Average Mass Scan of the Total Ion Chromatogram *versus* Percentage Chemical Composition in Multivariate Statistical Comparison of Complex Volatile Mixtures

Niko S. Radulović,[*,a] *Polina D. Blagojević*[a] *and Danielle Skropeta*[b]

[a]*Department of Chemistry, Faculty of Science and Mathematics, University of Niš, Višegradska 33, 18000 Niš, Serbia*

[b]*School of Chemistry, University of Wollongong, Wollongong NSW 2522, Australia*

A análise de misturas voláteis complexas usando cromatografia gasosa-espectrometria de massa (CG-EM) é um processo demorado. Este processo envolve a separação e identificação de componentes baseados em seus tempos de retenção e padrões de fragmentação, seguido da determinação de suas porcentagens relativas a partir da integração das áreas dos picos. No presente trabalho mostramos que análise estatística multi-variada das abundâncias relativas dos valores *m/z* obtidos a partir das varreduras médias de massa (VMM) da mistura complexa é um método mais rápido e potencialmente mais confiável de avaliação destas misturas. Foram preparadas 15 misturas modelo complexas contendo quantidades variadas de 10 constituintes diferentes. O perfil VMM e a composição química de cada mistura foram comparados uns aos outros utilizando análise de cluster hierárquica aglomerante e análise de componentes principais. Os resultados obtidos indicam fortemente que análise estatística multi-variada dos perfis VMM é uma ferramenta promissora, eficiente e confiável para analisar misturas voláteis complexas, em particular óleos essenciais.

The analysis of complex volatile mixtures by gas chromatography-mass spectrometry (GC-MS) is a time-consuming process. It involves separation and identification of the components based on their retention times and fragmentation patterns, followed by determination of their relative percentages from integration of their peak areas. Herein we show that multivariate statistical analysis of the relative abundances of the *m/z* values obtained from the average mass scans (AMS) of the complex mixture is a faster and potentially more reliable method of assessing these mixtures. To achieve this, 15 model complex mixtures, were prepared comprising varying amounts of 10 different constituents. The AMS profile and chemical composition of each mixture were compared to one another using agglomerative hierarchical cluster analysis and principal component analysis. The results obtained strongly suggest that multivariate statistical analysis of AMS profiles is a promising, time saving and reliable tool for analyzing complex volatile mixtures, in particular essential oils.

Keywords: gas chromatography-mass spectrometry, multivariate statistical analysis, complex mixtures, average mass scan, percentage composition, total ion chromatogram

Introduction

Over the last two decades, multivariate statistical analysis (MSA) has become an important tool for the evaluation of a wide variety of complex samples, enabling the comparison of large data sets. In recent years, MSA has been successfully employed in the analysis of air pollution components,[1] in the area of food control,[2] and in the analysis of different biological[3] and geological samples.[4] Moreover, it has been shown that automated MSA tools are capable of rapidly converting complicated non-mass selected data sets into a handful of chemical components that are much easier to interpret.[5] MSA of mass spectra has also been used as a tool for the clarification of the main humic substances according to their structural and conformational features.[6] Mass to charge values (*m/z*) from the mass spectra of different amber samples were also used as variables in MSA to provide useful information on amber age.[7]

Multivariate statistical comparison of volatile plant secondary metabolites could be used as a promising tool in a variety of areas including revealing evolutionary relationships among different plant species[8,9] and for tracking storage effects in the case of economically and/or pharmacologically important essential oils. However, prior to performing a MSA comparison, time-consuming interpretation of the results of

*e-mail: vangelis0703@yahoo.com

chemical analysis of the complex volatile mixtures (typically from GC and GC-MS) is required. In some cases, the identity of certain volatiles cannot be ascertained and in other cases, constituents may be misidentified. There are further difficulties when comparing data obtained by different researchers due to varying experimental conditions.[10] Although theoretically a peak eluting at a given retention time (R_t) on specified equipment should represent the same compound, practically this is not always the case when comparing data from different GC instruments, and even less so when comparing literature data. In addition, the unambiguous quantification of components of a mixture based on integration of peak areas is often not possible due to incomplete separation of the component peaks.

Multivariate analysis of the mass spectra of complex volatile mixtures typically uses the percentages of the individual constituents of the mixture obtained from integration of the peak areas from the GC chromatogram as one of the variables.[8] However, this method suffers from the drawbacks mentioned above. An alternative method is to use the relative abundances of the *m/z* values obtained from the average mass scans (AMS) of the total GC ion chromatogram of the mixture. These represent the average response of the MS detector in a given timeframe. The relative abundances of the AMS *m/z* values correspond to the arithmetic mean for a given timeframe and account for both the relative abundances of ions in individual mass spectra, as well as the relative percentages of the corresponding mixture components. It should be noted that this is not an average mass spectrum of the mixture, as this would result in a loss of the information about the relative percentages of the mixture components. The AMS method described above has the potential to greatly facilitate the multivariate analysis of complex volatile mixtures, making it both more reliable and faster.

