Electronic Supplementary Information

Multivariate Curve Resolution of GC-MS measurements to thorough analysis of the volatile chemical constituents in Citrus aurantium L. peel

F. Azimi, M. H. Fatemi*
Department of Analytical Chemistry, Faculty of Science, University of Mazandaran, Babolsar,Iran.

1. Continued solving equation 1; (section 3.1)

MCR-ALS, solves (Eq.(1)) for C and S^T, based on two linear least-squares steps derived from Eq. (1), depicted by Eqs. (2) and (3) find matrix S^T and C, alternatively using minimization of them in constant C or S^T respectively:

$$\min_{S^T} \|X_{PCA} - CS^T\| \text{ , C constant} \quad \text{(S2)}$$

$$\min_{C} \|X_{PCA} - CS^T\| \text{ , } S^T \text{ constant} \quad \text{(S3)}$$

X_{PCA} is estimated matrix by principal components analysis (PCA) solutions, for more details about PCA and other factor analysis methods see.[1] Modus operandi to obtain solutions of Eqs. (2) and (3) by least squares are that:

1 .Suppose we have at first, an initial estimate of pure spectra profile (S^T),

2. Therefore we can easily calculate the corresponding concentration profiles of the experimental data (X_{PCA}) and S^T during the unconstrained least square calculation from the expression:

$$C\hat{} = (X_{PCA} \cdot S)(S^T \cdot S)^{-1} = X_{PCA}(S^T)^+ \quad \text{(S4)}$$

3. Using the estimated C ($C\hat{}$), we can estimate the corresponding spectra:

$$(S^T)\hat{} = (C\hat{}^T \cdot C\hat{})^{-1} \cdot (C^T \cdot X_{PCA}) = (C\hat{})^+ X_{PCA} \quad \text{(S5)}$$

Where $(S^T)^+$ and C^+ are the pseudoinverses of S^T and C matrices, respectively, when S^T and C are of full rank.[2-4] With the implementation two separate linear regression steps (eqs. 4,5) in an alternating least squares cycle, new matrices of $C\hat{}$ and $(S^T)\hat{}$ are then achieved in each iteration.

4. Reproduction of (X) from constrained solutions $C\hat{}$ $(S^T)\hat{}$. If reproduction is acceptable, end of the process. If not, repeat steps 1, 2, 3 until one obtained optimal solution fulfills the constraints and the postulated convergence criteria.[5]

2. Definition of Lack of fit (LOF) and R^2 in MCR-ALS Model (section 3.1):

Lack of fit (LOF), is defined as the difference between the original data X and the data reproduced by MCR-ALS:

$$LOF = \sqrt{\frac{\sum_{1}^{i=1}\sum_{j}^{j=1}\left(x_{i,j} - x\hat{d}_{i,j}\right)^2}{\sum_{1}^{i=1}\sum_{j}^{j=1}(x_{i,j})^2}} \times 100 \quad \text{(S7)}$$

$$R^2 = \left(1 - \frac{\sum_{1}^{i=1}\sum_{j}^{j=1}(x_{i,j} - \hat{x_{i,j}})^2}{\sum_{1}^{i=1}\sum_{j}^{j=1} x_{i,j}^2}\right) \times 100 \quad \text{(S8)}$$

Where $x_{i,j}$ and $x_{i,j}\hat{}$ represents an element of the input data matrix X in the experimental and reproduced matrices using MCR-ALS or PCA, respectively. These quantities allow a simple comparison between different methods and models in the description of the same data set. [6, 7]

3. **Figures:**
   (section 3. 2)

