



Collaborative validation of the quantification method for suspected allergens and test of an automated data treatment

Alain Chaintreau^{a,*,1}, Esmeralda Cicchetti^a, Nathalie David^b, Andy Earls^c, Pascal Gimeno^d, Béatrice Grimaud^e, Daniel Joulain^f, Nikolai Kupfermann^g, Gryta Kuroпка^h, Frédéric Saltronⁱ, Christine Schippa^j

^a Firmenich SA, Corporate R&D Division, 1 Route des Jeunes, PO Box 239, CH-1211 Geneva-8, Switzerland

^b Chanel, 135 Avenue Charles de Gaulle, F-92521 Neuilly-sur-Seine Cedex, France

^c LGC Ltd., Queens Road, Middlesex, Teddington TW11 0LY, UK

^d AFSSAPS, 635 rue de la Garenne, F-34740 Vendargues, France

^e Yves-Rocher, 101 quai du Président Roosevelt, F-92444 Issy les Moulineaux Cedex, France

^f PRODAROM, 48 avenue Riou Blanquet, BP 21017, 06131 Grasse Cedex, France

^g Eurofins Consumer Product Testing GmbH, D-22761 Hamburg, Germany

^h TÜV Rheinland LGA Products GmbH, Am Grauen Stein, D-51105 Köln, Germany

ⁱ SCL-DGCCR-F-DGDDI, 146 traverse Charles Susini, F-13388 Marseille Cedex-13, France

^j V. MANE FILS, 620 route de Grasse, F-06620 Le-Bar-Sur-Loup, France

ARTICLE INFO

Article history:

Received 6 June 2011

Received in revised form 22 August 2011

Accepted 23 August 2011

Available online 31 August 2011

Keywords:

Quantification

Fragrance

Allergen

Validation

Accuracy profile

GC/MS

ABSTRACT

Previous publications investigated different data treatment strategies for quantification of volatile suspected allergens by GC/MS. This publication presents the validation results obtained on “ready to inject” samples under reproducibility conditions following inter-laboratory ring-testing. The approach is based on the monitoring of three selected ions per analyte using two different GC capillary columns. To aid the analysts a decisional tree is used for guidance during the interpretation of the analytical results. The method is evaluated using a fragrance oil concentrate spiked with all suspected allergens to mimic the difficulty of a real sample extract or perfume oil. At the concentrations of 10 and 100 mg/kg, imposed by Directive 76/768/EEC for labeling of leave-on and rinse-off cosmetics, the mean bias is +14% and −4%, respectively. The method is linear for all analytes, and the prediction intervals for each analyte have been determined. To speed up the analyst’s task, an automated data treatment is also proposed. The method mean bias is slightly shifted towards negative values, but the method prediction intervals are close to that resulting from the decisional tree.

© 2011 Elsevier B.V. All rights reserved.

1. Introduction

Directive 2003/15/EC amending Council Directive 76/768/EEC relating to cosmetic products, regulates the obligation to inform consumers of the presence of potentially allergenic fragrance substances identified as likely to cause allergic reactions in cosmetic products [1]. It requires that these substances must be listed in the ingredients of cosmetic products if they exceed 0.001% “in leave-on products” or 0.01% in “rinse-off products”. Among the list of 26 regulated compounds, 24 are volatile and can be analyzed by gas-chromatography. Quantification represents a challenge, due to the complex composition of fragrance oils: taking into account all

constituents, their isomers, and the constituents of natural ingredients used in a formulation, more than 100 GC peak responses can be detected in a fragrance oil. This problem has given rise to many publications and most of them involve the use of GC/MS using electron impact mode (EI) [2–7], and in few cases chemical ionization mode (CI) [8]. Fast GC, interfaced to quadrupole MS has also been investigated [9]. As co-elutions are frequent, either the selectivity of the detector can be increased by using tandem mass spectrometry [10], or the separation resolution is enhanced using comprehensive two-dimensional GC (GC × GC) [11–13]. The “targeted mode” of GC × GC has also been proposed [14].

Additional analytical techniques have been investigated to quantify the suspected allergens. Comprehensive GC × GC with a flame ionization detector (FID) was shown not to be very selective in avoid overlapping peaks, and to unambiguously assign the target analytes in complex fragrance oils [11]. Using HPLC diode array detection, the low resolution of LC columns compared to capillary GC columns is questionable, as many peaks overlap during analysis

* Corresponding author. Tel.: +41 22 780 3471; fax: +41 22 780 3334.

E-mail address: alain.chaintreau@firmenich.com (A. Chaintreau).

¹ Chairman of the Working Group on Fragrance Allergens of the “Comité Européen de Normalisation” (CEN).

of the 24 analytes [15]. GC/FTIR was applied to the quantification in various cosmetic products, but only 10 suspected allergens were quantified [16]. A recent paper reports the use of electrokinetic chromatography in a microemulsion phase, and using UV detection [17]. Good resolutions and linearities were achieved but the method could not determine all of the 24 compounds and was only applied to rinse-off cosmetics.

The papers dedicated to the quantification of suspected allergens in real consumer products are variants of the initial paper of the International Fragrance Association (IFRA) [2]: Niederer et al. proposed a size-exclusion clean-up to remove the non-volatiles from cosmetic products [4]. To analyze fragranced baby bathwater, Lamas et al. recovered the volatiles using solid-phase microextraction (SPME) [6]. The same SPME technique was also applied in the evaluation of perfumes [18].