Herein we describe the use of the relative abundances of the AMS *m/z* values from the GC total ion chromatograms compared to the use of the percentages of the individual mixture constituents as variables in the MSA of complex volatile mixtures. To achieve this, the chemical composition of 15 model complex mixtures consisting of 10 different constituents of varying percentage compositions, were individually compared using agglomerative hierarchical cluster (AHC) analysis and principal component analysis (PCA).

Experimental

Preparation of the model complex mixtures

The following 10 compounds, obtained from either Merck (Darmstadt, Germany) or Sigma-Aldrich (St. Louis,

Missouri, USA), were used for the preparation of the model complex mixtures: α-pinene, camphor, methyl anthranilate, (*E*)-α-ionone, eugenol, 1-bromonaphthalene, anthracene, hexadecanoic acid, sulfur (S_8) and cholesterol. Except for 1-bromonaphthalene, anthracene and sulfur, all substances are representatives of different (volatile) natural product classes (*e.g.*, terpenoids, phenylpropanoids) and are often identified in real samples.

Saturated solutions of anthracene and sulfur in 10 mL of chloroform (Sigma-Aldrich, St. Louis, Missouri, USA) were employed as the stock solutions. For all other compounds, stock solutions were prepared by dissolving 1 g of the corresponding substance in 10 mL of chloroform. The 15 model complex mixtures, designated as M1 to M15, were prepared with similar percentage compositions of the constituents by combining different volumes (*e.g.* 0, 250, 500, 750 or 1000 μL) of the stock solutions, which were subsequently diluted to 50 mL with chloroform. All of the model complex mixtures were analyzed by both GC and GC-MS.

GC and GC-MS

The GC-MS analyses of the model complex mixtures were carried out in triplicate using a Hewlett-Packard 6890N gas chromatograph equipped with a fused silica capillary column HP-5MS (5% phenylmethylsiloxane, 30 m × 0.25 mm, film thickness 0.25 μm, Agilent Technologies, USA) and coupled with a 5975B mass selective detector from the same company. The injector and interface were operated at 250 and 300 ºC, respectively. The oven temperature was raised from 80 to 290 ºC at a heating rate of 10 ºC min⁻¹ and then held isothermally for 10 min. Helium (99.999%, Messer Tehnogas, Serbia) at 1.3 mL min⁻¹ was used as the carrier gas. The samples (1 μL of the mixtures prepared as described above) were injected in a pulsed split mode with the flow at 1.5 mL min⁻¹ for the first 0.5 min and then set to 1.3 mL min⁻¹ throughout the remainder of the analysis (split ratio 31:1). The MS (electron impact) conditions were as follows: ionization voltage of 70 eV, acquisition mass range 35-500, scan time 0.32 s. GC (FID) analyses were carried out under the same experimental conditions using the same column as described for GC-MS. The percentage compositions of the constituents of the model complex mixtures were computed from the GC peak areas without any corrections.

EI-MS *m/z* (relative intensity in %) of the compounds used for the preparation of the model complex mixtures are as follows:

α-Pinene: 136(8, [M⁺]), 121(12), 119(3), 107(5), 94(10), 93(100), 92(38), 91(43), 81(3), 80(9), 79(24),

78(5), 77(30), 68(3), 67(8), 65(6), 55(5), 53(7), 51(5), 43(4), 41(12), 39(12).

Camphor: 152(29, [M⁺]), 137(5), 110(11), 109(32), 108(39), 97(5), 96(10), 95(100), 93(15), 83(29), 82(12), 81(68), 80(13), 79(9), 77(8), 70(3), 69(31), 68(14), 67(26), 65(5), 55(29), 53(12), 51(5), 43(6), 42(4), 41(38), 39(21).

Methyl anthranilate: 151(61, [M⁺]), 120(32), 119(100), 93(5), 92(55), 91(9), 90(3), 66(4), 65(24), 64(8), 63(9), 62(3), 52(5), 39(8).

Eugenol: 164(100, [M⁺]), 163(4), 150(3), 149(33), 147(5), 137(20), 133(18), 132(9), 131(29), 122(8), 121(16), 115(3), 107(4), 105(9), 104(19), 103(30), 102(3), 94(8), 93(5), 92(3), 91(26), 89(3), 79(3), 78(9), 77(31), 66(5), 65(11), 63(6), 62(5), 55(16), 53(7), 52(5), 51(11), 50(4), 43(3), 41(4), 39(10).

(*E*)-α-Ionone: 192(15, [M⁺]), 177(8), 149(4), 137(8), 136(42), 135(5), 134(3), 122(10), 121(100), 119(5), 109(20), 107(7), 105(5), 95(4), 94(7), 93(65), 92(11), 91(30), 81(5), 79(9), 78(5), 77(23), 69(3), 67(4), 65(6), 55(6), 53(6), 51(4), 43(31), 41(13), 39(8).

