directives [5, 6, 37]), the type of aroma of the added fruit-flavors (according to the label information) and the coded brands are given in Table 1.

Sample preparation and physicochemical parameters assessment

The pH and conductivities of all water samples were measured at 20 °C. pH values of all water samples were measured using a microprocessor pHMeter (pH 211 from Hanna Instruments) with a pH combined electrode. The system was calibrated with pH buffer solutions (pH 4.00, 7.00 and 9.00 from Panreac). The conductivities of the water samples were measured using a multi-parameter analyzer (Consort C861) with a conductivity probe (CS SK10B), which was previously calibrated using KCl standard solutions (0.001 mol/L–146.9 µS/cm; 0.01 mol/L–1413 µS/cm; and, 0.1 mol/L–12.880 mS/cm). For both type of measurements, still waters were analyzed directly and sparkling waters were firstly degassed (Elma Transsonic 460/H ultrasound bath) during 10 min. Fruit-flavored waters were also analyzed regarding the main sugars (i.e., fructose, glucose and sucrose) and organic acids (i.e., citric, malic and ascorbic acids) contents using a high performance liquid chromatography (HPLC) method previously in-house validated by the research team [36, 38]. Briefly, a HPLC Varian ProStar equipped with a 220 pump (Varian, Inc.), a 7725i Rheodyne manual injector, provided with a 20 µL loop, a 7981 Jones Chromatography column oven and a UV detector (Varian, model 9050) coupled to a RI detector (Varian, model RI-4) was used to simultaneously separate and quantify the main organic acids (ascorbic, citric and malic acids by UV at 210 nm) and sugars (glucose, fructose and sucrose by RI) contained in the flavored water samples. The separation of the main sugars and organic acids was achieved in a thermostated (45 °C) Supelcogel C-610H size-exclusion column (SEC: C-610H model, 30 cm, 7.8 mm id), using an isocratic elution, with a mobile phase free of organic solvents (0.1 % orthophosphoric acid aqueous solution). For chromatographic assays, water samples were analyzed as purchased, except sparkling waters, which were degassed during 5 min in an ultrasonic bath (Elma Transsonic 460/H, Singen, Germany). When needed, samples were diluted with deionized water. All water samples were filtered through a 0.2-µm nylon filter (Whatman, Buckinghamshire, UK) and stored at −5 °C until analysis.

Sensory indexes of fruit-flavored drinking waters

Still and sparkling fruit-flavored drinking waters contain different amounts of sugars and organic acids contributing to different sensory perceptions that can be quantitatively represented by the sweetness perception index (SWP_{Index}), total acid flavor (TA_{Flavor}) and the well-balanced flavor (WB_{Flavor}) index, which are used as consumer's beverage acceptability or overall taste indicators. These sensory indexes may be calculated using the individual sugars and/or organic acids contents as previously described [36]. In this work, these three overall taste attributes were calculated and further applied to infer about the differences among the studied drinking waters as well as to tentatively explain the E-tongue performance.

Sample preparation and E-tongue analysis

All independent water samples were electrochemically analyzed in the same day, totalizing 34 electrochemical assays. For each assay, 50 mL of water were directly removed from each bottle and analyzed, at 20 °C, after a 10 min stabilization period without any pre-treatment, with the exception of sparkling waters, which were firstly, degassed using an ultrasound bath (Elma Transsonic 460/H), during a 5 min period. Between assays, the E-tongue device was washed with deionized water and the excess of water was removed carefully using absorbent paper.

The E-tongue included two print-screen potentiometric arrays with different cross-sensitivity lipid membranes (diameter: 3.6 mm; thickness: 0.3 mm) used as chemical sensors: 4 lipidic additives (octadecylamine, oleyl alcohol, methyltrioctylammonium chloride and oleic acid from Fluka; corresponding to approximately 3 %), 5 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 %), as been previously described in detail [36, 39]. The type of sensors and polymeric membrane compositions (relative percentage of additive, plasticizer and PVC) were selected based on a previous work of Dias et al. [40] taking into account their satisfactory signal stability over time (%RSD < 5 %) and repeatability (0.5 % < %RSD < 15 %) for basic standard taste compounds (e.g., sweet, acid, bitter, salty and umami). Also, lipid polymeric membranes were used since they may interact with tastant substances through electrostatic or hydrophobic interactions Kobayashi et al. [41]. 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).

Table 1 Label information regarding the composition of the studied bottled drinking waters