The data analysis strategy varies strongly from one publication to another. When quantifying target compounds in a complex mixture, it is advisable to check the identity of analytes. When a monodimensional detector is used, this identification is not feasible and so, a risk of confusion with the other perfume or cosmetic ingredients exists [15,17]. When using an MS detector and only a few mass fragments, the identity should be assessed using qualifiers and ion ratios between the qualifiers and the quantifier [19]. However, this precaution does not overcome the over-estimation of co-eluted peaks. To avoid this problem, some authors prefer the scan mode to allow the selection of another quantification ion that does not appear in the overlapping peak [3,9]. Unfortunately, the sensitivity resulting from ions extracted from a full scan does not reach that of ions monitored in SIM mode, and the GC-peak area reproducibility is significantly greater with the latter. To minimize the over-evaluation caused by a co-elution, we previously proposed a “play-down strategy”: it consists in successively considering each of the three ions monitored in SIM as the quantifier, and using the two other ions as qualifiers. Selecting the minimum amount from the six values resulting from an analysis two different GC columns, in combination with a computerised peak recognition, was shown to be efficient [2].

Some of the above-cited papers were submitted to a partial or extensive validation. The fast-GC/MS method was only tested for its limit of quantification (LOQ) and relative standard deviations (RSD) of GC peak areas using a standard solution [9]. The SPME-GC/ion-trap-MS and the GC $\times$ GC/MS methods were tested for linearity, precision and LOQ [6,13]. However in the three cases the LOQs were determined from blank samples, which is not representative of the complexity of a real perfume sample for which a baseline free of peaks is an exception. The IFRA method based on a GC/MS determination using the above-mentioned play-down strategy, was evaluated for its reproducibility in different fragrance concentrates, using 10–15 compounds among the 24 suspected allergens, but the precision and the LOD were not determined under reproducibility conditions [20].

Since these 24 volatile allergens are now regulated in the European Union, a working group was created by a European Committee for Standardisation (CEN), to select and validate a method for use as a European standard. However, the analysis of suspected allergens in consumer products requires two steps: isolation from a representative sample matrix that is compatible with GC/MS analysis, and quantification of the analytes from the sample extract. The alcoholic fragrances of fine perfumery and the fragrance oil concentrates produced by the fragrance industry are usually directly compatible with this second step. Therefore, this work only reports the validation of the later performed by the CEN working group, i.e. the GC/MS quantification of ready-to-inject samples. The consumer products requiring a sample preparation will be considered in another publication. The present approach starts from the IFRA method based on GC/MS in selected-ion monitoring mode (SIM)

[2], as such equipment is available in most laboratories dealing with the analysis of these types of analytes. In addition, the risk of over-estimation was evaluated thanks to a theoretical calculation estimating the probability that an analyte co-elutes with another fragrance constituent exhibiting isobaric fragments [20]. The objective of this work is the validation of the method under reproducibility conditions and the investigation of an automated data treatment procedure to reduce the analyst's time on this task.

2. Experimental

2.1. Materials

Suspected allergens were supplied by Firmenich, with the following purities: limonene: 97.6%, linalool: 99.8%, methyl 2-octynoate: 99.5%, citronellol: 99.2%, citral quantified as the sum of its two isomers, neral: 37.1%; geranial: 62.9%, cinnamaldehyde: 98.6%, anisyl alcohol: 100%, hydroxycitronellal: 98.4%, cinnamic alcohol: 98.7%, eugenol: 100%, coumarine: 100%, isoeugenol: 96.0%, α -amylcinnamic alcohol: 42.0% (+ benzylheptanol), benzyl alcohol: 99.7%, α -amylcinnamaldehyde: 99.5%, α -hexylcinnamaldehyde: 91.7% (sum of both isomers), benzyl benzoate: 100%, benzyl cinnamate: 100%, farnesol: 100% (quantified using its two main isomers: 50.9% and 42.4%), geraniol: 99.4%, linal[®]: 98.3%, lyral[®]: 81.3%, alpha isomethylionone: 87.6%, benzyl salicylate 100%. All participants were supplied with the suspected allergen analytes from the same batches to minimize the variability inherent in different suppliers, the final results were corrected for purity. The spiking of samples was completed using the same batches. The internal standards, 1,4-dibromobenzene and 4,4'-dibromobiphenyl, had a purity of 98% and were supplied by Aldrich (Steinheim, Germany).

The above-listed analytes were spiked at concentration levels between 10 and 100 mg/kg in a perfume made of volatile constituents representing about 150 GC peaks (ingredients + isomers + impurities). This non-spiked perfume was included in the ring test to check that it did not contain any of the 24 analytes. Neral and geranial were calibrated and quantified separately, but only their sum was taken into account, in agreement with the European Directive.

2.2. Standard solutions

The ring-test protocol required preparation of a stock standard solution of all analytes at a concentration of 5 g/l in an inert and non-volatile solvent (methyl pivalate, *o*-cresol-free *o*-fluorotoluene). This was stored for less than 1 month in the absence of light at a temperature below -18°C . Alternatively, two different stock standard solutions were prepared, one for carbonyl compounds, another for non-carbonyl compounds and store in the dark below $+4^{\circ}\text{C}$ for a maximum of 2 months.

The internal standard solution of 1–4 dibromobenzene and 4–4'-dibromobiphenyl, was prepared at 1 g/l in the same solvent as the stock standard solutions and stored at about $+4^{\circ}\text{C}$ for a maximum of 2 months. From the stock standard solution(s), calibration standards were prepared. Each containing 10 mg/l of internal standards, and 1, 5, 10, 20, 30 or 40 mg/l of analytes.

2.3. GC/MS

The collaborative validation was performed by the organizations/companies that are listed in the authors of this paper, except LGC. Prodarom's experimental results were provided by Robertet SA (Grasse, F). Chanel was represented by two contributing laboratories. The MS instruments used in this ring test are listed in Table 1.

Table 1

Mass spectrometers and column phases used in the ring-test.