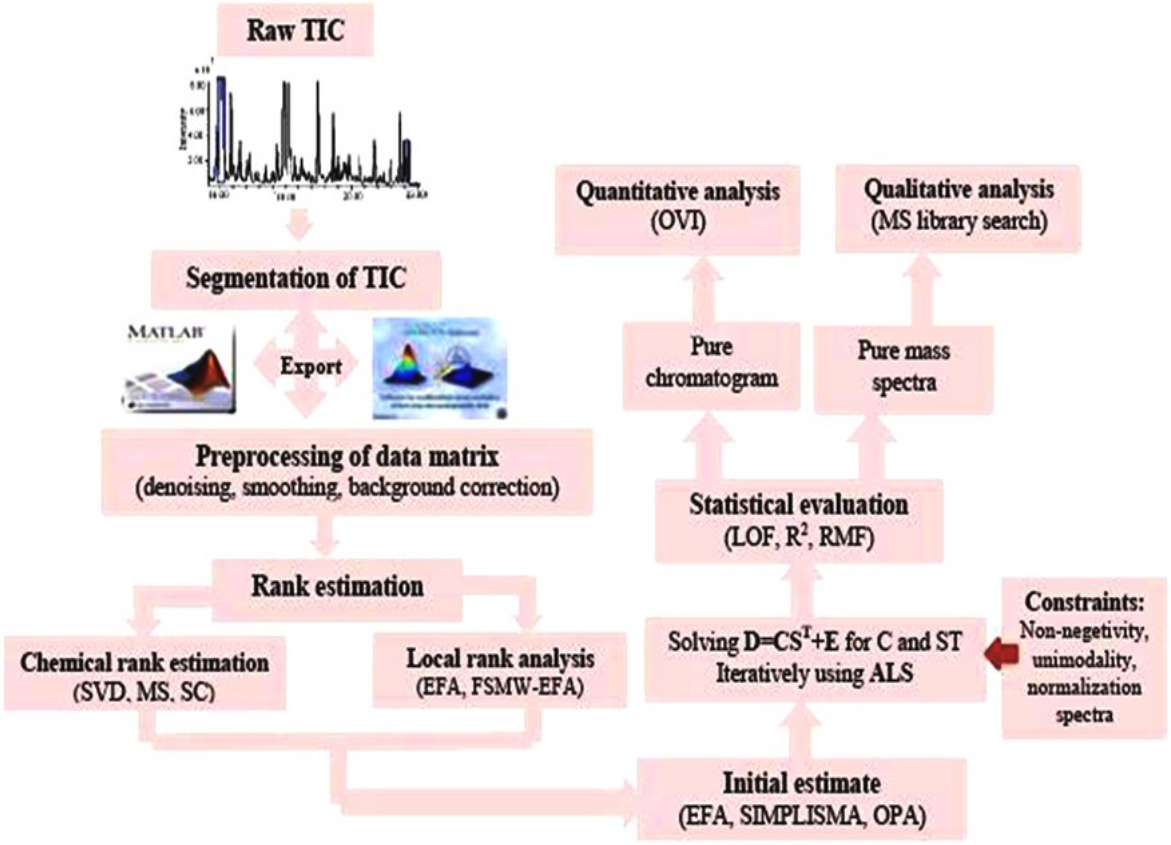

Fig. S1. The flow chart given is a representation of the entire strategy to resolve and quantify the two-way data achieved from GC-MS fingerprint.

(section 4. 1)