No.	Commercial brand	Origin	Class of water	Abbreviation	Label information
1	A	Portugal	Fruit-flavored sparkling water	FSpW	Raspberry, apple, pear, citric acid
2	A	Portugal	Fruit-flavored sparkling water	FSpW	Lemon, apple, citric acid
3	B	Portugal	Fruit-flavored sparkling water	FSpW	Apple, citric acid
4	B	Portugal	Fruit-flavored sparkling water	FSpW	Apple, gooseberry, citric acid
5	B	Portugal	Fruit-flavored sparkling water	FSpW	Sugar, fruit aroma, citric acid
6	C	Portugal	Fruit-flavored sparkling water	FSpW	Lemon, lime, apple, citric acid, ascorbic acid
7	D	Portugal	Fruit-flavored sparkling water	FSpW	Lemon
8	D	Portugal	Fruit-flavored sparkling water	FSpW	Lime
9	E	Portugal	Fruit-flavored still water	FStW	Apple, lemon, fruit-aroma, citric acid
10	E	Portugal	Fruit-flavored still water	FStW	Fruit-aroma, coconut, citric acid
11	E	Portugal	Fruit-flavored still water	FStW	Apple, grape, carrot, strawberry, citric acid
12	E	Portugal	Fruit-flavored still water	FStW	Fruit-aroma, apple, caramel syrup, citric acid
13	F	Portugal	Fruit-flavored still water	FStW	Fruit-aroma, citric acid
14	F	Portugal	Fruit-flavored still water	FStW	Sugars, lemon, apple, malic acid
15	F	Portugal	Fruit-flavored still water	FStW	Sugars, lemon, apple, citric acid
16	B	Portugal	Effervescent natural mineral water-sparkling	SpW	Natural mineral water fortified with gas from the spring
17	D	Portugal	Effervescent natural mineral water-sparkling	SpW	Naturally carbonated natural mineral water
18	D	Portugal	Effervescent natural mineral water-sparkling	SpW	Naturally carbonated natural mineral water
19	G	Portugal	Effervescent natural mineral water-sparkling	SpW	Natural mineral water fortified with gas from the spring
20	H	France	Effervescent natural mineral water-sparkling	SpW	Carbonated natural mineral water
21	I	Portugal	Effervescent natural mineral water-sparkling	SpW	Naturally carbonated natural mineral water
22	J	Portugal	Effervescent natural mineral water-sparkling	SpW	Carbonated natural mineral water
23	A	Portugal	Effervescent natural mineral water-sparkling	SpW	Naturally carbonated natural mineral water
24	K	Portugal	Effervescent natural mineral water-sparkling	SpW	Naturally carbonated natural mineral water
25	L	Portugal	Effervescent natural mineral water-sparkling	SpW	Naturally carbonated natural mineral water
26	M	Italy	Effervescent natural mineral water-sparkling	SpW	Naturally carbonated natural mineral water
27	N	Portugal	Natural mineral water-still	StW	Natural water
28	F	Portugal	Natural mineral water-still	StW	Natural water
29	O	Portugal	Natural mineral water-still	StW	Natural water
30	P	Portugal	Natural mineral water-still	StW	Natural water
31	E	Portugal	Natural mineral water-still	StW	Natural water
32	Q	France	Natural mineral water-still	StW	Natural water
33	R	Portugal	Effervescent natural mineral water-sparkling	SpW	Naturally carbonated natural mineral water
34	E	Portugal	Fruit-flavored sparkling water	FSpW	Apple, orange, fruit-aroma, citric acid

Statistical analysis

One-way analysis of variance (one-way ANOVA) was used to infer about the existence of statistical significant differences, at a 5 % significance level, of the mean pH and conductivities among the four classes of drinking mineral waters evaluated (i.e., StW, SpW, FStW and FSpW). When a statistical difference was found (P value <0.05) the analysis was complemented with the Tukey's post hoc multi-comparison test. For the other physicochemical parameters (main sugars and organic acids concentrations and related sensory indexes), unilateral or bilateral Welch's t tests (or unequal variances t tests) were carried out for comparing the two classes of fruit-flavored mineral waters (FStW and FSpW).

The LDA was used as a supervised pattern recognition tool to verify if the E-tongue signals contained information capable of correctly classifying mineral water samples according to four different classes (StW, SpW, FStW and FSpW). Regarding the potential of the E-tongue device to quantitatively estimate pH and conductivity of the mineral water samples, regardless the type of water (natural or effervescent natural mineral water, fruit-flavored or not) was also tested using MLR models. Both techniques correspond to mathematical models consisting of linear combinations of independent variables established with the aim of optimizing the separation of two or more groups of samples (LDA) or to quantitatively estimate the concentrations of physico-chemical parameters (MLR model), with predictive purposes. Detailed information regarding these two common multivariate statistical tools can be found in the literature [42, 43]. Briefly, the LDA sets the relationship between a qualitative dependent variable (groups of data) and a set of quantitative independent variables (X_j). The methodology of this supervised recognition technique involves the establishment of one or more discriminant functions D_i, that are linear combinations of the predictor independent variables X_j (being X the potential signal of each j sensor recorded for the water samples analyzed), in order to maximize the distance between different groups. The first (D_1) function has the greatest discrimination between groups, being the other D_i functions orthogonal and with decreasing discrimination capability, but that also established to maximize the differences between groups. Considering g groups and n independent variables, it is possible to set m functions given by,

$$D_i = \beta_0 + \sum_{j=1}^{n} (\beta_j X_j) + \varepsilon_i, \qquad (1)$$
$$\text{with } 1 \leq i \leq m = \min(g - 1, n)$$

where, the β_k are the discriminant coefficients calculated in such a way, that the ratio of the sum of the square errors

between groups and within groups is as large as possible, and ε_i is the related error term for each function.

Finally, each function gives a discriminant score that allows classifying each sample as belonging to the group with the nearest centroid.

