Talanta 125 (2014) 248–256

Contents lists available at ScienceDirect

Talanta

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

Statistical analysis for improving data precision in the SPME GC–MS analysis of blackberry (*Rubus ulmifolius* Schott) volatiles

M.F. D'Agostino [a], J. Sanz [b], I. Martínez-Castro [b], A.M. Giuffrè [a], V. Sicari [a], A.C. Soria [b,*]

[a] *Università degli Studi Mediterranea di Reggio Calabria – Dipartimento AGRARIA, Contrada Melissari, 89124 Reggio Calabria, Italy*
[b] *Instituto de Química Orgánica General (CSIC), Juan de la Cierva 3, 28006 Madrid, Spain*

ARTICLE INFO

Article history:
Received 22 October 2013
Received in revised form
22 January 2014
Accepted 22 February 2014
Available online 2 March 2014

Keywords:
Blackberry (*Rubus ulmifolius* Schott)
Solid-phase microextraction (SPME)
Gas chromatography–mass spectrometry (GC–MS)
Volatiles
Statistical analysis
Precision

ABSTRACT

Statistical analysis has been used for the first time to evaluate the dispersion of quantitative data in the solid-phase microextraction (SPME) followed by gas chromatography–mass spectrometry (GC–MS) analysis of blackberry (*Rubus ulmifolius* Schott) volatiles with the aim of improving their precision. Experimental and randomly simulated data were compared using different statistical parameters (correlation coefficients, Principal Component Analysis loadings and eigenvalues). Non-random factors were shown to significantly contribute to total dispersion; groups of volatile compounds could be associated with these factors. A significant improvement of precision was achieved when considering percent concentration ratios, rather than percent values, among those blackberry volatiles with a similar dispersion behavior.

As novelty over previous references, and to complement this main objective, the presence of non-random dispersion trends in data from simple blackberry model systems was evidenced. Although the influence of the type of matrix on data precision was proved, the possibility of a better understanding of the dispersion patterns in real samples was not possible from model systems.

The approach here used was validated for the first time through the multicomponent characterization of Italian blackberries from different harvest years.

1. Introduction

Analysis of blackberry (*Rubus ulmifolius*) volatiles represents a valuable approach not only to study its flavor but also for its objective characterization, as volatile concentration depends on different factors such as blackberry type, origin or processing (freezing, drying, etc.). Whereas the gas chromatography–mass spectrometry coupling (GC–MS) is the technique of choice for the analysis of the complex mixtures of volatiles of different functionalities present at low concentration in blackberries, the required volatile fractionation and preconcentration step can be addressed by different procedures [1–6].

Solid-phase microextraction (SPME) is a fast, simple, affordable and solvent-free technique which provides a fraction suitable for GC–MS analysis [7]. In the headspace sampling mode, the SPME fiber consisting of a fused-silica fiber coated with a polymeric film is exposed to the sample headspace to fractionate and preconcentrate volatiles by partition, adsorption or mixed mechanisms. Desorption of the SPME fiber into the injection port of a GC or GC–MS system allows the chromatographic analysis of this volatile fraction without matrix interferences. Although SPME has been widely used for the fractionation/preconcentration of food volatiles prior to their chromatographic analysis [7,8], it has scarcely been applied to the study of blackberry aroma [3,5,6,9].

Despite the above mentioned advantages of SPME, the changeable precision when multicomponent mixtures are to be analyzed has frequently questioned the quantitative performance of SPME [8], particularly for studies on characterization of samples where precision is the most important analytical parameter. Thus, few papers on application of SPME for fractionation of food volatiles show the dispersion of quantitative data [10,11] or refer data precision for only a very limited number of volatiles, as those included in model systems used for optimization of SPME operating conditions [3,12]. Although the matrix effect is known to play a role on volatile recovery and Multiple Headspace SPME approaches have been used to estimate the effect [13], the matrix effect's influence on data precision has not yet been considered in papers on SPME.

Previous studies carried out in our laboratory on fractionation of volatiles from different food samples (honey and cumin seeds) by means of diverse techniques such as SPME and Purge and Trap (P&T) have shown an improvement of data precision when considering relative data (percentage of total volatiles) rather

---

[*] Corresponding author. Tel.: +34 912587559; fax: +34 915644853.
E-mail address: acsoria@iqog.csic.es (A.C. Soria).

http://dx.doi.org/10.1016/j.talanta.2014.02.058
0039-9140

than absolute data (ng, ng g^{-1}, etc.) for quantitation [14,15]. Furthermore, application of statistical analysis to relative SPME and P&T data has evidenced the existence of non-random patterns in experimental data, affecting in a similar way to groups of compounds, which could be used to improve their precision. Calculation of concentration ratios for those volatiles with similar fractionation properties has been shown to significantly reduce data dispersion.

As a further step in the development of advanced SPME methodologies for fractionation of blackberries volatiles, the main aim of this work was to evaluate for the first time if the study of the compound-depending patterns in the dispersion of quantitative data evidenced by statistical analysis could be used to improve the precision of data. In addition, and as novelty over previous references [14,15], the proposed approach was also applied to data from different blackberry volatile model systems in the search for a better understanding of the influence of sample matrix on non-random dispersion of data. Furthermore, validation of this statistical approach was also possible for the first time from the results obtained in the precise multicomponent characterization of Italian blackberries from different harvest years.

2. Materials and methods

2.1. Blackberry samples

Wild blackberries (*R. ulmifolius* Schott) from La Cuerva (Cáceres, Spain) collected in 2012 were selected for the evaluation of patterns in the dispersion of quantitative data obtained by SPME followed by GC–MS. For sample homogeneity, whole blackberries were freeze-dried, powdered and sieved (<0.5 mm) prior to analysis (sample WBLACK).

Validation of the improved SPME procedure was done using two additional sets of Italian blackberries collected in 2011 and 2012 in different locations all over Calabria (Table 1).

