Talanta 162 (2017) 1–9

Contents lists available at ScienceDirect

Talanta

journal homepage: www.elsevier.com/locate/talanta

Predicting consumer preferences for mineral composition of bottled and tap water

Stefan Platikanov[a,c], Alejandra Hernández[b], Susana González[c], Jose Luis Cortina[c], Roma Tauler[a], Ricard Devesa[b,*]

[a] Department of Environmental Chemistry, IDAEA-CSIC, Jordi Girona, 18-26, 08026 Barcelona, Spain
[b] Aigües de Barcelona (AGBAR) Laboratory, General Batet, 5-7, 08028 Barcelona, Spain
[c] CETAQUA Water Technological Center, Ctra. Esplugues, 75, 08940, Cornellà de Llobregat, Barcelona, Spain

ARTICLE INFO

Keywords:
Taste
Consumer preference
Bottled and Tap
Chemometrics

ABSTRACT

The overall liking for taste of water was correlated with the mineral composition of selected bottled and tap waters. Sixty-nine untrained volunteers assessed and rated twenty-five different commercial bottled and tap waters from. Water samples were physicochemical characterised by analysing conductivity, pH, total dissolved solids (TDS) and major anions and cations: HCO_3^-, SO_4^{2-}, Cl^-, NO_3^-, Ca^{2+}, Mg^{2+}, Na^+, and K^+. Residual chlorine levels were also analysed in the tap water samples. Globally, volunteers preferred waters rich in calcium bicarbonate and sulfate, rather than in sodium chloride. This study also demonstrated that it was possible to accurately predict the overall liking by a Partial Least Squares regression using either all measured physicochemical parameters or a reduced number of them. These results were in agreement with previously published results using trained panellists.

1. Introduction

The topic of the taste has a long tradition in the potable water sector. A new specialised group called “Tastes, Odours, and Algal Toxins in Drinking Water Resources and Aquaculture” has been recently established [1] at the International Water Association (IWA) following the mission developed by the Off Flavours in the Aquatic Environment group since 1980. The water sector has its own sensory wheel and a flavour profile analysis method [2,3]. Interest in the taste of both tap and bottled water has been increasing significantly in recent years, and different methodologies have been used for its characterisation [4]. Water utilities have implemented new advanced treatments, particularly membranes, to expand water resources and also to improve the general quality, particularly the taste, of distributed water. The utilities have also focused on other relevant issues, such as remineralisation of membrane-treated waters [5], minimisation of disinfectant residuals and evolution of the water in the distribution system [6]. The bottled water sector has promoted its product based on purity and health considerations.

The total dissolved solids (TDS) has become the most common parameter in water taste studies [7–11]. International regulations and recommendations establish maximum levels for TDS with important differences: 500 mg/L in the USA [12] and Canada [13], 1000 mg/L in the WHO Guidelines [14], and 1600 mg/L in Europe, which corresponds to a 2500 μS/cm conductivity at 25 °C [15]. Several studies have shown that high levels of minerals are not well liked [16,17]. Other papers have focused on the role of specific salts [5,7,17,18]. These studies have demonstrated that the interaction between cations and anions introduces great complexity to the subject. A critical discussion about the USEPA standards that influence the taste of water has been recently published [19].

Some papers have been published specifically for bottled water [20].

In recent years, the water supply company for the metropolitan area of Barcelona (3 million inhabitants), Aigües de Barcelona, has performed several studies on the influence of minerals on the taste of tap water because one of its main sources of freshwater, the Llobregat river, presents high levels of minerals due to the mining area in the upper part of its basin. Reducing the mineral contents via reverse osmosis (RO) and electrodialysis reversal (EDR) has been demonstrated to improve the drinking water [8,10,21].

Multivariate data analysis provides useful tools and techniques to explore and model panellists’ taste and preference data related to physicochemical properties of different waters [22].

Principal Component Analysis (PCA) is a commonly used technique in sensory analysis [23]. When data contains hidden information and

* Corresponding author.
E-mail addresses: stefan.platikanov@idaea.csic.es (S. Platikanov), rdevesa@aiguesdebarcelona.cat (R. Devesa).

http://dx.doi.org/10.1016/j.talanta.2016.09.057
Received 17 June 2016; Received in revised form 16 September 2016; Accepted 23 September 2016
Available online 24 September 2016
0039-9140/ © 2016 Elsevier B.V. All rights reserved.

consists of multicorrelated, physicochemical parameters as variables, PCA may prove useful to reduce data complexity into a small number of orthogonal variables (principal components) enabling a better interpretation [24].

Multiple factor analysis (MFA) is similar method to PCA [25] that can handle concatenated multiple data sets, e.g., when two blocks of variables share similar information, such as physicochemical parameters and volunteers' ratings for a common set of water samples. The advantage of MFA is that it also reduces the predominant influence of the data block with the larger number of variables. In this study, the variance of the block containing the 69 individual volunteers' ratings as variables would dominate over the variance of the data block with only 14 variables, i.e., the physicochemical parameters.

Moreover, the relationship between the overall taste liking (i.e., the average from the 69 individual ratings) and the fourteen physicochemical parameters can be modelled using a Partial Least Squares regression, PLS [26,27]. In addition to its good predictive ability in sensory studies [28], PLS has also been shown to provide useful tools to analyse, visualise and interpret the physicochemical parameters' contribution and importance to the taste of water samples [17].

The aim of this study is to investigate the most influential physicochemical parameters from selected mineral and tap waters associated with the global (overall) and individual liking of untrained volunteers. The results will be compared to previously published results using trained panellists for the experiment. This paper also aims to develop a predictive regression model for the global liking based on the physicochemical composition.

2. Materials and methods

2.1. Water samples

Twenty-five bottled and tap waters were chosen to cover a wide range of total dissolved solids (TDS) and physicochemical compositions, chiefly cations and anions. Thus, the sample selection included waters with very different characteristics to explain precisely how and to what extent the physicochemical composition influences the liking of the volunteers.

Bottled mineral water samples were selected from the most common brands available in Spain and France. Tap waters were collected from various locations in Catalonia, Spain. Some water samples were mixtures of bottled and tap water, or they were diluted with deionised Milli-Q water or spiked with various amounts of NaCl. Overall, the study included 25 samples, 13 bottled mineral waters and 12 tap waters.

The results of the physicochemical analysis of all the water samples used in this work are presented in Table 1. The analysis was conducted at the Aigües de Barcelona laboratory. Water blends and dilutions were allowed to equilibrate for 48 h before analysis. Sodium, potassium, calcium, magnesium and silica levels were analysed by inductively coupled plasma optical emission spectrometry (ICP-OES) (Perkin Elmer Optima 4300 DV, Wellesley, MA, USA). The conductivity at 20 °C, pH and bicarbonate levels were determined by a robotic titrosampler (Metröhm modules 855 and 856). Chloride, nitrate and sulfate concentrations were analysed by ion chromatography (Dionex ICS-2000). The TDS (dry residue at 180 °C) levels were measured by gravimetry. Free residual chlorine was analysed by the classical DPD colourimetric method. All these analytical determinations are ISO17025-certified by ENAC, the accreditation body in Spain, except for the bicarbonates and TDS, which are not included in the EC Drinking Water Directive (98/83/EC).