The MLR models may be used when more than one predictor is available, allowing correlating a quantitative dependent variable, Y_i (e.g., where Y would represent the pH or conductivity data) with two or more independent variables, $X_{i,j}$ (representing the letter X the potential signal of each sensor recorded for water samples) according to the equation:

$$Y_i = \alpha_0 + \sum_{j=1}^{n} (\alpha_j X_{i,j}) + \varepsilon_i \qquad (2)$$

where, α_0 is the intercept, α_j are the coefficients and ε_i is the error for each MLR model established.

To achieve the best prediction performance appropriate variable selection algorithms must be used to enable identifying the most informative set of predictors. Among these algorithms, heuristic techniques may be implemented and a robust cross-validation methodology should also be applied for establishing a suitable predictive linear model. In this work, for both qualitative and quantitative approaches, the best subsets of k independent predictors, among the 40 E-tongue potentiometric signals recorded, were chosen using a meta-heuristic simulated annealing (SA) variable selection algorithm [44–46]. The algorithm searches, iteratively, for a global minimum that optimizes a system with k ($\subseteq$K) variables. Globally, a maximum of 10,000 attempts is fixed to select the best subset of variables (best model), starting the process of selecting the best subsets of variables on each trial, thus ensuring a greater confidence in finding a true optimal solution. In the present study, for each sub-set of sensors under evaluation (possible combinations of 2–39 sensors), being the sub-set of sensors chosen the one that allowed reaching the maximum value of a pre-defined quality criterion, used as a measure of the goodness of fitting [44]. In this work, tau2 criterion and the Pearson determination coefficient (R^2) were used for LDA and MRLM, respectively. To evaluate LDA and MLR models, a leave-one-out cross-validation (LOO-CV) procedure was used as a first approach. Since this internal-validation technique may lead to over-optimistic results a repeated K-fold cross-validation strategy was also implemented for the sub-sets of sensors selected by the SA algorithm. For this purpose, data was divided into K subsets that allowed obtaining K models, each one fitted considering K-1 subsets, as the training set, leaving out one of the subsets for the internal validation, to compute the predictive error for the obtained model [47]. The number of K-folds was set equal to 6, enabling the formation of testing subset containing at least 15 % of the initial data and ensuring the presence of at least

Table 2 Effect of drinking waters' class (still, sparkling, flavored still and flavored sparkling waters) on physicochemical quality parameters (pH and conductivity), based on one-way ANOVA (the results are given as mean $\pm$ SD)

Bottled drinking water	No. of samples	pH	Conductivity (mS/cm)
StW	6	6.0 ± 0.7^a	3 ± 4^c
SpW	12	6.3 ± 0.3^a	$1.4 \pm 0.9^{a,b}$
FStW	8	3.2 ± 0.3^c	$0.8 \pm 0.3^{b,c}$
FSpW	8	4.4 ± 1.2^b	1.7 ± 0.8^a
P value[1]		<0.0001	0.0019

In each column and for parameters, different letters mean statistically significant differences (P value < 0.05 for Tukey's post hoc multi-comparison test)

StW natural mineral still water, *SpW* natural mineral sparkling water, *FStW* fruit-flavored still water, *FSpW* fruit-flavored sparkling water

[1] *P* value from one-way ANOVA

one sample of each group and thus permitting bias reduction. Also, by applying a repeated K-fold cross-validation procedure (with 10 repetitions), the uncertainty of the estimates could be significantly reduced. To normalize the weight of each variable in the final linear classification model, variable scaling and centering procedures were applied. The possibility of using the E-tongue method as tool for quantifying the classical quality parameters (pH and conductivity) was further checked by testing if the slope and intercept values for LOO-CV or repeated K-folds procedures were equal to the theoretical expected values (one and zero, respectively), from a statistical point of view [35], when representing the predicted values, estimated by the MLR-SA models versus the experimental data. All statistical analysis were performed at a 5 % significance level using the Subselect [44] and MASS [48] packages of the open source statistical program R (version 2.15.1).

Results and discussion

Physicochemical parameters and sensory indexes evaluation

All drinking waters studied were evaluated regarding pH and conductivity quality parameters. As can be verified in

Table 2, the determined range of those parameters are in accordance with legal requirements [4–6, 37], which established a maximum pH of 9 and 9.5 for natural still and sparkling mineral waters, respectively, and a conductivity of 2500 µS/cm at 20 °C). Fruit-flavored drinking water (still or sparkling) were further analyzed regarding main sugars (i.e., fructose, glucose and sucrose) and main organic acids (i.e., citric, malic and ascorbic acids) contents, which enabled the calculation of sensory indexes that are usually used as consumers' acceptability indicators [36], namely sweet perception index (SWP$_{Index}$), total acidity flavor (TA$_{Flavor}$) and well-balanced flavor (WB$_{Flavor}$).