2.2. Blackberry model systems

2.2.1. Standard mixture

A stock standard solution in water including 17 blackberry volatiles (2-heptanone, 1-heptanol, 1-octanol, terpinen-4-ol from Fluka Chemie (Buchs, Germany); 3-methyl-1-butanol, 2-heptanol, 6-methyl-5-hepten-2-one, 1-hexanol, *trans*-3-hexen-1-ol, 2-hexen-1-ol, linalool, myrtenol, benzyl alcohol, 1-phenylethanol from Sigma-Aldrich (St. Louis, MO, USA); 1-pentanol, benzaldehyde, 1-nonanol from PolyScience Corp. (Niles, IL, USA)) in the concentration range 0.1–0.3 mg mL^{-1} was prepared. Compounds were chosen based on the literature [3,6, etc.] and experimental results previously obtained in our laboratory in the SPME GC–MS analysis (under identical conditions) of a commercial blackberry sample. Compounds were selected to be in different concentrations so that they simulate the variability in volatile concentration of a real blackberry sample. Dilution 1:20 of this stock volatile mixture was used for preparation of the model systems detailed below.

Table 1
Location in Calabria and Harvest year of the Italian blackberries under analysis.

Location	Harvest year	
	2011	2012
Nicotera 38°32′02″N 15°56′23″E	NIC11	NIC12
Bovalino 38°9′0″N 16°10′0″E	BOV11	BOV12
Cicerna 38°27′34″N 15°55′25″E	CIC11	CIC12
Filadelfia 38°46′36″N 16°17′25″E	FIL11	FIL12
Granatara 38°29′13″N 15°57′03″E	GRA11	GRA12
Rosarno 38°29′08″N 15°58′47″E	ROS11	ROS12
Cosoleto 38°16′33″N 15°55′42″E	COS11	COS12

2.2.2. Model systems

Model system MSSUG comprised 0.2 g of a carbohydrate matrix including major blackberry carbohydrates as described by Souci et al. [16] (44.8% glucose, 44.7% fructose, 3.0% sucrose and 7.5% pectin) and 40 μL of the standard mixture. Model system MSCOT was prepared by adding 20 μL of the standard mixture to 0.2 g of cotton previously conditioned overnight in an oven at 50 °C. Model system MSGW consisted of 0.2 g of glass wool (Quimicen, Madrid Spain) and 20 μL of the standard mixture detailed above.

2.3. Solid-phase microextraction (SPME)

Fractionation of volatiles from blackberry samples and model systems was carried out according to D'Agostino et al. [9]. Headspace sampling was done using a manual SPME holder equipped with a 85 μm Carboxen™-Polydimethylsiloxane StableFlex fiber (both from Supelco, Bellefonte, PA, USA). The experimental procedure was as follows: 0.2 g of freeze-dried blackberries/model systems were weighted into a 5-mL vial sealed with a screw cap provided with a predrilled Teflon-faced septum. After an equilibrium time of 20 min at 66 °C [9], fiber was exposed to the headspace of the blackberry sample/model system for 16 min.

2.4. Gas chromatography–mass spectrometry (GC–MS) analysis

GC–MS analyses ($n=9$–13 for sample WBLACK and model systems; $n=2$ for blackberries listed in Table 1) were performed on a 6890 gas chromatograph coupled to a 5973 quadrupole mass detector (Agilent Technologies, Palo Alto, CA, USA). The SPME fiber was desorbed into the injection port at 250 °C in splitless mode (3 min), using a 0.75 mm narrow bore liner (Supelco). Compounds were resolved on a Supelcowax capillary column (27.2 m × 0.25 mm i.d. × 0.25 μm film thickness, Supelco) using helium as carrier gas. The oven was temperature programmed from 40 °C (3 min) to 220 °C (60 min) at 3 °C min^{-1}. Mass spectra were recorded in electron impact (EI) mode at 70 eV within the mass range *m/z* 35–450. The transfer line and ionization source were thermostated at 280 and 230 °C, respectively. Acquisition was done using a HPChem Station software (Agilent Technologies, Palo Alto, CA, USA).

Qualitative analysis was based on the comparison of experimental spectra with those of the Wiley mass spectral library [17] and with published data. Identifications were further confirmed by using available standards and linear retention indices (I^T) calculated, according to [18], from data of a *n*-alkane (C_{10}–C_{20}) mixture analyzed under identical experimental conditions. Semiquantitative data (percentage of total volatiles) were directly calculated from peak areas of total ion current (TIC) profiles.

2.5. Statistical data analysis

2.5.1. Experimental data

Quantitative percent data for individual replicates were organized as data matrices in Tables S1–S4 in Supplementary data: WBLACK (62 compounds × 11 replicates, Table S1), MSCOT

Table 2
Experimental percent data (mean and relative standard deviation for $n=11$ replicates) and linear retention indices (I^T) of volatiles determined in the SPME followed by GC–MS analysis of blackberry sample WBLACK.