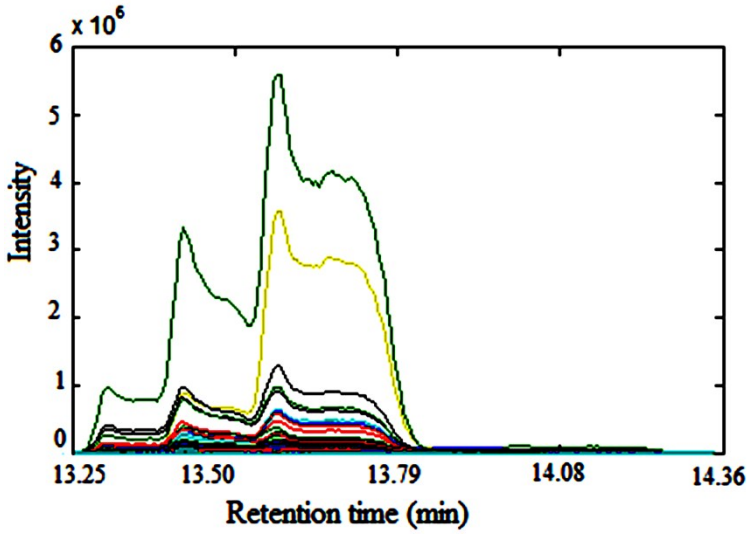

Fig. S2. The exported data matrix of peak clusters A

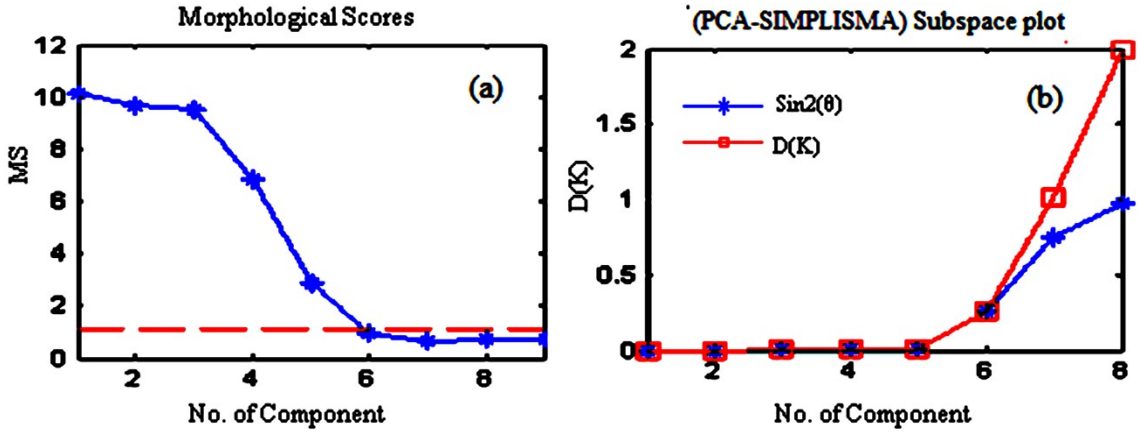

Fig. S3. Morphological score plot for in chromatographic segment A by applying OPA method (a). Subspace plot for this region (b). Comparison between key selected factors of methods PCA and SIMPLISMA.

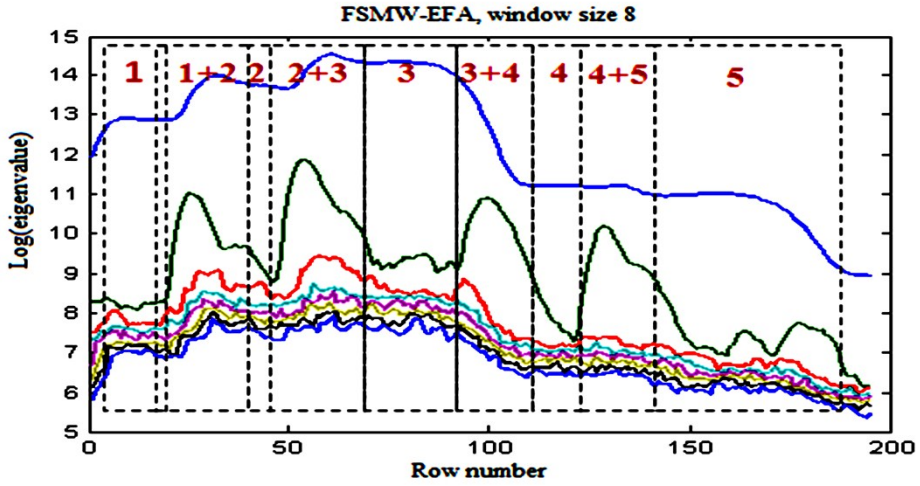

Fig. S4. FSMWEFA plot with a window size of 8 for analyzing peak cluster A.

