Talanta 219 (2020) 121208

Contents lists available at ScienceDirect

Talanta

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

Characterisation of complex perfume and essential oil blends using multivariate curve resolution-alternating least squares algorithms on average mass spectrum from GC-MS

Leo Lebanov[a,b], Laura Tedone[a,b], Alireza Ghiasvand[a,c], Brett Paull[a,b,*]

[a] Australian Centre for Research on Separation Science (ACROSS), School of Natural Sciences, University of Tasmania, Hobart 7001, Australia
[b] ARC Industrial Transformation Research Hub for Processing Advanced Lignocellulosics (PALS), School of Natural Sciences, University of Tasmania, Hobart, Tasmania, 7001, Australia
[c] Department of Chemistry, Lorestan University, Khoramabad, Iran

ARTICLE INFO

Keywords:
Essential oil blend
GC-MS
Ylang-ylang
MCR-ALS
Average mass spectrum

ABSTRACT

Quality control of essential oil blends and the discovery of potential adulterations and product fraud is a significant challenge within the natural oil and perfume industry. In this research, total chromatogram average mass spectra (TCAMS), created from the GC-MS three-way raw data, were employed for the characterisation of complex samples of perfumes and essential oil blends. A multivariate approach for curve resolution was used to resolve the TCAMS of pure essential oils within such perfume and essential oil blends. Resolved TCAMS, in combination with unsupervised pattern recognition approaches revealed the distillation grade and origin of used ylang-ylang oils in perfume mixtures. TCAMS resolved from the essential oil blends were used with a supervised machine learning classification model to identify oils, used in creating the blends. Quantification was performed using a multivariate curve resolution approach, resulting in relative errors of prediction lower than 17.84% with root mean square errors of prediction smaller than 3.43.

1. Introduction

Qualitative and quantitative characterisation of blends created from complex natural products, represents a significant analytical challenge, especially when the 'measurand' of interest is itself a complex mixture of compounds, such as the case of specific essential oil (EOs). Gas chromatography hyphenated with mass spectrometry (GC-MS) is the most widely used analytical technique for this purpose. In the conventional GC-MS analysis, analytes are identified from the total ion current chromatogram (TICC), through comparison of their retention indexes and mass spectrum similarity with the standard library matches. For the quantitative analysis of a single analyte, a univariate calibration model is typically created, where the instrumental response is peak area, which is integrated and correlated with the known concentration. However, for highly complex samples such as EO blends, this traditional approach can result in incorrect conclusions due to high incidences of co-eluting compounds [1]. One approach to solve coelution is to increase the chromatographic separation power, using for example, multidimensional gas chromatography (MDGC) [2,3]. An alternative approach is the application of multiway algorithms, such as parallel factor analysis 2 (PARAFAC2) [4], the PARAFAC2-based deconvolution and identification system (PARADISe) [5], or the MS-assisted resolution of signals (MARS) [6]. Within these various strategies, the sample chemical composition is first determined using a conventional approach. Then, the markers for each of the individual components (e.g. EOs) inside of the blend are identified. These markers are used for confirmation of the quality and possibly to quantify the target component within the blend. However, this process is very tedious.

Another way to approach this problem is to analyse each EO individually and use the full instrumental (e.g. GC-MS) dataset as a fingerprint for that EO. In this way, the analytical information regarding the EO of interest can be extracted from the full dataset for the complete blend using mathematical approaches. Multivariate curve resolution (MCR) techniques [7] and PARAFAC-like methods (PARAFAC, PARAFAC2, PARADISe, etc) have recently received considerable attention due to their capacity to address the issue of chromatographic co-

* Corresponding author. Australian Centre for Research on Separation Science (ACROSS), School of Natural Sciences, University of Tasmania, Hobart, 7001, Australia.

E-mail addresses: leo.lebanov@utas.edu.au (L. Lebanov), laura.tedone@utas.edu.au (L. Tedone), alireza.ghiasvand@utas.edu.au (A. Ghiasvand), brett.paull@utas.edu.au (B. Paull).

https://doi.org/10.1016/j.talanta.2020.121208
Received 1 April 2020; Received in revised form 20 May 2020; Accepted 21 May 2020
Available online 27 May 2020
0039-9140/

List of acronyms:

ALS	alternating least squares
EO	essential oil
GC	gas chromatography
GUI	graphical user interface
HCA	hierarchical cluster analysis
LOF	lack of fit
MARS	MS-assisted resolution of signals
MCR	multivariate curve resolution
MCR-BANDS	MATLAB set of programs for the evaluation of rotational ambiguity in multivariate curve resolution analysis
MDGC	multidimensional gas chromatography
OOBE	out-of-bag error
PARADISe	PARAFAC2 based deconvolution and identification system
PARAFAC	parallel factor analysis
PCA	principal component analysis
qMS	quadrupole mass spectrometer
REP	relative error of prediction
RF	random forests
RMSEP	root mean square error of prediction
SEP	standard error of prediction
SIMPLISMA	simple-to-use interactive self-modelling mixture analysis
SVD	singular vector decomposition
TCAMS	total chromatogram average mass spectrum
TICC	total ion current chromatogram

elution, without the need to increase chromatographic resolution. One of the most beneficial features of MCR and PARAFAC-like methods is the second order advantage, in which the calibration can be built upon relatively few standards, instead of a large calibration set, with the additional potential to quantify the compounds of interest even in the presence of interferences, not originally present in the standard calibration set. This is performed through the decomposition of the experimental data matrix into two blocks, one related to the concentration profiles and the other one related to the instrumental profile for the pure standard. These matrices can be iteratively adjusted to the data set [8]. PARAFAC requires trilinearity, while PARAFAC2 and MCR do not. Compared to the PARAFAC-like methods, MCR can handle big data arrays without slowing down the calculations [9]. Even though MCR offers several solutions which are mathematically correct, compared to a single solution offered by PARAFAC2, selection of the initial state and application of appropriate natural restrictions, constraints, enables the delivery of a single solution [10]. Additionally, it has been proven that PARAFAC2 has lower flexibility towards the variability in the chromatographic data, due to its specific algorithmic restrictions, especially when the number of coeluting compounds increases in the sample. Due to the higher similarity of the MCR–ALS underlying model to the real variations in the chromatographic data, resolved spectra of co-eluting analytes have higher agreement with the true spectra, compared to the ones resolved with PARAFAC2 [9,10].

Complex natural extracts such as EOs may contain up to several hundred compounds [11]. Due to their fragrant characters, EOs have been used in perfume blending for centuries, and today still generate significant market value from their application within the perfume and cosmetic industries, with the value of the natural cosmetic products' market estimated to be $36.3 billion in 2019, and with estimated growth to roughly $54.5 billion expected by the year 2027 [12]. The supply market for these raw natural products and EOs is also currently experiencing substantial growth, with Ylang-ylang (YY, *Cananga odorata* [Lam] Hook. f. & Thomson forma *genuina*) EO, being a particularly highly valuable natural extract used in many high-end perfume and cosmetic products. YY-EO is distilled from fresh-mature flowers using fractional distillation, producing four different grades, depending upon distillation time. The fraction that is produced in the shortest distillation time has the highest content of fragrant volatile compounds, and it is of the highest value to the perfume industry [13]. Many EOs have also found application in commercial therapeutics, health remedies, lotions and balms. In many of the above products, EO blending is usually performed to deliver potential synergistic therapeutic effects, in which the pharmacological effect of the blend is proclaimed to be greater than the sum of the effects of the individual EOs [14].

EO blends and associated perfume products are typically highly complex mixtures, the quality control of which is of considerable importance commercially, including for identification of potential adulterations and fraud. However, the rapid and reliable characterisation of these complex blends still presents a challenge for analytical chemistry [15]. Therefore, the aim of this work was to evaluate MCR-ALS for both the qualitative and quantitative characterisation of perfume mixtures and EO blends, using total chromatogram average mass spectrum (TCAMS) data sets, created from the three-way output data array (obtained from GC-MS analysis). Compared to the TICC, TCAMS is much faster to derive, where the chemical composition of each EO is fully represented within the m/z intensity data. In this way, a more comprehensive data set is created which increases accuracy in multivariate statistical approaches and the possible discrimination among different EO classes, as already shown [16,17]. To identify the grades and origins of YY in perfume mixtures, its 'blend resolved' TCAMS were compared with pure samples of different grades using principal component analysis (PCA). In an alternative approach, the resolved TCAMS of individual components within the EO blends were subjected to a Random Forests (RF) classification model to identify the individual EOs that have been used to create the blends.