Regarding the pH values measured and based on the results of the one-way ANOVA (Table 2), it is clearly there are significant statistical differences between the water class groups analyzed (P value < 0.0001). In fact, flavored waters (FStW and FSpW) are significantly more acidic than not flavored waters (StW and SpW) (P value $\leq$ 0.0006, for Tukey's post hoc multi-comparison test), flavored still waters (FStW) showed a higher acidity than flavored sparkling waters (FSpW) (P value = 0.0035, for Tukey's post hoc multi-comparison test) and finally, natural still and sparkling waters (not flavored, i.e., StW and SpW) had statistical similar acidities (P value = 0.8785, for Tukey's post hoc multi-comparison test). The drinking waters evaluated show quite

Table 3 Effect of fruit-flavored drinking waters' class (still or sparkling) on sugars (sucrose, glucose and fructose), organic acids (citric and malic acids) and sensory indexes (sweetness perception index, total acidity flavor and well-balanced flavor), based on Welch's *t* test (the results are given as mean $\pm$ SD)

Bottled drinking water	No. of samples	Sugars			Organic acids (g/L)		Sensory indexes		
		Fructose (g/L)	Glucose (g/L)	Sucrose (g/L)	Citric acid (g/L)	Malic acid (g/L)	SWP$_{Index}$ (g/L)	TA$_{Flavor}$ (g/L)	WB$_{Flavor}$
FStW	8	13 ± 6	19 ± 7	61 ± 10	0.82 ± 0.49	0.24 ± 0.20	44 ± 21	1.6 ± 0.2	37 ± 17
FSpW	8	1.8 ± 2.7	2.5 ± 4.3	5.9 ± 7.8	1.8 ± 1.2	0.03 ± 0.02	11 ± 14	1.9 ± 1.3	12 ± 13
P value[a]		$0.0004^†$	$0.0019^†$	$0.2853^‡$	$0.0104^†$	$0.0050^†$	$0.0020^†$	$0.1279^‡$	$0.0053^†$

FStW fruit-flavored still water, *FSpW* fruit-flavored sparkling water

[a] *P* value from Welch's *t* test: †unilateral or ‡bilateral test depending if it is found or not a statistical significant difference at a 5 % significance level
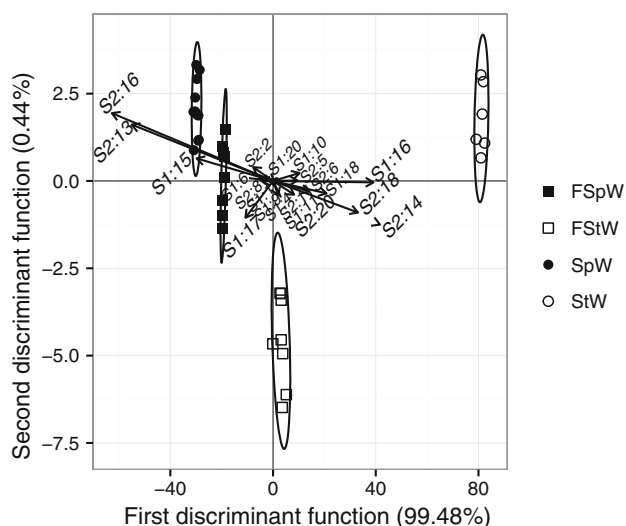
Fig. 1 Drinking waters classification (original data grouping) using LDA coupled with SA variable selection algorithm, based on the best sub-set of E-tongue sensors (21 sensors; 1st sensor-array: S1:1, S1:4, S1:6, S1:9 to S1:11, S1:15 to S1:18 and S20, 2nd sensor-replica-array: S2:2, S2:5, S2:6, S2:8, S2:11, S2:13, S2:14, S2:16, S1:18 and S2:20). *StW* natural still waters; *SpW* natural sparkling waters; *FStW* fruit-flavored still waters; *FSpW* fruit-flavored sparkling waters

higher conductivity variability, even within the same water class group, which is clear from the magnitude of the standard deviations (Table 2). Globally, it can be concluded from the one-way ANOVA that there mean conductivity values significantly differ with the water class groups evaluated (P value = 0.0019). Indeed, sparkling waters (flavored and not flavored, i.e., FSpW and SpW, respectively) had similar conductivities (P value = 0.05 for Tukey's post hoc multi-comparison test). Still waters (FStW and StW) also showed similar conductivities (P value = 0.05 for Tukey's post hoc multi-comparison test). Finally, a trend can be inferred from the results (Table 2) showing that sparkling waters normally have higher conductivities than still waters.

The fruit-flavored drinking waters (still or sparkling mineral waters) were analyzed and the main sugars (fructose, glucose and sucrose) and organic acids (ascorbic, citric and malic acids) concentrations were quantified, with the exception of ascorbic acid, which was never detected, enabling the calculation of sensory indexes, namely sweetness perception index (SWP$_{Index}$), total acidity flavor (TA$_{Flavor}$) and well-balanced flavor (WB$_{Flavor}$), which values were further statistically compared (Table 3). As can be inferred by the magnitude of the standard deviation (SD) values, in general, a great variability was found for each parameter evaluated within the same fruit-flavored water group. These variations may be attributed to the inclusion of water samples from different brands and with quite different added fruit-flavors in the same group of waters (Table 1). Based on the Welch's t test it could be

concluded that the contents of sucrose in the fruit-flavored still or sparkling water (FStW or FSpW, respectively) were not statistically different (P value = 0.2853 for Welch's t test) but the fructose and glucose levels of FSpW were significantly lower than the respective contents found in FStW (P value $\leq$ 0.0019 for Welch's t test). On the other hand, for the two organic acids found in the fruit-flavored drinking waters, it could be concluded that FSpW had significantly higher concentrations of citric acid than FStW (P value = 0.0334 for Welch's t test) but lower contents of malic acid (P value = 0.0104 for Welch's t test). These observed trends lead to the conclusion that FSpW had global SWP$_{Index}$ and WB$_{Flavor}$ lower than FStW (P value $\leq$ 0.0020 for Welch's t test), but statistical similar TA$_{Flavor}$ values (P value = 0.1279 for Welch's t test).