A single batch was used for each of the bottled waters tested. Bottles were stored at room temperature and in a dark place free of odours. Tap water samples were taken from established sampling points in the routine monitoring of supply systems. Samples were taken in glass amber bottles, previously cleaned with a procedure appropriate for sensory analysis. Bottles were refrigerated until several hours before the tasting sessions. Tap water samples were not de-chlorinated, and no other treatment was applied.

All water samples were assigned a label according to the major ions. Ten different labels were established.

- Low mineralised waters, LM (TDS < 60 mg/L)
- Sodium, bicarbonate type, Na/HCO_3
- Calcium, bicarbonate type, Ca/HCO_3
- Calcium, bicarbonate and sulfate type, $Ca/HCO_3/SO_4$
- Sodium, calcium, magnesium, bicarbonate, sulfate and chloride type
- $Na/Ca/Mg/HCO_3/SO_4/Cl$
- Sodium, calcium, bicarbonate and chloride type, $Na/Ca/HCO_3/Cl$
- Sodium, calcium and bicarbonate type, $Na/Ca/HCO_3$
- Sodium and chloride type, Na/Cl
- Sodium, chloride and bicarbonate type, $Na/Cl/HCO_3$
- High mineralisation, HM (TDS > 750 mg/L)

2.2. Sensory analysis

The sensory analysis was conducted with 69 untrained volunteers, 29 of whom were male, and 40 of whom were female, and the volunteers were between 23 and 55 years of age. A rating test was used to assess the overall liking of the water samples. Volunteers were asked to rate the taste of water samples on a 0–10 scale (0 corresponding to a very bad flavour and 10 to an excellent one). The volunteers were not allowed to eat, drink (except water) or smoke for one hour before the tasting. The tasting took place in a specifically designed room, comfortable and free from interfering odours. Water samples were kept at 25° C prior the tasting and presented in codified plastic glasses.

The study was composed of three rounds. All 25 samples were tasted in each round, and each round included five tasting sessions (one per week). Thus, each volunteer tasted each water three times. When the experiment was designed, the characteristics of the five waters presented in every session were balanced with respect to their origin (tap or bottled) and mineral level. The order of presentation of samples was also balanced. After the initial tasting of five waters, panellists were allowed to taste the waters again in any order. A commercial mineral water with TDS=132 mg/L previously defined as "neutral" by the trained panel of the water supply company [29] was offered to the participants as an optional blank sample to rinse their mouth at the beginning of the sessions or between samples.

To assess the reproducibility of the volunteers' rating (separately from the main study), 17 volunteers were asked to taste fifteen waters twice (two replicates per sample) with one hour between replications.

Three data sets were used in the chemometric analysis:

- The data set with the 13 physicochemical parameters [X0 matrix with dimensions (25, 13)] alone and with the global mean liking scores vector (y) of all volunteers for all 25 water samples [X1 (25,14) matrix]
- The data set with the 13 physicochemical parameters plus the 69 individual volunteers' mean liking vectors for all 25 water samples [X2 (25,82) matrix]
- The data set with the 17 individual volunteers liking vectors for 15 water samples in two replicates [X3 (15,34) matrix]

2.3. Chemometrics

Preliminary exploratory data analysis was performed using univariate descriptive statistics. For each variable, the average, minimum and maximum values were calculated (please see Table 1). Pair-wise correlations among all the variables in matrix $\mathbf{X_1}$ were also evaluated and are visualised in Fig. 1..

Analysis of variance for one factor (one-way ANOVA) and for two

Table 1
Identification and mineral contents of the bottled and tap water samples.

samples[a]	Water Lables[b]	Conductivity (μS/cm)	TDS (mg/L)	Cl (mg/L)	SO4 (mg/L)	NO3 (mg/L)	HCO3 (mg/L)	Ca (mg/L)	Mg (mg/L)	Na (mg/L)	K (mg/L)	pH (mg/L)	Si (mg/L)	Cl2 (mg/L)	Mean liking taste score
1 bot	LM(26)[c]	38.0	26.0	0.7	1.3	3.5	18.0	5.0	1.0	1.3	0.3	6.7	8.9	0.00[d]	7.0
2 bot	Na/HCO3(193)	278.0	193.0	17.0	11.6	0.4	145.4	12.4	0.8	51.5	1.5	7.3	16.0	0.00	7.1
3 bot	Na/HCO3(347)	539.0	347.0	27.4	21.3	0.7	285.1	22.8	1.6	87.7	4.0	7.9	32.2	0.00	5.8
4 bot	Ca/Mg/HCO3(262)	375.0	262.0	7.8	21.9	1.9	285.0	56.9	25.5	5.3	1.1	8.0	7.5	0.00	6.7
5 bot	Ca/Mg/HCO3(131)	201.0	131.0	4.2	12.2	0.9	124.2	21.3	12.6	2.7	0.5	8.1	5.0	0.00	7.2
6 bot	Ca/HCO3/SO4(124)	188.3	124.3	1.5	41.0	0.0	88.2	36.0	7.7	2.7	1.7	7.7	1.4	0.00	7.4
7 bot	Ca/HCO3/SO4(373)	565.0	372.9	5.0	123.0	0.0	264.7	108.0	23.0	8.0	5.0	7.7	4.2	0.00	6.6
8 bot	Na/Ca/Mg/HCO3SO4/Cl(248)	375.2	247.6	52.8	77.4	12.4	77.8	42.2	18.0	17.6	2.2	7.7	0.9	0.00	6.9
9 bot	Na/HCO3/Cl(195)	289.0	195.0	49.6	6.5	1.8	85.6	9.6	1.0	52.8	0.9	7.2	13.0	0.00	7.1
10 bot	Na/HCO3/Cl(373)	591.0	373.0	112.2	12.1	2.5	164.4	12.2	1.3	116.0	1.7	7.2	20.8	0.00	5.4
11 bot	Na/Cl(165)	274.0	165.0	74.3	1.2	3.5	25.9	6.3	1.0	121.6	0.4	6.9	10.9	0.00	6.8
12 bot	Na/Cl(389)	635.0	389.0	193.9	1.5	3.9	20.8	6.2	0.9	31.3	0.4	7.0	9.6	0.00	5.2
13 bot	HM(1238)	1876.0	1238.2	264.0	387.0	62.0	389.2	211.0	90.0	88.0	11.0	7.7	4.6	0.00	4.0
14 tap	LM(54)	82.3	54.3	20.5	9.0	1.9	28.2	7.3	1.0	8.0	2.5	7.5	3.4	0.48	5.8
15 tap	Na/Ca/HCO3(228)	345.7	228.1	22.9	38.3	4.1	147.6	50.0	9.1	14.3	2.5	7.8	1.9	0.82	5.4
16 tap	Ca/HCO3(291)	440.7	290.8	14.0	23.7	11.9	278.6	88.3	7.2	11.0	2.5	7.8	6.2	0.45	5.4
17 tap	Ca/Mg/HCO3/SO4(693)	1050.7	693.4	18.4	387.0	14.1	305.8	179.3	51.0	10.3	2.5	7.9	4.4	0.57	4.0
18 tap	Na/Ca/HCO3/Cl(742)	1125.7	742.9	169.0	121.0	10.5	309.9	112.3	25.7	109.7	15.7	7.6	2.4	0.58	3.0
19 tap	Na/Ca/HCO3/Cl(172)	260.0	171.6	20.6	29.0	2.6	104.3	34.3	6.3	13.0	2.5	7.8	1.2	0.51	6.0
20 tap	Na/Ca/HCO3/Cl(600)	909.0	599.9	104.7	144.3	2.8	263.9	92.7	25.7	72.7	15.7	7.8	3.5	0.39	4.4
21 tap	Na/Cl(72)	110.3	72.8	30.4	6.0	3.3	20.7	2.5	0.5	18.7	2.5	6.2	2.5	0.50	5.6
22 tap	HM(765)[c]	1158.7	764.7	212.3	130.3	4.9	222.9	100.7	24.0	119.3	25.3	8.0	1.0	0.08	3.9
23 tap	HM(911)	1380.3	911.0	235.3	171.0	6.8	299.7	111.7	35.3	141.0	29.3	7.8	3.3	0.55	2.5
24 tap	HM(1799)	2725.3	1798.7	539.3	501.7	8.9	360.7	193.0	73.3	369.7	15.3	7.7	4.2	0.74	1.6
25 tap	HM(1595)	2417.0	1595.2	403.3	585.7	25.8	316.4	245.3	76.7	243.0	10.7	7.5	5.2	1.01	1.8