1-Bromonaphthalene: 208(98, [M⁺]), 206(100), 128(11), 127(99), 126(29), 125(3), 104(8), 103(8), 101(9), 100(3), 99(4), 98(5), 87(4), 86(3), 77(12), 76(6), 75(12), 74(12), 63(18), 62(5), 61(3), 51(6), 50(7).

Anthracene: 178(100, [M⁺]), 177(9), 176(18), 175(3), 152(7), 151(7), 150(5), 89(11), 88(7), 76(9), 75(3), 63(3).

Hexadecanoic acid: 256(23, [M⁺]), 227(4), 213(24), 199(5), 185(13), 171(14), 157(14), 143(6), 129(41), 115(16), 111(8), 101(10), 99(7), 98(12), 97(22), 96(4), 87(20), 85(23), 84(11), 83(24), 82(4), 81(4), 74(9), 73(100), 71(36), 70(10), 69(34), 68(5), 67(7), 61(20), 60(78), 57(57), 56(15), 55(57), 53(3), 45(9), 43(69), 42(13), 41(55), 39(9).

Sulfur (S₈): 256(72, [M⁺]), 224(3), 194(8), 192(29), 162(12), 160(53), 130(11), 128(57), 98(4), 96(26), 66(11), 65(3), 64(100).

Cholesterol: 386(100, [M⁺]), 371(44), 369(22), 368(51), 353(46), 301(68), 275(81), 255(39), 247(20), 231(27), 213(55), 201(10), 199(22), 187(17), 185(18), 163(35), 161(45), 159(49), 145(75), 144(13), 143(35), 137(10), 136(10), 135(34), 134(19), 133(48), 132(14), 131(45), 130(11), 129(20), 128(7), 125(11), 124(11), 123(24), 122(13), 121(46), 120(41), 119(54), 107(71), 105(81), 95(66), 93(66), 91(66), 85 81(75), 80(8), 79(56), 71(29), 69(42), 67(46), 57(62), 55(74), 43(97), 41(56).

Results and Discussion

The percentage compositions of the 15 model complex mixtures (M1-M15) obtained from GC and GC-MS analyses are given in Table 1. The mixture components and their relative composition were carefully selected to comprise non-overlapping peaks that covered almost the entire span of the chromatogram. This would avoid any ambiguous quantification based on the integration of non-resolved peaks. The identity of the component itself was unimportant but its retention time, chromatographic behavior and mass spectral fragmentation characteristics were taken into consideration. For example, constituents displaying peak tailing such as palmitic acid and cholesterol were included, in order to assess how this would affect the integration of the peak areas. The extent of fragmentation of a compound could also have an impact on the AMS. Thus, aromatics and S₈ which have intense molecular ions, but a relatively small number of fragment ions were included. Conversely, aliphatics such as palmitic acid and cholesterol, which have extensive fragment ions, were also employed. The 10 components that were selected also differed in their number of oxygen atoms and the presence of elements that impact mass spectra. For example, nitrogen-containing compounds such as methyl anthranilate were included which give rise to even numbered fragment ions, along with compounds displaying intense isotopic ions such as 1-bromonaphthalene. Compounds exhibiting abundant ions with the same *m/z* values in their mass spectra such as cholesterol / α-pinene (*m/z* 93) and cholesterol / camphor (*m/z* 81) were chosen. Furthermore, 1-bromonaphthalene was employed as an example with intense fragment ions with the same *m/z* value as those of column bleed peaks, *e.g. m/z* 208, which has the same value as the ¹³C isotope of the column bleed fragment at *m/z* 207. Finally, since the percentage composition of the various constituents in related environmental and other natural source samples are usually quite similar to one another, the mixtures were prepared to reflect this as well. Low concentrations of the injected mixture solutions were intended to demonstrate the impact of column bleed and other contaminants that are often observed by GC.

The AMS of all 10 mixtures (M1-M15) were obtained directly from the ChemStation as an average of 2.15 to 24.80 min and present the arithmetic average value of the abundances of each ion recorded by the mass selective detector in the given timeframe, rounded to a nominal mass (35-500 amu). Large solvent peaks appearing up to a R$_t$ of 2 min were not recorded. The duration of a single run was 31 min with the last peak apex appearing at R$_t$ 23.92 min (cholesterol). After R$_t$ 24.80 min no further ions corresponding to cholesterol were detected and the interval between R$_t$ 24.80-31.00 mins was not taken into account to lessen the effect of column bleed peaks. The relative abundances of the AMS *m/z* peaks are given in percentages, with 100% assigned to the most abundant peak in every AMS, and the percentages of all other peaks

Table 1. Percentage compositions of 15 model complex mixtures (M1-M15) obtained as a result of GC and GC-MS analyses