2. Experimental

2.1. Essential oil samples

HPLC-grade dichloromethane (DCM) was purchased form Thermo Fisher Scientific (Waltham, MA, USA). Standard *n*-alkane mixture C_7–C_{30} (1000 μg mL^{-1} in hexane) was purchased from Sigma-Aldrich (St. Louis, MO, USA). Essential oils of allspice (Als), clary sage (Cls), cajeput (Caj), sweet fennel (Swf), cinnamon (Cin), myrrh (Myr), peppermint (Pep), spearmint (Spe), mandarin (Man), geranium bourbon (Geb), cedarwood (Cew), juniper berry (Jub), dill weed (Diw), cypress (Cyp), lavender (Lav), sweet orange (Swo), bergamot (Ber), jasmine (Jas), rosemary (Ros), neroli (Ner), angelica (Ang), clove bud (Clb), eucalyptus globulus (Eug), grapefruit (Grf) frankincense serrata (Frs), tea tree (Tet), vetiver (Vet), and four grades of YY (YExtra, YI, YII, and YIII) were obtained from Plant Therapy Inc. (Twin Falls, ID, USA), Ahimsa Oils (Mitchelton, QLD, Australia) and Albert Vieille (Vallauris, France). All YY EOs used in the prepared blends and in the PCA are presented in Table 1.

Three different commercial perfumes (PI, PII, and PIII) were obtained from local stores in Hobart (TAS, Australia). For the evaluation of the YY grade, four different matrices were created, as show in Table 2. Matrix #1 was composed of PI, rosemary and neroli EOs, blended with all four grades of YY. Matrix #2 was created from PII, jasmine and geranium bourbon mixed with YExtra and YIII. PI, bergamot and lavender were blended with YI and YII into matrix#3, while matrix #4 consisted of YExtra, YIII, PII, sweet orange and cedarwood EOs. Additionally, a matrix (matrix #5) was created for the evaluation of origin by mixing PIII, myrrh, clary sage and YY from five different origins (Madagascar, Comoros, Congo, Indonesia, and Australia), as presented in Table 3. All perfume-EO blends were prepared by pipetting

Table 1
List of ylang-ylang samples used in this research for the grade and origin determination.

Code	Grade	Origin	Code	Grade	Origin
YExtraMad01[a]	Extra	Madagascar	YIIIMad01[a]	III	Madagascar
YExtraMad02[a]	Extra	Madagascar	YIIIMad02[a]	III	Madagascar
YExtraMad03[a]	Extra	Madagascar	YIIIMad03[a]	III	Madagascar
YExtraNep04[a]	Extra	Nepal	YIIIMad04[a]	III	Madagascar
YExtraMad05[a]	Extra	Madagascar	YIIIMad05[a]	III	Madagascar
YExtraCom06[a]	Extra	Comoros	YIIIMad06[a]	III	Madagascar
YExtraMad07[a]	Extra	Madagascar	YIIIMad07[a]	III	Madagascar
YExtraCom08[b]	Extra	Comoros	YIIICom08[a]	III	Comoros
YExtraCom09[b]	Extra	Comoros	YComplMad02[a]	Complete	Madagascar
YExtraCom10[b]	Extra	Comoros	YComplMad03[a]	Complete	Madagascar
YExtraCom11[b]	Extra	Comoros	YComplMad04[a]	Complete	Madagascar
YIFra01[a]	I	France	YComplMad05[a]	Complete	Madagascar
YIMad02[a]	I	Madagascar	YComplMad06[a]	Complete	Madagascar
YIMad03[a]	I	Madagascar	YComplMad07[a]	Complete	Madagascar
YIMad04[a]	I	Madagascar	YComplAus08[b]	Complete	Australia
YICom05[a]	I	Comoros	YComplAus09[b]	Complete	Australia
YIMad06[a]	I	Madagascar	YComplAus10[b]	Complete	Australia
YIMad07[a]	I	Madagascar	YComplAus11[b]	Complete	Australia
YIMad08[a]	I	Madagascar	YComplCng12[b]	Complete	Congo
YICom09[c]	I	Comoros	YComplCng13[b]	Complete	Congo
YIIFra01[a]	II	France	YComplCng14[b]	Complete	Congo
YIIMad02[a]	II	Madagascar	YComplCng15[b]	Complete	Congo
YIIMad03[a]	II	Madagascar	YComplInd17[b]	NA	Indonesia
YIICom04[a]	II	Comoros	YComplInd18[b]	NA	Indonesia
YIIMad05[a]	II	Madagascar	YComplInd19[b]	NA	Indonesia
YIIMad06[a]	II	Madagascar	YComplInd20[b]	NA	Indonesia

Complete: Mixture of all 4 grades of YY; NA: Not available; Mad: Madagascar; Aus: Australia; Nep: Nepal; Com: Comoros; Fra: France; Cng: Congo; Ind: Indonesia.
[a] Plant Therapy Inc.
[b] Ahimsa Oils.
[c] Albert Vieille.

Table 2
Qualitative composition of the perfume-essential oil blends used in the grade and origin determination.

Matrix	Perfume	EO#1	EO#2	YY
#1	I	Rosemary	Neroli	Extra, I, II, III
#2	II	Jasmine absolute	Geranium bourbon	Extra, III
#3	I	Lavender	Bergamot	I, II
#4	II	Orange sweet	Cedarwood Atlas	Extra, III
#5	III	Myrrh	Clary sage	Mad, Con, Com, Aus, Ind

Table 3
Quantitative composition of perfume-essential oil blends used in the grade and origin determination.

Type	No	Perfume (μL)	EO#1 (μL)	EO#2 (μL)	YY (μL)
Calibration	1	40	10	10	0
	2	20	20	0	5
	3	10	0	30	10
	4	0	30	15	20
	5	10	10	10	30
Validation	6	15	15	15	15
	7	0	10	15	25
	8	10	15	10	20
	9	25	5	15	10
	10	30	20	20	5

predetermined volumes and diluted to 1 mL with DCM. For the RF classification model, the following EOs were used (the numbers in the brackets indicate the number of samples obtained from different batches, for each EO class): allspice (4 samples), angelica (3 samples), cajeput (4 samples), cedarwood (4 samples), cinnamon (3 samples), clary sage (4 samples), clove bud (4 samples), cypress (4 samples), dill weed (6 samples), eucalyptus globulus (4 samples), grapefruit (3 samples), frankincense serrata (5 samples), geranium bourbon (4 samples), juniper berry (4 samples), lavender (4 samples), mandarin (5 samples), myrrh (6 samples), neroli (3 samples), peppermint (8 samples), spearmint (5 samples), sweet fennel (4 samples), tea tree (4 samples), vetiver (4 samples) and ylang-ylang complete (7 samples). Consequently, 24 classes and a total of 106 EOs were used for the RF classification model (17 classes were used in creating the EO blends). EO blend sets were created by choosing six and eleven classes and prepared by pipetting predetermined amounts of different EO in DCM and diluting to 1 mL, as presented in Table 4.

2.2. GC-qMS analysis

All EOs and blends were analysed using a GC-MS-QP2010 Plus (Shimadzu, Kyoto, Japan) equipped with a quadrupole-MS (qMS) detector. The chromatographic separations were performed using a Mega-XMLB capillary column (30 m × 0.25 mm × 0.25 μm) from BGB Analytik (Alexandria, VA, USA). The injector was set at 250 °C and 1 μL of each sample injected directly in split mode (split ratio of 50:1). The initial oven temperature was set at 50 °C and increased to 250 °C at a rate of 4.3 °C min^{-1}. The temperature was then quickly ramped (15 °C min^{-1}) to 280 °C and held constant for 5 min, resulting in a total run time of 53.5 min. Ultra-high purity helium from BOC (Guildford, Surrey, UK) was used as the carrier gas at a constant flowrate of 1.54 mL min^{-1}. The transfer line temperatures and ion source and were set at 250 and 220 °C, respectively. The mass fragments were recorded over the range of 40–400 Da. The acquisition rate was 0.2 s with an acquisition rate of 2000 to provide satisfactory number of points per peak for effective resolution.

2.3. Data processing

TCAMS of each sample was taken from 2.5 min until 42 min with subtraction of baseline average mass spectrum (AMS) calculated from 41 to 42 min, using GCMSsolution software (version 4.30). The baseline showed its highest intensity at 41–42 min with very few eluting peaks, which provided for easy calculation of its AMS. After 42 min, there were no further peaks present in the chromatograms. Raw data was processed by MetaboAnalyst 4.0 online software (www.metaboanalyst.ca) [18], based on XCMS code written in R software, by aligning the *m/z* fragments and processing for the missing values (by importing with half of the minimal value for the fragments). The MCR-ALS GUI 2.0 algorithm is freely available on the internet (www.mcrals.info) and described in the Jaumot et al. report [19]. It was run on MATLAB R2017a 9.2 (MathWorks, Natick, MA, USA). PCA and hierarchical cluster analysis (HCA) were also performed using MetaboAnalyst software.