Peak no.	Compound	I^T	Mean (%)	RSD (%)
1	3-Methylbutanal	1036	6.1	4.8
2	Ethanol	1041	23.9	13.1
3	2,3-Butanedione	1056	5.3	8.5
4	Hexanal	1111	0.9	34.0
5	2-Methyl-2-butenal	1118	2.2	71.3
6	3-Penten-2-one	1141	1.1	34.7
7	1-Butanol	1158	5.6	14.0
8	2-Heptanone	1188	1.1	18.6
9	Methyl hexanoate	1194	0.1	16.9
10	Limonene	1197	0.1	37.2
11	3-Methyl-1-butanol	1213	2.1	14.4
12	*trans*-2-Hexenal	1219	4.0	29.3
13	Ethyl hexanoate	1238	0.1	19.3
14	1-Pentanol	1255	0.5	6.7
15	*p*-Cymene	1268	tr[a]	13.6
16	3-Hydroxy-2-butanone	1289	1.4	12.3
17	2-Methylbutyl-3-methylbutanoate	1298	tr	19.1
18	4-Methyl-1-pentanol	1318	0.3	9.4
19	2-Heptanol	1325	0.6	7.6
20	6-Methyl-5-hepten-2-one	1339	0.2	10.8
21	1-Hexanol	1357	19.6	3.7
22	*trans*-3-Hexen-1-ol	1368	0.4	8.5
23	Methyl octanoate	1391	tr	21.7
24	Nonanal	1393	1.6	5.6
25	2-Hexen-1-ol	1411	0.1	13.5
26	Hexyl butanoate	1417	0.1	12.7
27	Ethyl octanoate	1437	0.1	21.1
28	α-Cubebene	1449	0.1	60.4
29	Unknown (43, 58, 84, 69)[b]	1456	0.2	6.3
30	1-Heptanol	1460	0.6	5.6
31	2-Furancarboxaldehyde	1467	0.1	19.5
32	α-Ylangene	1470	tr	67.7
33	5,5-Dimethyl-2-cyclopenten-1-one	1474	tr	62.6
34	α-Copaene	1478	0.3	56.0
35	2,4-Heptadienal	1491	0.1	16.3
36	Decanal	1497	0.2	9.6
37	3-Ethyl-4-methylpentanol	1512	0.3	11.3
38	Benzaldehyde	1517	0.3	5.8
39	Epizonaren	1536	0.1	60.6
40	Linalool	1554	0.9	17.3
41	1-Octanol	1562	8.7	3.9
42	5-Methylfurfural	1572	tr	18.9
43	Methyl decanoate	1596	tr	12.0
44	Hexyl hexanoate	1611	0.2	31.5
45	Dihydro-2(3H)-furanone	1617	0.6	10.8
46	Phenylacetaldehyde	1638	0.6	13.1
47	Ethyl decanoate	1640	0.1	14.4
48	1-Nonanol	1664	0.9	5.3
49	2-Furanmethanol	1667	0.1	4.8
50	α-Terpineol	1693	0.1	18.0
51	Methyl butanoic acid	1710	3.2	7.8
52	δ-Cadinene	1746	tr	61.9
53	1-Decanol	1767	0.6	23.2
54	Myrtenol	1788	0.1	7.3
55	Methyl dodecanoate	1804	0.1	9.8
56	Ethyl dodecanoate	1845	0.1	19.3
57	Butyl benzoate	1854	tr	23.9
58	Benzyl alcohol	1874	1.6	5.4
59	2-Phenylethanol	1906	2.1	6.0
60	2-Methyl-3-phenyl-1-propanol	1988	0.1	20.8
61	4-Ethyl-2-methoxy-phenol	2026	tr	47.0
62	Benzenepropanol	2040	0.2	10.0

[a] tr=trace ($<0.05\%$).
[b] Characteristic m/z ratios.

(17 compounds × 9 replicates, Table S2), MSSUG (17 compounds × 13 replicates, Table S3) and MSGW (17 compounds × 12 replicates, Table S4). Mean and relative standard deviation (RSD, %) of the percent concentration values for each volatile compound in the blackberry sample (Table 2) and model systems (Table 3) were calculated.

Table 3
Experimental percent data (mean and relative standard deviation for $n=9$–13 replicates) and linear retention indices (I^T) of volatiles determined in the SPME followed by GC–MS analysis of three blackberry model systems.

Peak no.	Compound	I^T	MSCOT		MSGW		MSSUG	
			Mean (%)	RSD (%)	Mean (%)	RSD (%)	Mean (%)	RSD (%)
1	2-Heptanone	1188	3.7	17.1	3.0	17.6	0.4	20.7
2	3-Methyl-1-butanol	1213	6.3	9.9	5.7	10.5	11.2	8.2
3	1-Pentanol	1255	7.9	8.2	7.4	6.4	10.5	3.2
4	2-Heptanol	1325	5.6	5.0	4.2	10.0	5.0	4.0
5	6-Methyl-5-hepten-2-one	1339	5.0	19.3	3.8	20.2	4.5	8.6
6	1-Hexanol	1357	10.3	2.2	8.4	5.5	10.3	2.8
7	*trans*-3-Hexen-1-ol	1368	9.8	4.4	8.6	3.5	10.3	5.4
8	2-Hexen-1-ol	1411	5.9	7.1	5.2	3.8	5.9	5.6
9	1-Heptanol	1460	6.5	1.3	5.4	9.0	4.6	4.6
10	Benzaldehyde	1517	6.5	8.9	5.7	10.2	6.9	5.7
11	Linalool	1554	8.5	3.5	7.9	5.6	3.9	6.5
12	1-Octanol	1562	4.3	3.9	5.0	4.3	3.1	9.4
13	Terpinen-4-ol	1614	6.5	3.2	7.3	4.6	3.4	6.6
14	1-Nonanol	1664	1.8	8.0	2.3	9.4	1.5	18.3
15	Myrtenol	1788	5.7	7.5	8.8	7.9	6.3	10.5
16	1-Phenylethanol	1833	2.8	10.1	4.9	15.5	5.2	10.0
17	Benzyl alcohol	1874	3.0	14.2	6.6	18.3	6.9	18.4

2.5.2. Simulated data

Ten sets of simulated data for blackberry sample and model systems (SIMWBLACK1–10, SIMMSSUG1–10, SIMMSCOT1–10, SIMMSGW1–10) were obtained using the RND function of a VisualBasic program developed at our laboratory. Random data having a normal distribution and the same mean and RSD values as those of experimental data were obtained as described in [19].

2.5.3. Volatile concentration ratios

For every experimental and simulated data matrix, ratios (vx/vy) between percent concentrations of volatiles x and y were calculated for all possible pairs of volatile components quantified. RSD values were calculated for these ratios.

2.5.4. Data processing

Statistical analysis (correlation coefficients, Principal Component Analysis (PCA), Stepwise Discriminant Analysis and one-sample t-test for significance of differences) was carried out by using the Statistica software [20].

3. Results and discussion

3.1. Quantitative data

Although an external or internal standard calibration, standard addition method or multistep headspace procedure might be used for quantitation of volatile compounds [8,13,21], these methods are difficult to implement for multicomponent SPME determinations of food volatiles because of the high number of standards required and the difficulties for homogenization of standards in complex food matrices such as blackberries. Therefore, and for the differentiation of blackberries according to their harvest years here intended, semiquantitative percent data calculated as described in Section 2.4 were used.

Table 2 lists the experimental results obtained in the SPME followed by GC–MS analysis of blackberry sample WBLACK. Qualitative data (assignation and linear retention indices) and percent quantitative data (mean and RSD) for a total of 62 volatiles of different functionality (alcohols, ketones, esters, etc.) in a wide range of both polarity and concentration were included. In a similar format, Table 3 summarizes the experimental results for the seventeen volatiles included in the three blackberry model systems under study (MSCOT, MSGW, MSSUG). For all these samples, data for individual replicates are compiled in Supplementary data section (Tables S1–S4).