Lab.	MS	Column 1 ^a	Column 2 ^a
1	Varian MS-1200	5% Ph-95% Me	50% Ph-50% Me
2	Shimadzu QP2010	PEG	100% diMe
3	Agilent 5975C	5% Ph-95% Me	50% Ph-50% Me
4	Agilent 5973	PEG	100% diMe
5	Agilent 5973	50% CN-50% Ph	50% Ph-50% Me
6	Agilent 5973	5% Ph-95% Me	PEG
7	Agilent 5973	PEG	50% Ph-50% Me
8	Agilent 5975C	PEG	50% Ph-50% Me
9	Agilent 5973	PEG	100% diMe
10	Agilent 5973	100% diMe	PEG

^a Me-, Ph-, CN- are methyl-, phenyl-, cyanopropyl-polysiloxane, respectively; PEG: polyethyleneglycol.

Quantification was achieved in SIM mode; the fragment ions listed in Table 2 allow quantification by using one of the three ions as the quantification ion and the other two as qualifiers. Each of these three ions becomes the quantifier and the two other ones the qualifiers.

2.4. GC columns

A number of GC columns with different phases were available to complete the ring test (Table 1). Participants selected phases that significantly differed in polarity, the mean resolution $\bar{R}$ between all analytes had to differ from zero and be preferably greater than five. In the presence of co-eluting peaks, samples analyzed in full scan

mode were examined, to conclude whether the ion profile of the coeluted compounds were similar to those of the target allergen used for quantification.

2.5. Computation

2.5.1. Mean resolution

The mean resolution was calculated as the geometric mean $\bar{R}$ of the resolution between all peaks of analyte [2]:

$$\bar{R} = 1.18 \left[\prod_{i=1}^{i=n-1} \left(\frac{t_{R,i+1} - t_{R,i}}{w_{h,i} + w_{h,i+1}} \right) \right]^{1/(n-1)}$$

with $t_{R,i}$ is the total retention time of the i th peak, $w_{h,i}$ is the width at half height of the i th peak, and n is the number of analyte peaks in the chromatogram.

2.5.2. Peak recognition

The identity of GC/MS peaks detected in SIM mode were assessed either by calculating the Q -value [2]:

$$Q = 100 - \frac{\sum_{i=1}^{i=n} (100 \times |r_i - r'_i|)(\ln[100r_i + 1])^2}{21.3 \sum_{i=1}^{i=n} r_i}$$

with n is the number of ions per compound, r_i is the reference peak area ratio, and r'_i is the observed peak area ratio; or using the tolerances of abundance ratios defined by the European Union [19].

2.5.3. Statistics

The standard deviation (σ) was calculated as:

$$\sigma = \sqrt{\frac{1}{l} \sum_{i=1}^{i=l} (m_i - \bar{m})^2}$$

with l is the the degree of freedom (number of laboratories minus one) and m_i and $\bar{m}$ are the determined amount, the mean of determinations, respectively.

Confidence interval:

$$\bar{m} \pm \frac{t\sigma}{\sqrt{l}}$$

with σ is the standard deviation.

Prediction interval:

$$\bar{m} \pm t\sigma \sqrt{1 + \frac{1}{n}}$$

with n is the the number of laboratories.

The mean standard deviation ($\bar{\sigma}$) was calculated such as σ , using $\bar{m}$, the mean over all laboratories $\times$ 24 analytes, and l' the (number of laboratories $\times$ number of analytes), minus one.

Mean confidence interval (all analytes together):

$$\bar{m} \pm \frac{t\bar{\sigma}}{\sqrt{l'-1}}$$

with t is the Student's constant: at a confidence level of 90%, $l' = (\text{number of laboratories} \times \text{number of analytes})$ minus one, and $\bar{\sigma}$ is the mean standard deviation (all laboratories, all analytes).

3. Results and discussion

3.1. Data treatment and ring-test organization

Although mass spectrometry is a very specific detection technique, the complexity of fragrance oils requires the use of two

Table 2

Selected ions used for the quantification of analytes in SIM.

Compound	CAS RN	Ion 1 (m/z)	Ion 2 (m/z)	Ion 3 (m/z)
α -Amylcinnamic alcohol	101-85-9	133	115	204
α -Amylcinnamaldehyde	122-40-7	202	201	129
Anisyl alcohol	105-13-5	138	137	109
Benzyl alcohol	100-51-6	108	79	107
Benzyl benzoate	120-51-4	105	212	194
Benzyl cinnamate	103-41-3	131	192	193
Benzyl salicylate	118-58-1	91	228	65
Butylphenyl methyl propional (Lilial®)	80-54-6	189	147	204
Cinnamic alcohol	104-54-1	92	134	115
Cinnamaldehyde	104-55-2	131	132	103
Citral: Neral	106-26-3	69	94	109
Citral: Geranial	141-27-5	69	84	94
Citronellol	106-22-9	95	69	81
Coumarine	91-64-5	146	118	89
Eugenol	97-53-0	164	103	149
Farnesol	4602-84-0	69	93	81
Geraniol	106-24-1	69	123	93
α -Hexylcinnamic aldehyde	101-86-0	216	215	129
Hydroxycitronellal	107-75-5	59	71	43
Hydroxyisohexyl-3-cyclohexene carboxaldehyde (Lylal®)	31906-04-4	192	136	149
Isoeugenol	97-54-1	164	149	131
α -Isomethylionone	127-51-5	206	135	150
Limonene	5989-27-5	68	93	67
Linalool	78-70-6	93	71	121
Methyl 2-octynoate	111-12-6	95	123	79
1,4-Dibromobenzene ^a	106-37-6	236	234	238
4,4'-Dibromobiphenyl ^a	92-86-4	312	310	314

^a Internal standard.