2.4. TCAMS and TICC

If m is the total number of mass scans (S_i) recorded during a single GC-MC run in time t, with Δt as the time necessary to record one mass scan, then, $m = {}^{t}/_{\Delta t}$. If n is considered as the number of m/z fragments (40–400) and y_{ij} as the response for each m/z fragment in each scan, where "$i = 1, 2, \ldots, m$ and $j = 1, 2, \ldots, n$", then, the masses scan S_i ($i = 1, 2, \ldots, m$) could be defined as a set of m/z fragments with their corresponding responses from the MS detector as:

$$S_i = \left\{ \left[y_{i1}, \left({}^{m}/_{z} \right)_1 \right], \left[y_{i2}, \left({}^{m}/_{z} \right)_2 \right], \ldots, \left[y_{in}, \left({}^{m}/_{z} \right)_n \right] \right\}. \quad \text{(Eq 1)}$$

The TICC can be defined as:

Table 4
Composition of the EO blends, prepared for the evaluation of the MCR-ALS model.

Blend No	Volume of EO (μL)																
	Six-component blend set						Eleven-component blend set										
	Als	Cls	Caj	Swf	Grf	Cin	Yly	Myr	Pep	Spe	Man	Geb	Cew	Jub	Diw	Cyp	Lav
1	70	10	10	10	0	0	10	10	10	0	10	10	0	0	0	50	0
2	10	70	5	5	5	5	0	0	0	5	0	0	80	10	0	0	5
3	0	0	70	5	5	20	5	5	10	0	0	20	0	0	50	5	5
4	20	5	5	70	0	0	0	0	20	30	0	0	40	0	0	10	0
5	5	20	0	0	70	5	20	0	0	0	30	0	10	40	0	0	0
6	0	0	5	5	20	70	0	5	5	20	0	0	0	0	0	0	70
7	40	10	10	10	10	20	0	0	0	0	10	60	0	10	15	5	0
8	30	50	0	0	20	0	30	0	0	15	0	0	20	0	20	0	15
9	0	0	40	0	30	30	0	78	0	0	0	5.5	0	5.5	5.5	5.5	0
10	10	10	10	50	10	10	9	0	28	9	45	0	0	9	0	0	0
11	15	15	15	5	40	10	0	30	0	0	20	15	0	0	15	0	20
12	20	15	0	15	0	50	2	2	2	2	2	2	2	2	70	10	4
13	20	20	15	30	15	0	80	0	5	0	0	5	5	0	0	0	5
14	10	15	30	15	15	15	0	10	60	0	0	0	0	0	15	15	0
15	20	30	0	10	0	40	5	5	0	70	0	5	0	5	0	5	5
16	0	0	20	0	20	60	0	0	5	0	5	0	5	0	5	80	0
17	0	10	10	10	60	10	0	5	0	5	0	10	0	70	0	0	10
18	10	0	60	15	0	15	5	0	5	0	80	0	5	0	0	5	0
19	5	60	5	5	5	20	5	15	0	5	0	50	15	0	5	0	5
20	60	0	10	10	10	10	10	0	40	10	10	0	0	10	10	0	10
21	35	15	0	0	25	25	15	0	15	0	15	10	15	0	15	0	15
22	5	15	50	10	15	5	0	25	0	25	0	0	0	25	0	25	0
23	0	10	0	20	40	30	5	5	50	5	5	5	5	5	5	5	5
24	20	0	5	50	0	25	10	10	0	0	0	5	5	0	60	0	10
25	15	15	15	15	20	20	0	0	30	15	10	0	0	20	0	25	0
26	0	65	25	10	0	0	10	10	10	10	5	5	5	5	0	0	40
27	10	10	10	0	10	60	50	0	0	0	0	0	0	10	20	20	0
28	60	0	25	15	0	0	0	5	5	40	10	20	15	0	0	0	5
29	20	5	5	5	60	5	10	0	0	0	0	0	0	10	10	50	20
30	0	25	25	0	35	15	0	10	10	10	10	0	40	0	10	10	0

$$\mathrm{TICC} = \{[(y_{11} + y_{12} + \ldots + y_{1n}), \Delta t], [(y_{21} + y_{22} + \ldots + y_{2n}), 2\Delta t], \ldots, [(y_{m1} + y_{m2} + \ldots + y_{mn}), m\Delta t]\} \quad \text{(Eq 2)}$$

A TCAMS can be defined as an average response of each *m/z* fragment during the whole GC analysis as:

$$\mathrm{TCAMS} = \left\{\left[\bar{y}_1, \left({}^{m}\!/_{z}\right)_1\right], \left[\bar{y}_2, \left({}^{m}\!/_{z}\right)_2\right], \ldots, \left[\bar{y}_n, \left({}^{m}\!/_{z}\right)_n\right]\right\} = \left\{\left[\frac{(y_{11,} + y_{21} + \ldots + y_{m1})}{m}, \left({}^{m}\!/_{z}\right)_1\right], \left[\frac{(y_{12,} + y_{22} + \ldots + y_{m2})}{m}, \left({}^{m}\!/_{z}\right)_2\right], \ldots, \left[\frac{(y_{1n,} + y_{2n} + \ldots + y_{mn})}{m}, \left({}^{m}\!/_{z}\right)_n\right]\right\} \quad \text{(Eq 3)}$$

$\bar{y}_j$ is the average response of the MS for a single $(m/z)_j$ ion current over the whole scanning period (t), where all recorded scans were taken into account.

As can be observed, the TICC, which is the summation of all the recorded *m/z* fragment intensities within a single mass scan, is a time dependent function. Thus, the TICC response is dependant to the abundance of compounds producing those fragments. On the other hand, the TCAMS is the average response of the MS detector during the whole GC-MS chromatogram (Δt - $m\Delta t$) as a function of *m/z*. The average response of each mass fragment is proportional to the abundance of the compounds producing these ions [16,17].

2.5. Dimensionality reduction algorithms

2.5.1. Singular value decomposition (SVD)

The rank of a data matrix is the minimum number of linear independent variables. In the context of this work, rank can be defined as the number of pure EOs which exist in each chemical system (EO blend). The SVD algorithm can be used for chemical rank determination by performing decomposition of the experimental data matrix (**D**) into **U**, **S** and **V** matrices. **U** and **V** matrices are eigenvectors of the square matrices $\mathbf{DD^T}$ and $\mathbf{D^TD}$, while **S** is a diagonal matrix with the singular values arranged in descending order (the values are the positive square roots of the eigenvalues). The chemical rank of the blend can be determined by plotting its eigenvalues (Y-axis) against the chemical rank (X-axis) and pointing the last rank before the eigenvalues start to converge. In this process, it is crucial to be able to differentiate between the variations, arising from the chemical components and the instrumental noise. From the mathematical point of view, this corresponds to the determination of the cut-off between small eigenvalues (describing the chemistry of the system) and the noise, when the eigenvalues start to converge to small values [20].

2.5.2. Principal component analysis (PCA)

PCA is an unsupervised statistical pattern recognition technique that is used for dimensionality reduction, where trends and patterns among the sample set can be more readily identified in a space of lower dimension, e.g., two or three-dimensional space. PCA is explained in detail elsewhere [21–23]. PCA has been widely used for the dimensionality reduction for complex instrumental responses, such as GC-MS data [24]. In this work, PCA was applied to identify trends among instrumental profiles of the pure YY EOs and the profiles recovered from the perfumes using the MCR-ALS approach. In this way, it was possible to determine the grade and origin of the YY-EOs used in creating the perfume mixtures.

2.6. MCR-ALS

The MCR method can extract information of the pure components in a mixture system, through a bilinear model decomposition. The bilinear MCR model assumes that individual responses of each component are additive and the experimental data matrix (**D**) can be decomposed into product of matrices, **C** and $\mathbf{S}^T$, that contain pure profiles of the components connected to the row mode (concentration) and the column mode (spectra). In this study, the data were recorded as the three-way data arrays, as depicted in Fig. 1B [25], where one dimension is the retention time, another one is *m/z* and the third one is intensity of each *m/z*. Accordingly, TCAMS for each EO was created by averaging the responses of each *m/z* over all mass scans (whole GC run) and subtracting the average spectrum of the baseline noise (Fig. 1C). Such TCAMS were compiled into the experimental data matrix (**D**) (Fig. 1D), which, in the case of perfume characterisation was sized 13 × 200 (sample × variable), with the rows containing the samples' spectra and the columns the *m/z* variables. The experimental data **D** matrix for the EO blends qualitative analysis was sized 30 × 200, while for the quantitative analysis after adding TCAMS of the pure oils, was sized 41 × 200 and 36 × 200 for the eleven- and six-component blends, respectively. This bilinear model can be expressed in a matrix form as:

$$D = CS^T + E \qquad \text{(Eq. 4)}$$

where, **D** ($J \times K$) is the matrix of experimental data, in which J is the number of samples and K is the number of *m/z* fragments. **C** (size $J \times N$) is the matrix that its columns contain the concentration profiles and N is the chemical rank (number of pure EOs in the blend). $\mathbf{S}^T$ ($N \times K$) is the matrix whose rows contain the resolved TCAMS of pure EOs, used to create the blend. Matrix **E** ($J \times K$) is the matrix of the experimental errors (Eq. (4)).