	A (2.61)	B (4.71)	C (7.11)	D (7.27)	E (8.15)	F (8.87)	G (12.41)	H (13.94)	I (14.95)	J (23.92)
	Compound[a] (R$_t$ min)									
	Composition %									
M1	23.5	16.6	12.9	12.1	13.4	17.4	3.2	0.0	0.4	0.0
M2	13.8	20.7	9.4	9.0	9.5	23.8	6.9	0.0	1.0	5.7
M3	18.3	9.0	16.5	16.4	17.0	11.1	3.3	2.7	0.7	4.5
M4	6.4	16.7	6.5	6.5	7.0	22.8	8.5	0.8	1.7	21.6
M5	8.4	15.4	16.6	8.1	8.7	20.4	6.3	2.5	0.6	12.0
M6	11.4	10.4	10.6	10.1	11.0	13.4	9.3	1.3	1.7	20.0
M7	14.5	12.1	12.5	11.8	12.9	15.9	9.1	1.0	1.8	7.9
M8	11.7	11.9	12.6	12.6	12.9	15.8	2.8	4.3	1.1	12.8
M9	9.1	14.8	14.1	13.4	14.1	17.1	5.1	1.4	0.9	8.9
M10	13.7	12.3	12.6	12.3	13.5	16.7	5.5	1.5	2.0	9.0
M11	6.1	13.6	15.2	5.6	6.0	18.7	7.8	6.0	1.5	18.3
M12	10.0	10.0	10.5	10.4	11.0	13.4	8.9	2.3	0.8	21.9
M13	5.3	12.5	12.6	11.7	12.4	15.2	5.6	3.6	1.0	18.7
M14	20.2	15.8	13.4	13.2	14.4	16.8	4.1	0.0	2.1	0.0
M15	11.0	15.7	14.2	7.0	7.4	17.3	8.4	1.0	2.2	15.8

[a]A = α-pinene; B = camphor; C = methyl anthranilate, D = eugenol, E = (*E*)-α-ionone, F = 1-bromonaphthalene, G = anthracene, H = palmitic acid, I = sulphur, J = cholesterol; R$_t$ = retention time in minutes.

given as relative to the AMS base peak. The *m/z* values that corresponded to column bleed peaks (*m/z* 281 and 207), carbon dioxide (*m/z* 44), and argon (*m/z* 40, contaminant from the carrier gas) were excluded from the final table used for the MSA, as well as those of peaks of less than 10% relative abundance. The relative abundances of the isotopic peak (*m/z* 208) corresponding to one ^{13}C atom isotopologue of the fragment *m/z* 207 was subtracted from the total abundance of *m/z* 208 (in the amount of one fifth of the relative abundance of *m/z* 207). In order to test if this subtraction influences the MSA we have chosen 1-bromonaphthalene (with *m/z* 208 as one of the dominant MS fragment ions) as one of the mixture constituents. In order to simplify the discussion, Table 2 lists only the *m/z* values and their relative abundances that correspond to the characteristic peaks from the AMS of the total GC chromatograms of the analyzed mixtures.

Principal component analysis (PCA) and agglomerative hierarchical clustering (AHC) on the 15 model complex mixtures were both performed using the Excel program plug-in XLSTAT version 2008.6.07. Both methods were applied utilizing two different sets of variables: the original variables based on the mean values of the percentage composition of the mixture components obtained from

the GC experiments (Table 1; AHC1, PCA1) and the AMS variables based on the relative abundances of the AMS *m/z* values from the total ion chromatogram (TIC) (Table 2; AHC2, PCA2). AHC was determined using Pearson dissimilarity where the aggregation criterion were simple linkage, unweighted pair-group average and complete linkage and Euclidean distance where the aggregation criterion were weighted pair-group average, unweighted pair-group average and Ward's method. PCA of the Pearson (n) type was performed.
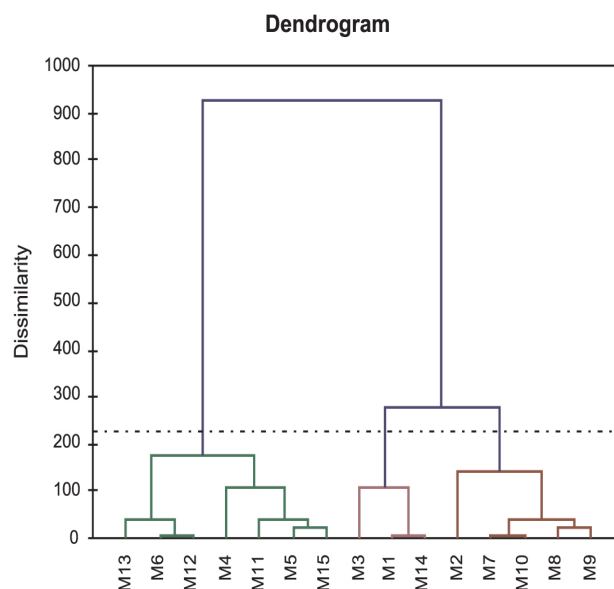

Figure 1. Dendrogram (AHC analysis) representing chemical composition (the original variables) dissimilarity relationships of 15 model complex mixtures (observations) obtained by Euclidian distance dissimilarity (dissimilarity within the interval [0, 930], using aggregation criterion-Ward's method). Three groups of mixtures, further divided to subclades and subsubclades, were found (from left to right): C1 (subclades SC1.1 and SC1.2 (subsubclades SC1.2.1 and SC1.2.2)), C2 and C3 (subclades SC3.1 and SC3.2).