E-tongue analysis of drinking water: qualitative and quantitative evaluation

The results previously discussed clearly identify trend of the drinking waters' physicochemical parameters evaluated (i.e., pH, conductivity, sugars and organic acids concentrations as well as calculated sensory indexes) demonstrating the four classes of drinking waters had different characteristics namely acidity and sweetness levels which are related with basic taste sensations and so, a successful discrimination of drinking waters could be envisaged by applying the E-tongue device, which potential to differentiate food samples with different sensory attributes has been already demonstrated [36, 49].

LDA-SA plus E-tongue: qualitative classification tool

Among the 40 sensors of the E-tongue, 21 potentiometric lipid sensors (1st sensor-array: S1:1, S1:4, S1:6, S1:9 to S1:11, S1:15 to S1:18 and S20, 2nd sensor-replica-array: S2:2, S2:5, S2:6, S2:8, S2:11, S2:13, S2:14, S2:16, S1:18 and S2:20) were selected by the SA algorithm as the best sub-set of electrochemical sensors. So a LDA-SA model was established enabling the correct classification of 100 % of the original grouped drinking water samples according to their specifications (i.e., still and sparkling mineral waters, natural or fruit-flavored: StW, SpW, FStW and FSpW), as can be seen from Fig. 1. The referred model had two significant discriminant functions (that explained (99.48 and 0.44 % of the data variability, respectively). For LOO-CV procedure, the referred LDA-SA model also allowed 100 % of correct classifications. The predictive performance of the mentioned LDA-SA model (with the same 21 sensors) was further demonstrated since it allowed an average correct classification accuracy of 96 ± 8 % (ranging from a minimum sensitivity of 71 % up to

Fig. 2 MLR models based on the best E-tongue sub-set of sensors (25 sensors; 1st sensor-array: S1:1, S1:3, S1:4, S1:6, S1:10, S1:11, S1:13 to S1:17 and S1:19, 2nd sensor-replica-array: S2:1 to S2:4, S2:7 to S2:11, S2:16 to S2:19) selected applying SA variable selection algorithm for drinking water pH values prediction. **a** leave-one-out cross-validation data ($R^2_{LOO-CV} = 0.98$); **b** K = 6 and 10 repeats cross-validation data (average $R^2_{repeatedK-folds} = 0.97 \pm 0.04$). *StW* natural still waters; *SpW* natural sparkling waters; *FStW* fruit-flavored still waters; *FSpW* fruit-flavored sparkling waters
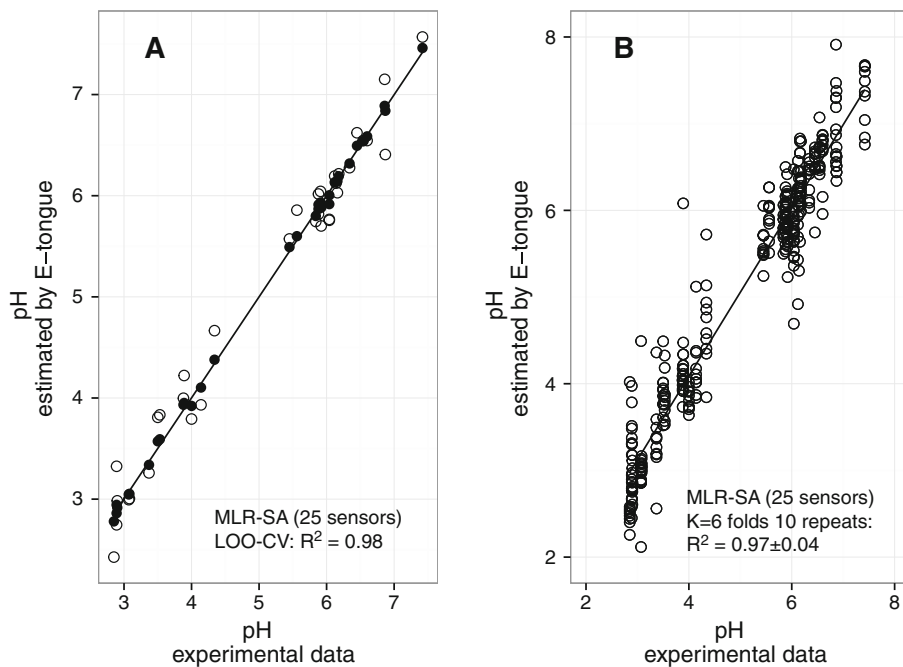