The first step in MCR-ALS-based studies is obtaining the chemical ranks using the SVD algorithm (Fig. 1E). The MCR-ALS model is applied using a proposed number of components using the initial estimates of the pure EOs ($\mathbf{S}^T$), as depicted in Fig. 1F. During these iterative recalculations of **C** and $\mathbf{S}^T$, only a non-negativity constraint is applied, for both concentration and spectra. Iterations are continued until an optimal solution is obtained, which fulfils the postulated constraints and the established convergence criteria, and at the same time decreasing the residual matrix **E** [26]. The quality and reliability of the MCR-ALS solutions are assessed using the explained data variance, R^2, and the lack of data fit (LOF), as follow:

$$R^2 = 100\sqrt{\frac{\sum_{i,j}^{n} d_{ij}^2 - \sum_{i,j}^{n} e_{ij}^2}{\sum_{i,j}^{n} d_{ij}^2}} \qquad \text{(Eq. 5)}$$

$$LOF(\%) = 100\sqrt{\frac{\sum_{i,j}^{n} e_{ij}^2}{\sum_{i,j}^{n} d_{ij}^2}} \qquad \text{(Eq. 6)}$$

where, d_{ij} is an element of the experimental data matrix and e_{ij} is the related residual value obtained from the difference between the experimental data (**D**) and the reproduced data matrix ($\mathbf{CS}^T$), obtained by MCR-ALS. For the quantification purposes, correlation constraints were applied. With the correlation constraint, built into the MCR-ALS graphical user-friendly interface (GUI), an internal calibration was constructed by relating the known concentration values of the EOs in calibration samples to those provided by the ALS optimisation procedure. This algorithm rescaled the concentration values found by the MCR-ALS algorithm (in calibration and test samples) to the real concentration units, during the iterative optimisation [27]. The correlation constraints were used to create a calibration model for the quantification of YY-EOs and EOs within the blends (Fig. 1G).

Several parameters were used to evaluate the multivariate calibration models' predictive capability on the validation set, including root mean square error of prediction (RMSEP), bias, standard error of prediction (SEP), and relative error of prediction (REP), as presented in equations (7)–(10):

$$RMSEP = \sqrt{\frac{\sum_{i=1}^{n} (C_i - C_i')^2}{n}} \qquad \text{(Eq. 7)}$$

$$bias = \frac{\sum_{i=1}^{n} (C_i - C_i')}{n} \qquad \text{(Eq. 8)}$$

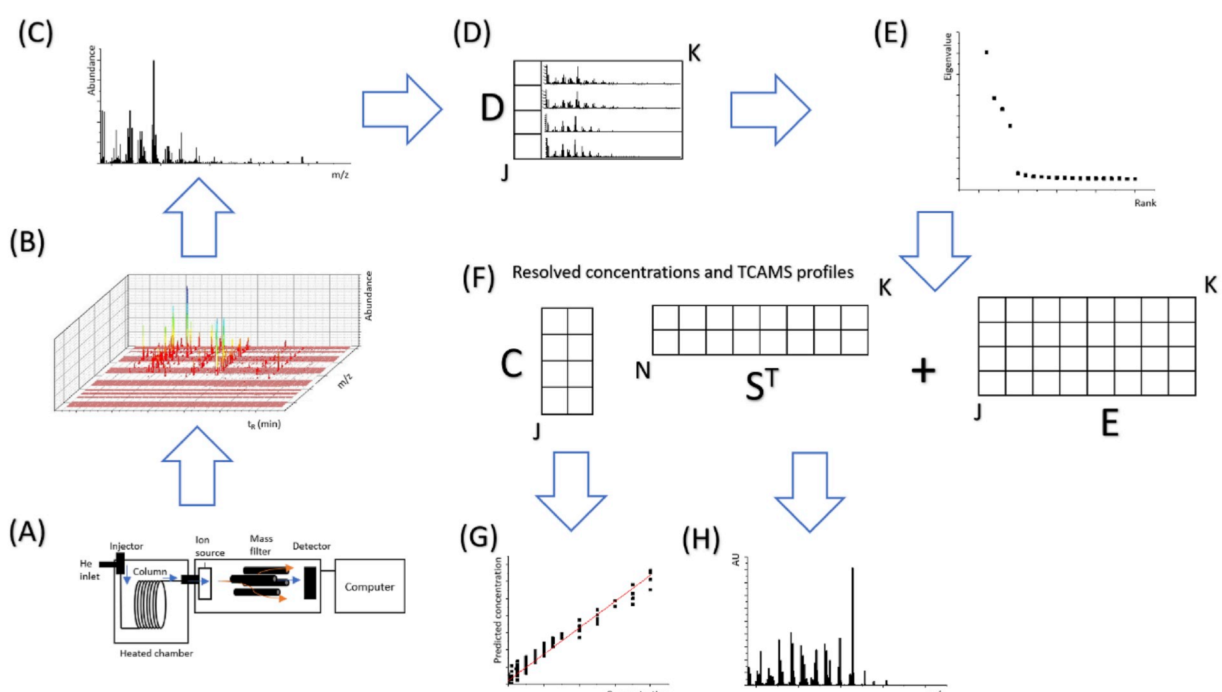

Fig. 1. Schematic representation of the pathway for the quality assessment of YY-EOs in the perfume mixtures, (A) GC-qMS, (B) three-way data array recorded for an EO blend, (C) averaging response for each *m/z* fragment over the chromatogram to create TCAMS, (D) three-way data matrix **D** containing TCAMS of all blend samples, (E) determination of chemical rank using SVD algorithm, (F) MCR-ALS, decomposition of **D** matrix into **C** and $\mathbf{S}^T$ matrices with TCAMS of each pure EO and residual matrix of **E**, (G) quantification of the pure EO in the blends using the external calibration and correlation constraints, and (H) resolved TCAMS for each EO blend constituent.

$$\mathrm{SEP} = \sqrt{\frac{\sum_{i=1}^{n}(C_i - C_i' - \mathrm{bias})^2}{n-1}} \quad \text{(Eq. 9)}$$

$$\mathrm{REP}\ (\%) = 100\sqrt{\frac{\sum_{i=1}^{n}(C_i - C_i')^2}{\sum_{i=1}^{n} C_i^2}} \quad \text{(Eq. 10)}$$

where, c_i and c_i' are the actual and predicted analyte concentration in sample i, respectively, and n is the total number of samples used in the validation set [27].

The MCR-BANDS, incorporated within MCR-ALS GUI 2.0, were employed to assess the rotational ambiguities associated with the obtained concentrations [28]. The concentrations and spectral profiles, resolved by MCR-ALS, were loaded directly, while the same constraints that have been previously used in the ALS optimisation, were chosen.

2.7. Random Forests classification model

RF is a classification statistical tool that generates many decision trees, each of which has been constructed using different sets of randomly selected input variables, to make prediction or class allocation. RF model uses a bootstrapping technique, in which the original data sets are randomly sampled by using 67% of the samples for the model training (*in-bag*). The rest of the data (*out-of-bag* set of 33%) is used for internal cross-validation and evaluating the performance of RF model through *out-of-bag error* (OOBE). In this way, each decision tree is independently produced without any pruning and each node is split using a defined number (*mtry* defined by user), selected randomly. By growing the forest up to the pre-defined number of trees (*ntree* defined by the user), through the bootstrap aggregation, multiple trees (all fitted on separate bootstrapped samples) are combined by averaging