AHC analyses

The AHC analysis obtained using the original percentage composition variables (AHC1) is shown in Figure 1, while the analysis using the relative abundances of the AMS *m/z* values as variables (AHC2) is shown in Figure 2. In both dendrograms, three different classes of mixtures can be observed, *e.g.*, classes C1-C3 in Figure 1 and C4-C6 in Figure 2. From the two AHC analyses performed, a significant level of similarity of the model complex mixture clustering is observed. For example, both dendrograms place the mixtures M5, M15 and M11 within the same clades (*i.e.,* subsubclade SSC1.2.2 in Figure 1 and subsubclade SSC6.2.1 in Figure 2), with samples M5 and M15 characterized by the low Euclidian

Table 2. Some of the characteristic AMS *m/z* values (relative intensities %) used as variables for PCA and AHC analyses

| *m/z*	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12	M13	M14	M15
39	25	32	26	29	30	28	28	32	34	29	29	30	35	26	33
41	27	41	29	44	40	38	33	42	43	35	44	42	50	30	45
43	17	24	27	33	30	36	28	38	36	30	36	41	49	22	33
77	45	52	49	47	44	51	50	55	57	51	39	52	56	47	50
81	26	50	22	56	43	38	32	38	45	33	45	41	51	29	51
91	56	60	64	53	54	66	62	67	68	64	50	68	70	58	64
93	100	100	100	76	79	100	100	100	100	100	68	100	91	100	100
95	36	67	26	73	57	46	40	47	58	42	58	49	64	38	66
109	16	28	18	32	26	25	22	26	31	23	26	27	35	19	28
119	52	57	71	53	100	72	69	78	92	71	100	77	100	59	99
121	45	49	60	45	47	60	56	64	71	60	39	64	75	50	48
127	42	86	28	100	76	51	51	58	67	55	78	55	71	42	73
164	28	31	41	27	28	37	37	44	51	39	22	41	52	33	28
206	42	85	28	99	76	50	51	57	66	55	77	54	70	42	73
208	35	79	26	93	71	47	47	53	62	50	73	50	65	38	67
368	0	3	1	11	5	9	3	6	4	4	9	11	11	1	8
386	0	3	2	13	6	11	4	7	5	5	11	13	13	1	10
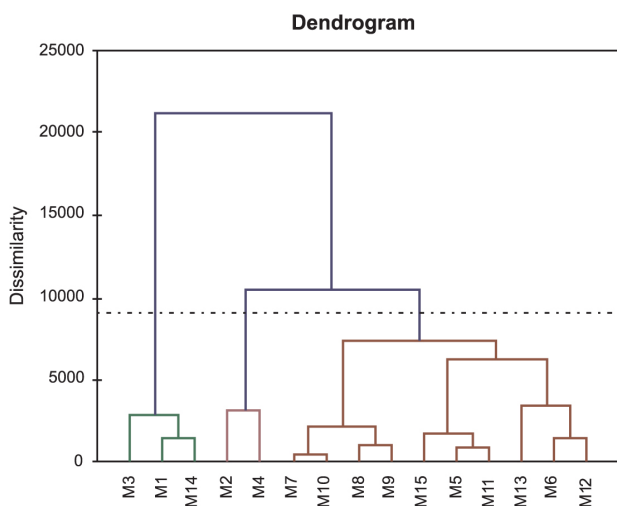
Figure 2. Dendrogram (AHC analysis) representing the average mass scans of the TIC (AMS relative abundances of m/z values as variables) dissimilarity relationships of 15 model 'complex mixtures' (observations) obtained by Euclidian distance dissimilarity (dissimilarity within the interval [0, 21200], using aggregation criterion-Ward's method). Three groups of mixtures, further divided to subclades and subsubclades, were found (from left to right): C4, C5 and C6 (subclades SC6.1 and SC6.2 (subsubclades SSC6.2.1 and SSC6.2.2)).

distance. The percentage composition of the mixtures M5 and M15 was very similar, with 1-bromonaphthalene as the main constituent (20.4% for M5 and 17.3% for M15), and comparable amounts of the other dominant constituents: methyl anthranilate, cholesterol, α-pinene and camphor (Table 1). The ratios of the representative constituents of M5 and M15 are also highly similar, for example in

both mixtures the ratio of 1-bromonaphthalene to methyl anthranilate is 1.2:1. Interestingly, 1-bromonaphthalene, methyl anthranilate, cholesterol and camphor were also the dominant contributors in M11 (Table 1). However, in M11 the relative amount of cholesterol was higher and the level of α-pinene lower than that of the M5 and M15 mixtures, with the amount of α-pinene almost halved in comparison to M15. These differences may well account for the higher degree of dissimilarity of M11 to that of M5 and M15.