[a] Thirteen bottled water samples and twelve tap water samples.
[b] Water type was identified by its most characteristic mineral content.
[c] Labels LM and HM identify bottled or tap water samples with very low or very high concentrations of TDS. In brackets is presented the concentration of TDS.
[d] Chlorine concentrations below the detection limit (< 0.15 mg/L).

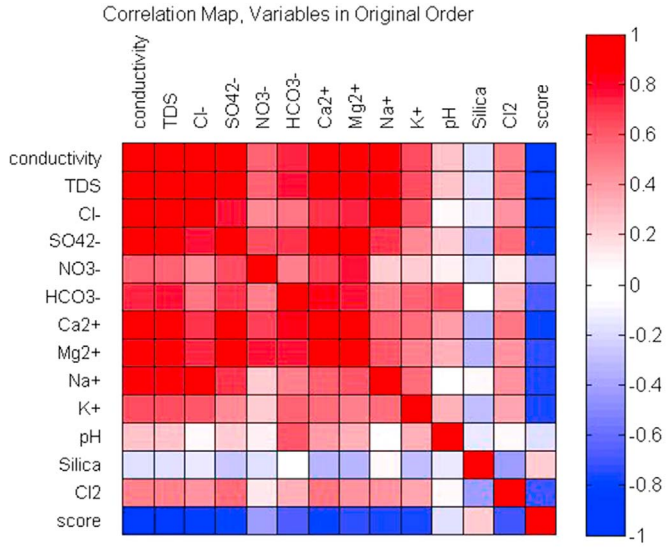


Fig. 1. Pair-wise correlation coefficient map between the global mean liking vector (i.e., the score) and thirteen physicochemical parameters for the 13 bottled and 12 tap water samples. More intense red and blue colours indicate higher positive or negative correlations, respectively.

factors (two-way ANOVA) were used to test the following questions:

- Whether one volunteer differs significantly within his evaluations for the 15 water samples during two repetitions on the same day.
- Whether the physicochemical content of the different water samples affected the volunteers' evaluations.
- Whether volunteers (among them) differ in their ratings for water samples.
- Whether a significant synergistic effect exists between the physicochemical content of waters and the volunteers' ratings.

2.3.1. Principle Component Analysis

Principal Component Analysis, PCA, [24] is probably the most widely used multivariate method for analysing environmental data sets. PCA is a procedure that examines the relationships among a set of correlated variables.

Using either a singular value decomposition of original data or an eigenvector decomposition of the covariance matrix, PCA transforms a number of correlated variables into a smaller number of uncorrelated, orthogonal variables called principal components (PCs) that adequately explain the data structure. Essentially, PCs are found by linear combination of the original variables. In addition, PCA discovers which samples are "similar to" or "different from" other samples and which variables contribute most to these differences. The obtained results can be visualised graphically by projecting the samples (scores) or the original variables (loadings) onto the space defined by the selected PCs. The first PC explains the maximum variance in the data, and the orthogonal PCs explain uncorrelated variance in decreasing order. More details about PCA are widely described in the literature [24]. Because the measured variables were on different scales and had different units, they were autoscaled to give them the same importance in the data analysis.

2.3.2. Multiple factor analysis

MFA is a suitable method to analyse several blocks of variables observed for the same samples [30]. In MFA, a similar method to PCA, the variables of different data blocks are weighted adequately to correct for differences in size. This method guarantees that all variables have a similar effect on the first principal component in the global solution of the MFA [31], which avoids the predominance of particular groups of variables. Thanks to its specific weighting scheme, MFA emphasises the similarity of the information in different groups of variables and is especially useful in data fusion procedures [25].

2.3.3. Partial least square regression

The partial least squares regression method [26] models systems using a series of predictors (in this case, the physicochemical parameters from the $\mathbf{X_O}$ matrix) to correlate them with a response variable (in this case, the global volunteers mean liking score vector $\mathbf{y}$). This factor analysis method transforms the original variables into a smaller number of new variables called latent or hidden variables, LVs. Both $\mathbf{X}$ and $\mathbf{y}$ are simultaneously decomposed to maximise the covariance. A new matrix of weights (reflecting the covariance structure) is calculated to determine the factor space most congruent to both $\mathbf{X}$ and $\mathbf{y}$. This matrix expresses how the information in each X variable relates to the variation of $\mathbf{y}$. Thus, variables with large weight value for a specific LV are considered important for the prediction of $\mathbf{y}$.

The use of Variable Importance in Projection (VIP) scores has been shown to be useful for interpreting the more relevant variables in PLS models [32]. VIP scores summarise the variance in the predictor variables that globally contributes most to the $\mathbf{y}$ variance explanation. VIP scores are then compared visually, and the largest ones are considered to indicate the most important variables in the PLS model.

The number of LVs in this study was optimised via internal full cross-validation. One sample was removed from the calibration set at a time, and the value for that sample was predicted by the model using the rest of the included samples. The number of PCs was selected by adopting the minimum prediction error in the full cross-validation.

The quality of the prediction models was assessed by evaluating the goodness of the fit for predicted $\mathbf{y}$ values vs. observed $\mathbf{y}$ scores during the calibration and internal validation. The figures of merit considered in this study were the coefficient of determination of model fitting, or R^2, and the Root Mean Squared Error of Calibration and of Cross-Validation (RMSEC and RMSECV). The relative prediction error of the volunteers' global mean liking score, which is given as a percentage for the calibration and for the internal validation, was calculated by Eq. (1) as follows:

$$Rel.\ error\ in\ \% = \sqrt{\frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{\sum_{i=1}^{n}(y_i)^2}} \times 100 \quad (1)$$

All calculations for two-way ANOVA, PCA, and PLS were performed using PLS Toolbox 7.8 (Eigenvector Research, Manson, WA, USA) and MATLAB 8.1 with the Statistical Toolbox (MathWorks Inc., Natick, MA, USA).

3. Results and discussion

3.1. Descriptive statistics

Fig. 1 presents the colour map of the pair-wise correlation analysis for the physicochemical parameters and the volunteers' global mean liking (matrix $\mathbf{X_1}$) vector of scores.