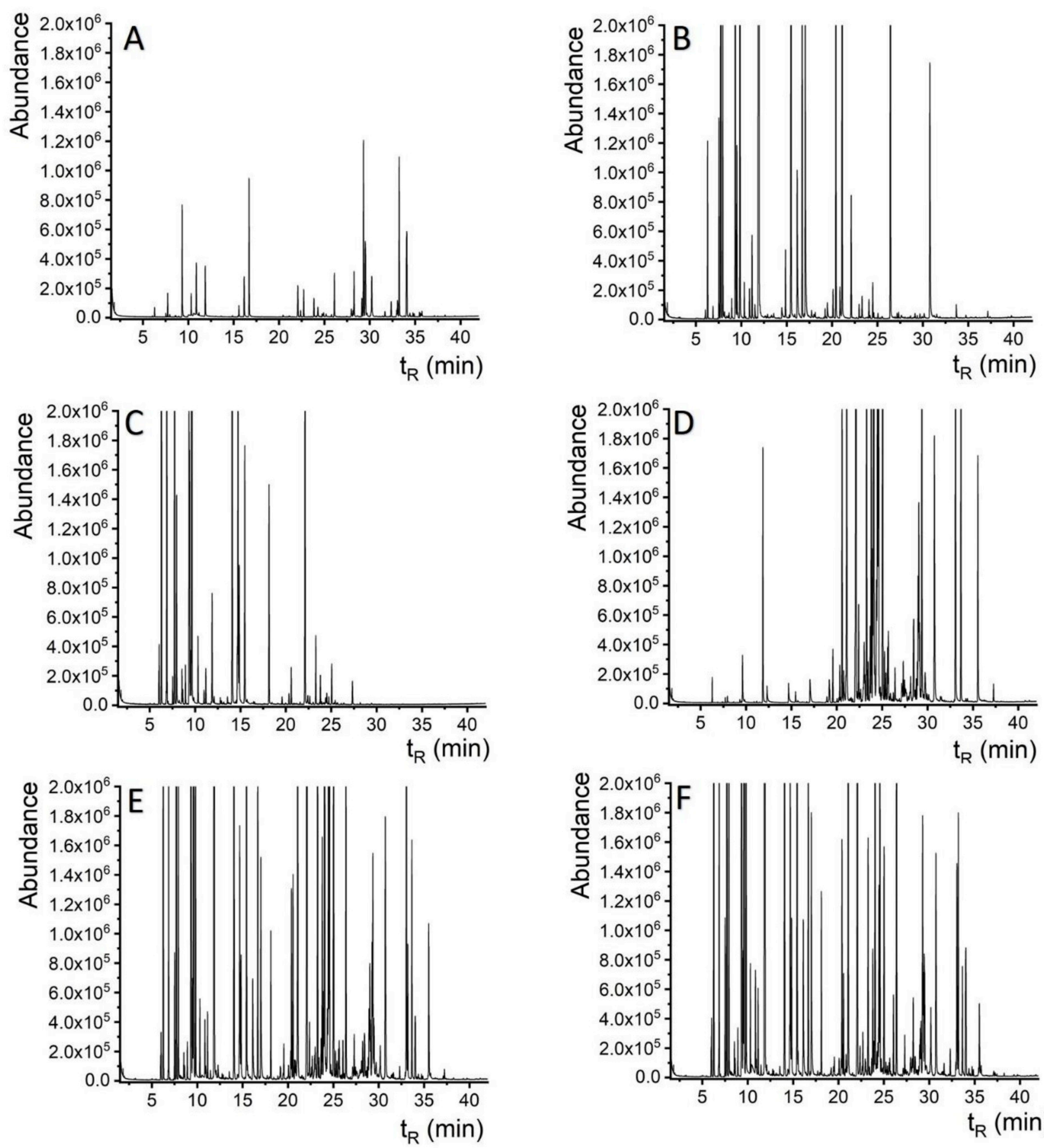

Fig. 2. TICC for the (A) perfume 1, (B) neroli, (C) rosemary, (D) ylang-ylang III, (E) Mix 6, and (F) Mix 10 (please see Table 3 for Mix 6 and 10).

their predictions, which reduces the overall variance of OOBE [29].

Herein, the RF approach was applied for the identification of resolved TCAMS, obtained from the MCR-ALS analysis of EOs blends. A RF model was created from 24 different classes of EOs (106 EO samples) by setting the number of variables used in each node (*mtry*) to 12 and the total number of trees (*ntree*) to 1000. The number of variables used in each node was determined as the square root of the total number of variables, which is the standard approach for classification of RF models [30]. The quality of the model was evaluated using OOBE, which is reported to be a good indicator of the predictive capabilities of RF models, due to the application of the bootstrapping technique. OOBE, as presented in this paper, was calculated as the average value of 50 repetitions.

3. Results and discussion

EOs are complex mixtures of volatile compounds and blending them with perfume or other EOs will create more complex mixtures (Fig. 2). As is clear from Fig. 2E and F, perfume-EO blends have many co-eluting regions, which cannot be resolved in one-dimensional GC-MS. This fact can be also seen in the chromatograms of six- and eleven-component EO blends (see Fig. S1). In this case, specific marker-compounds for the quality control purposes cannot easily be determined. Therefore, in this study, the whole instrumental signal (represented in the TCAMS) for each EO was used as a fingerprint for the characterisation of perfumes and EO mixtures.

Chemical rank of the mixtures was determined using the SVD algorithm and as it is clear from Fig. S2A, for each of the perfume

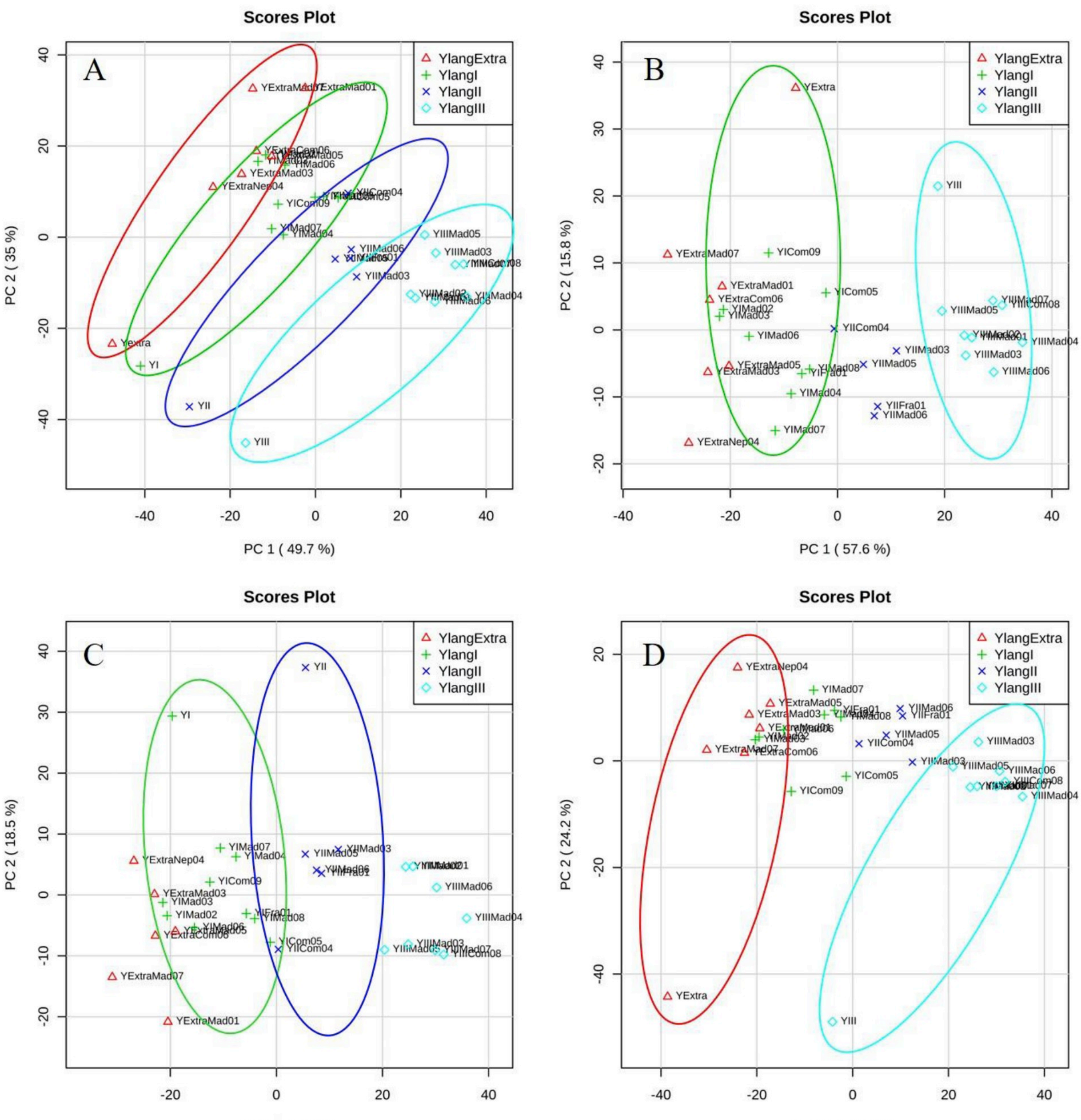

Fig. 3. PCA plot of different YY samples and their resolved grades from perfume-essential oil mixtures, (A) matrix #1, (B) matrix #2, (C) matrix #3, and (D) matrix #4.

mixtures was 4, while for the EO blends was 6 and 11 (see Figs. S2B and C). Initial estimate of the TCAMS of the perfume blend constituents was performed using a SIMPLISMA like algorithm [31,32], with the allowed noise level set at 5%. Initial TCAMS were, in the ALS process, iteratively adjusted through 50 steps to the experimental dataset **D**, obtaining final TCAMS used for the further quality assessment and classification.

