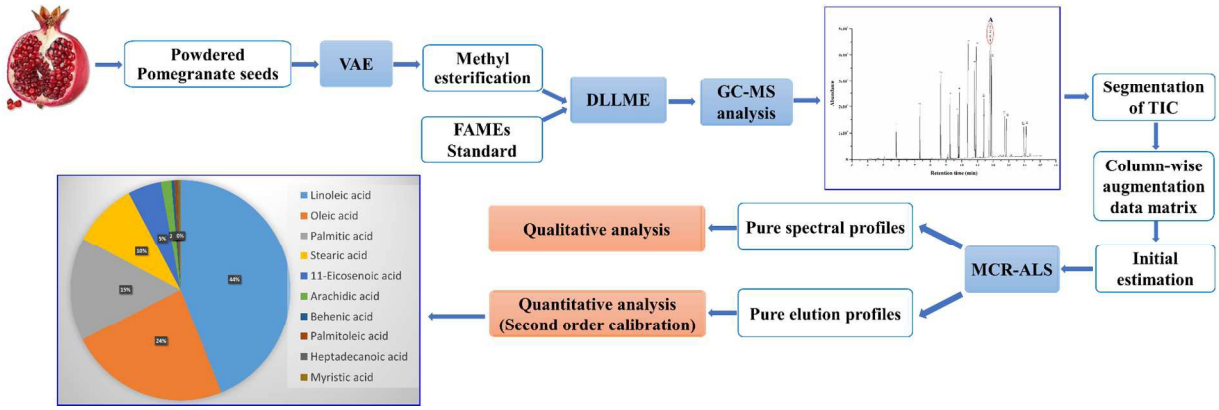
RSC Advances

ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/advances

Graphical Abstract



Second-order calibration for determination of fatty acids in pomegranate seeds by vortex assisted-extraction dispersive liquid-liquid micro-extraction and gas chromatography-mass spectrometry

Mohammad Ahmadvand[a], Hassan Sereshti[a], Hadi Parastar[b,*]

[a]*Department of Chemistry, Faculty of Science, University of Tehran, Tehran, Iran*

[b]*Department of Chemistry, Sharif University of Technology, P.O. Box 11155-3516, Tehran, Iran*

* Author to whom correspondence should be addressed;

Tel.: +98- 21-66165306; fax: +98- 21-66005718

E-mail: h.parastar@sharif.edu, h.parastar@gmail.com

Abstract

Multivariate curve resolution-alternating least squares (MCR-ALS) as a second-order calibration algorithm was proposed for simultaneous analysis of eighteen fatty acid methyl esters (FAMEs) in a standard mixture and pomegranate seed sample using vortex-assisted extraction-dispersive liquid-liquid microextraction (VAE-DLLME) followed by gas chromatography–mass spectrometry (GC-MS). Chemometric resolution, identification and quantification of the target FAMEs in the standard mixture and real sample (pomegranate seed) were carried out successfully in the presence of some uncalibrated interferences. The lack of fit (LOF) and reverse match factor (RMF) were used for evaluation of the MCR-ALS results of calibration samples. The LOF (%) and RMF values were in the ranges 7.24-24.36 and 708-977, respectively. In addition, regression coefficients (R^2) and relative errors (REs, %) of calibration curves of different FAs were in the satisfactory range of 0.9934-0.9989 and 3.70-7.45, respectively. Application of the proposed strategy to pomegranate seed extract showed that linoleic acid, oleic acid, palmitic acid, stearic acid and cis-11-eicosenoic acid were respectively the main fatty acids of pomegranate seed with concentration of 9098.0, 4873.0, 3147.0, 1960.0 and 1019.0 mg kg^{-1} and relative standard deviations (RSD, %) between 0.15-11.57. It is concluded that MCR-ALS combined to VAE-DLLME-GC-MS is a fast and simple strategy to be used for qualitative and quantitative analysis of complex samples such as natural products.

***Keywords*:** Second-order calibration; Multivariate curve resolution-alternating least squares; Vortex-assisted extraction-dispersive liquid-liquid microextraction; Gas chromatography-mass spectrometry; Fatty acids; Pomegranate seeds.

1 Introduction

Determination and quantification of fatty acids (FAs) in different sample matrices including blue crab[1], fish oil[2], beer[3], alcoholic beverages and tobaccos[4], raw spirits[5] and wild mushroom[6] have attracted the attention of health and nutrition researchers owing to their significant role in biological tissues. Pomegranate is one of the native fruits of Iran with annual production of about 700,000 tons. Regarding the high content of FAs and other bioactive components, pomegranate seeds can be used as a rich source of FAs. Hence, extraction and characterization of pomegranate seed compounds have attracted considerable attention in recent years.[7,8] The amount of some FAs in pomegranate seeds has been obtained by different extraction methods such as soxhlet extraction, supercritical fluid extraction (SFE) and ultrasonic-assisted solvent extraction (UASE).[9-11] However, quantification of FAs has always been a challenging task due to insufficient sensitivity, the presence of a variety of polyunsaturated fatty acids (PUFA) isomers, and coeluted components.[12,13]

Among different extraction techniques, dispersive liquid-liquid microextraction (DLLME) has been introduced by Assadi and co-workers in 2006 as a simple and efficient extraction/preconcentration technique.[14] In general, the DLLME technique consists of two simple steps: (1) rapid injection of appropriate mixture of extractor and disperser solvent into aqueous sample containing the sought analyte(s); and (2) centrifugation of cloudy solution to separate the phases. In the first step, a stable cloudy solution (containing very fine droplets of extraction solvent dispersed into the aqueous phase) is formed. Owing to the large surface area between the extractor solvent and the aqueous sample, equilibrium state is quickly achieved. In the second step, after centrifugation, the organic phase is separated and is analyzed by an appropriate instrumental technique.[15] Only in less than one decade after introduction of DLLME,

it has been frequently used for extraction and preconcentration of broad range of organic and inorganic compounds from different sample matrices, such as foods, environmental, and biological samples.[16-18]

Gas chromatography coupled to mass spectrometry (GC-MS) is the most frequently used instrument for identification and quantification of FAs in animal and plant samples.[19] The derivatization of the FAs to their methyl esters (FAMEs) is a primary step in GC separation to increase the resolution and sensitivity of the FAs analysis. Very often, overlapped regions in GC profiles of FAMEs (specially unsaturated FAMEs) can be observed due to the complexity of FAMEs matrices. Therefore, finding proper conditions for comprehensive separation, identification, and quantification of saturated and unsaturated FAMEs particularly in the presence of other interferences is a very difficult task.

Fortunately, over the past decades, second-order calibration methods have attracted great attention of analytical chemists due to their great achievements in improving the sensitivity, increasing the selectivity and modeling of analyte contribution in the presence of uncalibrated interferences.[20] Various second-order multivariate calibration algorithms have been developed in the recent decade and they have been reviewed in refs.[21-23]

Among different second-order calibration methods, multivariate curve resolution-alternating least squares (MCR-ALS) and parallel factor analysis 2 (PARAFAC2)[24] have been proposed and extensively used to resolve multiple pure responses and concentrations of the components present in unknown mixtures. However, Different studies have been done to show the superiority of MCR-ALS over PARAFAC2 when fundamental chromatographic problems exist.[25] PARAFAC2 is a variant of the PARAFAC technique which has been applied to chromatographic data. Strict trilinearity in the PARAFAC2 model is not necessary, although it does not allow for

significant shape changes in the elution peak across different runs (the extent of the allowed peak deformation in practice is still unknown). PARAFAC2 allows for analysis of data where there are moderate elution time shifts in chromatographic dimension due to, for example, temperature programming or misalignment across samples. However, PARAFAC2 is computationally more complex and expensive, and it does not allow for constraints (e.g., non-negativity and/ or unimodality) in the chromatographic direction and therefore negative values and multimodal peaks may appear in the results. Also, applying constraints selectively only to some selected components is not possible.

The MCR-ALS has been applied to data collected from multi-component liquid chromatography-mass spectrometry (LC-MS),[26] gas chromatography-mass spectrometry (GC-MS),[2] and high performance liquid chromatography-diode array detection (HPLC-DAD)[27] and fluorescence excitation–emission matrices (EEM).[28] In MCR-ALS, the measured analytical signals are assumed to follow a generalized bilinear additive model and the contribution of each component to the measured signal depends on its concentration and on its spectral profile. MCR-ALS can also be used to obtain quantitative analytical information because any type of constraints can be easily applied to the solutions.