Fig. 3 MLR models based on the best E-tongue sub-set of sensors (21 sensors: 1st sensor-array: S1:1 to S1:3, S1:6, S1:9 to S1:11, S1:13 to S1:15, S1:17 and S1:20; 2nd sensor-replica-array: S2:3, S2:4, S2:8, S2:10 to S2:13, S2:16 and S2:18) selected applying SA variable selection algorithm for drinking water conductivity (mS/cm) values prediction. **a** Leave-one-out cross-validation data ($R^2_{LOO-CV} = 0.997$); **b** K = 6 and 10 repeats cross-validation data (average $R^2_{repeatedK-folds} = 0.99 \pm 0.02$). *StW* natural still waters; *SpW* natural sparkling waters; *FStW* fruit-flavored still waters; *FSpW* fruit-flavored sparkling waters
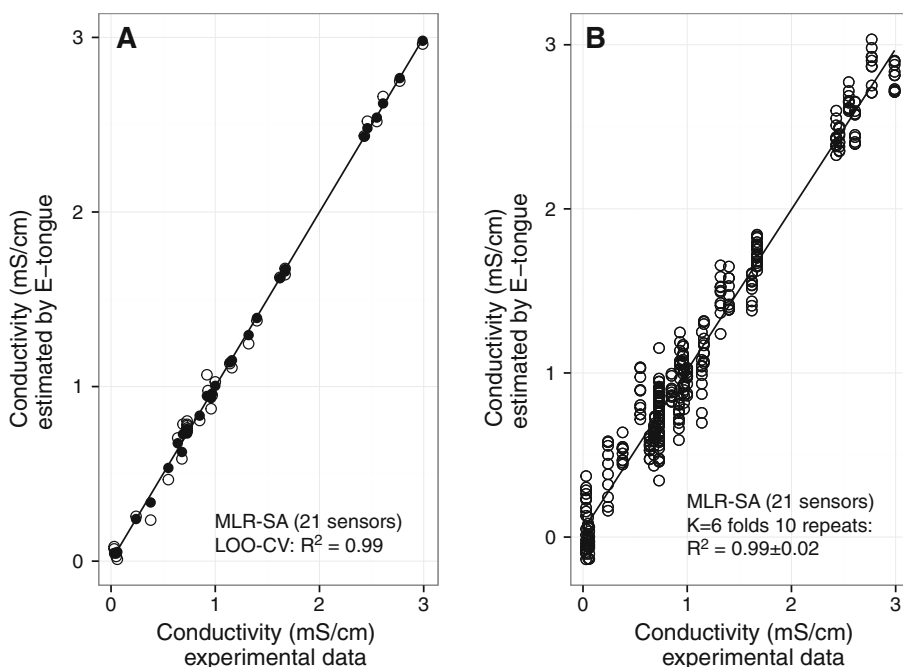
100 %), for the repeated K-folds cross-validation procedure (K = 6, meaning that at least 15 % of the independent data samples, i.e., one to two drinking water samples of each group, were used for internal-validation purposes; being this process repeated 10 times). So, the very satisfactory overall results clearly indicate that the E-tongue coupled with a LDA-SA tool may be seen as a valuable water testing system with great discriminating potential enabling to correctly classify natural and effervescent mineral waters, fruit-flavored or not, regardless their commercial brand, geographical origin, natural or flavored with fruit-aromas.

MLR-SA plus E-tongue: pH and conductivity quantitative estimation

Finally, the capability of using the E-tongue device coupled with MLR-SA approach for quantifying two classical

Table 4 Linear regression lines obtained for the representation of predicted pH or conductivity of drinking waters (using MLR-SA models plus E-tongue models) versus experimental data for LOO-CV and repeated K-folds internal-validation procedures: slope, intercept values and respective 95 % confidence intervals (CI)

Regression line parameters	Drinking water physicochemical quality parameters			
	pH		Conductivity (mS/cm)	
	LOO-CV	Repeated K-folds	LOO-CV	Repeated K-folds
R^2	0.977	0.919	0.9947	0.968
Slope	0.985	0.981	0.999	0.995
Slope CI[a]	[0.930, 1.040]	[0.950, 1.012]	[0.973, 1.026]	[0.981, 1.008]
Intercept	0.088	0.151	0.001	0.003
Intercept CI[b]	[−0.201, 0.378]	[−0.014, 0.316]	[−0.035, 0.038]	[−0.016, 0.021]

[a] 95 % slope confidence interval

[b] 95 % intercept confidence interval

drinking water quality parameters (pH and conductivity) was assessed. So, two MLR models based on sub-sets of sensors, selected by applying the SA algorithm, were established, one for each parameter. The results showed that it was possible to satisfactorily estimate the pH values of the different drinking waters studied, regardless their specific group, by using a MLR-SA model with 25 sensors (1st sensor-array: S1:1, S1:3, S1:4, S1:6, S1:10, S1:11, S1:13 to S1:17 and S1:19; 2nd sensor-replica-array: S2:1 to S2:4, S2:7 to S2:11, S2:16 to S2:19). The satisfactory MLR-SA model performance was checked for LOO-CV procedure ($R^2_{LOO-CV} = 0.99$) and further demonstrated using a repeated K-folds internal cross-validation technique (average $R^2_{repeatedK-folds} = 0.97 \pm 0.04$, ranging from 0.838 to 0.999). For the latter, the dataset was split using sixfolds (K = 6), meaning that at least 15 % of independent data was used for internal validation during each one of the 10 repetition cycles). Similarly, it could be concluded that a MLR-SA model based on 23 E-tongue signals (1st sensor-array: S1:1, S1:2, S1:7 to S1:11, S1:14, S1:16, S1:18 and S1:19; 2nd sensor-replica-array: S2:1, S2:3, S2:4, S2:7, S2:8, S2:10, S2:12, S2:13, S2:15, S2:17, S2:18 and S2:20) could be used to satisfactorily estimating drinking waters' conductivities. As previously, the satisfactory MLR-SA model performance towards conductivity prediction in different water samples was checked for LOO-CV procedure ($R^2_{LOO-CV} = 0.997$) and further demonstrated using a repeated K-folds internal cross-validation technique (average $R^2_{repeatedK-folds} = 0.99 \pm 0.02$, ranging from 0.932 to 0.999). In the same way, the dataset was also split into sixfolds, allowing to use one fold (containing at least 15 % of independent data) for internal-validation purposes, being this procedure repeated 10 times to ensure the accuracy of the estimates and to minimize bias effect.