Grade determination of YY-EOs, used in creating the perfume blends, was performed using PCA. PCA, as an unsupervised method, allowed that trends among YY-EOs to be easily observed in two- or three-dimensional space by calculation of variance among samples. Each perfume blend set was evaluated in a separate PCA. Resolved YY-TCAMS from each blend set, was joined with the TCAMS from the pure YY-EO. In this way, it was observed that resolved and pure TCAMS follow the same pattern. The Scree plots for all four matrices can be found in the supplement material (Fig. S3). Fig. 3 depicts the resolved TCAMS of the YY-EOs from the complex mixtures which follows the same pattern as pure oils in PCA. As can be seen for the matrix #1 (where all four grades were used, see Table 2) resolved YY profiles followed the same pattern observed between pure samples, allowing accurate grade identification (Fig. 3A). In case of the matrix #2, resolved lower grade of YY showed similarity with I grade, instead of YYExtra that has been used in the blend. Resolved YYIII followed the pattern of pure YYIII samples (Fig. 3B). In case of matrix #3 and #4, all of four YY grades (Extra, I, II, and III) were assigned to the appropriate grades based on the PCA plots, as have been depicted in Fig. 3C and D.

For the origin determination, ten YY-EOs (from 5 origins) were blended with PIII, myrrh and clary sage EOs (see Table 3, matrix #5). Pure TCAMS of perfume, myrrh and clary sage were added to the experimental matrix **D** in order to resolve the pure YY-EO TCAMS. To evaluate origin of unknown YY-EOs, the resolved pure TCAMS were analysed by the unsupervised PCA and HCA in order to observe clustering among the oils. The resolved TCAMS showed certain variances compared to the pure EOs, coming from the matrix used for creating the perfume blends. As shown in Fig. S4, comparison of the resolved TCAMS with the TCAMS of the pure EOs could not reveal the origin. To address this problem, one EO from each origin was blended to the

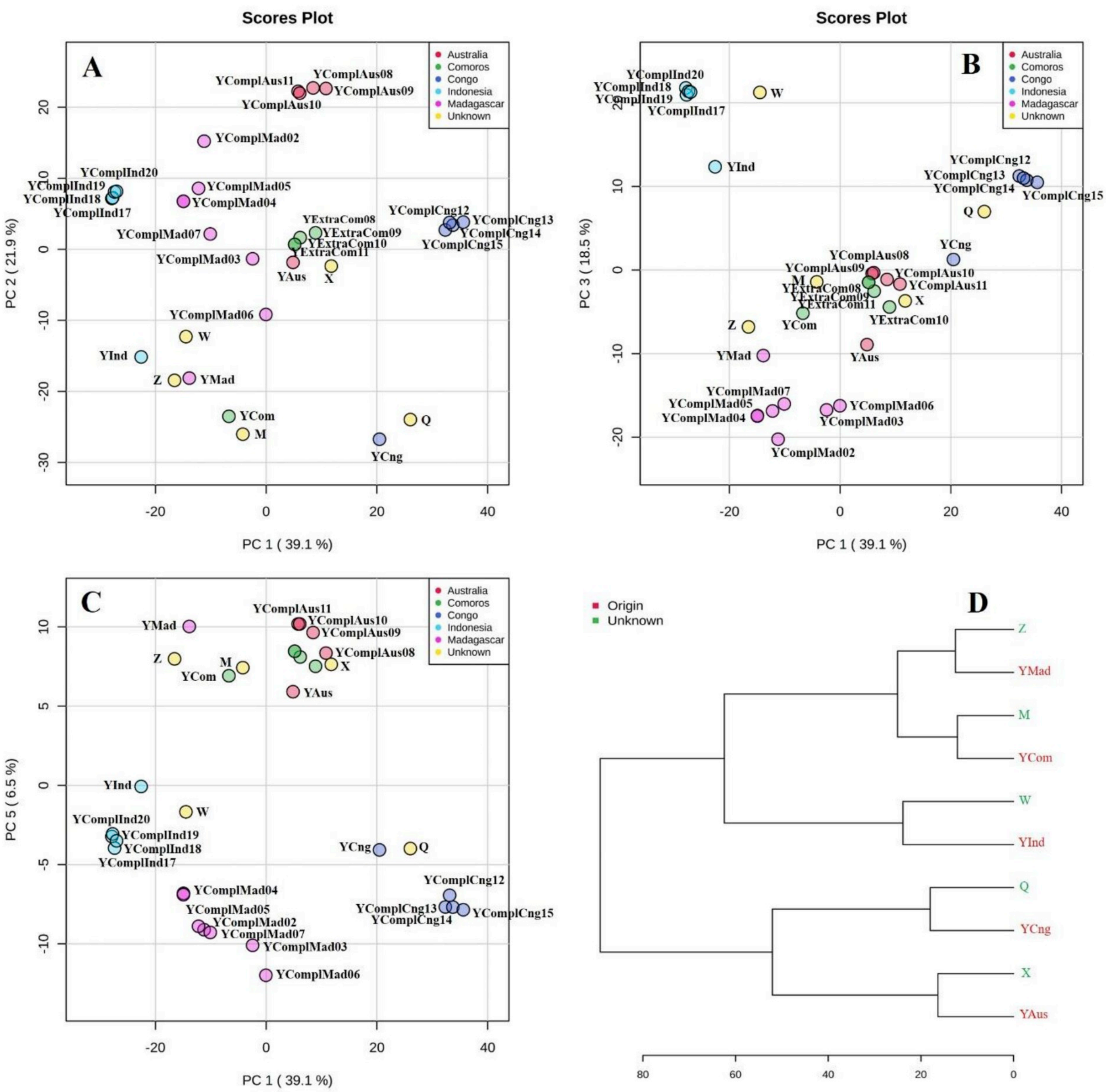

Fig. 4. PCA plots for the evaluation of origin of the resolved YY samples from perfume mixtures by using distances between resolved TCAMS (with known origin) and samples with unknown origin, (A) PC 1–2, (B) PC 1–3, (C) PC 1–5, and (D) HCA. Red marks are the reference samples and green marks are unknown samples. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

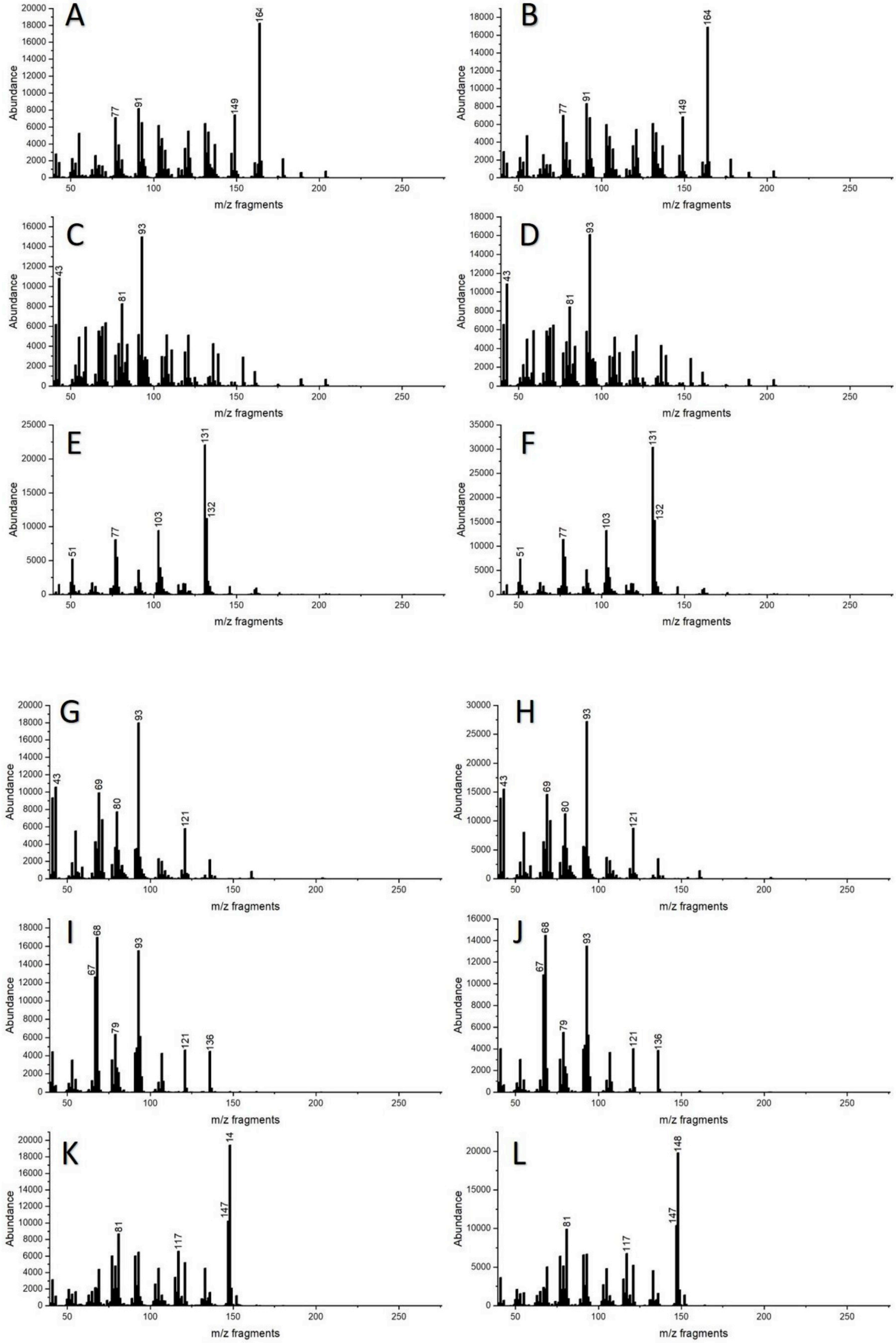

Fig. 5. The resolved TCAMS obtained from the experimental data set of the EO blends for allspice (A), cajeput (C), cinnamon (E), clary sage (G), grapefruit (I), and sweet fennel (K); and TCAMS of the pure EOs allspice (B), cajeput (D), cinnamon (F), clary sage (H), grapefruit (J), and sweet fennel (L).