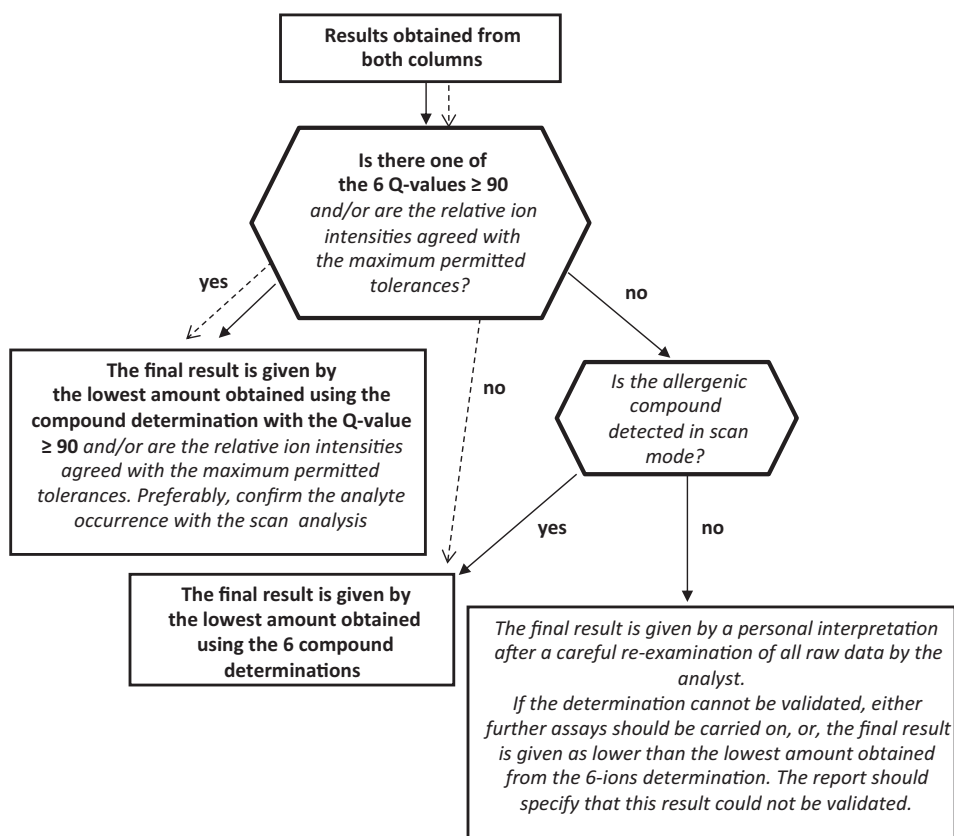


Fig. 1. Decisional tree (whole text and continuous arrows) and automated data treatment (text in bold only and dashed arrows) used for the data treatment.

different columns to minimize the risk of overestimation inherent in the possible co-elution of analytes with fragrance ingredients [20]. Therefore, the samples were analyzed by GC/MS using two GC capillary columns of different phase polarity. The suspected allergens were quantified in SIM mode using three ions: one quantifier, and two qualifiers. Each of the three ions was successively considered as the quantifier, and the remaining two as qualifiers. A resulting set of 6 quantitative results per analyte was calculated from the three ions using the results from both columns. The analyte identification was achieved, either by calculating the *Q*-value using the three ions [2], or by applying the European guidelines concerning the performance of analytical methods and the interpretation of results [19].

The first ring test was organized for training purposes to ensure that all participants gained experience on both the experimental approach and data treatment. A standard method was drafted close to the final protocol that will be issued by CEN. A second ring test was performed following the draft standard to identify any discrepancy. A third ring test was performed after correction of the draft standard, to generate the data presented in this paper and validate the method for “ready to inject” samples. The 24 suspected allergens were spiked in a perfume representing a medium complexity sample (150 GC peaks). Two data treatment approaches were investigated on the results produced from the third ring-test. The first one involved the use of a decisional tree to unify the interpretation within the participants (Fig. 1, whole text). The second approach investigated the possible use of an automated data treatment. This automation was based on the same principle as the decisional tree (Fig. 1, text in bold). Only the data obtained using the *Q*-values were considered in the automated treatment, because it required that the quality of the peak recognition was expressed by a single numerical value, which was not feasible with the criteria of the European Directive.

For both procedures, the decisional tree and the automated treatment, the principle is based on a “play-down” strategy. A *Q*-value higher than 89, or ion ratios within the tolerances of the European Directive indicate a positive analyte recognition. Outside the tolerances, if its characteristic ions are detected, the analyte can be present but co-eluted with another fragrance ingredient. In both cases, the lowest concentration calculated using two GC columns and 3 analyte ions is taken into account to lower the possible bias resulting from the co-elution.

The validation was completed using the accuracy profiles as recommended by the International Conference on Harmonisation (ICH) [21]. However, the intra-day repeatability could not be evaluated, due to the length of time taken to complete all the analysis. Performing the repeat analysis would have taken more than one day and would be altered by other causes of variability.

3.2. Decisional-tree results

3.2.1. Response function

The occurrence of a quadratic calibration curve can be a sign of poor suitability of the GC/MS system, due to injector pollution by non-volatile materials [2]. As a first test, and just after the calibration, the participants injected a standard solution at a concentration corresponding to the middle of the calibration curve. The analytical system was considered to be suitable if the determined concentration of all analytes deviated by less than 20% from the nominal concentration (data not shown).

3.2.2. Method linearity

Calibration of the GC/MS is performed using linear regression. The linearity of the method was checked by verifying that the mean of analyte quantifications from the GC/MS procedure plus the application of the decisional tree gave values proportional to