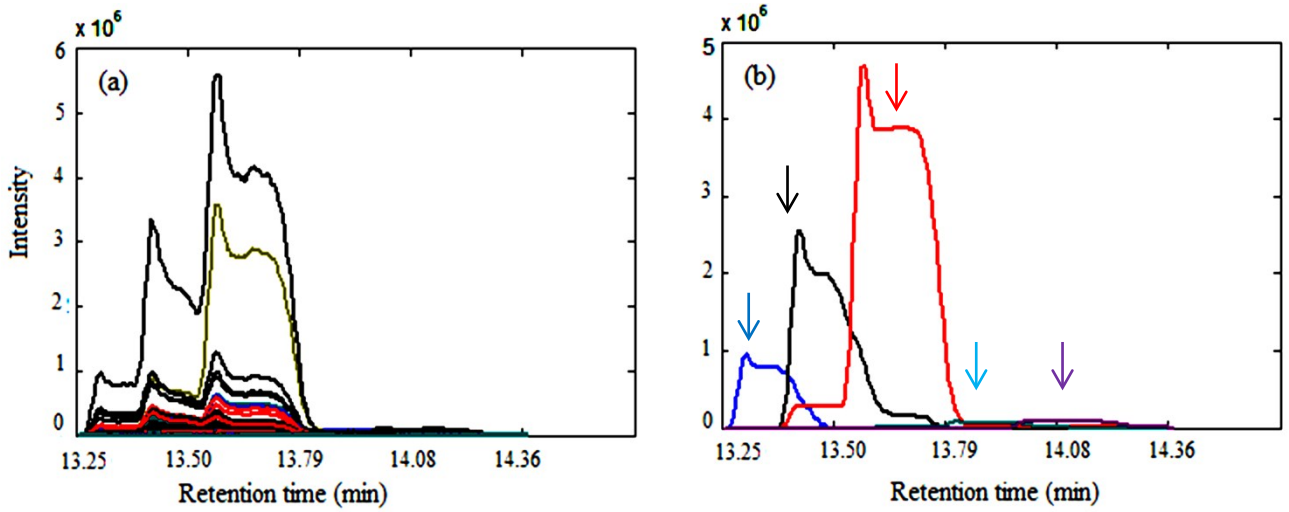

Fig. S5. Two dimension plot of peak cluster A (a) ; Resolved MCR-ALS chromatographic profiles (b) (1–5).