As supported by the average relative standard deviation (20.3%) calculated for the 62 volatiles quantified in WBLACK sample, the precision of the SPME GC–MS here used was good for a multicomponent determination. Ibáñez et al. [3], in the quantitative analysis by headspace SPME of a model system, reported RSD values in the range 4–16% for normalized percent areas corresponding to eight fruit volatiles. In a study on the characterization by SPME of five Chinese sweet cherry cultivars, Sun et al. [10] described relative standard deviations in the range 10–13% for 52 aroma-active compounds. Similarly, Alvarez et al. [11] reported RSD values as high as 24% in the multicomponent (61 compounds) SPME GC-MS determination of volatiles from mandarin juice processed by different squeeze technologies.

As shown in Table 2, experimental data dispersion was highly variable: 20 volatiles had RSD $<$ 10% and 44 volatiles showed RSD $<$ 20%. As expected, the highest dispersion values were generally associated with blackberry volatiles present at concentrations lower than 0.1% (e.g. methyl octanoate, ethyl octanoate, α-cubebene, α-ylangene, 5,5-dimethyl-2-cyclopenten-1-one, δ-cadinene, etc.). However, it is also worth noting that several low-retained major blackberry volatiles such as 2-methyl-2-butenal and *trans*-2-hexenal showed a relative standard deviation much higher than that of other volatiles present at similar concentrations.

In blackberry model systems (Table 3), most volatiles showed dispersion values lower than 10% (average standard deviation of 7.9%, 9.5% and 8.7% for MSCOT, MSGW and MSSUG, respectively). This dispersion value match well within the range previously reported by Ibáñez et al.[3] for relative data obtained by SPME GC–MS from volatile model systems. In general, compounds showing the lowest precision were common to all model systems (e.g. 2-heptanone, 6-methyl-5-hepten-2-one, 1-phenylethanol and benzyl alcohol).

As evidenced by PCA, no correlation was found between the dispersion determined for volatiles in model systems and in blackberry sample WBLACK. Regarding the dispersion results for the three model systems assayed here, MSGW and MSSUG were the most similar (higher scores for PC1 in the PC plot of cases). As supported by the PC plots of variables, compounds with higher scores for PC1 were associated to the volatiles previously mentioned with the highest relative standard deviation (2-heptanone, 6-methyl-5-hepten-2-one and benzyl alcohol).

3.2. *Study of patterns in the dispersion of quantitative data*

Taking into account the results obtained in Section 3.1, and in order to evaluate the existence of non-random patterns in the dispersion of quantitative data, experimental (WBLACK, MSCOT, MSGW and MSSUG) and simulated (SIMWBLACK1–10, SIMMSCOT1–10, SIMMSGW1–10 and SIMMSSUG1–10) data matrices were statistically compared by studying their correlation coefficients (r), PCA loadings and eigenvalues.

3.2.1. *Correlation coefficients*

Analytical conditions might affect in a different way the response of groups of compounds sharing a similar characteristic. In quantitative data matrices, calculation of correlation coefficients ($-1 \leq r \leq 1$) between compounds is the simplest measurement of their similar dispersion behavior. A significance level can be assigned to these values, by estimating their probability (*p-level*, p) of being caused by random effects. Significant values indicate that two compounds are affected in the same way by one or more non-random effects.

For n compounds, the number of correlation coefficients is $n(n-1)/2$. In the case of random data, a few of them might have significant values: their percentage can be calculated from statistical considerations, or from simulated data.

Fig. 1A plots the distribution within a given interval of the 1891 correlation coefficients calculated from experimental (white bars) and simulated (gray bars) data matrices in the SPME followed by GC–MS analysis of the blackberry sample WBLACK. For simulated data, average values and error bars representing the standard deviation for the ten sets of simulated data are shown in this figure.

In a similar format, Fig. 1B shows as an example of the results obtained for model systems, the distribution of the 136 correlation coefficients calculated from the 17 volatiles included in MSGW and SIMMSGW1–10 data matrices. Similar plots for model systems MSCOT and MSSUG are included in Supplementary data section (Figs. S1 and S2).

As expected, the frequency (n) of correlation coefficients for randomly simulated WBLACK data showed a normal distribution, with predominance of r values with low absolute value ($-0.3 < r < 0.3$). On the contrary, r values from experimental data showed a very distinctive pattern: the number of significant (absolute value >0.6) correlation coefficients was higher than

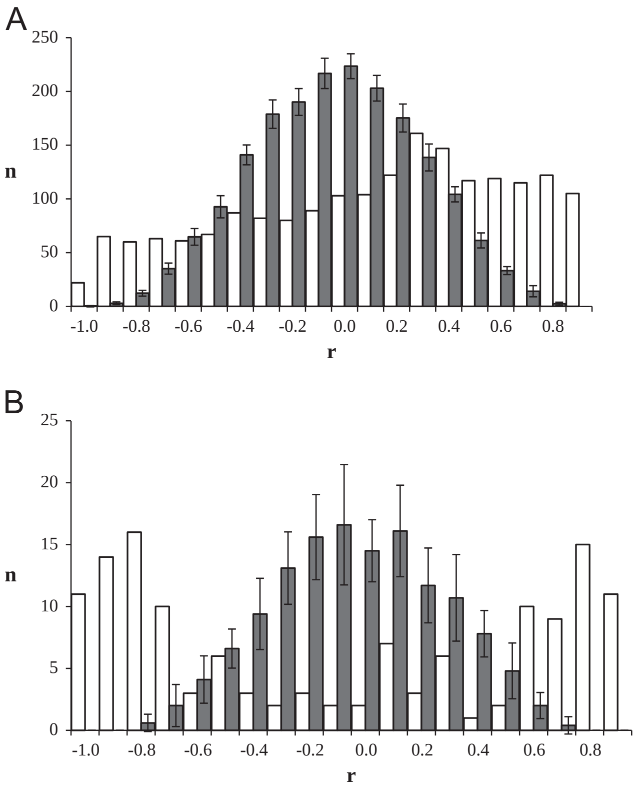

Fig. 1. Distribution (n: frequency) of the correlation coefficients (r) for experimental (white bars) and simulated (gray bars) data matrices from (A) WBLACK sample and (B) model system MSGW.

that of simulated data (particularly for positive correlations), whereas the number of r values with low absolute value (center of the plot) was lower. In order to evaluate the significance of these differences, the one-sample *t*-test for the differences in frequencies was applied to experimental and simulated data. For correlation coefficients in the 0.6–1.0 interval and lower than -0.6, t values in the range (-61.7, -279.7) and (-17.0, -152.6) were obtained, respectively, showing a probability ($p < 0.01$) of being due to random causes.