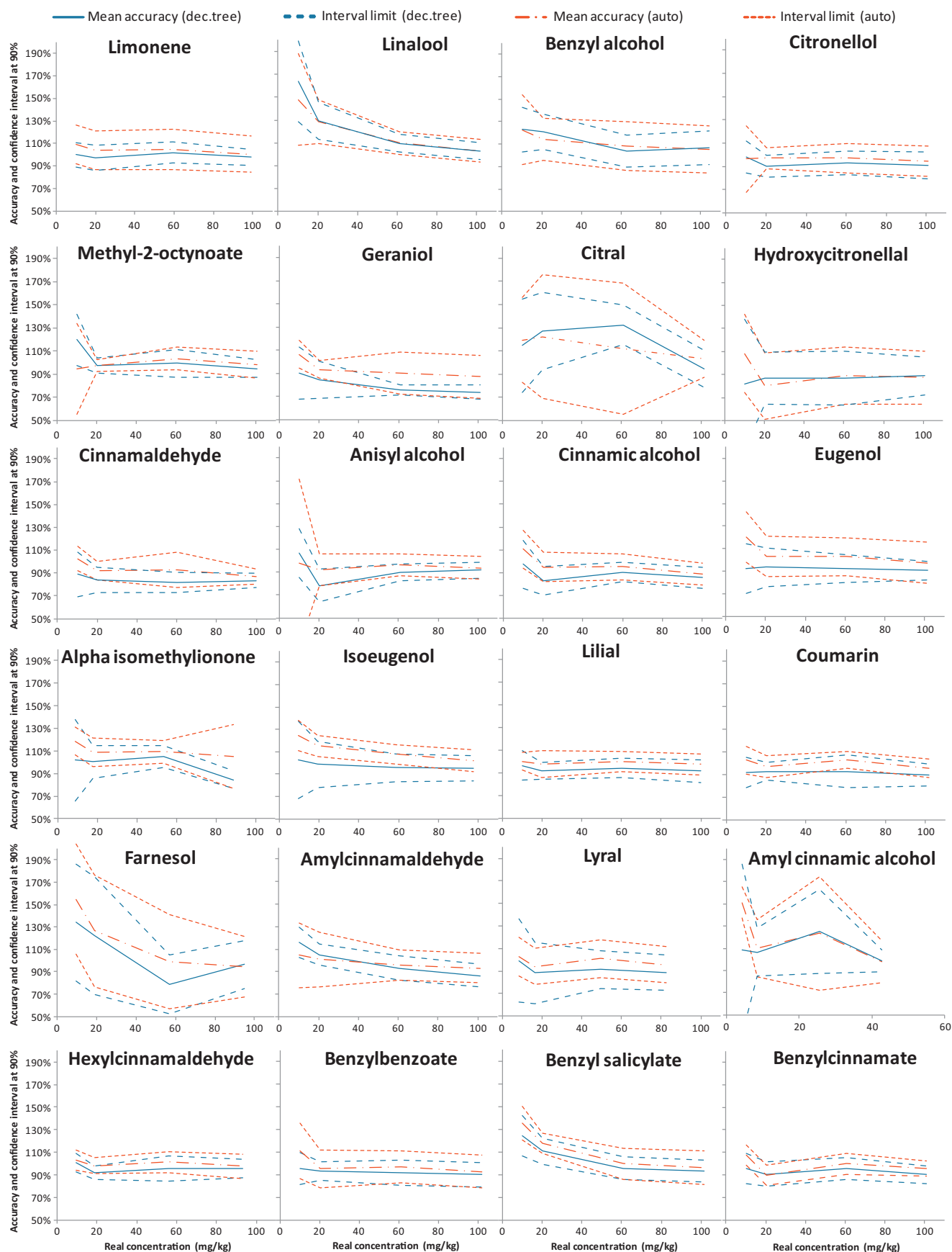


Fig. 2. Accuracy profiles resulting from the decisional tree (continuous and long-dotted blue lines) and the automated data treatment (dot-dash-dot and short-dotted red lines). The dotted lines represent the confidence interval at 90%.

Table 3

Method linearity after application of the decisional tree or automated data treatment.

	Decisional tree		Autom. treat.	
	Slope	r^2	Slope	r^2
Limonene	1.002	0.9993	0.982	0.9990
Linalool	0.980	0.9993	0.967	0.9993
Benzyl alcohol	1.035	0.9999	1.036	0.9985
Citronellol	0.942	0.9996	0.906	0.9996
Methyl-2-octynoate	0.992	0.9985	0.934	0.9990
Geraniol	0.855	0.9997	0.718	0.9999
Citral (geranial)	1.013	0.9972	0.949	0.9339
Hydroxycitronellal	0.871	0.9983	0.892	0.9999
Cinnamaldehyde	0.860	0.9984	0.828	0.9997
Anisyl alcohol	0.940	0.9995	0.919	0.9977
Cinnamic alcohol	0.876	0.9983	0.858	0.9982
Eugenol	0.961	0.9990	0.905	0.9999
α -Isomethylionone	1.044	0.9995	0.845	0.9752
Iso-eugenol	1.000	0.9993	0.944	0.9993
Lilial®	0.977	0.9996	0.917	0.9996
Coumarin	0.962	0.9996	0.894	0.9994
Farnesol	0.869	0.9998	0.882	0.9702
α -Amylcinnamaldehyde	0.919	0.9998	0.832	0.9987
Lylal®	0.960	0.9984	0.885	0.9996
α -Amyl cinnamic alcohol	0.965	0.9721	1.011	0.9682
α -Hexylcinnamaldehyde	0.975	0.9993	0.958	0.9998
Benzyl benzoate	0.918	0.9990	0.896	0.9999
Benzyl salicylate	0.913	0.9999	0.894	0.9999
Benzyl cinnamate	0.962	0.9984	0.907	0.9984
Median	0.961	0.9993	0.905	0.9993
Minimum	0.855	0.9721	0.718	0.9339
Maximum	1.044	0.9999	1.036	0.9999

their nominal concentrations. As an illustration of the global performance, the median values of slopes and determination coefficients were close to unity (0.961 and 0.9993, respectively), with a small dispersion (see the minima and maxima, Table 3).

3.2.3. Accuracy profiles

In a validation protocol based on the accuracy profile, the LOQ is usually determined as the lowest concentration that can be obtained with a confidence interval within a tolerance range [22]. In the present case, the European Directive restricts the use of 24 suspected allergens at a minimal concentration of 10 and 100 mg/kg in “leave-on” and “rinse-off” products, respectively [1]. Any method should be capable of quantifying these analytes at these concentrations. Therefore, the question becomes: what is the confidence interval of the quantification at these pre-defined minimal levels? The accuracy profiles were constructed for all analytes at a confidence level of 90% from the results of the 10 laboratories (Fig. 2).