Many of the parameters gave high positive (intense red colours) pair-wise correlations. The physicochemical parameters in this large cluster of positive correlations were conductivity, TDS, and the concentrations of chlorides, sulfates, nitrates, bicarbonates, calcium, magnesium, sodium and potassium. Silicon and residual chlorine presented a strong negative pair-wise correlation, given that they are characteristic parameters of mineral and tap water, respectively.

It was also revealed that the global mean liking (the *score* variable) of the volunteers correlated negatively with most of the physicochemical parameters. Fig. 1 shows that Si was positively correlated to the global mean liking, while it was negatively correlated to the other parameters. The pH was found to be negatively correlated (although not very strongly) to the global mean liking, but it exhibited positive correlation with many of the physicochemical parameters.

Table 2
Results from One-way and Two-way ANOVA.

Analysis of variance for one factor (one-way ANOVA)					
Source of variance	SS[a]	df	MS	F	Prob > F
volunteers	0.3	1	0.3	0.03	0.861
Error	269.867	28	9.6381		
Total	270.167	29			
Analysis of variance for two factors (two-way ANOVA)					
Source of variance	SS[a]	df	MS	F	Prob > F
water type	1884.59	14	134.613	80.46	< 0.001
volunteers	245.73	16	15.358	9.18	< 0.001
interaction	1179.03	224	5.264	3.15	< 0.001
Error	426.63	255	1.673		
Total	3735.97	509			

[a] SS –sum of squares; df – degree of freedom; MS – mean square; F- F-statistic; Prob > F – p-value associated with the F-statistic.

The pair-wise correlation map suggested that volunteers did not like waters with a high TDS level. The highly correlated structure of the analysed data also suggested that projection-based methods for multivariate data analysis such as PCA or PLS can be useful to handle the multivariate correlation present in this data set.

3.2. Analysis of variance

Table 2 presents the results of one- and two-way ANOVA for the analysed datasets.

First, differences between the evaluations for two repetitions of a volunteer in one day were statistically tested to determine the reliability of the experiment (matrix $\mathbf{X_3}$). Because two volunteers were missing in one of the two repetitions, they were omitted from this analysis, and the data included fifteen volunteers. The results of the one-way ANOVA revealed (Table 2, upper section) that there were no significant differences between the overall evaluations among the two repetitions per day for each of the fifteen volunteers. The volunteers were rather consistent in their overall evaluation in the two repetitions, and the data could be considered reliable and representative for further PCA and PLS modelling.

The two-way ANOVA results revealed that the water type (i.e., the differences in mineral composition) and the average evaluations of the 17 volunteers (individually) differed statistically at a significance level of 0.001 (Table 2, bottom section). In other words, the type of water had a strong effect on the volunteers' preference. The personal factor was also statistically significant, meaning that different volunteers showed different preferences for various waters. Finally, the interaction between these two factors was also found to be significant, which highlights the complexity of sensory data and the usefulness of chemometrics tools to reveal hidden information.

3.3. Principal component analysis of the $\mathbf{X_1}$ matrix containing 13 physicochemical parameters and the global mean liking vector of volunteers for all 25 water samples

First, data $\mathbf{X_1}$ were analysed via PCA to investigate the distribution of selected water samples and to explore the relationships between the measured physicochemical parameters. PCA on the autoscaled data revealed a model with four principal components (PCs) with eigenvalues larger than 1, explaining as much as 89% of the data variance (information). The percentage distribution of the captured variance among PCs was as follows: PC1 62.5% > PC2 10.2% > PC3 9.1% > PC4 7.8%. Therefore, it was possible to conclude that there was a dominant factor in the data variance (PC1), plus up to three minor factors that needed further investigation for meaningful results.

The variance captured by the PCA model for each original variable is shown in Fig. 2. The stacked bar chart shows that the PCA model captures from ~72% to ~98% of the variance in all the variables. The plot also shows how each PC captures the variance..

For example, PC1 (dark blue) captures most of the (average) variation in all parameters, except pH, Si and chlorine. In other words, PC1 explains variance in the global mineralisation (conductivity and TDS) and the major ions: Cl^-, SO_4^{2-}, HCO_3^-, Ca^{2+}, Mg^{2+}, Na^+ and K^+. Additionally, the largest part of the explained variance in the volunteers' evaluations is captured by PC1.

PC2 (light blue) captures most of the variation in pH and a significant part of the variation in Na^+, Ca^{2+}, Si, Cl^- and HCO_3^-. This result can be attributed to the influence of typical water properties such as alkalinity (pH and HCO_3^-), hardness (Ca^{2+}) and salinity (Na^+ and Cl^-).

PC3 (yellow) captures the largest part of the variance in Si and residual chlorine. Both parameters are characteristic of the waters' origin, i.e., bottled mineral (Si) or tap water (Cl_2).

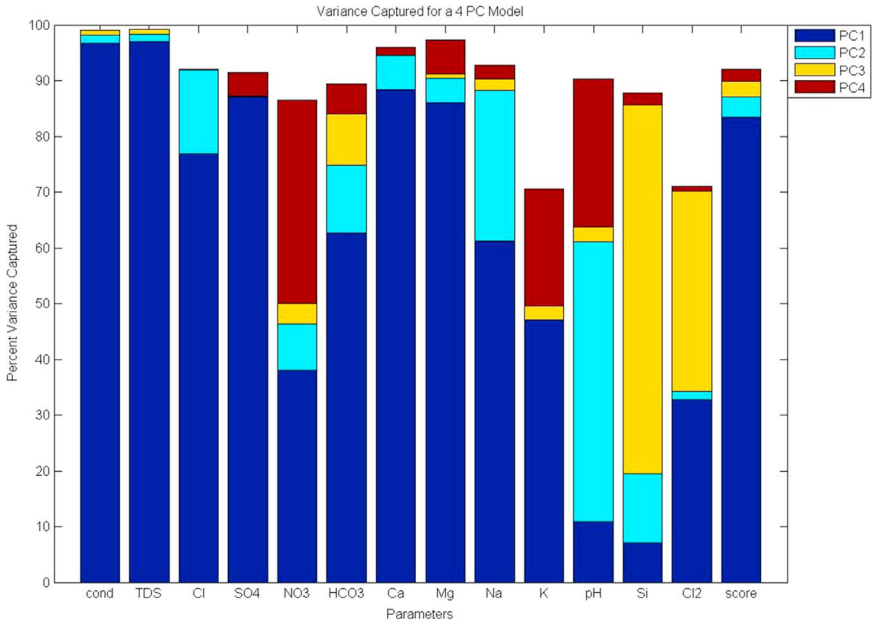


Fig. 2. Variance captured for each variable in a PCA model with four principal components.