A similar trend in the distribution of correlation coefficients was found for experimental and simulated data for model system MSGW. Application of sample *t*-test to experimental and simulated correlation coefficients with absolute value higher than 0.6 showed a significant probability ($p < 0.01$) for these differences to be caused by random factors. Similar results in the ranges $-1.0 < r < -0.7$ and $r > 0.6$ were obtained for model systems MSCOT ($p < 0.03$) and MSSUG ($p < 0.01$).

As dispersion of experimental data has been proved to be affected by non-random factors, and in order to characterize the significance of the correlations among blackberry volatiles shown above, PCA was applied to both experimental (WBLACK, MSCOT, MSGW and MSSUG) and simulated (SIMWBLACK1–10, SIMMSCOT1–10, SIMMSGW1–10 and SIMMSSUG1–10) data matrices.

3.2.2. Principal Component Analysis

In Principal Component Analysis, the dimensionality of a data set is reduced by defining several mathematical factors (principal components) which are a linear combination of the original variables. PCA eigenvalues are used as a measurement of the amount of variance explained by each of these factors, while PCA loadings afford information on the associated variables (volatile compounds) and its importance for each principal component. These mathematical factors can be the result of the combination of different physical/chemical (systematic) and random effects.

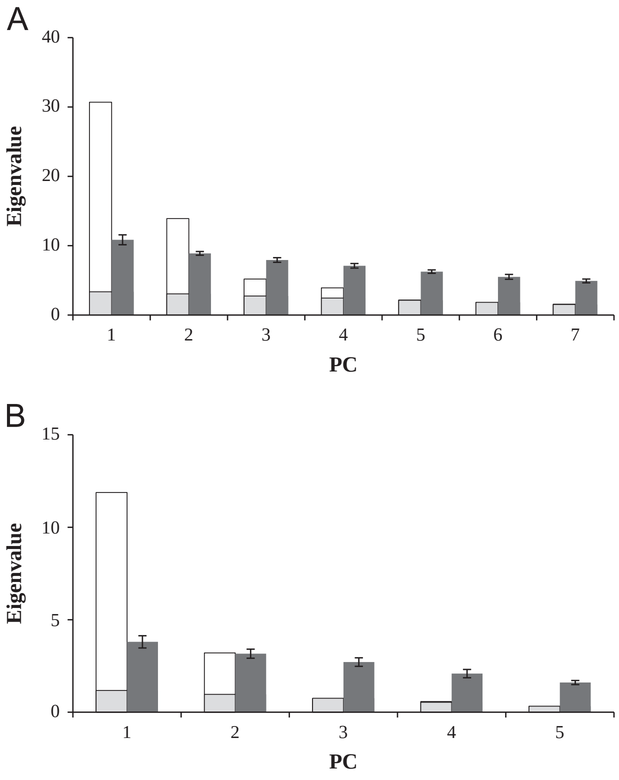

Fig. 2. Eigenvalues obtained for the first principal components in the PCA analysis of experimental (white bars) and simulated (dark gray bars) data matrices from (A) WBLACK sample and (B) model system MSGW. Contribution of random factors to experimental eigenvalues is shown by light gray bars.

Fig. 2A plots the eigenvalues of the first seven principal components obtained in the PCA of experimental and simulated data for sample WBLACK. Bar height represents the variance explained by each principal component for experimental (white bars) and simulated (dark gray bars). As for correlation coefficients, error bars of simulated data represent the standard deviation of the ten sets of simulated data. In a similar format, Fig. 2B and Figs. S3, S4 in Supplementary data section compile the experimental and simulated eigenvalues for the first principal components calculated for model systems MSGW, MSCOT and MSSUG, respectively.

The most remarkable feature in Fig. 2A is the different trend shown for experimental and simulated eigenvalues for principal components 2 and 3. Differences for experimental and simulated data were highly significant ($p < 0.01$, one sample *t*-test) for the first seven eigenvalues. Simulated eigenvalues (dark gray bars) showed a continuous linear decrease from principal component 1–7, and a similar trend was found for experimental eigenvalues of high order (PC > 5), where the contribution of random causes to data variance predominates. Assuming the linear contribution of random factors to data variance extrapolated from PC 5–7 to principal components of first order (Fig. 2A, light gray bars), the non-random (systematic) contribution to PC1 and PC2 was 89 and 78%, respectively.

As regards the model system MSGW (Fig. 2B), the gap between experimental eigenvalues of second and third principal components, which does not appear in simulated data, is the most striking feature. Differences between experimental and simulated eigenvalues were significant for the first and third principal components: t values ranging between -77.2 and 27.2 ($p < 0.01$). Application of a similar procedure to that previously mentioned for experimental WBLACK data, estimated that the contribution of non-random factors to the variance of experimental MSGW data was 90% and 70% for PC1 and PC2, respectively. Similar results were obtained in the comparison of experimental and simulated eigenvalues of model systems MSCOT and MSSUG (Figs. S3–S4, respectively). In this case, the average proportion of non-random variance for MSCOT and MSSUG data was 82% for PC1–PC2 and 68% for PC1–3, respectively.

As demonstrated above, systematic factors seem to be the most important contribution to data variability. In order to identify the factors potentially contributing to the non-random variance of experimental data, compounds sharing a particular dispersion behavior can be grouped from the study of their loading coefficients. As a first step, and given the high number of volatiles considered, comparison of the distribution of significant loadings for experimental and simulated data is also required to draw conclusions of statistical validity.

Loadings values for the first two principal components obtained in the PCA analysis of WBLACK and SIMWBLACK1–10 data matrices are summarized in the plot of Fig. 3A. For each principal component, the number (n) of blackberry volatiles with absolute value higher than 0.8 is represented by the bar height. As before, white bars and gray bars are used for experimental and simulated data, respectively. Similar bar plots are shown in Fig. 3B and Figs. S5, S6 in Supplementary data for comparison of experimental and simulated loadings of model systems MSGW, MSCOT and MSSUG, respectively.