In the present contribution, MCR-ALS was used as a second-order calibration algorithm for identification and quantification of FAMEs in the standard mixture and pomegranate seed samples. For this purpose, the simple microextraction method of VAE-DLLME followed by GC-MS was carried out on a standard mixture of FAMEs. Then, the data were analyzed by MCR-ALS algorithm to overcome the chromatographic problems, to build calibration curves and to obtain analytical figures of merit. Finally, the pomegranate seed FAMEs were investigated as the test sample for evaluation of the proposed calibration method.

116 **Abbreviations**

117 **COW**: correlation optimized warping; **DLLME**: dispersive liquid-liquid micro-extraction;
118 **DTW**: dynamic time warping; **EEM**: excitation–emission matrices; **EFA**: evolving factor
119 analysis; **FA**: fatty acid; **FAME**: fatty acid methyl ester; **GC-MS**: gas chromatography-mass
120 spectrometry; **HPLC-DAD**: high performance liquid chromatography-diode array detection;
121 **LC-MS**: liquid chromatography-mass spectrometry, **LOF**: lack of fit; **MCR-ALS**: multivariate
122 curve resolution-alternating least squares; **OVI**: overall volume integration, **PUFA**:
123 polyunsaturated fatty acid; **RE**: relative error; **RMF**: reverse match factor; **RSD**: relative
124 standard deviation; **SFE**: supercritical fluid extraction; **SIMPLISMA**: simple-to use interactive
125 self-modeling mixture analysis; **SVD**: singular value decomposition; **TIC**: total ion
126 chromatogram; **UASE**: ultrasonic-assisted solvent extraction; **VAE**: vortex assisted-extraction.
127

128 ## 2 Experimental

129 ### 2.1 Chemicals and reagents

130 Fatty acid methyl esters (FAMEs) mixture (Grain FAME mix. 10 mg mL^{-1} in CH_2Cl_2) was
131 purchased from Sigma Aldrich (St. Louis, MO, USA). Pentadecanoic acid, chloroform, methanol
132 and sodium chloride with the purity higher than 99.0% were purchased from Merck (Darmstadt,
133 Germany). Pure helium (99.999%) was obtained from Roham gas company (Tehran, Iran,
134 http://www.roham.com).
135
136

137 ### 2.2 Sample preparation: extraction, esterification and preconcentration of FAs

138 Fresh pomegranates were obtained from a market in Karaj (Alborz province, Iran). The
139 pomegranates were crushed and their seeds were separated from other parts. The dewatered
140 seeds were washed with tap water several times for the removal of residuals and then were dried
141 at room temperature for 48 h. The dried seeds were powdered by a household grinder and sieved

using a 60-mesh sieve (pore size 0.3 mm). The obtained powder was stored in a glass vessel at 4 °C for subsequent analyses.

A portion of the seed powder (0.3 g) was placed into a 5 mL screw cap glass test tube and then 2 mL methanol (extraction solvent) was added to it. Afterward, the mixture was subjected to vortexing using a vortex-mixer (Velp Scientifica, Milan, Italy) for 3 min. After the extraction, the solid particles were separated from the mixture by centrifugation, and the supernatant was transferred to another test tube. Then, pentadecanoic acid was added to it as internal standard (5 mg L^{-1}). In the next step, a simple esterification procedure was performed by adding one drop of concentrated H_2SO_4 to 1 mL of the supernatant fixed in a water bath (at 70 °C for 60 min). Then, 0.5 mL of the esterified solution was placed into a test tube and 28 µL chloroform (preconcentration solvent) was added to it. After that, the solution was injected rapidly into a conic bottom test tube containing 2 mL aqueous NaCl solution (5% w/v). Accordingly, a cloudy solution containing tiny droplets of chloroform dispersed into the aqueous phase was formed. In this step, the analytes were extracted from the aqueous sample into the chloroform droplets. Then, the organic extraction phase was separated by centrifugation at 4000 rpm for 3 min. Finally, 0.5 µL of the lower phase (chloroform phase) was injected into GC-MS using a 1.0 µL microsyringe.

2.3 GC-MS analysis

A 6890 GC system coupled with a 5973 network mass selective detector (Agilent Technologies, Santa Clara, CA, USA) and equipped with a HP5-MS capillary fused silica column (30 m length; 0.25 mm I.D.; 0.25 µm film thicknesses, methyl 5% phenyl polysiloxane) was used for analysis of the samples. The temperature program started at 50 °C for 1 min then

increased with the rate of 20 °C min^{-1} to 270 °C, held for 3 min. The total GC run time was 15 min. In this regard, the chromatographic conditions were chosen to achieve a reasonable chromatographic resolution and a short analysis time. The carrier gas (helium, 99.999%) was maintained at a constant pressure of 36 psi with a flow rate of 1 mL min^{-1} (4 min solvent delay). Other operating conditions were as follows: injection volume, 0.5 µL; split ratio, 1:20; interface temperature, 250 °C; and ion source temperature, 230 °C. Mass spectra were taken at 70 eV ionization energy and full scan mode. The scanned mass range was set at 50-450 m/z. An enhanced ChemStation software package (G1701 DA-MSD, Rev. D.00.01.27) was used for the data collection and analysis in GC-MS.

2.4 Chemometric analysis

Hyphenated systems such as HPLC-DAD, LC–MS, and GC-MS produce a large volume of data that can be stored in a rectangular data arrays called second-order data matrices. This mixed data matrix contains information from two different data directions (i.e., rows and columns) and can be decomposed into the contribution of pure component profiles of the constituents, by means of a simple bilinear data decomposition, which is defined as follow:

$$\mathbf{X} = \mathbf{C}\mathbf{S}^{T} + \mathbf{E} \quad (1)$$

where $\mathbf{X}$ (I,J) is the raw experimental data matrix, $\mathbf{C}$ (I,N) is the factor matrix which contains the resolved concentration profiles of the N components present in the data matrix, and $\mathbf{S}$ (N,J) is the factor matrix of their corresponding pure spectral profiles. The part of data which is not explained by the model is in the error $\mathbf{E}$ (I,J) matrix. Indices I and J are the numbers of row and column variables, respectively, spectral channels (e.g., wavelengths or m/z ratios), and time points (e.g., elution times). In addition, N is the number of eluted components in the analyzed

188 data matrix **X**. The number of chemical components in each chromatographic region was
189 determined using singular value decomposition (SVD). Also, evolving factor analysis (EFA) was
190 used to confirm the number of present components in each region. Furthermore, MCR-ALS
191 model was checked with fewer and more components to make sure about the correct number of
192 chemical components in each chromatographic region. In this regard, the change in lack of fit
193 (LOF) of MCR-ALS model was used as a criterion for decision about the presence or absence of
194 a component in the model.

195 Eq. (1) is solved for **C** and $\mathbf{S}^T$, using an iterative algorithm based on two constrained linear
196 least squares steps. It requires an initial estimation of the concentration, **C**, or of the spectra, $\mathbf{S}^T$,
197 profiles, which can be easily obtained using different methods or from the ‘‘purest’’ data
198 samples or variables (e.g., simple-to use interactive self-modeling mixture analysis
199 (SIMPLISMA)[29]). Although, finding unique solutions of Eq. (1) is impossible when only the
200 information about the data matrix **X** is provided, the use of constraints (e.g., non-negativity or
201 any other previously known property or constraint about the nature of the component profiles)
202 can significantly decrease this indeterminacy and eventually totally eliminate it. A more detailed
203 discussion about the MCR-ALS method can be found in previous works.[30, 31]

204 One of the most important features of MCR-ALS is its potential to extend to the analysis of
205 higher-order data. The extended MCR-ALS bilinear decomposition is as follow:

206 $$\mathbf{X}_{aug} = [\mathbf{X}_1; \mathbf{X}_2; ...; \mathbf{X}_k] = [\mathbf{C}_1; \mathbf{C}_2; ...; \mathbf{C}_k]\mathbf{S}^T + [\mathbf{E}_1; \mathbf{E}_2; ...; \mathbf{E}_k] = \mathbf{C}_{aug}\mathbf{S}^T + \mathbf{E}_{aug} \quad (2)$$

207 where $\mathbf{X}_1$, $\mathbf{X}_2$, ..., $\mathbf{X}_K$ are respectively, the data matrices corresponding to the *k=1,...,K*
208 chromatographic runs having components in common to be analyzed simultaneously collected in
209 column-wise data matrix $\mathbf{X}_{aug}$, $\mathbf{C}_1$, $\mathbf{C}_2$, ..., $\mathbf{C}_k$ in augmented form $\mathbf{C}_{aug}$ are the concentration
210 matrices with the elution profiles of the resolved components in the *k=1,..., K* runs, $\mathbf{S}^T$ is the

matrix of spectral profiles of these eluted components, and $\mathbf{E}_1$, $\mathbf{E}_2$, ..., $\mathbf{E}_k$ in augmented form $\mathbf{E}_{aug}$ are the corresponding error matrices containing the part of the measured data unexplained by the proposed bilinear model.