Table 5
MCR-ALS parameters for each prediction model.

Parameters	Matrix/Grade										Origin					EO blend	
	1st				2nd		3rd		4th							6-component	11-component
	Extra	I	II	III	Extra	III	I	II	Extra	III	Aus	Ind	Con	Com	Mad		
REP	8.73	12.37	8.62	8.96	4.51	12.68	10.32	8.76	11.58	11.86	11.79	8.61	13.97	10.61	6.44	12.20	17.84
RMSEP	1.45	2.05	1.43	1.49	0.75	2.10	1.71	1.45	1.92	1.97	1.95	1.43	2.32	1.76	1.07	3.22	3.43
Bias	−1.04	−0.83	0.68	0.37	0.36	0.33	0.39	−0.04	1.10	1.55	−1.83	−0.88	−1.94	0.43	1.02	−1.54	−0.68
SEP	1.13	2.10	1.41	1.61	0.84	2.35	1.91	0.31	1.94	1.53	2.19	1.60	2.59	1.97	1.19	2.85	3.40

Aus: Australia; Ind: Indonesia; Con: Congo; Com: Comoros; Mad: Madagascar; REP: relative errors of prediction; RMSEP: root mean squared error of prediction; SEP: squared error of prediction.

perfume mixtures and resolved TCAMS from each origin were used as the reference spectra. In this way, the YY-EOs with unknown origins were determined by measuring the distance between reference and unknown oils in the PCA plot. As it can be seen from Fig. 4A, unknown YY-EO Q is in the vicinity of the reference oil from Congo, while sample X is close to the reference one from the Australia. The YY-EO W was discriminated by PCs 1 and 3 (Fig. 4B), showing a high similarity with the reference oil from Indonesia. Sample M showed a great similarity with Comoros reference oil, while Z sample was closest to the Madagascar EO (Fig. 4C). By using Euclidean distance [33] and Ward's linkage [34], as the HCA methods, the results of the PCA plots were confirmed (see Fig. 4D).

The RF algorithm was used to identify the classes of EOs used in creating the blends, since through resolved pure TCAMS data obtained from the MCR-ALS analysis. In comparison with the RF model, partial least squares discriminatory analysis (PLS-DA) could not discriminate the EO classes, having the sensitivity = 0. EO classification model has already been applied for the classification and quality control of EOs [16]. First, a RF classification model was created to differentiate between 24 EO classes, by setting the number of trees (*ntree*) to 1000 and using 12 variables for each node (*mtry*). This process was performed 50 times resulting in an OOBE of 2.34 ± 0.81%. Resolved TCAMS of two EO blend sets were used to determine their class from the RF model, created from the pure EO samples. The resulting recovered TCAMS and the TCAMS obtained from the analysis of pure EOs are compared in Fig. 5. Excellent similarities between the two can be clearly seen. The similarity between pure and resolved TCAMS allowed positive identification of all 17 EOs used in creating the two blend sets, through the RF EO classification model.

The quantitation of YY-EOs in perfume-EO blends was performed using MCR-ALS analysis enforced with correlation constraints. For this purpose, TCAMS obtained in the analysis of the pure oils were added to the experimental data matrix **D** and chemical rank was again determined through the SVD algorithm. The pure EOs were manually selected from the experimental data set. The quantification was accomplished by applying the correlation constraints, which build internal univariate calibration models between the reference concentration values in the calibration samples and the analogous values in the MCR concentration profiles. For the perfume mixtures, 5 samples were used to prepare the internal calibration curve, while in case of EO blends 20 and 22 samples for six-component and eleven-component EO blends were used, respectively (see Tables S1–3). The relative errors of prediction (REP) for the grade determination were 4.51–12.68%, while in the case of origin determination, they found to be 6.44–13.97% (Table 5). REP for the six-component blend validation sample set was 12.20%, while for the eleven-component blend was 17.84%. The root mean squared error of prediction (RMSEP) for all predictions was in the range of 0.75–3.43 (see Table 5). This clearly showed the ability of the model to deliver accurate quantification of YY-EO in complex blends. It can be concluded from the results that this model is a fast and efficient way to determine the quality of the perfume blends based on the quality of the EOs used in creating the blends. Additionally, the rotational ambiguity was assessed by applying the same constraints used in the ALS iteration step. Results of the MCR-BANDS method clearly showed that the extent of the rotational ambiguities, measured by the difference between fmax and fmin values were negligible (lower than 10^{-7}).

4. Conclusion

The MCR-ALS strategy was applied on the easily generated TCAMS data obtained from the GC-MS analysis, for the quality control of complex perfume mixtures and EO blends. The ability of the SVD algorithm for accurate identification of number of individual EOs, used for creating the blends, was demonstrated. MCR-ALS showed a great ability in the qualitative and quantitative characterisation of perfume and EO blends. Resolved pure TCAMS were successfully applied with PCA to identify grade of YY EOs. It was necessary to establish reference EOs with known origins for the origin determination, in order to determine origin of unknown set of oils through the comparison with the reference EOs in the PCA and HCA models. Additionally, by utilising the RF classification model, all 17 EOs used to create six- and eleven-component blends, were positively identified.

The quantification was performed using MCR-ALS, enforced with the correlation constraints. It was demonstrated that REP values in the validation sample sets were smaller than 18%, while by application of correlation constraints and RMSEPs were smaller than 3.43. These findings clearly showed that GC-MS combined with MCR-ALS, PCA and HCA can be successfully used in the quality control of high-value perfumes and EO blends.

Ethical approval

This article does not contain any studies with human participants or animals performed by any of the authors.

Funding

This work was supported (grant number: IH170100020) by the ARC Industrial Transformation Research Hub for Processing Advanced Lignocellulosics (PALS).

CRediT authorship contribution statement

Leo Lebanov: Investigation, Conceptualization, Methodology, Writing - original draft. **Laura Tedone:** Methodology, Investigation. **Alireza Ghiasvand:** Conceptualization, Methodology, Writing - review & editing. **Brett Paull:** Conceptualization, Methodology, Writing - review & editing.

Declaration of competing interest

The authors have declared that there is no conflict of interest.