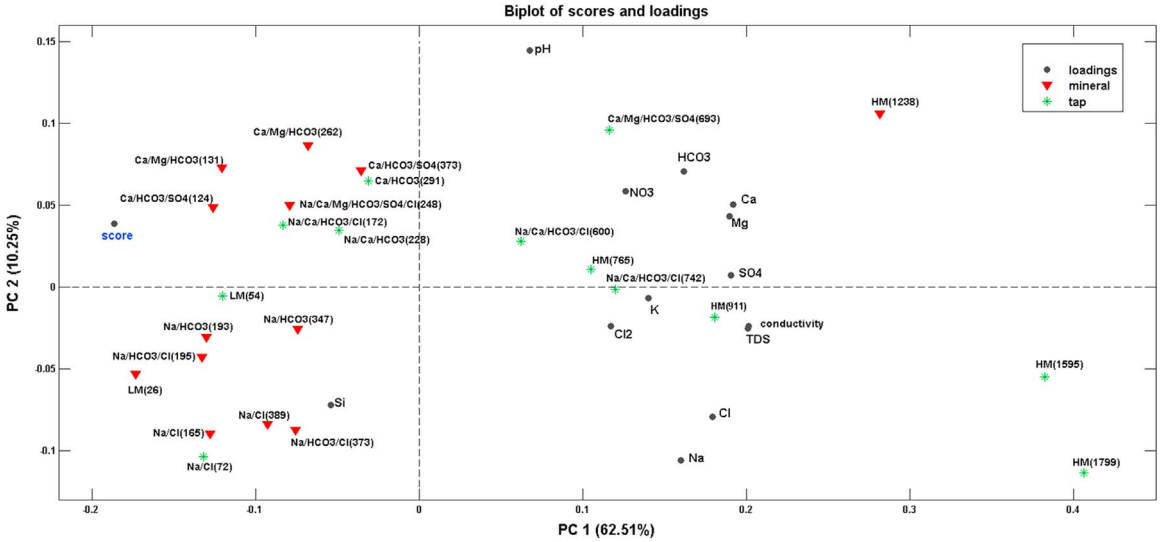


Fig. 3. PCA results of the autoscaled X_1 data matrix: PC1 vs. PC2 biplot showing bottled (red triangles) and tap waters (green asterisks) scores, as well as the loadings of the physicochemical parameters (black dots) and the volunteers' global mean liking (labelled in blue). For the definitions of the water labels, please refer to Table 1.

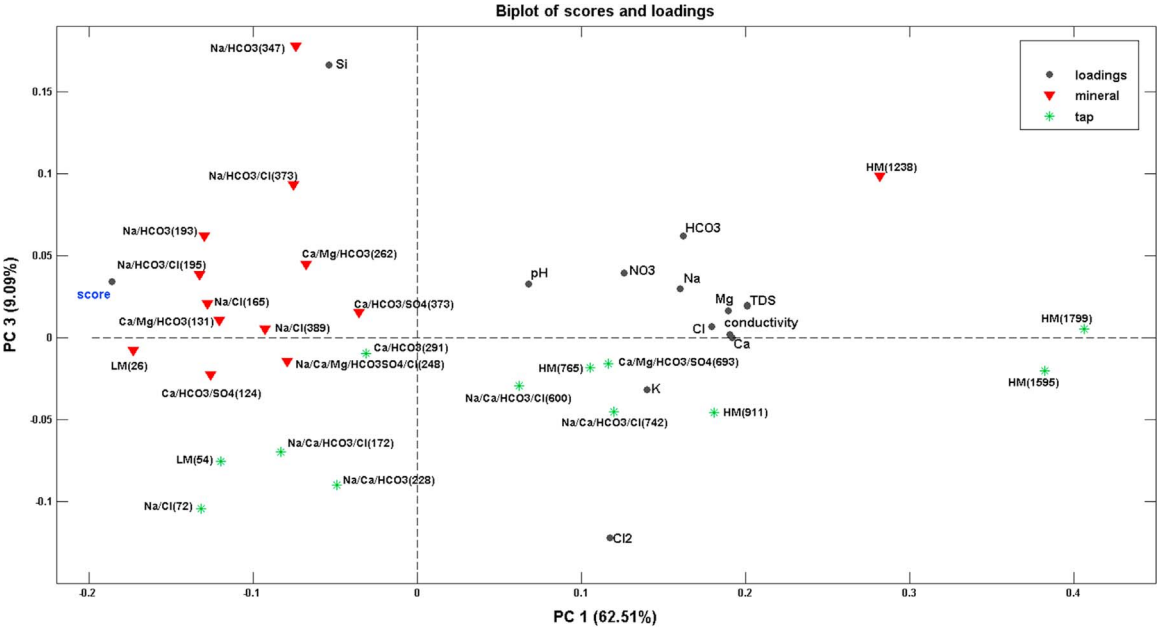


Fig. 4. PCA results of the autoscaled X_1 data matrix: PC1 vs. PC3 biplot showing bottled (red triangles) and tap (green asterisks) water scores, as well as the loadings of the physicochemical parameters (black dots) and the volunteers' global mean liking (labelled in blue). For the definitions of the water labels, please refer to Table 1.

Finally, PC4 (dark brown) captures variation in NO_3^-, K^+ and pH. It was not possible to assign a characteristic pattern for this PC because NO_3^- and K^+ did not present significant concentrations to affect the taste [7]. Therefore, for the sake of the data interpretation below, the final number of retained PCs in the model was 3, and the total explained variance was equal to 82%.

Water samples (*scores*) together with the physiochemical parameters and the global mean liking (*loadings*) are projected onto the spaces spanned by the first two PCs and shown in the biplot of Fig. 3..

Water samples containing high mineralisation (right side of the plot) were differentiated from the water samples with low to medium TDS (left side) along PC1.

With respect to PC2, sample differentiation was attributed to the chemical composition of the waters. Those with typical high alkalinity (high values of pH and HCO_3^-) and strong hardness (Ca^{2+}) were distributed in the upper part of the plot, while waters with typically high salinity (high concentrations of Na^+ and Cl^-) were distributed at the bottom of the plot.

Almost all physicochemical parameters, including TDS, had large positive loadings on PC1. On the contrary, the loading of the mean liking (score label) was distributed in the negative part of PC1. Thus, it can be concluded that the volunteers liked water samples with overall mineralisation below the average of all water samples, i.e., a TDS of 480 mg/L (see Table 1).

PC2 had positive loadings from pH, Mg^{2+}, Ca^{2+}, HCO_3^-, SO_4^{2-} concentrations and the global mean liking score vector, while K^+, Si, Na^+ and Cl^- concentrations had negative loadings on PC2.

Thus, the volunteers' preference was for water samples containing high alkalinity and hardness with typical profiles of $CaSO_4$ and $Ca(HCO_3)_2/CaSO_4$, whereas the salty waters (NaCl and NaCl/$NaHCO_3$) were rated lower.