The results from MCR-ALS when applied to chromatographic data analysis are the pure elution and spectral profiles of the constituents of the analyzed samples. Resolved spectral profiles can be used to identify these components by comparing the resolved spectra with those of authentic standards or standard spectra in available libraries. On the other hand, to obtain quantitative chromatographic information, the areas of the resolved elution profiles can be exploited for quantification, especially in the case of simultaneous analysis of several chromatographic runs. One of the interesting aspects of MCR-ALS model in this work is its potential to model baseline/background contribution instead of its correction before analysis. In addition, elution time shift can be handled by MCR-ALS. In other words, presence of elution time shifts cannot affect the MCR-ALS solutions because of its bilinear model assumption.

2.5 Quantitative analysis

In order to prepare the calibration samples, the original standard of FAMEs diluted to 20, 50, 100, 250, 500 and 1000-fold of the original in methanol. Then, pentadecanoic acid methyl ester as internal standard (with the final concentration of 5.0 mg L^{-1}) was added to each solution and were treated in accordance with the proposed procedure (section 2.2). The obtained total ion chromatogram (TIC) of the standards was divided into sixteen segments (based on emerging chromatographic peaks of standard FAMEs) for simplifying the calculations. The matrices for each analytes in different standards were arranged in a column-wise augmented data matrix. The augmented data matrix was analyzed using MCR-ALS under the application of non-negativity,

234 unimodality and spectral normalization constraints. Also, SIMPLISMA was used to estimate the
235 initial spectral profiles.

236 Quantification was performed based on the summation of the peak areas of the resolved
237 profiles by MCR-ALS. Thus, relative quantitative information for one target compound can be
238 then directly derived from the comparison of MCR-ALS resolved elution profiles for different
239 samples under the assumption of linear relation between relative peak areas of the resolved
240 elution profiles and their relative concentrations. It should be pointed out that overall volume
241 integration (OVI) was considered as analyte signal for peak area calculation.[32] The OVI is
242 preferred to total peak area because of all mass spectral intensities are taken into account in the
243 calculation. The calculated OVI for each of FAMEs was divided to the OVI of internal standard
244 (pentadecanoic acid methyl ester) and was used to build calibration curves. Fig. 1 shows the
245 general procedure used in this study.

246 Data analyses were performed on an Intel (R) Core (TM) 2 Duo-based DELL (Vostro)
247 personal computer with 2.50 GHz CPU and 4 GB RAM. All calculations were carried out using
248 MATLAB 7.10.0 (The Mathworks Inc., MA, USA). MCR-ALS toolbox was freely available
249 from the homepage of MCR at http://www.mcrals.info/. MCRC software[33] was used for data
250 preprocessing and local rank analysis. The library searches and spectral matching of the resolved
251 pure components were conducted on the NIST MS database.[34]

252 **Fig. 1 near here**

253 ## 3 Results and discussion

254 ### 3.1 Optimization of the extraction procedure

255 In order to achieve the maximum efficiency of the proposed method, the effective parameters
256 including type of extraction solvent, extraction time, type and volume of preconcentration

solvent, and salt concentration were investigated and optimized. In addition, the total peak areas of the sought FAs were considered as the response for optimization.

The extraction solvent was chosen based on the properties such as ability to extract the sought analytes (FAs), and good miscibility with organic preconcentration solvent and aqueous sample solution (because it plays the role of disperser in DLLME step). Therefore, methanol and ethanol were tested as the potential extraction solvents in accordance with the proposed procedure in section 2.2. The results showed that the maximum total peak area was produced using methanol (Fig. 2a), thus it was used as the extraction solvent in the subsequent experiments.

The extraction time was the next parameter that its influence was investigated in the range of 0.5-4 min. According to the results presented in Fig. 2b, 3 min was selected as the optimum extraction time for the subsequent experiments.

The appropriate preconcentration solvent was selected based on the characteristics such as extraction capability for the sought compounds, immiscibility with water, higher density than water, and to be suitable for gas chromatography analysis. Considering these properties, several solvents consisting of chloroform, chlorobenzene, trichloroethylene and tetrachloroethylene were evaluated for this purpose. As shown in Fig. 2c, among these solvents, the highest total peak area was generated using chloroform as preconcentration solvent. Therefore, it was chosen as the most suitable preconcentration solvent in the proposed procedure. In the next step, the influence of volume of the selected preconcentration solvent (chloroform) was studied in the range of 25-45 μL. Inspection of the results in Fig. 2d (blue line) shows that with increasing the volume of chloroform (extraction solvent), the response decreases due to the reducing the concentration of FAMEs. However, as the highest total peak areas were obtained in the region 25-30 μL, therefore, 28 μL was chosen as optimized preconcentration solvent volume.

Finally, the effect of salt concentration on the preconcentration of the target analytes (FAs) was investigated in the range of 0 to 15% (w/v). Fig. 2d (red line) shows increasing the efficiency with increasing salt concentration up to 5%. However, higher salt concentrations cause decrease in the efficiency. The addition of salt in the first region (0-5%) decreased the solubility of analytes in the aqueous solution and enhanced the extraction efficiency by salting-out effect. At the concentrations above 5%, the increased viscosity of aqueous solution overcame the salting-out effect, led to difficult mass transfer and low extraction efficiency.[35]

Fig. 2 near here

3.2 Resolution and quantification of FAMEs in the standard mixture solution

Under the optimized conditions, the standard FAMEs were extracted from the standard mixture, preconcentrated and then analyzed with GC-MS. The obtained chromatogram (Fig. 3 a) shows that most of the peaks are well separated except for the region marked with "A". In this region, three peaks related to the components 11, 12 and 13 have been overlapped. The GC-MS TIC of pomegranate seed extract depicted in Fig. 3b is very complex consisting of a large number of peaks together with different chromatographic problems such as baseline/background contribution, low S/N signals and peak overlap. The black rectangles in Fig. 3b marked with B (12.7-13.1 min) and C (13.8-14.3 min) highlight two problematic regions in the TIC. These chromatographic regions will then be analyzed to demonstrate the potential of the proposed strategy in this work.

Fig. 3 near here

At first, the GC-MS data of standard mixture samples of eighteen FAMEs were analyzed using the proposed strategy presented in Fig. 1. For this purpose, the second-order data obtained

from chromatographic regions containing the target analytes were exported to MATLAB and then the GC-MS data for different concentrations and replicates of sought analyte were arranged in a column-wise augmented data matrix. The augmented data matrix was then analyzed using MCR-ALS under non-negativity, unimodality, spectral normalization and component correspondence constraints. In the analysis of standard mixture samples, some of the components are overlapped. For example, region A in Fig. 3a which is also shown in Fig. 4a depicts a chromatographic region where some FAMEs are heavily overlapped in both chromatographic and mass spectrometric dimensions. The singular values from singular value decomposition (SVD)[36] were used to find the number of chemical components in this region. As can be expected, three chemical components were recognized (singular values for these three components were 1.53×10^7, 6.85×10^6 and 2.56×10^6, respectively). The SIMPLISMA was then used to obtain an initial estimate of the spectral profiles for these three components to start ALS optimization. In the next step, MCR-ALS was applied under application of non-negativity (to chromatographic and spectral profiles), unimodality (to chromatographic profiles), and normalization (to spectral profiles) constraints. The MCR-ALS resolved elution profiles for this chromatographic region are shown in Fig. 4b. The values of lack of fit (LOF, %) and explained variance (R^2, %) were equal to 18.83% and 96.45%, respectively. The resolved mass spectral profiles for each component (Fig. 4c, 4d & 4e) were matched with the stored mass spectra in NIST MS library database, and oleic acid methyl ester, linoleic acid methyl ester and linolenic acid methyl ester were respectively identified (Fig. 4f, 4g & 4h).

Fig. 4 near here

The same procedure was used for the MCR-ALS resolution of all other FAMEs and their calibration curves were built up. Table 1 shows the retention times, LOF, RMF, calibration

326 equation, linear dynamic range, regression coefficient and relative errors (REs) of the calibration
327 concentrations for 18 FAMEs. Two parameters of LOF and RMF were explored for evaluation
328 of the MCR-ALS results which were in the range of 7.24-24.36% and 708-977, respectively. It is
329 important to note that in this study, MCR-ALS was performed by using raw data without any
330 data preprocessing method, such as baseline correction and denoising. In addition, the R^2 and
331 REs were in the satisfactory ranges of 0.9934-0.9989 and 3.70-7.45%, respectively. All of these
332 results confirm the validity of the proposed strategy.