Acknowledgement

Authors would like to thank the University of Tasmania and PALs Research Hub for their generous provision of co-funded PhD scholarship and project support for Leo Lebanov. Authors would also like to thank Plant Therapy, Ahimsa Oils, and Albert Vieille companies for providing the essential oil samples.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.talanta.2020.121208.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] L.A. de Godoy, L.W. Hantao, M.P. Pedroso, R.J. Poppi, F. Augusto, Quantitative analysis of essential oils in perfume using multivariate curve resolution combined with comprehensive two-dimensional gas chromatography, Anal. Chim. Acta 699 (1) (2011) 120–125, https://doi.org/10.1016/j.aca.2011.05.003.
[2] P.J. Marriott, S.-T. Chin, B. Maikhunthod, H.-G. Schmarr, S. Bieri, Multidimensional gas chromatography, Trends Anal. Chem. 34 (2012) 1–21, https://doi.org/10.1016/j.trac.2011.10.013.
[3] L. Lebanov, L. Tedone, M. Kaykhaii, M.R. Linford, B. Paull, Multidimensional gas chromatography in essential oil analysis. Part 1: technical developments, Chromatographia 82 (1) (2019) 377–398, https://doi.org/10.1007/s10337-018-3649-3.
[4] J.M. Amigo, T. Skov, R. Bro, J. Coello, S. Maspoch, Solving GC-MS problems with PARAFAC2, Trends Anal. Chem. 27 (8) (2008) 714–725, https://doi.org/10.1016/j.trac.2008.05.011.
[5] L.G. Johnsen, P.B. Skou, B. Khakimov, R. Bro, Gas chromatography - mass spectrometry data processing made easy, J. Chromatogr. A 1503 (2017) 57–64, https://doi.org/10.1016/j.chroma.2017.04.052.
[6] P. Ma, Z. Zhang, X. Zhou, Y. Yun, Y. Liang, H. Lu, Feature extraction from resolution perspective for gas chromatography-mass spectrometry datasets, RSC Adv. 6 (115) (2016) 113997–114004, https://doi.org/10.1039/c6ra17864b.
[7] R. Tauler, Multivariate curve resolution applied to second order data, Chemometr. Intell. Lab. Syst. 30 (1) (1995) 133–146, https://doi.org/10.1016/0169-7439(95)00047-X.
[8] S.A. Bortolato, J.A. Arancibia, G.M. Escandar, Non-trilinear chromatographic time retention – fluorescence emission data coupled to chemometric algorithms for the simultaneous determination of 10 polycyclic aromatic hydrocarbons in the presence of interferences, Anal. Chem. 81 (19) (2009) 8074–8084, https://doi.org/10.1021/ac901272b.
[9] A. de Juan, R. Tauler, Comparison of three-way resolution methods for non-trilinear chemical data sets, J. Chemom. 15 (10) (2001) 749–771, https://doi.org/10.1002/cem.662.
[10] S.A. Bortolato, A.C. Olivieri, Chemometric processing of second-order liquid chromatographic data with UV-vis and fluorescence detection. A comparison of multivariate curve resolution and parallel factor analysis 2, Anal. Chim. Acta 842 (2014) 11–19, https://doi.org/10.1016/j.aca.2014.07.007.
[11] J.J. Filippi, E. Belhassen, N. Baldovini, H. Brevard, U.J. Meierhenrich, Qualitative and quantitative analysis of vetiver essential oils by comprehensive two-dimensional gas chromatography and comprehensive two-dimensional gas chromatography/mass spectrometry, J. Chromatogr. A 1288 (2013) 127–148, https://doi.org/10.1016/j.chroma.2013.03.002.
[12] https://www.statista.com/statistics/673641/global-market-value-for-natural-cosmetics/ , Accessed date: 1 April 2020.
[13] M. Brokl, M.L. Fauconnier, C. Benini, G. Lognay, P. du Jardin, J.F. Focant, Improvement of ylang-ylang essential oil characterization by GCxGC-TOFMS, Molecules 18 (2) (2013) 1783–1797, https://doi.org/10.3390/molecules18021783.
[14] A. Orchard, S. van Vuuren, Commercial essential oils as potential antimicrobials to treat skin diseases, Evid. Based Complement, Alternative Med. 2017 (2017) 4517971, https://doi.org/10.1155/2017/4517971.
[15] P. Rubiolo, B. Sgorbini, E. Liberto, C. Cordero, C. Bicchi, Essential oils and volatiles: sample preparation and analysis. A review, Flavour Fragrance J. 25 (5) (2010) 282–290, https://doi.org/10.1002/ffj.1984.
[16] L. Lebanov, L. Tedone, A. Ghiasvand, B. Paull, Random Forests machine learning applied to gas chromatography – mass spectrometry derived average mass spectrum data sets for classification and characterisation of essential oils, Talanta 208 (2020) 120471, https://doi.org/10.1016/j.talanta.2019.120471.
[17] L. Lebanov, S. Chatterjee, L. Tedone, S.C. Chapman, M.R. Linford, B. Paull, Comprehensive characterisation of ylang-ylang essential oils according to distillation time, origin, and chemical composition using a multivariate approach applied to average mass spectra and segmented average mass spectral data, J. Chromatogr. A 1618 (2020) 460853, https://doi.org/10.1016/j.chroma.2020.460853.
[18] J. Chong, O. Soufan, C. Li, I. Caraus, S. Li, G. Bourque, D.S. Wishart, J. Xia, MetaboAnalyst 4.0: towards more transparent and integrative metabolomics analysis, Nucleic Acids Res. 46 (W1) (2018) W486–W494, https://doi.org/10.1093/nar/gky310.
[19] J. Jaumot, A. de Juan, R. Tauler, MCR-ALS GUI 2.0: new features and applications, Chemometr. Intell. Lab. Syst. 140 (2015) 1–12, https://doi.org/10.1016/j.chemolab.2014.10.003.
[20] L.W. Hantao, H.G. Aleme, M.P. Pedroso, G.P. Sabin, R.J. Poppi, F. Augusto, Multivariate curve resolution combined with gas chromatography to enhance analytical separation in complex samples: a review, Anal. Chim. Acta 731 (2012) 11–23, https://doi.org/10.1016/j.aca.2012.04.003.
[21] S. Chatterjee, B. Singh, A. Diwan, Z.R. Lee, M.H. Engelhard, J. Terry, H.D. Tolley, N.B. Gallagher, M.R. Linford, A perspective on two chemometrics tools: PCA and MCR, and introduction of a new one: pattern recognition entropy (PRE), as applied to XPS and ToF-SIMS depth profiles of organic and inorganic materials, Appl. Surf. Sci. 433 (2018) 994–1017, https://doi.org/10.1016/j.apsusc.2017.09.210.
[22] J.E. Jackson, A User's Guide to Principal Components, John Wiley & Sons, Inc., New York1991.
[23] R. Bro, A.K. Smilde, Principal component analysis, Anal. Methods 6 (9) (2014) 2812–2831, https://doi.org/10.1039/c3ay41907j.
[24] B.N. Nsuala, G.P. Kamatou, M. Sandasi, G. Enslin, A. Viljoen, Variation in essential oil composition of Leonotis leonurus, an important medicinal plant in South Africa, Biochem. Systemat. Ecol. 70 (2017) 155–161, https://doi.org/10.1016/j.bse.2016.11.009.
[25] M. Jalali-Heravi, H. Parastar, Recent trends in application of multivariate curve resolution approaches for improving gas chromatography-mass spectrometry analysis of essential oils, Talanta 85 (2) (2011) 835–849, https://doi.org/10.1016/j.talanta.2011.05.045.
[26] V. Boeris, J.A. Arancibia, A.C. Olivieri, Determination of five pesticides in juice, fruit and vegetable samples by means of liquid chromatography combined with multivariate curve resolution, Anal. Chim. Acta 814 (2014) 23–30, https://doi.org/10.1016/j.aca.2014.01.034.
[27] R.R. de Oliveira, K.M. de Lima, R. Tauler, A. de Juan, Application of correlation constrained multivariate curve resolution alternating least-squares methods for determination of compounds of interest in biodiesel blends using NIR and UV-visible spectroscopic data, Talanta 125 (2014) 233–241, https://doi.org/10.1016/j.talanta.2014.02.073.
[28] J. Jaumot, R. Tauler, MCR-BANDS: a user friendly MATLAB program for the evaluation of rotation ambiguities in Multivariate Curve Resolution, Chemometr. Intell. Lab. Syst. 103 (2) (2010) 96–107, https://doi.org/10.1016/j.chemolab.2010.05.020.
[29] V. Svetnik, A. Liaw, C. Tong, J.C. Culberson, R.P. Sheridan, B.P. Feuston, Random Forest: A classification and regression tool for mompound mlassification and QSAR modeling, J. Chem. Inf. Comput. Sci. 43 (6) (2003) 1947–1958, https://doi.org/10.1021/ci034160g.
[30] M. Belgiu, L. Drăguţ, Random forest in remote sensing: a review of applications and future directions, ISPRS J. Photogrammetry Remote Sens. 114 (2016) 24–31, https://doi.org/10.1016/j.isprsjprs.2016.01.011.
[31] W. Windig, J. Guilment, Interactive self-modeling mixture analysis, Anal. Chem. 63 (14) (1991) 1425–1432, https://doi.org/10.1021/ac00014a016.
[32] M. Wasim, R.G. Brereton, Determination of number of significant components and key variables using genetic algorithms in liquid chromatography-nuclear magnetic resonance spectroscopy and liquid chromatography-diode array detection, Chemometr. Intell. Lab. Syst. 81 (2) (2006) 209–217, https://doi.org/10.1016/j.chemolab.2005.12.006.
[33] D. Granato, J.S. Santos, G.B. Escher, B.L. Ferreira, R.M. Maggio, Use of principal component analysis (PCA) and hierarchical cluster analysis (HCA) for multivariate association between bioactive compounds and functional properties in foods: a critical perspective, Trends Food Sci. Technol. 72 (2018) 83–90, https://doi.org/10.1016/j.tifs.2017.12.006.
[34] J.H. Ward, Hierarchical grouping to optimize an objective function, J. Am. Stat. Assoc. 58 (301) (1963) 236–244, https://doi.org/10.1080/01621459.1963.10500845.