A different distribution of the water samples scores occurs on the PC1 vs. PC3 biplot (see Fig. 4). Whereas the effect of the total mineralisation factor on PC1 was already shown, two new clusters of

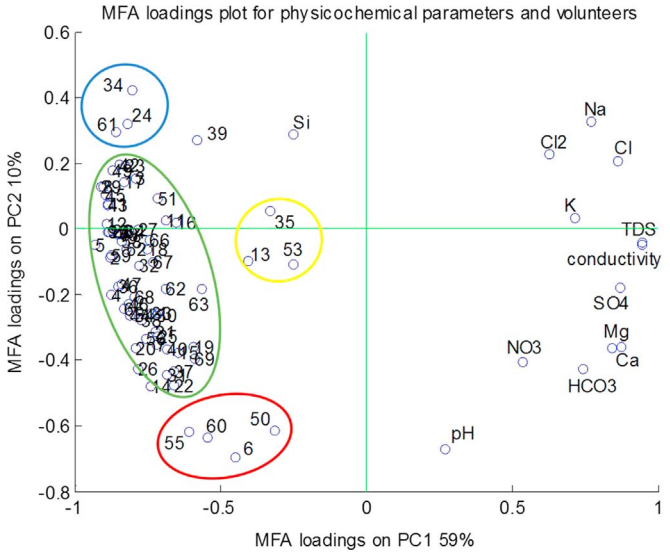


Fig. 5. MFA exploratory analysis of the thirteen physicochemical parameters and the sixty-nine individual volunteers' mean liking.

water sample scores are now observed on PC3. In The bottled mineral water samples (red triangles) with positive scores are in the upper part of the plot. In contrast, the tap waters (green asterisks) with negative scores are at the bottom. PC3 is determined mainly by two parameters: the concentration of residual chlorine (negative loading) and Si (positive loading). The volunteers' mean liking is distributed with a small positive loading on PC3 (i.e., positive but close to the 0-axis). Untrained panellists showed a light preference for bottled waters., They could probably detect the characteristic disinfectant flavour, and they may have used it to distinguish between the two types of water. However, French untrained tasters were not able to discriminate between tap and bottled water after chlorine was allowed to disappear [16]..

Note that the results of the present paper on the influence of the major dissolved species (sodium, calcium, bicarbonate, chloride and sulfate) and the role of residual chlorine are in agreement with the main conclusions of a previous work with trained panellists [17].

3.4. Multiple factor analysis of the X_2 matrix containing 13 physicochemical parameters and sixty-nine volunteers mean liking vectors for all 25 water samples

MFA was shown to be an appropriate technique to reveal common relationships between the two data blocks. MFA was applied to the data matrix $\mathbf{X_2}$, which included the autoscaled data sets in two blocks, the physicochemical parameters data all together and the 69 individual volunteers' mean liking score vectors.

The first two PCs explain 69% of the total data variance. Fig. 5 shows PC1 vs. PC2 loadings obtained in the MFA analysis of the $\mathbf{X_2}$ augmented data matrix. Along PC1, almost all volunteers' individual liking vectors (with positive loadings) were segregated strongly from the physiochemical parameters (with negative loadings). However, a small group of volunteers in the yellow circle (13, 35, and 53) liked waters with a high TDS. Thus, PC1 implied that the majority of the volunteers liked waters with medium to low level of overall mineralisation..

The segregation along PC2 corresponded to the differences in the chemical composition. Loadings of pH, Mg^{2+} Ca^{2+}, HCO_3^- and SO_4^{2-} were distributed on the negative side of PC2, while loadings of K^+, Si, Na^+ and Cl^- were on the positive side. Along PC2, there was an obvious differentiation among the volunteers' loadings. Up to four groups of volunteers can be distinguished according to their preferences for the chemical content of water samples. There were two groups of volunteers with strongly contrasting preferences for the water mineral content. The group including volunteers 6, 50, 55 and 60 (in the red circle) showed preferences for waters with higher concentrations of Mg^{2+} Ca^{2+}, HCO_3^- and SO_4^{2-} and higher pH. However, the volunteers in the blue circle (24, 34, and 61) preferred waters with low pH and high amounts of sodium and chloride. The largest group of volunteers (green ellipse) was distributed in the middle part on PC2, corresponding to a preference for intermediate waters.

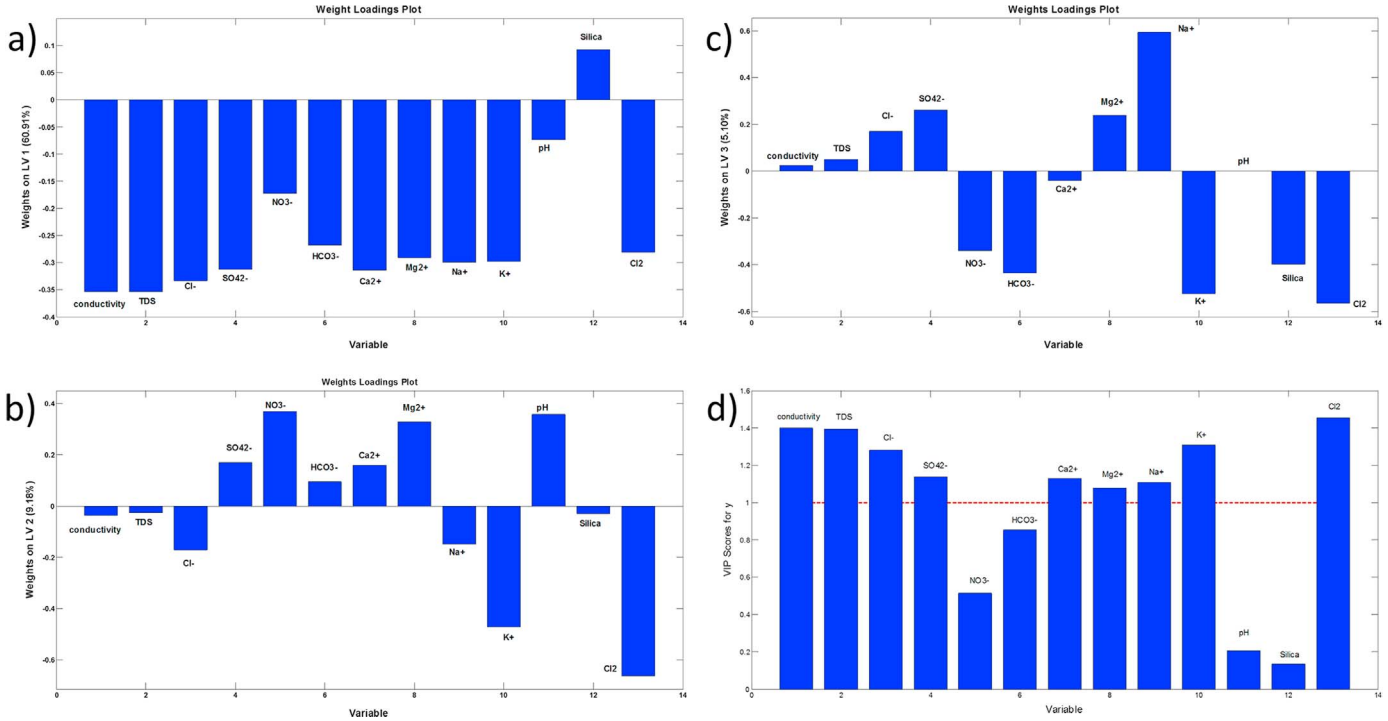


Fig. 6. PLS weight loadings and VIP scores of physicochemical parameters for the first three latent variables in the prediction of the volunteers' global mean liking: a) weights on LV1; b) weights on LV2; c) weights on LV3; d) VIP scores. Parameters with values above the threshold VIP score of one (red dotted line) were selected as the most significant for the PLS model.