Compounds showing similar loadings are supposed to share similar fractionation characteristics (e.g. similar structure, polarity, volatility, etc.) and the extent of this similarity is related to the magnitude of these loadings. As for wild blackberry data, the number of loadings with absolute value > 0.8 for PC1 and PC2 was significantly higher for experimental data ($p < 0.01$). Table 4 lists

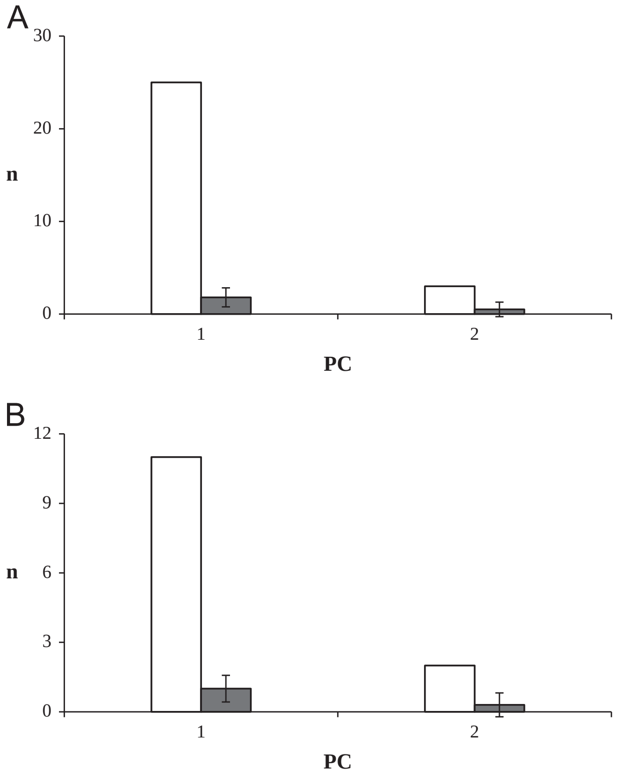

Fig. 3. Number (n) of loadings with absolute value higher than 0.8 in the PCA analysis of experimental (white bars) and simulated (gray bars) data matrices from (A) WBLACK sample and (B) model system MSGW.

Table 4
Loading coefficients with absolute value higher than 0.8 for the two first principal components in the statistical analysis of WBLACK data matrix.

Peak no.	Compound	PC1	PC2
2	Ethanol	−0.891	
3	2,3-Butanedione	0.908	
5	2-Methyl-2-butenal		−0.869
7	1-Butanol	−0.901	
8	2-Heptanone	0.884	
10	Limonene	0.900	
11	3-Methyl-1-butanol	−0.860	
12	*trans*-2-Hexenal	0.987	
13	Ethyl hexanoate	0.891	
15	*p*-Cymene	0.871	
16	3-Hydroxy-2-butanone	0.855	
18	4-Methyl-1-pentanol		0.826
23	Methyl octanoate	0.857	
25	2-Hexen-1-ol	0.835	
26	Hexyl butanoate	0.885	
27	Ethyl octanoate	0.941	
33	5,5-Dimethyl-2-cyclopenten-1-one	0.862	
35	2,4-Heptadienal	0.802	
37	3-Methyl-2-pentene	0.968	
40	Linalool	0.956	
46	Phenylacetaldehyde	0.955	
47	Ethyl decanoate	0.926	
50	α-Terpineol	0.973	
51	Methyl butanoic acid	0.845	
53	1-Decanol	−0.973	
56	Ethyl dodecanoate	0.913	
60	Non identified (117, 91, 132,150)[a]		−0.856
61	4-Ethyl-2-methoxy-phenol	−0.894	

[a] Characteristic m/z ratios.

Table 5
Loading coefficients with absolute value higher than 0.8 for the two first principal components in the statistical analysis of MSGW data matrix.

COMPOUND	PC1	PC2
2-Heptanone	0.986	
1-Pentanol	−0.850	
2-Heptanol	0.972	
6-Methyl-5-hepten-2-one	0.990	
1-Hexanol	0.892	
trans-3-Hexen-1-ol		−0.861
2-Hexen-1-ol		−0.848
1-Heptanol	0.974	
Benzaldehyde	0.896	
Linalool	0.878	
Myrtenol	−0.959	
1-Phenylethanol	−0.958	
Benzyl alcohol	−0.920	

the assignation of the 28 wild blackberry volatiles showing significant loadings for principal components 1–2. As an example, compounds with a monoterpenic alcohol structure such as α-terpineol and linalool or with similar retention index (3-methyl-2-pentene, linalool, phenylacetaldehyde, α-terpineol; $I^T = 1512$–1693) showed high positive loadings (>0.96) for PC 1. Negative contributions to this principal component were associated with alcohols eluting early (ethanol, 1-butanol, 3-methyl-1-butanol) or late (1-decanol and 4-ethyl-2-methoxy-phenol) in the chromatogram ($-0.97 < \text{loading} < -0.86$). Only three volatiles showed significant loadings for PC2 (positive contribution of 4-methyl-1-pentanol, negative contribution of 2-methyl-2-butenal and 2-methyl-3-phenyl-1-propanol). As expected, compounds included in groups with similar SPME fractionation were highly correlated ($r > 0.92$).

As for model systems under study, the number of high loadings for the first principal components was also significantly ($p < 0.01$) higher for experimental rather than for simulated data. As an example, 11 over a total of 17 volatiles from MSGW showed high loadings (most of them positive) for PC1 (Table 5), and only isomeric alcohols (*trans*-3-hexen-1-ol and 2-hexen-1-ol) showed negative loadings for PC2. A lower number of significant loadings for PC1-2 was found for MSCOT and MSSUG data matrices (Tables S5 and S6). It is worth noting that, although the compounds significantly contributing to each PC were different for every model system, proving the relevance of the matrix on data dispersion, the negative/positive character of these contributions to every PC is kept irrespective of the blackberry model system considered (e.g. 2-heptanone and 6-methyl-5-hepten-2-one: positive loadings for PC1; 1-pentanol and myrtenol: negative loadings for PC1).

3.3. *Volatile concentration ratios*

As previously mentioned, when analytical results of multicomponent samples are to be used for characterization purposes, they can be expressed as relative values taking as reference (100%) the total amount of compounds quantified in the analysis. In these cases, specific response is usually supposed to be the same for all volatile compounds, and relative areas are used for estimation of their relative concentrations. Ratios between these areas can also be used for sample characterization.