Most of the analytes were quantified down to 10 mg/kg with an accuracy greater than, or equal to $\pm 30\%$, five compounds remained within this limit down to only 20 mg/kg, and four exceeded this limit at higher concentrations. All analytes were in the range of

Table 4

Frequency (%) of false negatives in ring test samples.

	Spiking level			
	10	20	60	100
Decisional tree	3.3	1.3	0.0	0.0
Automated treatment	0.5	0.5	0.5	0.5

70–110% (except benzyl alcohol) at 100 mg/kg. Such variability may look important, but it is representative of the real conditions for this type of analysis, because the present ring test was achieved by spiking a fragrance oil concentrate with the investigated analytes. In a previous paper, the frequency of analyte co-elution was shown to be non-negligible, co-elutions may lead to overestimations in spite of the selectivity and specificity of GC/MS, and the use of two columns [20]. However, these interferences can be expected to randomly occur for an analyte when the matrix composition changes. If we assume that these analytes are more or less equivalent, a global accuracy profile can be established as a mean over 240 results (10 laboratories $\times$ 24 suspected allergens). The method can be considered to be globally accurate, because the mean bias is less than 5% down to 20 mg/kg, and less than 15% at 10 mg/kg (Fig. 3, left). Such results are within the range expected from the repeatability estimation performed in the pre-validation (mean bias < 35% down to 20 mg/l and between –49 and 77% at 10 mg/l) [20].

3.2.4. Prediction intervals

For a given laboratory dealing with the quantification of suspected allergens, knowing the confidence of individual values is important when it delivers results. To answer this question, the prediction intervals for all analytes have been established (Fig. 4).

Such as for the accuracy profile, if we assume the equivalence of all analytes, the mean prediction intervals can be drawn between 10 and 100 mg/kg over 240 results (10 laboratories $\times$ 24 suspected allergens) (Fig. 5, left). With a confidence of 90%, this prediction remains in the range [–36%:+42%] down to 20 mg/kg and rapidly increases to [–53%:+81%] at 10 mg/kg.

3.2.5. False negatives and positives

When analyzing real fragrance samples, the occurrence of false negatives has already been mentioned [20]. It is inherent in the fact that an analyte may be hidden by a co-eluted perfume ingredient, or is shifted after the SIM window due to the elution of a major peak. In the present work, some false negatives were only observed at the lowest concentration (Table 4). As all suspected allergens were present in the test samples, the occurrence of false positives could not be investigated. However, in the second ring test, performed with blind fragrance samples spiked with 4–6 of the 24 suspected allergens, no false positive were identified, after the play-down strategy used in the data treatment (data not shown).

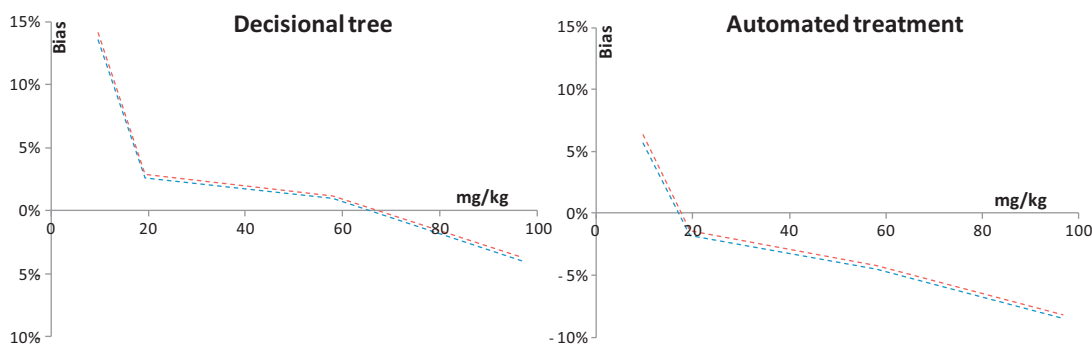


Fig. 3. Confidence interval of the mean bias at a confidence of 90% over the 24 suspected allergens and 10 (decisional tree) or 8 (automated treatment) laboratories.