The mixture M4 is placed within the same clade with M5, M11 and M15 (subclade SC1.2, Figure 1) but is distinguished from them by having a considerably lower amount of methyl anthranilate (Table 1). In comparison to all other mixtures, M4, M5, M11 and M15 had lower levels of α-pinene, eugenol and (*E*)-α-ionone. In regard to the mixtures M5, M15 and M11, the results of AHC2 are in agreement with those of AHC1. As expected, the dominant *m/z* values in the AMS of mixtures M5, M15, M4 and M11 correspond to those of the most abundant ions in the mass spectra of the major contributors of each mixture (Table 2). Thus, the AMS base peak of M5 and M11, and the second most abundant AMS peak of M15 was *m/z* 119, which corresponds to the base peak in the mass spectrum of methyl anthranilate. Other *m/z* values that corresponded to abundant fragment ions observed in the AMS of M5, M15 and M11 were 127, 206 and 208 from 1-bromonaphthalene; 81 and 95 from camphor; and 91 and 93 from α-pinene (Table 2). The AMS base peak of M15 at *m/z* 93 was only slightly more abundant than *m/z*

119 (100% *vs.* 99%), and corresponded to the base peak in the mass spectrum of α-pinene. This could be attributable to the relative percentage of α-pinene, which is higher in M15 than in both M5 and M11.

Mixtures M5, M15 and M11 represented a separate clade of the class C6 (subsubclade SSC6.2.1, Figure 2). The dendrogram obtained from the AHC1 analysis shows that the mixtures M4, M5, M11 and M15, although mutually still strongly related, are placed together within the same class with the mixtures M6, M12 and M13 (class C1, Figure 1). These latter samples had cholesterol as the main constituent at 20.0, 21.9 and 18.7% respectively for M6, M12 ad M13, and slightly lower percentages of 1-bromonaphthalene, methyl anthranilate, and camphor. They also had slightly higher amounts of eugenol, (*E*)-α-ionone and α-pinene (except for M13, in which the level of α-pinene was lower than in samples M5, M15, M4 and M11).

Based on the AHC2 analysis, samples M6, M12 and M13 were placed in a separate subsubclade of the class C6 (subsubclade SSC6.2.2, Figure 2), along with M5, M11 and M15. Once again, the AMS of M6, M12 and M13 were characterized by the high relative abundances of the *m/z* values that corresponded to the predominant peaks in the mass spectra of the main constituents (Table 2). Mixtures M6, M12 and M13 had amongst the highest relative abundances of the *m/z* values related to the cholesterol peaks (*m/z* 368, 386) compared to all the other mixtures. Incidentally, these peaks were not among the most abundant ones, probably due to the extensive fragmentation characteristic for the mass spectrum of cholesterol (Table 2). The separation of mixtures M6, M12 and M13 to a different subsubclade from that of samples M5, M11 and M15, within the same class (class C6, Figure 2) may also be attributable to the relative abundances of the following *m/z* values, which were more significant in the AMS of M6, M12 and M13: *m/z* 91 and 93 (α-pinene); *m/z* 164 (eugenol); and *m/z* 77 and 121 ((*E*)-α-ionone). In addition the relative abundances of the *m/z* values corresponding to 1-bromonaphthalene (*m/z* 127, 206, 208) were all lower for M6, M12 and M13 relative to those for M5, M11 and M15. Within the class C6, the samples M5, M11, M15, along with M6, M12 and M13 were all placed in same subclade SC6.2 (Figure 2).

The other two classes of mixtures resulting from the AHC1 analysis were mixtures M2, M7, M8, M9, and M10 in class C3 and mixtures M1, M3 and M14 in class C2 (Figure 1). All mixtures from the C2 class contained α-pinene as the main constituent (Table 1). Additionally, cholesterol and palmitic acid were completely absent from the compositional analysis of samples M1 and M14, which in turn were more similar to each other than to the third

member of the same class, M3. The dendrogram based on the AHC2 analysis also shows the samples M1, M3 and M14 represented as a separate class, C4 (Figure 2). In accordance with these results, the AMS of all of the class C2 samples were characterized by high abundances of the *m/z* values 91 and 93 related to α-pinene, and an almost complete lack of those corresponding to cholesterol and palmitic acid.

1-Bromonaphthalene was found to be the main constituent of the mixtures M2, M7, M8, M9, and M10 from class C3, as well as for the mixtures M4, M5, M11, and M15 from class C1 (Figure 1; AHC1). Inspite of that similarity, the subclade that is comprised of samples M4, M5, M11 and M15 was clearly separated from class C3 on the corresponding dendrogram (Figure 1), possibly due to the lower cholesterol levels in the class C3 mixtures compared to those in C1. Nevertheless, based on the results of the AHC2 analysis, M7, M8, M9 and M10 were grouped in the subclade SC6.1 of the same class (class C6) with other mixtures possessing high levels of both 1-bromonaphthalene and cholesterol (class C6, subclade SC2: M5, M11, M15, M6, M12 and M13) (Figure 2).