As described in Section 2, ratios (vx/vy) between percent concentrations for the 62 volatiles considered in WBLACK and SIMWBLACK1-10 data matrices were calculated. RSD values for these ratios were also obtained. Eighty-one experimental ratios showed RSD $< 6\%$ (Table 6). As compared to data listed in Table 2 (average RSD $= 20.3\%$), the use of volatile concentration ratios gave rise to a significant improvement in the precision of experimental data. The high number of ratios considered was ruled out as the reason for the improvement of precision, as only 2 simulated ratios

Table 6
Relative standard deviation (RSD, %) for percent concentration ratios (vx/vy) calculated from experimental WBLACK data. For identification of volatiles vx and vy, see Table 2.

vx	*vy*	RSD (%)
1	24	5.8
1	49	5.5
2	7	4.2
2	11	5.2
3	24	5.6
3	37	6.0
3	38	5.8
3	51	4.9
7	2	4.3
7	11	2.7
8	13	3.9
8	23	5.9
8	35	5.4
11	2	5.1
11	7	2.8
13	8	3.9
14	18	5.7
14	19	5.0
14	21	4.1
15	26	2.5
16	46	5.8
16	51	4.8
18	14	5.8
19	14	5.0
19	22	5.0
19	24	5.3
21	1	5.7
21	14	4.2
21	30	5.5
21	41	5.2
22	19	4.8
22	24	4.4
23	8	6.0
23	27	5.3
24	1	5.8
24	3	5.7
24	19	5.5
24	22	4.3
24	29	5.2
24	30	5.9
24	51	5.4
26	15	2.6
27	23	5.4
27	40	5.9
29	24	5.2
30	21	5.5
30	38	5.7
30	41	4.3
35	8	5.5
37	3	6.0
38	3	5.9
38	30	5.4
38	41	5.6
38	51	4.5
40	27	5.8
41	21	5.2
41	30	4.3
41	38	5.9
41	48	4.6
41	54	5.6
48	41	4.7
48	58	5.9
48	59	6.0
49	1	5.5
49	58	4.1
49	59	4.7
51	3	4.9
51	16	5.0
51	24	5.2
51	38	4.4
54	41	5.8
58	48	5.8
58	49	4.0
58	59	1.4
58	62	5.6
59	48	5.8
59	49	4.6
59	58	1.4
59	62	4.7
62	58	5.4
62	59	4.6

showed RSD values lower than 7% and 5 lower than 8% (significance of differences: $p < 0.01$).

Regarding the ratios with the lowest dispersion ($< 4\%$), they were associated with compounds showing high loadings for first principal components (e.g. $v7/v11$: 1-butanol/3-methyl-1-butanol with loadings < -0.86 for PC1; $v59/v58$: 2-phenylethanol/benzyl alcohol with loadings of -0.8 for PC2; $v15/v26$: *p*-cymene/hexyl butanoate with loadings > 0.87 for PC1; $v8/v13$: 2-heptanone/ethyl hexanoate with loadings > 0.88 for PC1, etc.). All these compounds were also highly correlated ($|r| > 0.98$).

Similar results were obtained for percent concentration ratios calculated for model systems. The number of ratios with RSD $< 6\%$ was significantly ($p < 0.01$) higher for experimental rather than for simulated data (MSGW: 38 experimental vs 13 simulated, MSSUG: 58 experimental vs 24 simulated), MSCOT (42 experimental vs 35 simulated).

As regard as concentration ratios for MSGW, the lowest dispersion was obtained for ratios including correlated compounds as shown in Tables 5 and 7 (2-heptanol/1-heptanol and 6-methyl-5-hepten-2-one/2-heptanone with loadings > 0.97 for PC1, linalool/1-hexanol with loadings > 0.88 for PC1, 2-hexen-1-ol/*trans*-3-hexen-1-ol with loadings < -0.85 for PC2). For model systems MSCOT and MSSUG, the number of ratios including compounds highly correlated was lower (Tables S7 and S8). Only the ratios 6-methyl-5-hepten-2-one/2-heptanone and 1-octanol/terpinen-4-ol (from MSCOT) and 2-hexen-1-ol/*trans*-3-hexen-1-ol (from MSSUG) showed loadings with absolute value higher than 0.84 for PC1-2.

In agreement with the previously mentioned results on matrix effect, no similarity was found for the volatile concentration ratios with the lowest dispersion calculated from data matrices of WBLACK sample and model systems. The noticeably higher number of volatiles considered in the analysis of WBLACK, as compared to that of the volatiles included in model systems (62 vs 17), should give rise to a higher number of precise ratios in wild blackberry sample. However, several ratios calculated including compounds which individually showed high dispersion, such as 6-methyl-5-hepten-2-one/2-heptanone, 2-hexen-1-ol/*trans*-3-hexen-1-ol, etc., showed a high precision in the three model systems with different matrices evaluated.

At the sight of the above results, it can be concluded that a simple model system cannot be used to explain the trends in data dispersion of blackberry samples. Although a much complex model system could be prepared and evaluated for this regard, the complexity of its preparation would question its usefulness with respect to a real sample selected as representative of those to be analyzed.

3.4. Use of selected volatile concentration ratios for characterization of Italian blackberries

Studies addressed to a multicomponent characterization of samples with metabolic fingerprinting purposes would benefit

Table 7
Relative standard deviation (RSD, %) for percent concentration ratios (vx/vy) calculated from experimental MSGW data. For identification of volatiles *vx* and *vy*, see Table 3.

vx	*vy*	RSD (%)
1	5	3.0
2	3	4.8
3	2	4.9
3	13	5.5
3	15	4.4
4	6	5.5
4	9	2.8
4	11	5.1
5	1	3.0
6	4	5.2
6	7	5.0
6	8	4.1
6	9	5.0
6	11	3.8
6	12	5.6
7	6	5.0
7	8	1.9
8	6	4.2
8	7	1.9
9	4	2.9
9	6	5.1
9	10	5.0
9	11	4.9
10	9	5.0
11	4	5.1
11	6	3.8
11	8	5.9
11	9	4.9
11	12	3.1
12	6	5.4
12	8	5.9
12	11	3.2
13	3	5.2
13	14	5.2
13	15	4.8
14	13	5.6
15	3	4.4
15	13	4.6

from a filtering of those variables with a higher precision before selecting those showing a higher correlation with the sample feature to be characterized. In this sense, and as an example of application of the approach here developed, statistical analysis was applied to volatile concentration ratios with the highest precision in order to select the most significant ratios for the characterization of Italian blackberries collected in different harvest years.