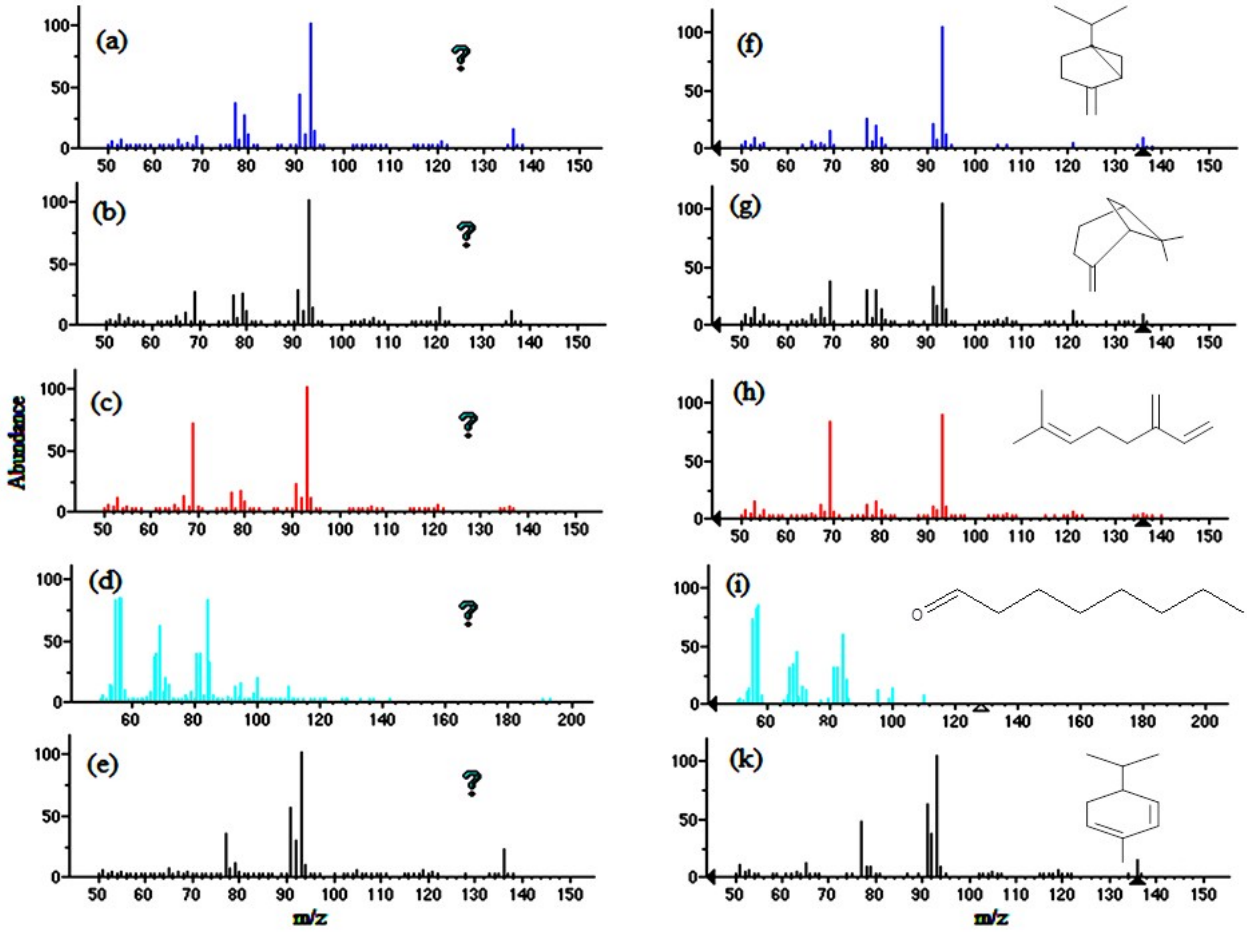

Fig. S6. Resolved mass spectra (standard mass spectra) for peak cluster A. Resolved mass spectra and their corresponding standard mass spectra for peak cluster C. Resolved (a) and standard (f) mass spectra of Sabinen; resolved (b) and standard (g) mass spectra of β-Pinene; resolved (c) and standard (h) mass spectra of β-Myrcene; resolved (d) and standard (i) of Octanal; resolved (e) and standard (k) mass spectra of α-Phellandrene.