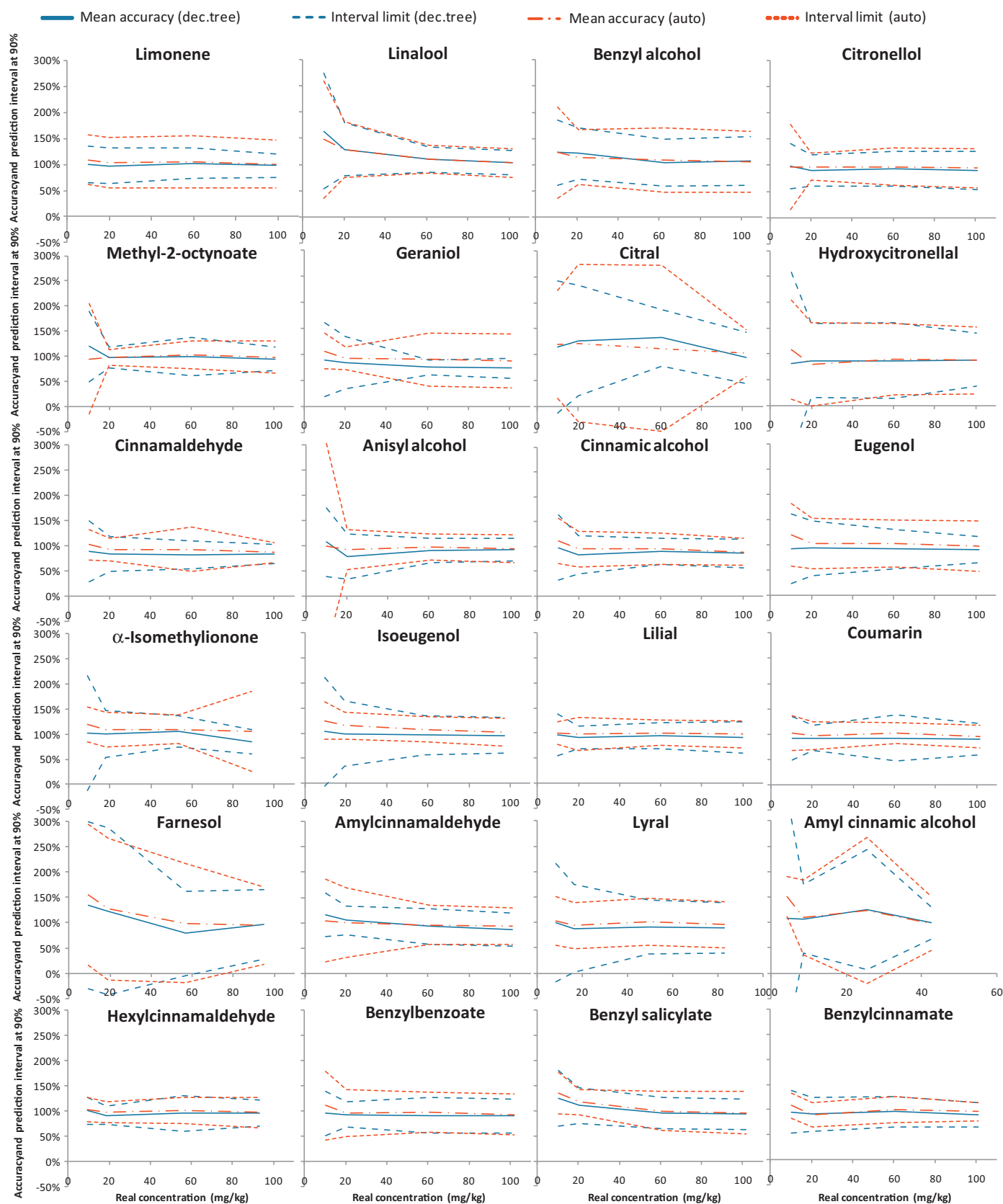


Fig. 4. Prediction profiles resulting from the decisional tree (continuous and long-dotted blue lines) and the automated data treatment (Dot-dash-dot and short-dotted red lines): Bias (continuous line and long-dotted blue lines) and prediction interval at 90% (dotted lines).

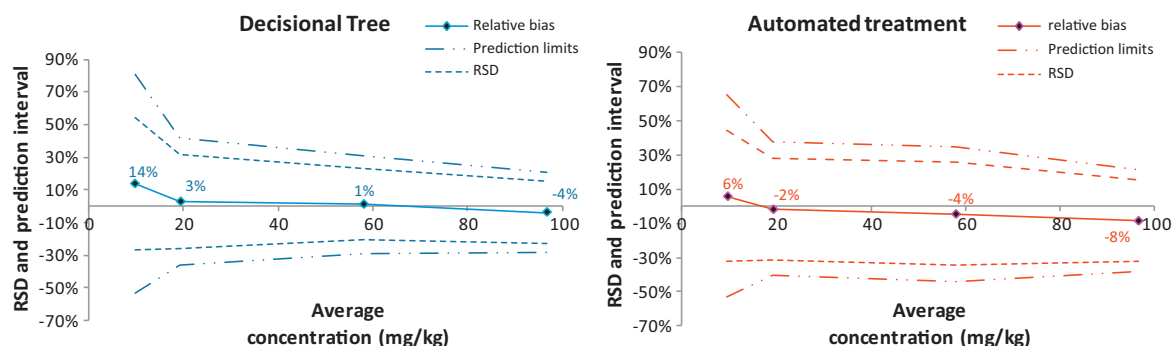


Fig. 5. Mean bias, RSD and prediction intervals at a confidence of 90% over the 24 suspected allergens and 10 (decisional tree) or 8 (automated treatment) laboratories.

3.3. Automated data treatment

If the analysis of suspected allergens requires a rigorous experimental procedure, the most time-consuming task remains the interpretation of data from the GC/MS workstation. For each of the 24 analytes, the most appropriate concentration must be deduced following the analysis on two GC columns and three ions as a function of either the *Q*-value or the tolerances of the European Directive [19]. To help the analyst and save time, this data treatment was automated on a spreadsheet based on a simplification of the decisional tree approach (Fig. 1, text in bold) and its result compared to the application of the decisional tree. As mentioned above, only the data provided by the participants together with a *Q*-value were considered in this automation (8 laboratories).

When the analyst had decided to manually re-integrate peak responses of analytes, the same integration result was used as a starting point for both the decisional tree and the automated data treatment. No further corrections were made to the results of the automated data procedure.

3.3.1. Method linearity

The method was tested using the automated strategy and the results were used just as they were to evaluate the linearity (Table 3). The latter was found to be excellent, with a median of the slope and determination coefficient: 0.905 and 0.9993, respectively.

3.3.2. Accuracy profiles

The accuracy profiles were generally similar to those coming from the decisional tree, with sometimes larger confidence intervals, in particular at the lowest concentration (Fig. 2). Such are for the results of the decisional tree, if the behaviour of the 24 analytes is considered to be equivalent, a mean accuracy profile can be drawn (Fig. 3, right). Here again, there is some similarity between the curves resulting from both data treatment, but the mean bias of the automated treatment was shifted towards negative values. This suggests that, in the absence of an additional interpretation by the analyst, the automated treatment under-estimates the concentrations due to the “play-down” strategy, *i.e.* choosing the lowest value among the 6 proposals (3 ions $\times$ 2 columns).