Moreover, for the two internal-validation procedures applied (LOO-CV and repeated K-folds), the representation of the predicted pH or conductivity values, calculated using the previous described MLR-SA models with 25 or 23 E-tongue sensors, respectively, versus the respective experimental data (measured using classical analytical techniques) gave linear straight lines (Figs. 2, 3), which slope and intercept parameters and the 95 % confidence intervals are gathered in Table 4. These results showed that, at 5 % of significance level, the respective slopes and intercept values (of each regression line) are statistically equal to the theoretical expected values (slope equal to one; intercept equal to zero), since the 95 % confidence intervals contain the values 1 and 0, respectively. These results confirmed the robustness of the proposed MLR-SA models and their possible practical application. Indeed, the overall satisfactory quantitative performance achieved is indicative that the proposed approach could be implemented for routine drinking water quality analysis, allowing an accurate estimative of pH and conductivity parameters.

Conclusions

The potentiometric E-tongue device was a versatile tool for discriminating natural mineral waters (still or sparkling) and fruit-flavored waters (still or sparkling), being achieved a 96 % (±8 %) mean correct classification rate, for a repeated K-folds cross-validation technique. The satisfactory predictive recognition capability of the LDA model proposed relied on the ability of the E-tongue to differentiate the different sweet and acid levels of the water samples analyzed as well as to the successful selection of the best sub-set of sensors, based on the SA algorithm, which gave a more reliable fingerprint of the different water matrices. Finally, the E-tongue device allowed the accurate prediction of pH and conductivity values, using MLR-SA technique (mean $R^2_{repeatedK-folds}$ equal to 0.97 ± 0.04 and 0.99 ± 0.02, respectively), showing their possible application for water quality monitoring. So, the present work demonstrated that a potentiometric E-tongue array with cross-sensitivity lipid membranes could be used to develop

LDA based water classifiers as well as MLR based water tools for bottled water quality assessment.

Compliance with ethical standards

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

Conflict of interest LG. Dias declares that he has no conflict of interest. Z. Alberto declares that she has no conflict of interest. A.C.A. Veloso declares that she has no conflict of interest. A.M. Peres declares that he has no conflict of interest.