333 **Table 1 near here**

334 Another interesting aspect of MCR-ALS is its proper performance in the presence of elution
335 time shifts. Elution time shifts is considered as one of the most important problems in
336 chromatography which can be caused due to factors such as column aging, minor changes in
337 mobile phase composition and instrumental drift. The common chemometric way to handle this
338 problem is its correction by the most applicable alignment methods such as dynamic time
339 warping (DTW),[37] correlation optimized warping (COW)[38,39] and multivariate curve resolution-
340 correlation optimized warping (MCR-COW).[40,41] However, these methods are based on selection
341 of a reference chromatogram and the optimization of other parameters which are not trivial to
342 choose in practice. Moreover, in the presence of interferences, alignment of chromatograms
343 becomes much more difficult. Therefore, elution time shifts modeling using MCR-ALS can be a
344 more efficient and better strategy to correct them. As an example, Fig. 5a shows overlaid TICs
345 for a single-component chromatographic region in different calibration samples where the
346 presence of elution time shifts and baseline/background contribution is evident. As it is
347 demonstrated in Fig. 5b, MCR-ALS can properly recover the elution profiles for the target
348 compounds in the presence of large amount of elution time shift. In addition, baseline

contributions in different standard samples are also properly modeled. The resolved mass spectral profile of the sought analyte and baseline are respectively shown in Fig. 5c and 5d. The comparison of the resolved mass spectrum of target analyte with standard one in NIST MS database confirmed the presence of arachidic acid methyl ester with acceptable RMF=961.

Fig. 5 near here

3.3 Resolution and quantification of FAMEs in pomegranate seed

In order to evaluate the performance of the proposed second-order calibration method, chemometric resolution, identification and quantification of FAMEs in pomegranate seed were performed. Extraction and methyl esterification of FAs from pomegranate seed were performed according to the procedure given in section 2.2. The GC-MS analysis of pomegranate seed sample was carried out at the same conditions as the standard sample. As it was shown earlier in Fig. 3b, the GC-MS TIC of the pomegranate seed seems to be very complex with a large number of components and different chromatographic problems. Two highlighted chromatographic regions in Fig. 3b (B (12.7-13.1 min) and C (13.8-14.3 min)) were used to discuss how MCR-ALS can resolve these problematic regions. The region B shows a very complex chromatographic region with one of the calibrated FAMEs and some uncalibrated-unknown interferences that have heavy overlap with the sought analyte. In addition, the peak cluster C illustrates the region of the pomegranate seed TIC that S/N is too low due to baseline drift.

The most challenging region in the GC-MS TIC of pomegranate seed is the region B. To the best of our knowledge, this region belongs to the main component of pomegranate seed i.e. punicic acid ((Z,E,Z) 9,11,13-Octadecatrienoic acid) which constitutes about 70% of total FAs of pomegranate seeds.[8,10] The GC-MS data of the desired chromatographic region was column-wise

372 augmented with the GC-MS data of the standard samples (S1-S6) of the target compound in this
373 region (11-eicosenoic acid methyl ester) (Fig. 6a). At first, the SVD was used for determination
374 of the number of chemical components involved in this region. The obtained singular values
375 confirmed the presence of four chemical components in this region. To make sure about the
376 number of chemical components in this data matrix, local rank analysis method of evolving
377 factor analysis (EFA)[42] was also used. The forward and backward EFA plots (for the sake of
378 brevity the results are not shown) confirmed the presence of four chemical components
379 previously estimated by SVD. SIMPLISMA was used to calculate the initial values of mass
380 spectral profiles to start ALS optimization. Non-negativity (in both chromatographic and mass
381 spectroscopic dimensions), unimodality (in chromatographic dimension), spectral normalization
382 and component correspondence constraints were applied during optimization to reduce the
383 effects of rotational ambiguity. Since the external calibration strategy was used in this work,
384 therefore, contribution of three unknown components in real sample was set to zero in calibration
385 set using component correspondence constraint which significantly reduces the extent of
386 rotational ambiguity. Fig. 6b shows the resolved MCR-ALS elution profiles for standard and real
387 samples. As can be seen, the pure contribution of 11-eicosenoic acid methyl ester in pomegranate
388 seed was successfully resolved by MCR-ALS (with LOF=18.72%) despite of its heavy overlap
389 with other components. The resolved mass spectral profile (Fig. 6c) of the sought analyte was
390 searched within the stored mass spectra of NIST MS library and the presence of 11-eicosenoic
391 acid methyl ester (Fig. 6d) was confirmed with RMF 882. Afterward, the resolved elution profile
392 was used for quantification of 11-eicosenoic acid methyl ester and the concentration (mg kg^{-1})
393 and RSD (%) values of 1019.36 and 4.98 were respectively obtained (Table 2).

394 **Fig. 6 near here**

The last example is devoted to show the potential of MCR-ALS as a second-order calibration method to model the baseline drift and therefore, to resolve the contribution of minor components (low S/N signals) (chromatographic region C in Fig. 3b). Fig. 7a shows the augmented data matrix for target compound in real sample along with corresponding data matrices in calibration samples. As can be seen, the low S/N signals are observed due to baseline drift for both standards (in low concentration levels, S1-S3) and pomegranate seed sample.

Fig. 7 near here

Fig. 7b illustrates the resolved MCR-ALS elution profiles in standard and real samples. The green line in Fig. 7b shows the contribution of baseline in different samples which emphasizes on one of the most important features of MCR-ALS for the modeling of baseline/background contribution instead of its correction before analysis. The GC-MS data (Fig. 7a) and resolved elution profiles (Fig. 7b) for S1 standard are enlarged for indicating the resolution of baseline and analyte elution profile (Fig. 7c and 7d). As can be seen, the analyte signal in Fig. 7c is indistinguishable from the baseline while after applying MCR-ALS its contribution is recognizable (Fig. 7d). The resolved spectral profile of analyte and baseline are shown in Fig. 7e and Fig. 7f, respectively. The resolved spectral profile was matched with NIST library search and the presence of behenic acid methyl ester with RMF equal to 959 was confirmed. Fig. 7g shows the calibration curve of behenic acid methyl ester with R^2 equals to 0.994 which was acceptable.

Table 2 represents the identified FAs in pomegranate seeds using MCR-ALS. The values of LOF, RMF, concentration (mg kg^{-1}) and RSD are also presented in this table. As can be seen, most of the FAs in standard samples were also detected in real sample with RMF higher than 900 which again confirms the reliability of the present method. However, the concentration of

418 caprylic acid, capric acid, lauric acid and tridecanoic acid were below the limit of quantification
419 (LOQ). In addition, myristoleic acid and linolenic acid were not detected in pomegranate seed
420 whereas linoleic acid, oleic acid, palmitic acid, stearic acid and cis-11-eicosenoic acid were
421 respectively the main fatty acids of pomegranate seed with the concentrations 9098.0, 4873.0,
422 3147.0, 1960.0 and 1019.0 mg kg^{-1} and the relative standard deviations (RSD%) between 0.15-
423 11.57.

424 **Table 2 near here**

425 ## 4 Conclusion

426 In the present study, the VAE-DLLME-GC-MS technique was used for determination of the
427 FAMEs in a standard mixture solution and in pomegranate seed sample in less than 20 min.
428 Then, the GC-MS data were column-wise augmented with retention times as rows and m/z
429 values as the columns of this data matrix. And finally, MCR-ALS was performed under
430 application of proper constraints to obtain pure elution and mass spectral profiles along with
431 calibration curves of the target FAMEs. Problems associated to the GC-MS analysis of FAMEs,
432 such as elution time shifts, baseline/background contribution, and peak overlaps were
433 successfully solved using MCR-ALS bilinear modeling of column-wise augmented data matrix.
434 Furthermore, proper qualitative (RMF higher than 800) and quantitative (RE below 7.5% and
435 RSD below 11.6%) results were obtained for both standard mixture and pomegranate seed
436 samples. Therefore, it can be concluded that the proposed combined strategy (VAE-DLLME-
437 GC-MS coupled with MCR-ALS) is applicable for complex natural products containing different
438 unknown interferences.
439
440

Notes and references

1 M. Çelik, C. Türeli, M. Çelik, Y. Yanar, Ü. Erdem, and A. Küçükgülmez, *Food Chem.*, 2004, **88**, 271.

2 M. Vosough and A. Salemi, *Talanta*, 2007, **73**, 30.

3 T. Horák, J. Culík, P. Cejka, M. Jurková, V. Kellner, J. Dvorák, and D. Hasková, *J. Agric. Food Chem.*, 2009, **57**, 11081.