Table S9 in Supplementary data lists the percent composition and precision (% RSD) data for a total of 74 volatiles determined in the SPME followed by GC–MS analysis of the fourteen Italian blackberries with different harvest year analyzed (for identification of samples, see Table 1).

Taking into account the volatile concentration ratios with the highest precision ($\text{RSD} < 7\%$) previously determined in the $n = 11$ replicates of the analysis of WBLACK sample (Table 6), these ratios were also calculated from data obtained for Italian blackberries (data matrix ITRATIOS (128 ratios × 14 samples) not shown). Stepwise Discriminant Analysis was applied to this data matrix considering the harvest year (2011 or 2012) as grouping variable. The ratio ethyl octanoate/methyl octanoate correctly classified 100% of samples (absolute value of residuals < 0.32; $r = 0.92$). Despite the low percent concentration of these two volatiles with similar structure, the precision of their ratio proved to be very useful for the differentiation of blackberry samples from different harvest years here intended.

4. Conclusions

Relative standard deviations, used as estimators of data precision, are generally assumed to be independent for the different compounds present in multicomponent mixtures. However, the statistical techniques applied in this work to SPME GC–MS semiquantitative data from blackberry volatiles, have evidenced the existence of non-random factors affecting in a similar way the dispersion of groups of compounds statistically correlated and, therefore, presumably showing a similar SPME fractionation behavior. This trend has also been confirmed for the first time by evaluating dispersion data from different model systems closely matching blackberry composition. Whereas from a theoretical point of view, studies with simple model systems could be appropriate to detect and characterize the effects involved in quantitative dispersion, the different results observed for blackberry samples and model systems appear to indicate that, for practical purposes, dispersion studies should be carried out on a sample representative of that to be analyzed.

Regarding the application of this improved SPME approach, the use of the percent concentration ratios between correlated compounds pointed out by statistical analysis has provided a better precision as compared to previous methodologies based on individual relative concentration values. On the other hand, in multicomponent GC analyses, the number of identified compounds usually exceeds to a great extent that of analyzed samples. In these cases, the number of variables (compounds) should be reduced in order to obtain significant results for discrimination or classification purposes. An objective way to “filter” non-relevant variables can be the use of those having a higher precision, followed by the selection among them of those more related with the grouping property (in this paper, harvest year of Italian blackberries). The use of this statistical approach, of general application to any sample type and/or volatile fractionation technique, may contribute to promote the use of SPME for precise quantitative applications.

Acknowledgments

This work has been funded by Ministerio de Economía y Competitividad (project CTQ2012-32957) and the Comunidad Autónoma of Madrid (Spain) and European funding from FEDER program (project S2009/AGR-1464, ANALISYC-II). A.C.S. thanks Ministerio de Economía y Competitividad of Spain for a Ramón y Cajal contract and M.F.D. thanks University of Reggio Calabria for a grant to stay abroad. Authors are also thankful to La Cuerva for providing wild blackberries.

Appendix A. Supplementary data

Supplementary data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.talanta.2014.02.058.

References

[1] D.N. Georgilopoulos, A.N. Gallois, Z. Lebensm, Unters. Forsch. 184 (1987) 374.
[2] A.L. Morales, D. Albarracín, J. Rodríguez, C. Duque, J. High Resol. Chromatogr. 19 (1996) 585.
[3] E. Ibáñez, S. López-Sebastián, E. Ramos, J. Tabera, G. Reglero, Food Chem. 63 (1998) 281.
[4] X.F. Du, A. Kurnianta, M. McDaniel, C.E. Finn, M.C. Qian, Food Chem. 121 (2010) 1080.
[5] G.P. Blanch, G. Flores, M.L. Ruiz del Castillo, Postharvest Biol. Technol. 62 (2011) 168.
[6] M. Meret, P. Brat, C. Mertz, M. Lebrun, Z. Günata, Food Res. Int. 44 (2011) 54.

[7] H. Kataoka, H.L. Lord, J. Pawliszyn, J. Chromatogr. A 880 (2000) 35.
[8] H.H. Jeleń, M. Majcher, M. Dziadas, Anal. Chim. Acta 738 (2012) 13.
[9] M.F. D'Agostino, J. Sanz, M.L.Sanz, A.M. Giuffrè, V. Sicari, A.C. Soria, 2014, Food Chem. (submitted for publication).
[10] S.Y. Sun, W.G. Jiang, Y.P. Zhao, Flavour Frag. J. 25 (2010) 206.
[11] R.Q. Alvarez, C.C. Passaro, O.G. Lara, J.L. Londono, Procedia Food Sci. 1 (2011) 1396.
[12] X. Yang, T. Peppard, J. Agric. Food Chem. 42 (1994) 1925.
[13] R. Costa, L. Tedone, S. De Grazia, P. Dugo, L. Mondello, Anal. Chim. Acta 770 (2013) 1.
[14] A.C. Soria, M.A. García, I. Martínez-Castro, J. Sanz, J. Chromatogr. A 1008 (2003) 105.
[15] A.C. Soria, I. Martínez-Castro, J. Sanz, J. Chromatogr. A 1216 (2009) 3300.
[16] S.W. Souci, W. Fachmann, H. Kraut, Food Composition and Nutrition Tables, Medpharm Scientific Publishers, Stuttgart (2000) 1057.
[17] F.W. McLafferty, D.B. Stauffe, The Wiley/NBS Registry of Mass Spectral Data, Wiley, New York, 1989.
[18] H. van Den Dool, P.D. Kratz, J. Chromatogr. A 11 (1963) 463.
[19] R.W. Hamming, Numerical Methods For Scientists and Engineers, McGraw-Hill, New York, 1962.
[20] StatSoft, Inc., 2005. STATISTICA (Data Analysis Software System), version 7.1. ⟨www.statsoft.com⟩.
[21] Supelco Bulletin 929, (2001) 1.