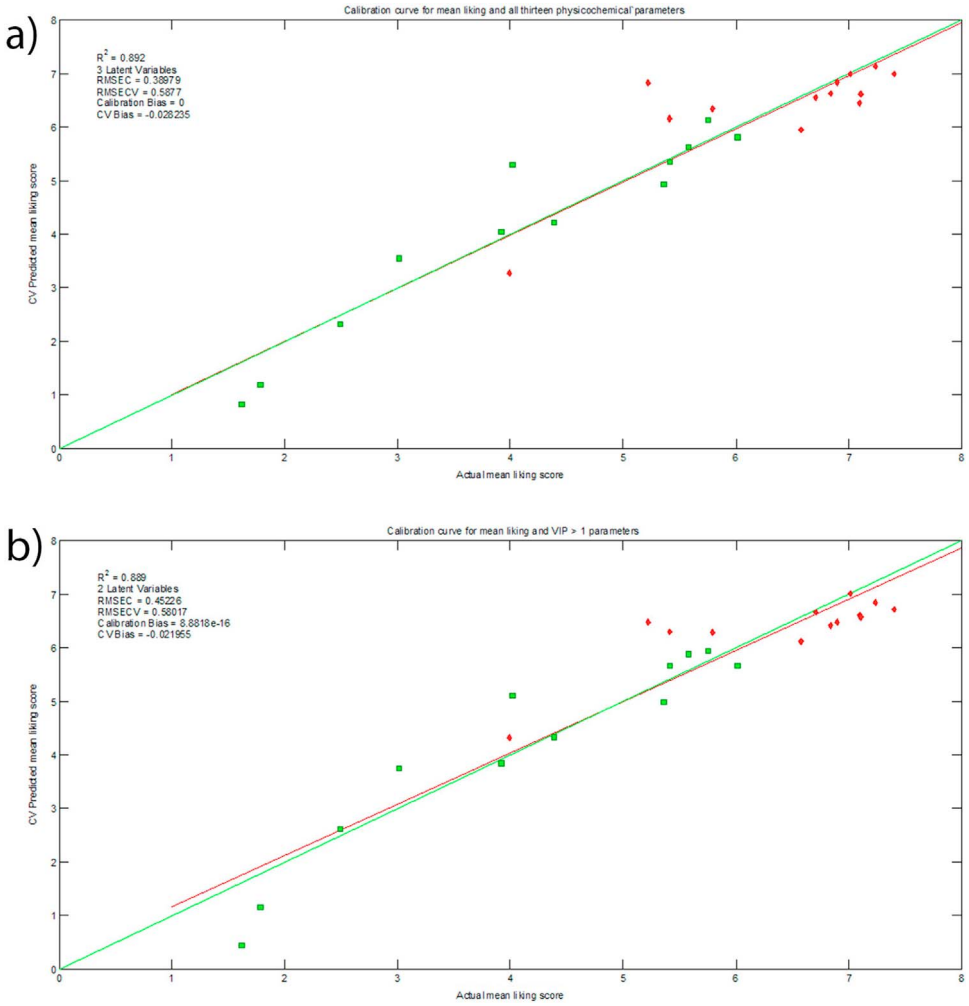


Fig. 7. PLS prediction results using a) thirteen physicochemical parameters and b) only the parameters with VIP scores above the threshold line of 1.

3.5. PLS prediction of global mean liking

PLS was applied to the model to predict the global volunteers' taste liking and to select a reduced number of the most important physicochemical parameters out of all the analysed data. The calibration model was developed using PLS analysis with internal cross-validation based on **y,** the global mean liking score vector, and $\mathbf{X_O}$, the corresponding thirteen physicochemical parameters,.

An excellent model with three latent variables was built to predict the volunteers' preference. A PLS model was built with these three latent variables, which captured a large part of the information (variance) in volunteers' preferences (approximately 95% of the **y** variance).

Weights on LV1 (accounting for 61% of the $\mathbf{X_O}$ variance and for 83% of the **y** variance) were positive only for Si, confirming the volunteers' preference for mineral water (see Fig. 6a). All other parameters had negative weights, which indicate that volunteers rated those waters with higher mineralisation with a low score..

LV2 accounts for 9% of the $\mathbf{X_O}$ variance and 9.5% of the **y** variance. Fig. 6b shows positive weights for pH, Mg^{2+}, Ca^{2+}, SO_4^{2-}, NO_3^-, HCO_3^- and negative weights for K^+, Na^+ and Cl^- and residual chlorine. This again shows that in general the volunteers preferred higher water hardness and alkalinity. The positive sign of the nitrate's weight is attributed to its spurious correlation with other parameters such as Ca^{2+}.

LV3 captures 5% of the $\mathbf{X_O}$ variance and explains only 2.5% of the **y** variance. LV3 reveals positive weights for Cl^-, Na^+ and SO_4^{2-} (Fig. 6c). This indicates that a small group of volunteers had preferred salty waters.

Fig. 6d shows the VIP score plot of the physicochemical parameters in the PLS model with the three latent variables. Parameters with VIP scores above the threshold of one (red dotted line) were considered to be the most influential in modelling the volunteers' preferences. This figure shows that TDS, conductivity, Cl^-, Na^+, K^+, SO_4^{2-}, Ca^{2+}, Mg^{2+} and chlorine were, in fact, the most important parameters for the construction of the model. Because Mg^{2+} and K^+ were present at very low concentrations, they should not have an effect on the taste, but they had significant correlation with calcium and sodium, respectively, and thus they were considered in the PLS model.

Fig. 7a shows the prediction results obtained in the internal cross-validation when all of the parameters were included in the PLS model. The calibration curve of the obtained model (red line) perfectly overlaps with the line of the best prediction (green line). The relative error of prediction was only 10.5%, and the determination coefficient (R^2 was 0.89, indicating good quantitative model..

When only nine parameters with VIP values above 1 (conductivity, TDS, Cl^-, SO_4^{2-}, Ca^{2+}, Mg^{2+}, Na^+, K^+ and chlorine) were considered and a new PLS model was re-calculated, the obtained relative error of prediction was again small (10.4%) and showed very similar figures of merit (see the legend of Fig. 7b). The two calibration curves on Fig. 7 show that both bottled (red diamonds) and tap (green squares) waters were estimated correctly along the two curves. These preliminary results are very promising and allow for the possibility to achieve a predictive model for water consumer preferences. The results showed

that physicochemical parameters can provide valuable qualitative and quantitative information for modelling consumer behaviour in a wide scenario.

4. Conclusion

Consumer preferences toward mineral composition of bottled and tap waters were successfully understood by applying various chemometric methods, such as PCA, MFA and PLS, to sensory data, physicochemical parameters and taste evaluations of untrained volunteers in an appropriately designed experiment. PCA and MFA revealed that overall, untrained volunteers rated water samples with high levels of mineralisation (i.e., TDS above 480 mg/L) with low scores. MFA, based on its specific weighting, made distinguishing among different groups of volunteers easier in relation to their individual preferences.

The majority of volunteers liked waters rich in $CaSO_4$ and $Ca(HCO_3)_2/CaSO_4$ with high alkalinity and pH. However, a small group of untrained volunteers preferred waters with NaCl and $NaHCO_3$ profiles and low pH. Minor constituents such as Mg^{2+}, K^+ and Si also appeared to be relevant for the taste evaluation of water samples, probably due to their spurious correlation with the major concentrations of Ca^{2+} and Na^+. These results are in agreement with the main conclusions obtained in a previous study using trained panellists to evaluate the role of major anions and cations in water taste [17].