4 L.-K. Ng, *Anal. Chim. Acta*, 2002, **465**, 309.

5 B. Plutowska and W. Wardencki, *Anal. Chim. Acta*, 2008, **613**, 64.

6 I. Marekov, S. Momchilova, B. Grung, and B. Nikolova-damyanova, *J. Chromatogr. B*, 2012, **910**, 54.

7 E. P. Lansky and R. A. Newman, *J. Ethnopharmacol.*, 2007, **109**, 177.

8 P. Jing, T. Ye, H. Shi, Y. Sheng, M. Slavin, B. Gao, L. Liu and L. (Lucy) Yu, *Food Chem.*, 2012, **132**, 1457.

9 A. Fadavi, M. Barzegar and M. H. Azizi, *J. Food Compos. Anal.*, 2006, **19**, 676.

10 H. Abbasi, K. Rezaei and L. Rashidi, *J. Am. Oil. Chem. Soc.*, 2008, **85**, 83.

11 Y. Tian. Z. Xu, B. Zheng and Y. M. Lo, *Ultrason. Sonochem.*, 2013, **20**, 202.

12 G. Wei and E. Y. Zeng, *Trends Anal. Chem.*, 2011, **30**, 1429.

13 A. Saba, F. Mazzini, A. Raffaelli, A. Mattei and P. Salvadori, *J. Agric. Food Chem.*, 2005, **53**, 4867.

14 M. Rezaee, Y. Assadi, M. R. Milani Hosseini, E. Aghaee, F. Ahmadi and S. Berijani, *J. Chromatogr. A*, 2006, **1116**, 1.

15 M. Rezaee, Y. Yamini and M. Faraji, *J. Chromatogr. A*, 2010, **1217**, 2342.

16 P. Viñas, N. Campillo, I. López-García and M. Hernández-Córdoba, *Anal. Bioanal. Chem.*, 2013, **406**, 2067.

17 S. Dadfarnia and A. M. H. Shabani, *Anal. Chim. Acta*, 2010, **658**, 107.

18 A. V. Herrera-Herrera, M. Asensio-Ramos, J. Hernández-Borges and M.Á. Rodríguez-Delgado, *Trends Anal. Chem.*, 2010, **29**, 728.

19 T. Seppänen-Laakso, I. Laakso and R. Hiltunen, *Anal. Chim. Acta*, 2002, **465**, 39.

20 K. S. Booksh and B. R. Kowalski, *Anal. Chem.*, 1994, **66,** 782A.

21 V. Gómez and M. P. Callao, *Anal. Chim. Acta*, 2008, **627**, 169.

22 A. C. Olivieri, G. M. Escandar and A. Peña, *Trend Anal. Chem.*, 2011, **30**, 607.

23 A. C. Olivieri, *Anal. Methods*, 2012, **4**, 1876.

24 R. Bro, C. A. Andersson and H. A. L. Kiers, *J. Chemom.*, 1999, **13**, 295.

25 S. A. Bortolato, A. C. Olivieri, *Anal. Chim. Acta.*, 2014, **842,** 11.

26 E. Pere´-Trepat, S. Lacorte and R. Tauler, *Anal. Chim. Acta*, 2007, **595**, 228.

27 M. Vosough and H. M. Esfahani, *Talanta*, 2013, **113**, 68.

28 M. Bravo. M, L. F. Aguilar, W. Quiroz. V, A. C. Olivieri and G. M. Escandar, *Microchem. J.*, 2013, **106**, 95.

29 W. Windig and J. Guilment, *Anal. Chem.*, 1991, **63**, 1425.

30 M. Jalali-Heravi and H. Parastar, *Talanta*, 2011, **85**, 835.

31 H. Parastar and R. Tauler, *Anal. Chem.*, 2014, **86**, 286.

32 F. Gong, Y. Liang, H. Cui, F. T. Chau and B. T. P. Chan, *J. Chromatogr. A*, 2001, **909**, 237.

33 M. Jalali-heravi, H. Parastar, M. Kamalzadeh, R. Tauler and J. Jaumot, *Chemom. Intell. Lab. Syst.*, 2010, **104**, 155.

34 NIST/EPA/NIH MS Library, v. 2.0d, Gaithersburg, MD, USA, 2005.

35 D. Ge and H. K. Lee, *J. Chromatogr. A*, 2012, **1263**, 1.

36 D. L. Massart, B. G. M. Vandeginste, L. M. C. Buydens, S. De Jong, P. J. Lewi and J. Smeyers-Verbeke, *Handbook of chemometrics and qualimetrics: Part A*, Elsevier, Amsterdam, 1997.

37 A. Kassidas, J. F. MacGregor and P. A. Taylor, *AIChE J.*, 1998, **44**, 864.

38 N. P. V. Nielsen, J. M. Carstensen and J. Smedsgaard, *J. Chromatogr. A*, 1998, **805**, 17.

39 G. Tomasi, F. van den Berg and C. Andersson, *J. Chemom.*, 2004, **18**, 231.

40 C. Tistaert and Y. Vander Heyden, *Anal. Chem.*, 2012, **84,** 5653.

41 H. Parastar and N. Akvan, *Anal. Chim. Acta*, 2014, **816**, 18.

42 M. Maeder and A.D. Zuberbuehler, *Anal. Chim. Acta*, 1986, **181,** 287.

Figure captions

Fig. 1: Flow chart explaining the different steps of the proposed strategy in this study.

Fig. 2: Optimization of VAE-DLLME parameters. Effect of (a) extraction solvent, (b) extraction time, (c) preconcentration solvent, (d) volume of preconcentration solvent (blue line) and salt concentration (red line) on the extraction efficiency of FAs. The error bars represent the standard deviation of the measurements (n=3).

Fig. 3: GC-MS TIC of FAMEs in (a) standard mixture and (b) pomegranate seed samples. Numbers are related to the components in Table 1. Regions A, B and C illustrate three investigated problematic regions due to coeluted components, uncalibrated interferences and low S/N ratio, respectively.

Fig. 4: (a) TIC of region A and (b) MCR-ALS resolved elution profiles. Left peak: oleic acid methyl ester, middle peak: linoleic acid methyl ester and right peak: linolenic acid methyl ester. (c) MCR-ALS resolved mass spectra of oleic acid methyl ester, (d) linoleic acid methyl ester, (e) linolenic acid methyl ester, and (f), (g), & (h) standard mass spectra of them, respectively.

Fig. 5: (a) Elution time shift occurred in different chromatographic runs at different standard samples, (b) resolved elution profile and baseline, (c) resolved mass spectral profile of target analyte (arachidic acid methyl ester), and (d) spectral background.

Fig. 6: (a) Augmented data matrix for region B in pomegranate seed TIC along with standard calibration samples. S1-S6 show the calibration samples (different concentration levels of standards). (b) Resolved elution profiles, (c) Resolved mass spectral profile for target compound, and (d) standard mass spectrum from NIST library.

Fig. 7: (a) Augmented data matrix for region C in pomegranate seed TIC and standard calibration samples (S1-S6 show the different calibration samples). (b) Resolved elution profiles of baseline (green line) and behenic acid methyl ester (blue peaks). (c) Enlarged S1 section of augmented data matrix and (d) resolved elution profiles. (e) Resolved mass spectral profiles of behenic acid methyl ester and (f) background. (g) Calibration curve of behenic acid methyl ester.