On the AHC2 dendrogram given in Figure 2, mixtures M2 (SSC3.2.1, Figure 1, AHC1) and M4 (SSC1.2.1, Figure 1, AHC1) were placed in a separate class (C5). Both mixtures are characterized by the highest percentage composition of 1-bromonaphthalene (23.8 and 22.8% respectively) and the lowest relative amount of methyl anthranilate (9.4 and 6.5% respectively) compared to all other model mixtures (Table 1). Accordingly, the AMS of both samples had the highest relative abundances of the *m/z* values related to the representative 1-bromonaphthalene peaks (*m/z* 127, 206, 208). However, the same type of correlation was not observed for *m/z* values related to methyl anthranilate.

Overall, the three classes of mixtures observed from the AHC1 dendrogram in Figure 1 could be defined as: the 1-bromonaphthalene-cholesterol class C1, with a subclade comprising samples with 1-bromonaphthalene as the main contributor (SC1.2) and another subclade with cholesterol as the main contributor (SC1.1); the α-pinene class C2; and the 1-bromonaphthalene-methyl anthranilate class C3. From the AHC2 analysis, one can statistically differentiate (at the given confidence level) three classes of mixtures: the α-pinene class C4; the 1-bromonaphthalene-low methyl anthranilate class C5; and the 1-bromonaphthalene-cholesterol-methyl anthranilate class C6, with the subclades SC6.1 (1-bromonaphthalene-methyl anthranilate) and SC6.2 (1-bromonaphthalene-cholesterol). Table 3 summarizes the correspondence of the classes (and subclasses) of the two different AHC analyses.

Table 3. Summarized results of AHC analyses using both sets of variables (original variables based on percentage composition - AHC1 and relative abundances of the *m/z* values obtained from average mass scans - AHC2)

Mixtures	Variables	
	Original variables (AHC1)	Average Mass Scans (AHC2)
M6, M12, M13	C1/SC1.1	C6/SC6.2/SSC6.2.2
M5, M11, M15	C1/SC1.2/SSC1.2.2	C6/SC6.2/SSC6.2.1
M7, M8, M9, M10	C3/SC3.2	C6/SC6.1
M1, M3, M14	C2	C4
M2, M4	C3/SC3.1 + C1/SC1.2/SSC1.2.1	C5

C-class; SC-subclade; SSC-subsubclade.

Principal component analysis (PCA)

The PCA results are in general agreement with those obtained by the corresponding AHC analyses. As expected, in the PCA correlation matrix obtained using AMS *m/z* values as variables, strong correlations between the variables (*m/z* values) originating from the fragmentation of a single compound were found, *e.g.* 1.00 coefficient of correlation for 1-bromonaphthalene peaks with *m/z* 127 and 206; 0.998 for 1-bromonaphthalene peaks with *m/z* 206 and 208; and 0.999 for cholesterol peaks with *m/z* 368 and 386, etc. However, this strong correlation is lost in the cases of those *m/z* values that originate from the fragmentation patterns of more than one compound. One of the reasons 1-bromonaphthalene was chosen as the constituent was the validation of the subtraction of certain *m/z* values (their relative contribution in %) from the AMS. In the case where the ^{13}C isotopic peak of the column bleed fragment (*m/z* 208) was not omitted from the total relative abundance of *m/z* 208, the very strong correlation described above between *m/z* 206 and 208 was reduced to 0.504. Furthermore, quite significant alterations to the net result of both the PCA and AHC analyses arise, highlighting the significance of this manipulation of variables.

Conclusions

In summary, two different data sets based on the percentage composition of mixture constituents and the relative abundances of AMS *m/z* values were used as variables in the multivariate statistical comparison of 15 model complex mixtures. The results obtained reveal a significantly high degree of similarity (Table 3). In both of the dendrograms from the AHC analyses, the model complex mixtures were mutually grouped in almost the same fashion. The only major difference between the two dendrograms was the separation of mixtures M2 and M4 into a separate class (C5). This is not entirely unexpected

since mixtures M2 and M4 show a certain degree of mutual similarity and are also very similar to the mixtures in classes C1 and C3, respectively.

These results demonstrate that multivariate analysis based on AMS data is a promising, time-saving tool for the comparison of complex mixtures such as essential oils. Moreover, the use of the relative abundances of the AMS *m/z* values as variables, rather than percentage compositions based on peak areas, has the potential to eliminate many of the shortcomings related to the direct application of data obtained from different research laboratories and/or instruments.[8] Multivariate analysis of complex mixtures based on R_t values and integration of peak areas, are hampered by the very frequent event of close peak elution (or co-elution), which can lead to erroneous integration results. This problem can be overcome by utilizing AMS, since it is not the elution time that is important, but rather the contributing fragmentation patterns of the different compounds. In addition, the *m/z* values that correspond to common contaminants such as butylated hydroxytoluene (*m/z* 205 and 220) and various phthalates (*m/z* 149) can be easily omitted and/or subtracted from the data sets to be used for MSA.