3.3.3. Prediction profiles

The prediction intervals were determined for each analyte at the different concentrations (Fig. 4). The mean prediction intervals were calculated as the average of all results (10 laboratories $\times$ 24 analytes) (Fig. 5, right). As expected, these intervals were slightly larger than those of the decisional tree (Fig. 5, left). Down to 20 mg/kg, the concentrations were predicted within an interval of [−41%;+37%], at a confidence level of 90%, and within an interval of [−53%;+65%] at 10 mg/kg.

3.3.4. False negatives

In contrast to the decisional tree results, the frequency of false negatives remained constant and low at all concentrations (Table 4). At the higher concentrations, doubt on the presence of an analyte is easily elucidated by visual inspection of the chromatogram and spectra, and so false negatives are less frequent than in the automated treatment. Conversely, the same reasoning leads to the opposite conclusion at low concentrations: when the *Q*-values are low and the peak undetected in scan, the peak identity cannot be ascertained, whereas this latter interpretation does not exist in the automated treatment.

4. Conclusion

The present work provides a validated method to the authorities, the fragrance industry and contract laboratories involved in the monitoring of suspected allergens, *i.e.* a method where the capabilities and limits are determined. The decisional tree has supported the data interpretation by the analysts. From the ring test results, the method appears to be unbiased in the range of 20–100 mg/kg (less than 4%), with a prediction interval of less than 42% at a confidence level of 90%. A slight overestimation appears at 10 mg/kg, due to the co-elution of compounds with isobaric ions that cannot be differentiated at this low level. The present prediction intervals represent an average over the 24 analytes fragrance ingredients may strongly interfere with some suspected allergens and give much larger biases and prediction intervals when compared with this mean. Because the interferences are matrix-dependent and investigating all matrices is not feasible, the present mean value, obtained using a “standard” complexity matrix fragrance oil, seems to a realistic estimation to take into account these random interferences.

The results of the automated data treatment were surprisingly good and close to those of the decisional tree. If they cannot replace the analyst’s interpretation, they provide an efficient tool to speed-up this data interpretation.

Acknowledgements

The authors wish to gratefully acknowledge V. Bernat (AFNOR) for the organizational support, and all analysts who contributed to the present ring test: P. Basty-Deloffre (SCL-DGCCRF-DGDDI), Claudine Bousquet (AFSSAPS), H. Brévard and N. Barat (Robertet), I. Dubrulle (Yves-Rocher), C. Peres (Chanel), E. Proal (Chanel), D. Smart (Firmenich).

References

- [1] European Parliament, Off. J. Eur. Commun. 46 (2003) 26.
- [2] A. Chaintreau, D. Joulain, C. Marin, C.O. Schmidt, M. Vey, J. Agric. Food Chem. 51 (2003) 6398.

- [3] H. Leijds, J. Broekmans, L. Van Pelt, C. Mussinan, J. Agric. Food Chem. 53 (2005) 5487.
- [4] M. Niederer, R. Bollhalder, C. Hohl, J. Chromatogr. A 1132 (2006) 109.
- [5] F. David, C. Devos, D. Joulain, A. Chaintreau, P. Sandra, J. Sep. Sci. 29 (2006) 1587.
- [6] P.J. Lamas, L. Sanchez-Prado, C. Garcia-Jares, M. Llompert, Anal. Bioanal. Chem. 394 (2009) 1399.
- [7] J. Pablo Lamas, L. Sanchez-Prado, C. Garcia-Jares, M. Llompert, J. Chromatogr. A 1217 (2010) 1882.
- [8] P.A. Cadby, M. Youssefi, A. Chaintreau, Perfum. Flavor. 28 (2003) 44.
- [9] L. Mondello, D. Sciarone, A. Casilli, P.Q. Tranchida, P. Dugo, G. Dugo, J. Sep. Sci. 30 (2007) 1905.
- [10] S. Kinani, S. Bouchonnet, A. Magne, Spectra Anal. 248 (2006) 28.
- [11] R. Shellie, P.J. Marriott, A. Chaintreau, Flavour Fragr. J. 19 (2004) 91.
- [12] C. Debonneville, A. Chaintreau, J. Chromatogr. A 1027 (2004) 109.
- [13] C. Cordero, C. Bicchi, D. Joulain, P. Rubiolo, J. Chromatogr. A 1150 (2007) 37.
- [14] M.S. Dunn, N. Vulic, R. Shellie, S. Whitehead, P. Morrison, P.J. Marriott, J. Chromatogr. A 1130 (2006) 122.
- [15] C. Villa, R. Gambaro, E. Mariani, S. Dorato, J. Pharm. Biomed. Anal. 44 (2007) 755.
- [16] J.-H. Jiang, S. Chen, J.-X. Xie, Zhongguo Weisheng Jianyan Zazhi 15 (2005) 1415.
- [17] S. Furlanetto, S. Orlandini, I. Giannini, B. Pasquini, S. Pinzauti, Talanta 83 (2010) 72.
- [18] W. Wei, University of Massachusetts Lowell (Editor), 2006.
- [19] 2002/657/EC, Off. J. Eur. Commun. L221 (2002) 8.
- [20] M. Bassereau, A. Chaintreau, S. Duperrex, D. Joulain, H. Leijds, G. Loesing, N. Owen, A. Sherlock, C. Schippa, P.-J. Thorel, M. Vey, J. Agric. Food Chem. 55 (2007) 25.
- [21] ICH, International Conference on Harmonisation of Technical Requirements of Pharmaceuticals for Human Use, 2005.
- [22] P. Hubert, J.J. Nguyen-Huu, B. Boulanger, E. Chapuzet, P. Chiap, N. Cohen, P.A. Compagnon, W. Dewe, M. Feinberg, M. Lallier, M. Laurentie, N. Mercier, G. Muzard, C. Nivet, L. Valat, STP Pharma Pratiques 23 (2003) 101.