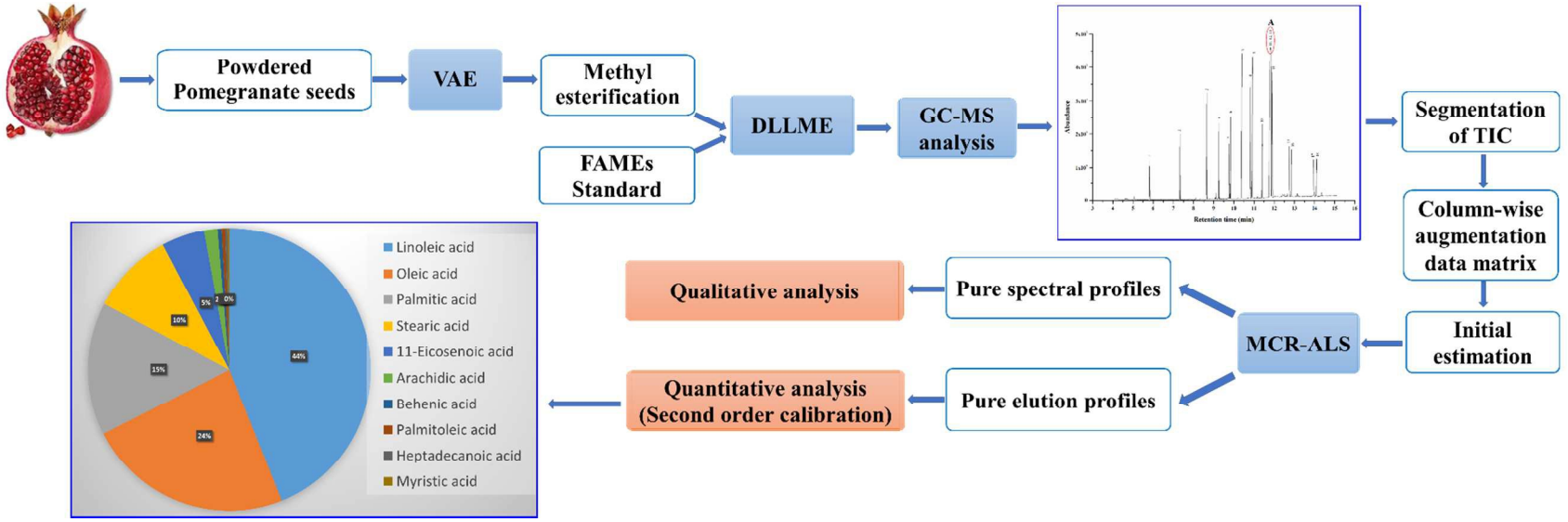


Fig. 1

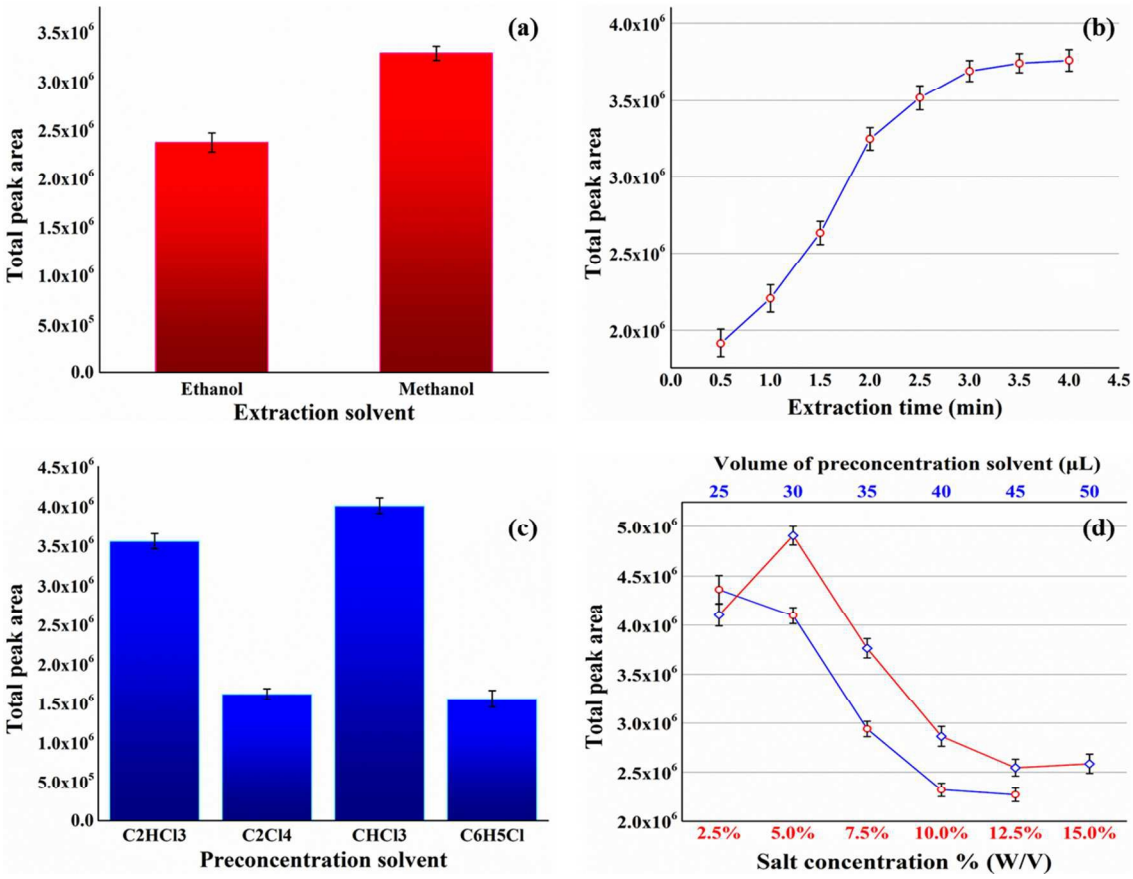


Fig. 2

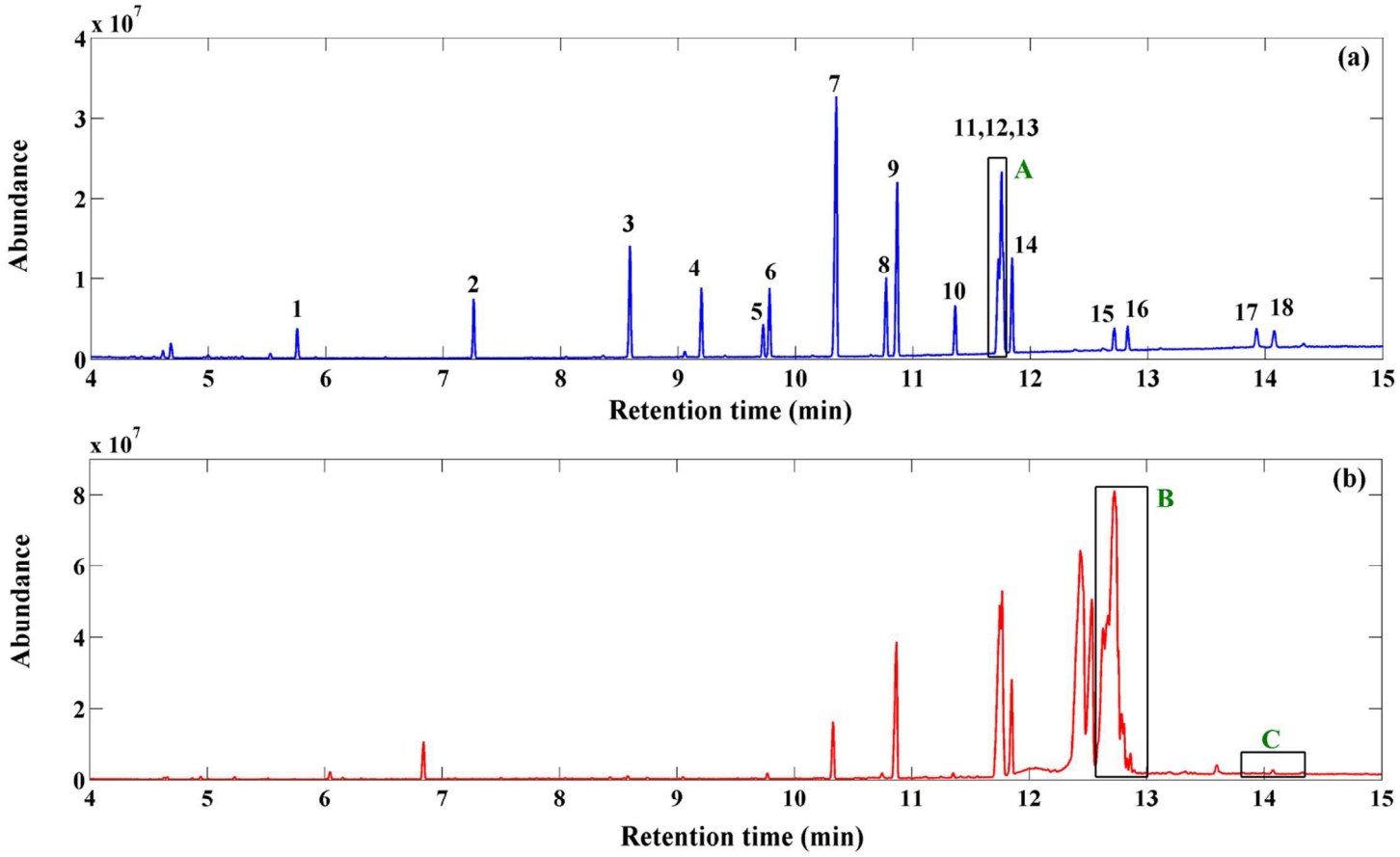


Fig. 3

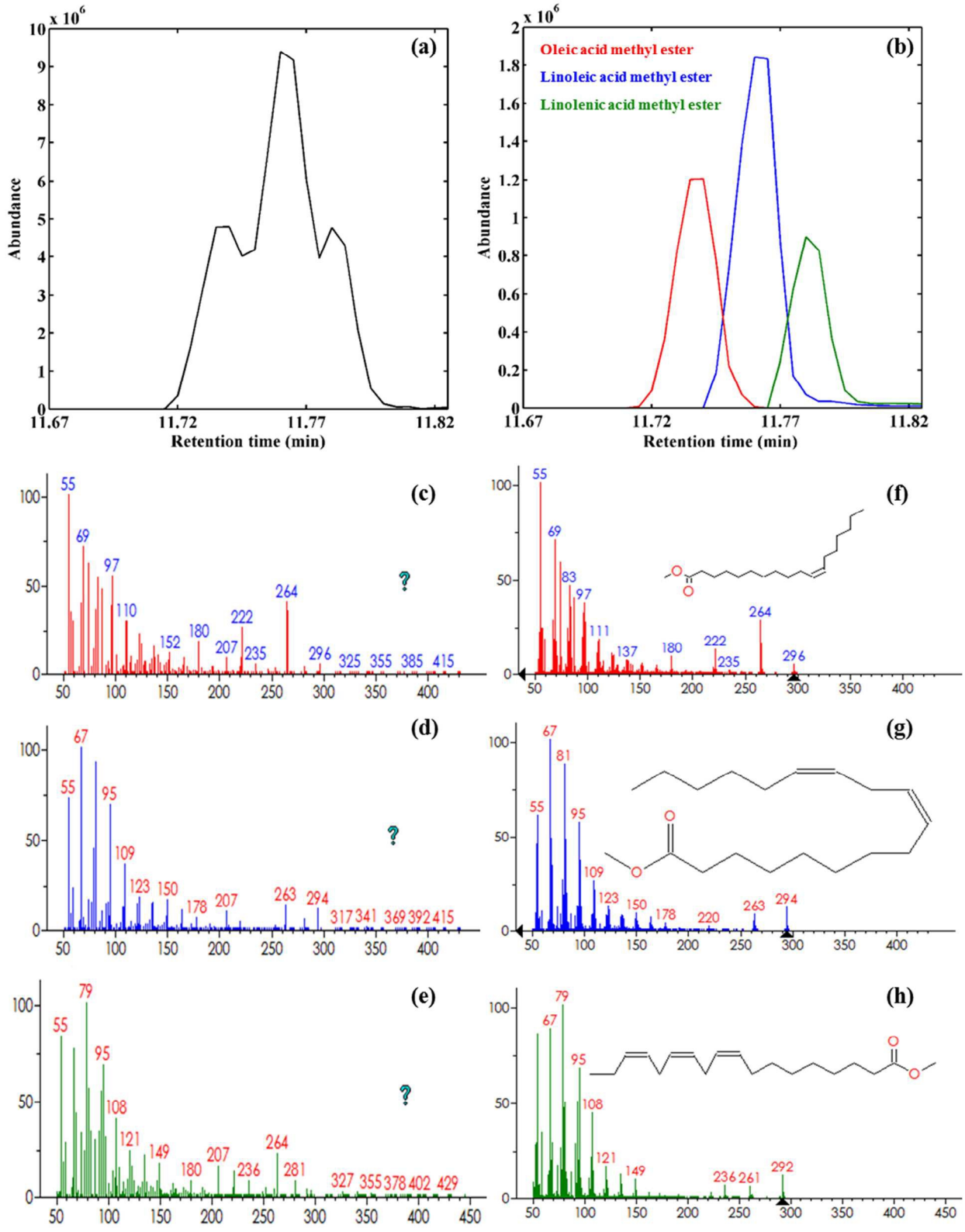


Fig. 4

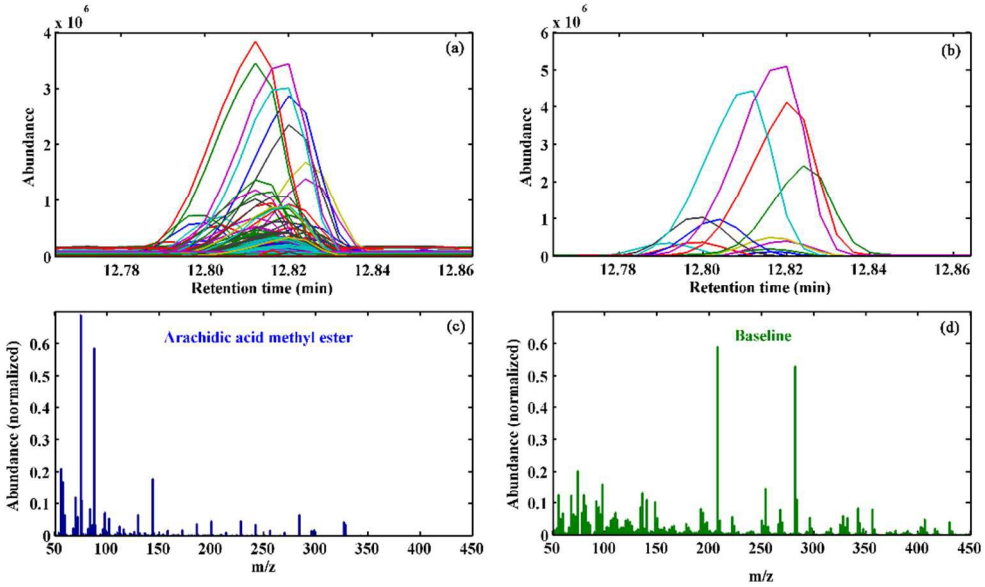


Fig. 5

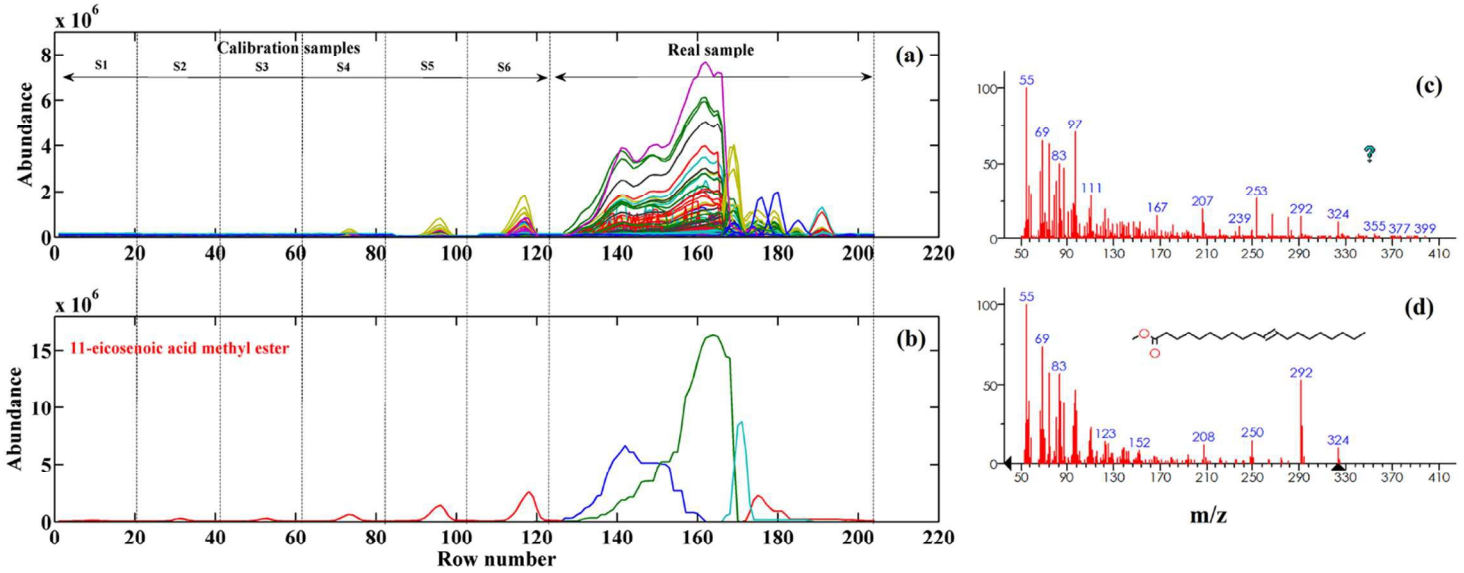


Fig. 6

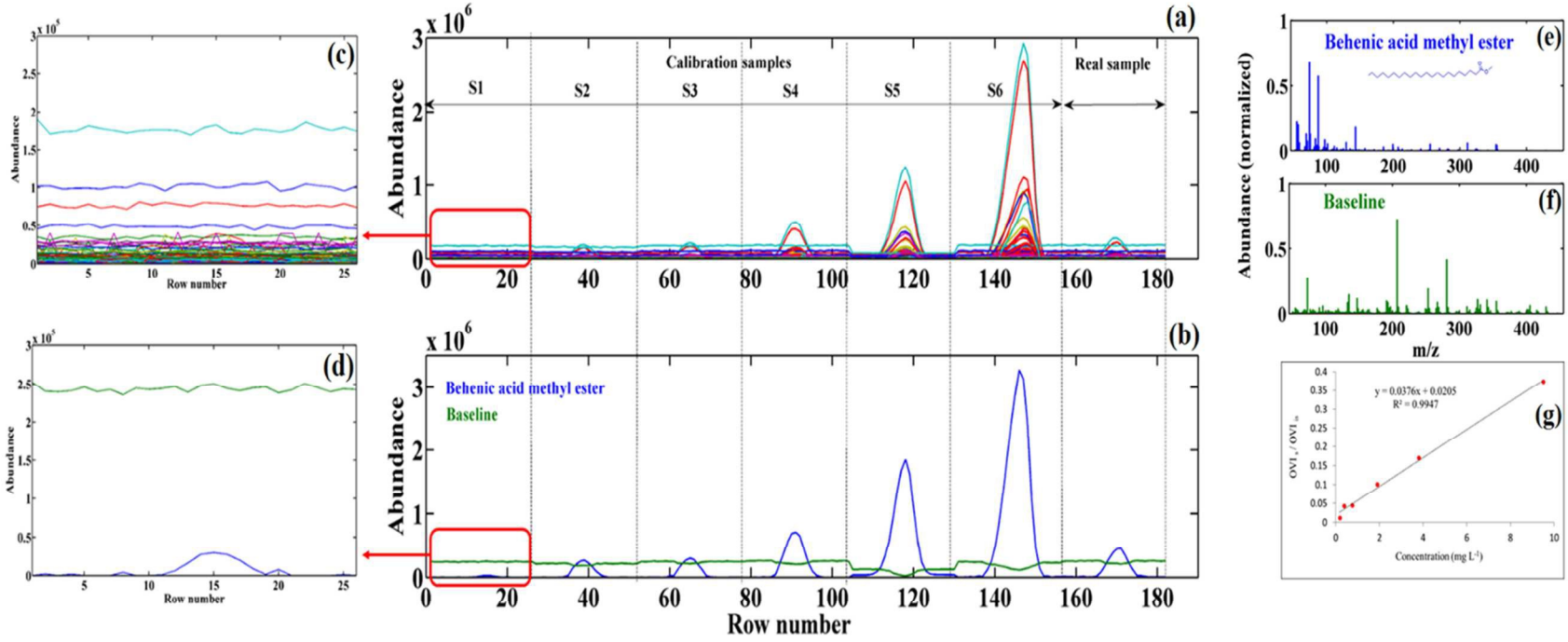


Fig. 7

Table 1. Characterization of FAMEs standard mixture in MCR-ALS and calibration steps.

No.	t_R[a]	Chemical name	Formula	LOF (%)[b]	RMF[c]	Calibration equation	Calibration range (mg L^{-1})	R^2 (calibration)	RE (%)[d]
1	5.75	Caprylic acid methyl ester	$C_9H_{18}O_2$	7.25	969	y=0.0285x+0.0139	0.19-9.5	0.9970	4.22
2	7.35	Capric acid methyl ester	$C_{11}H_{22}O_2$	8.87	971	y=0.0336x+0.0239	0.32-16	0.9971	4.14
3	8.65	Lauric acid methyl ester	$C_{13}H_{26}O_2$	11.35	975	y=0.033x+0.0577	0.64-32	0.9976	3.76
4	9.30	Tridecanoic acid methyl ester	$C_{14}H_{28}O_2$	8.99	964	y=0.0372x+0.0358	0.32-16	0.9977	3.71
5	9.77	Myristoleic acid methyl ester	$C_{15}H_{28}O_2$	22.69	931	y=0.0401x+0.0242	0.19-9.5	0.9955	5.10
6	9.84	Myristic acid methyl ester	$C_{15}H_{30}O_2$	10.93	955	y=0.0397x+0.0339	0.32-16	0.9967	4.43
7	10.51	Pentadecanoic acid methyl ester	$C_{16}H_{32}O_2$	11.36	932	Internal standard	-	-	-
8	10.80	Palmitoleic acid methyl ester	$C_{17}H_{32}O_2$	14.39	977	y=0.0263x+0.0795	0.64-32	0.9906	7.46