As the PCA analysis showed, some of the AMS ions that originate from the fragmentation of a single compound are well correlated. As these ions essentially convey the same sample information, they could be omitted from the AMS data set. This reduction step, by excluding the highly related ions, might represent a possible extension of this method. However, it also introduces an additional step that lengthens the method and compared to the speed of directly employing the AMS, it may be necessary only for very large data sets.

When considering the applicability of using the AMS of a set of mixtures possessing more significant variations in the ratios of the constituents, one can differentiate two cases. The first is an extension of the work described here, but where there is a much broader range of variation in the concentration of the constituents, provided they are qualitatively but not significantly different. Since we are assuming that a dependence between the percentage composition and the MS fragmentation patterns exists, this should not alter the outcome of the MSA (*i.e.*, the variable value, if not zero, does not modify the function itself). The second case is when this dependence changes due to the inclusion of additional compounds where their MS fragmentation may qualitatively transform the function in question. Since the MS fragmentation patterns of a number of different compounds are very similar, one could argue that there may be a significant loss of information due to the loss of the identity of the compound itself in the

MSA. However, it is this similarity in their mass spectra, especially when dealing with biological and natural product samples, that transmits information about the possible biogenetic resemblance of the substances in question. Information, which is not present in the original data set based on percentage compositions. Multivariate analysis is of greatest use when there is a need to compare data sets that are mutually very alike, when an at first glace inspection of the (dis)similarity is not feasible. In fact, this provided the impetus for choosing to perform this detailed study of strongly related mixtures.

The AMS approach to multivariate analysis does not appear to be applicable to mixtures consisting solely of homologues and/or isomers with very similar mass spectral fragmentation (*e.g.* mixtures of alkanes). But, these, if of a natural origin, share usually a common biosynthetic starting point (a class of compounds for example) and may still be useful since compound class and not only individual compounds have also been shown to be significant markers.[8] However, a potential solution to this problem is the use of several average scans of defined time intervals instead of a single total average scan, which will be explored elsewhere.

In the analysis of complex volatile mixtures, the inclusion of the AMS data of the total ion chromatogram, along side the tables of identified constituents and their relative percentages, would be of great assistance. It would facilitate the creation and comparison of large data sets and provide a way for reviewers to readily verify the identification of the constituents obtained from the complex mixture. A further benefit of this approach is that it is readily performed using standard GC-MS equipment and does not require any new or specialized equipment.

Supplementary Information

Supplementary data are available free of charge at http://jbcs.sbq.org.br, as PDF file.

Acknowledgements

References

1. Yetimoglu, E. C.; Ercan, O.; *J. Braz. Chem. Soc.* **2008**, *19*, 1399.

2. Antignac, J. P.; Marchand, P.; Gade, C.; Matayron, G; Qannari, E. M.; Le Bizec, B.; Andre, F.; *Anal. Bioanal. Chem.* **2006**, *384*, 271.

3. Lin, J.; Su, M.; Wang, X.; Qiu, Y.; Li, H.; Hao, J.; Yang, H.; Zhou, M.; Yan, C.; Jia, W.; *J. Sep. Sci.* **2008**, *31*, 2831; Gika, H. G.; Theodoridis, G. A.; Wilson, I. D.; *J. Chromatogr., A* **2008**, *1189*, 314; Chen, Y. Zhu, S.-B.; Xie, M.-Y.; Nie, S.-P.; Liu, W.; Li, C.; Gong, X.-F.; Wang, Z.-X.; *Anal. Chim. Acta* **2008**, *623*, 146.

4. Pe-Piper, G.; Triantafyllidis, S.; Piper, D. J. W.; *J. Sediment. Res.* **2008**, *78*, 595.

5. Smentkowski, V. S.; Ostrowski, S. G.; Kollmer, F.; Schnieders, A.; Keenan, M. R.; Ohlhausen, J. A.; Kotula, P. G.; *Surf. Interface Anal.* **2008**, *40*, 1176.

6. Baigorri, R.; Zamarreno, A. M.; Fuentes, M.; Gonzalez-Gaitano, G.; Garcia-Mina, J. M.; Almendros, G.; Gonzalez-Vila, F. J.; *J. Agric. Food Chem.* **2008**, *56*, 5480.

7. Tonidandel, L.; Ragazzi, E.; Roghi, G.; Traldi, P.; *Rapid Commun. Mass Spectrom.* **2008**, *22*, 630.

8. Radulovic, N.; Zlatkovic, B.; Palic, R.; Gordana, S.; *Nat. Prod. Commun.* **2007**, 2, 453.

9. Ramanoelina, P. A. R.; Viano, J.; Bianchini, J. P.; Gaydou, E. M.; *J. Agric. Food Chem.* **1994**, *42*, 1177; Bombayda, I.; Dupuy, N.; Van Da, J. P. L; Gaydou, E. M.; *Anal. Chim. Acta* **2008**, *613*, 31.

10. Hupp, A. M.; Marshall, L. J.; Campbell, D. I.; Smith, R. W.; McGuffin, V. L.; *Anal. Chim. Acta* **2008**, *606*, 159.