Most volunteers preferred bottled waters. However, small quantities of residual chlorine did not appear to have a strong effect on the volunteers' preferences. The fact that some volunteers showed preferences for tap waters over bottled ones suggested that the water mineral content has a stronger influence on taste than residual chlorine concentrations, especially if the chlorine concentrations are not high, i.e., disinfectant levels above 1 mg/L.

PLS regression models enabled the prediction of each water's liking from its physicochemical parameters. Modelling water taste quality and consumer preferences with PLS was shown to be feasible. Moreover, it was possible to recalculate the model with only the most important physicochemical parameters without losing predictive power. The VIP (variable importance in projection) variable selection method has shown to be very useful for this purpose. Using either all thirteen physicochemical parameters or only nine VIP pre-selected parameters, the relative prediction errors calculated by cross-validation were 10.5% and 10.4%, respectively.

Acknowledgments

The authors would like to thank Isabel Pérez and Verònica García for coordinating the tap water tastings and the volunteers from Aigües de Barcelona and CETaqua for participating in over fifteen sessions of the tasting program. The authors are also grateful to Dr. Ivana Stanimirova for providing the MFA MATLAB code. SP and RT also acknowledge the Ministerio de Economía y Competividad, Spain for the grant CTQ2015-66254-C2-1-P.

References

[1] IWA, International Water Association. ⟨http://www.iwa-network.org/specialist-groups.php⟩. Accesed September 28, 2015.
[2] American Public Health Association (APHA), AWWA, WEF. 2012. Standard Methods for the Examination of Water and Wastewater 21st ed. Washington DC, USA
[3] I. Suffet, J. Mallevialle, E. Kawczynski, Advances in Taste-and-Odour Treatment and Control, American Water Association Research Foundation - Lyonnaise des Eau, Denver, 1995.
[4] E. Teillet, P. Schlich, C. Urbano, S. Cordelle, E. Guichard, Sensory methodologies and the taste of water, Food Qual. Pref. 21 (2010) 967–976.
[5] M.H. Vingerhoeds, M.A. Nijenhuis-de Vries, N. Ruepert, H. van der Laan, W.L. Bredie, S. Kremer, Sensory quality of drinking water produced by reverse osmosis membrane filtration followed by remineralization, Water Res 94 (2016) 42–51.
[6] A. Dietrich, Aesthetic issues for drinking water, J. Water Health 4 (suppl.1) (2006) 11–16.
[7] G.A. Burlingame, A.M. Dietrich, A.J. Whelton, Understanding the basics of tap water taste, J. Am. Water Works Assoc. 99 (2007) 100–110.
[8] R. Devesa, V. García, L. Matía, Water flavour improvement by membrane (RO and EDR) treatment, Desalination 250 (2010) 113–117.
[9] D. Gallagher, A. Dietrich, Statistical approaches for analyzing customer complaint data to assess aesthetic episodes in drinking water, J Water Supply Res Technol AQUA 63 (2014) 358–367.
[10] J. Raich-Montiu, J. Barios, V. García, M. Medina, F. Valero, R. Devesa, Integrating membrane technologies and blending options in wáter production and distribution systems to improve organoleptic properties. The case of the Barcelona Metropolitan Area, J. Clean. Prod. 69 (2014) 250–259.
[11] A. Whelton, A. Dietrich, G. Burlingame, M. Schechs, S. Duncan, Minerals in drinking water: impacts on taste and importance to consumer health, Water Sci. Technol. 55 (2007) 283–291.
[12] USEPA, National Secondary Drinking Water Regulation. ⟨http://water.epa.gov/drink/contaminants/index.cfm#Secondary⟩, Accessed September 25, 2015.
[13] Health Canada, Guidelines for Canadian Drinking Water Quality, 2012
[14] World Health Organisation (WHO), Guidelines for drinking-water quality, (4th ed), Geneve, Switzerland, 2011.
[15] EC (European Community), 1988, Council Directive 98/83/EC of 3 November, on the Quality of Water Intended for Human Consumption.
[16] E. Teillet, C. Urbano, S. Cordelle, P. Schlich, Consumer perception and preference of bottled and tap water, J. Sens. Stud. 25 (2010) 463–480.
[17] S. Platikanov, V. García, I. Fonseca, E. Rullán, R. Devesa, R. Tauler, Influence of minerals on the taste of bottled and tap water: a chemometric approach, Water Res 47 (2013) 693–704.
[18] H. Marcussen, W. Bredie, S. Stolzenbach, W. Brüsch, P. Holm, H. Hansen, Sensory properties of Danish municipal drinking water as a function of chemical composition, Food Res. Int. 54 (2013) 389–396.
[19] A. Dietrich, G. Burlingame, Critical review and rethinking of USEPA secondary standards for maintaining consumer acceptability of organoleptic quality of drinking water, Environ. Sci. Technol. 20 (2015) 708–720.
[20] L. Rey-Salgueiro, A. Gosálbel-García, C. Pérez-Lamela, J. Simal-Gándara, E. Falque-Lopez, Training of panelists for the sensory control of bottled natural mineral water in connection with water chemical properties, Food Chem. 141 (2013) 625–636.
[21] R. Devesa, R. Cardeñoso, L. Matía, Contribution of the FPA tasting panel to decision making about drinking water treatment facilities, Water Sci. Technol. 55 (2007) 127–135.
[22] F. Sinesio, E. Moneta, A. Sabba, Comparison of multivariate methods of analysis to evaluate panelists' performance, Food Qual. Pref. 3 (1991/2) 201–208.
[23] T. Naes, E. Risvik, Multivariate analysis of Data in Sensory Science, Elsevier, Amsterdam, 1996.
[24] I. Jolliffe, Principal Component Analysis, 2nd ed, Springer Verlag, Berlin, 2002.
[25] I. Stanimirova, B. Walczak, D. Massart, Multiple factor analysis in environmental chemistry, Anal. Chim. Acta 545 (2005) 1–12.
[26] P. Geladi, B. Kowalski, Partial least-squares regression: a tutorial, Anal. Chim. Acta 185 (1986) 1–17.
[27] V. Lengard, M. Kermit, 3-way and 3 block PLS regressions in consumer preference analysis, Food Qual. Pref. 17 (2006) 234–242.
[28] H. Kermadec, J. Durand, R. Sabatier, Comparison between linear and nonlinear PLS methods to explain overall preferences from sensory characteristics, Food Qual. Pref. 8 (1997) 395–402.
[29] R. Devesa, C. Fabrellas, R. Cardeñoso, L. Matia, F. Ventura, N. Salvatella, The panel of Aigües de Barcelona: 15 years of history, Water Sci. Technol. 49 (2004) 145–151.
[30] J. Pagés, M. Tenenhaus, Multiple factor analysis combined with PLS path modelling. Application to the analysis of relationships between physicochemical variables, sensory profiles and hedonic judgements, Chem. Intell. Lab. Syst. 58 (2001) 261–273.
[31] I. Stanimirova, C. Boucon, B. Walczak, Relating gas chromatographic profiles to sensory measurements describing the end products of the Maillard reaction, Talanta 83 (2011) 1239–1246.
[32] I. Chong, C. Jun, Performance of some variable selection methods when multicollinearity is present, Chem. Intell. Lab. Sys 78 (2005) 103–112.