9	10.90	Palmitic acid methyl ester	$C_{17}H_{34}O_2$	8.43	959	y=0.0261x+0.1301	1.3-65	0.9959	4.97
10	11.37	Heptadecanoic acid methyl ester	$C_{18}H_{36}O_2$	14.25	948	y=0.0345x+0.0331	0.32-16	0.9934	6.22
11	11.75	Oleic acid methyl ester	$C_{19}H_{36}O_2$	18.83	708	y=0.0269x+0.0637	1.3-65	0.9986	2.21
12	11.75	Linoleic acid methyl ester	$C_{19}H_{34}O_2$	18.83	708	y=0.0142x+0.0363	2.22-111	0.9989	1.96
13	11.75	Linolenic acid methyl ester	$C_{19}H_{32}O_2$	18.83	718	y=0.0352x+0.0442	0.64-32	0.9989	1.98
14	11.87	Stearic acid methyl ester	$C_{19}H_{38}O_2$	18.13	968	y=0.0264x+0.0798	0.65-32.5	0.9953	5.29
15	12.73	11-Eicosenoic acid methyl ester	$C_{21}H_{40}O_2$	24.36	944	y=0.0375x+0.011	0.19-9.5	0.9964	4.61
16	12.84	Arachidic acid methyl ester	$C_{21}H_{42}O_2$	17.02	961	y=0.0375x+0.018	0.19-9.5	0.9964	4.57
17	13.94	Erucic acid methyl ester	$C_{23}H_{44}O_2$	24.13	945	y=0.0395x+0.0101	0.19-9.5	0.9956	5.07
18	14.10	Behenic acid methyl ester	$C_{23}H_{46}O_2$	17.70	959	y=0.0376x+0.0205	0.19-9.5	0.9947	5.58

[a] Retention time (min). [b] $Lack\ of\ fit\ (\%) = \sqrt{\sum_{ij} e_{ij}^2 / \sum_{ij} x_{ij}^2} \times 100$, where x_{ij} and e_{ij} are the elements of the matrices X and E, respectively. [c] Reverse match factor, the criterion of similarity between resolved and library search mass spectra. [d] $Relative\ error\ (\%) = \left(\sqrt{\sum_i (c_i - \hat{c}_i)^2} / \sqrt{\sum_i