References

1. C. Krantz-Rülcker, M. Stenberg, F. Winquist, I. Lundström, Anal. Chim. Acta **426**, 217 (2001)
2. L. Bityukova, V. Petersell, J. Geochem. Explor. **107**, 238 (2010)
3. W. Bertoldi, A.M. Gurnell, N.A. Drake, Water Resour. Res. **47**, W06525 (2011)
4. DL 156/98 of 6 June 1998, Diário da República - I Série-A, No. 131, 2593
5. Council Directive 80/777/EEC of 15 July 1980 on the approximation of the laws of the Member States relating to the exploitation and marketing of natural mineral waters. Official Journal of the European Communities, No L229/1 (http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:31980L0777)
6. Council Directive 96/70/EC of 28 October 1996 amending Council Directive 80/777/EEC on the approximation of the laws of the Member States relating to the exploitation and marketing of natural mineral waters. Official Journal of the European Communities, No L299/26 (http://eur-lex.europa.eu/legal-content/PT/TXT/?uri=CELEX:31996L0070)
7. F. Winquist, J. Olsson, M. Eriksson, Anal. Chim. Acta **683**, 192 (2011)
8. Portaria n° 703/96 of 6 December 1996, Diário da República - I Série-B, No. 282, 4387
9. Portaria n° 1296/2008 of 11 November 2008, Diário da República, 1.ª série, No. 219, 7870
10. P.K. Kundu, P.C. Panchariya, M. Kundu, ISA Trans. **50**, 487 (2011)
11. P. Ciosek, W. Wró blewski, Analyst **132**, 963 (2007)
12. B. Iliev, M. Lindquist, L. Robertsson, P. Wide, Fuzzy Set. Syst. **157**, 1155 (2006)
13. G.S. Braga, L.G. Paterno, F.J. Fonseca, AIP Conf. Proc. **1137**, 504 (2009)
14. G.S. Braga, L.G. Paterno, F.J. Fonseca, Sens. Actuat. B **171–172**, 181 (2012)
15. B. Adhikari, M. Mahato, T. Sinha, A. Halder, N. Bhattacharya, IEEE Sens. Proc. (2013)
16. M.Y. Vagin, F. Winquist, in *High Throughput Screening for Food Safety Assessment—Biosensor Technologies, Hyperspectral Imaging and Practical Applications*, ed. by A.K. Bhunia, M.S. Kim, C.R. Taitt (Elsevier, Cambridge, 2015), p. 265
17. I. Campos, L. Pascual, J. Soto, L. Gil-Sánchez, R. Martínez-Máñez, Sensors **13**, 14064 (2013)
18. L. Nuñez, X. Cetó, M.I. Pividori, M.V.B. Zanoni, M. del Valle, Microchem. J. **110**, 273 (2013)
19. O.A. Zadorozhnaya, D.O. Kirsanov, YuG Vlasov, V.D. Tonkopii, V.N. Rybakin, A.O. Zagrebin, A.V. Legin, Russ. J. Appl. Chem. **87**, 412 (2014)
20. S. Cavanillas, F. Winquist, M. Eriksson, Anal. Chim. Acta **859**, 29 (2015)
21. E.W. Nery, J.A. Guimarães, L.T. Kubota, Electroanalysis **27**, 2357 (2015)
22. D. Kirsanov, E. Legin, A. Zagrebin, N. Ignatieva, V. Rybakin, A. Legin, Anal. Chim. Acta **824**, 64 (2014)
23. J.- Gallardo, S. Alegret, M. del Valle, Talanta **66**, 1303 (2005)
24. R. Martínez-Máñez, J. Soto, E. Garcia-Breijo, L. Gil, J. Ibáñez, E. Llobet, Sens. Actuat. B **104**, 302 (2005)
25. L. Moreno, A. Merlos, N. Abramova, C. Jiménez, A. Bratov, Sens. Actuat. B **116**, 130 (2006)
26. K. Sghaier, H. Barhoumi, A. Maaref, M. Siadat, N. Jaffrezic-Renault, Sens. Lett. **7**, 683 (2009)
27. E. Garcia-Breijo, J. Atkinson, L. Gil-Sanchez, R. Matsot, J. Ibanez, J. Garrigues, M. Glanc, N. Laguarda-Miro, C. Olguin, Sens. Actuat. A **172**, 570 (2011)
28. R. Kumar, A.P. Bhondekar, R. Kaur, S. Vig, A. Sharma, P. Kapur, Sens. Actuat. B **171–172**, 1046 (2012)
29. P.K. Kundu, M. Kundu, J. Chemometr. **27**, 379 (2013)
30. H. Men, P. Zhang, C. Zhang, R. Wen, Z. Ge, J. Comput. **6**, 2692 (2011)
31. L. Sipos, Z. Kovács, V. Sági-Kiss, T. Csiki, Z. Kókai, A. Fekete, K. Héberger, Food Chem. **135**, 2947 (2012)
32. L. Sipos, A. Gere, D. Szöllosi, Z. Kovács, Z. Kókai, A. Fekete, J. Food Sci. **78**, S1602 (2013)
33. J.E. Oliveira, V. Grassi, V.P. Scagion, L.H.C. Mattoso, G.M. Glenn, E.S. Medeiros, IEEE Sens. J. **13**, 759 (2013)
34. A.M. Peres, L.G. Dias, T.P. Barcelos, J. Sá Morais, A.A.S.C. Machado, Procedia Chem. **1**, 1023 (2009)
35. L.G. Dias, A.M. Peres, T.P. Barcelos, J. Sá Morais, A.A.S.C. Machado, Sens. Actuat. B **154**, 111 (2011)
36. L.G. Dias, C. Sequeira, A.C.A. Veloso, M.E.B.C. Sousa, A.M. Peres, Anal. Chim. Acta **848**, 32 (2014)
37. DL n° 306/2007 of 27 August 2007, Diário da República, 1.ª série, 5747
38. L.G. Dias, C. Sequeira, A.C.A. Veloso, J. Sá Morais, M.E.B.C. Sousa, A.M. Peres, Chromatography **1**, 141 (2014)
39. M.E.B.C. Sousa, L.G. Dias, A.C.A. Veloso, L. Estevinho, A.M. Peres, A.A.S.C. Machado, Talanta **128**, 284 (2014)
40. L.G. Dias, A.M. Peres, A.C.A. Veloso, F.S. Reis, M. Vilas Boas, A.A.S.C. Machado, Sens. Actuat. B **136**, 209 (2009)
41. Y. Kobayashi, M. Habara, H. Ikezazki, R. Chen, Y. Naito, K. Toko, Sensors **10**, 3411 (2010)
42. J. Miller, J.C. Miller, *Statistics and Chemometrics for Analytical Chemistry*, 6th edn. (Prentice Hall, Harlow, 2010), pp. 231–235
43. A.J. Izenman, *Modern Multivariate Statistical Techniques: Regression, Classification, and Manifold Learning*, 2nd edn. (Springer, New York, 2008), pp. 107–122
44. J. Cadima, J.O. Cerdeira, M. Minhoto, Comput. Stat. Data Anal. **47**, 225 (2004)
45. S. Kirkpatrick, C.D. Gelatt, M.P. Vecchi, Science **220**, 671 (1983)
46. D. Bertsimas, J. Tsitsiklis, Stat. Sci. **8**, 10 (1993)
47. M. Kuhn, K. Johnson, *Applied Predictive Modeling* (Springer, New York, 2013)
48. W.N. Venables, B.D. Ripley, *Modern Applied Statistics with S (Statistics and Computing)*, 4th edn. (Springer, New York, 2002)
49. A.C.A. Veloso, L.G. Dias, N. Rodrigues, J.A. Pereira, A.M. Peres, Talanta **146**, 585 (2015)