c_i^2}\right) \times 100$, where c_i is the known concentration of standard i and $\hat{c}_i$ is its calculated value using the calibration equation obtained from the overall volume integration (OVI) in TIC mode.

Table 2. Characterization of fatty acids in pomegranate seeds.

No.	Component	LOF (%)	RMF	FAs amount[a]	RSD%[b]
1	Caprylic acid	7.28	969	trace[c]	0.22
2	Capric acid	8.88	971	trace	5.11
3	Lauric acid	11.39	956	trace	1.84
4	Tridecanoic acid	9.00	964	tarce	0.36
5	Myristoleic acid	-	-	n.d.[d]	-
6	Myristic acid	10.95	955	47.81	0.15
7	Pentadecanoic acid	11.24	943	IS[e]	-
8	Palmitoleic acid	24.61	969	84.20	11.57
9	Palmitic acid	12.22	957	3147.02	2.07
10	Heptadecanoic acid	14.35	948	51.92	3.33
11	Oleic acid	6.25	715	4873.71	5.78
12	Linoleic acid	6.25	717	9098.23	6.16
13	Linolenic acid	-	-	n.d.	-
14	Stearic acid	18.46	957	1960.16	1.97
15	11-Eicosenoic acid	18.72	882	1019.36	4.98
16	Arachidic acid	21.62	961	319.80	6.09
17	Erucic acid	23.98	946	16.60	11.09
18	Behenic acid	17.77	959	93.81	0.61

[a] Amount of fatty acids (mg kg^{-1}) in pomegranate seeds. [b] Relative standard deviation based on OVIs. $RSD = \frac{s}{\overline{x}} \times 100$, where s is the standard deviation and $\overline{x}$ is the mean calculated from the overall volume integration (OVI) in TIC mode. [c] Below the limit of quantification (LOQ) [d] Not detected. [e] Internal standard.