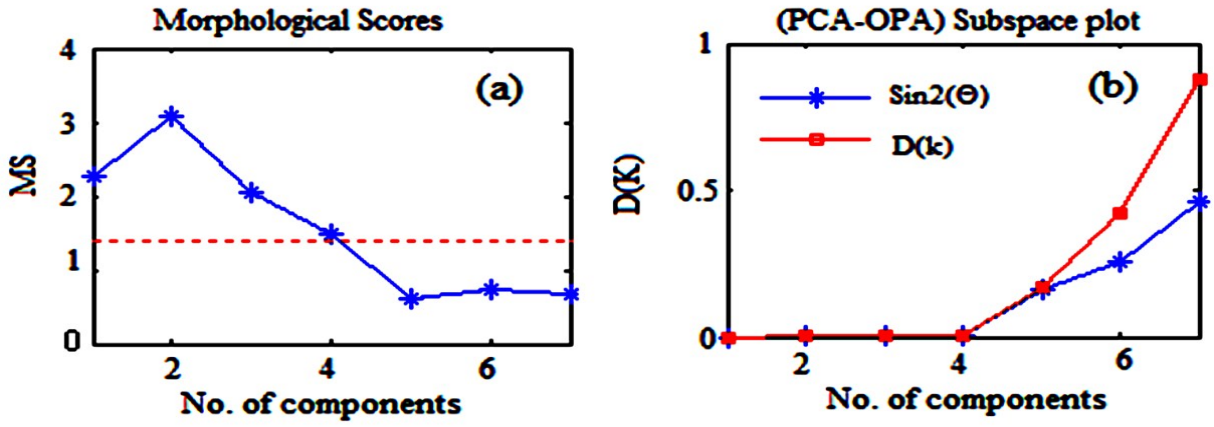

Fig. S7. Morphological score plot for in chromatographic segment C by applying OPA method (a). Subspace plot for this region (b). Comparison between key selected factors of methods PCA and OPA.

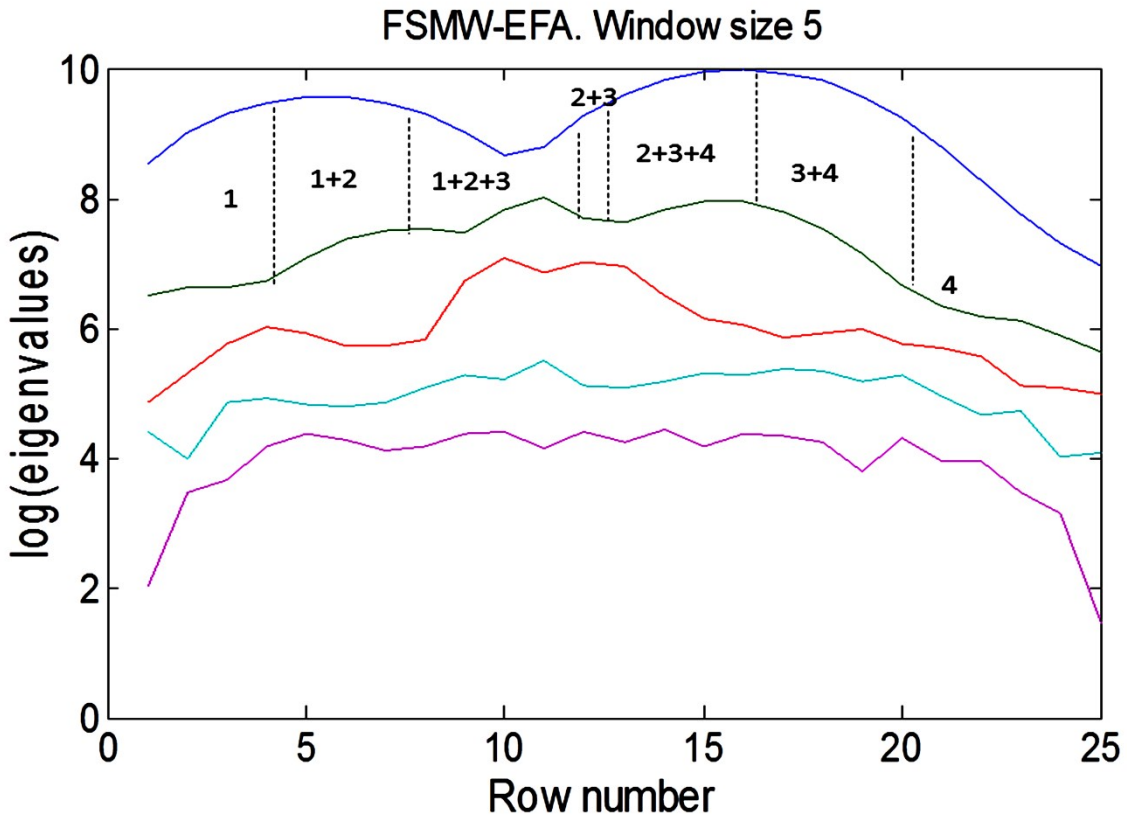

Fig. S8. FSMWEFA plot with a window size of 5 for analyzing peak cluster C.

.

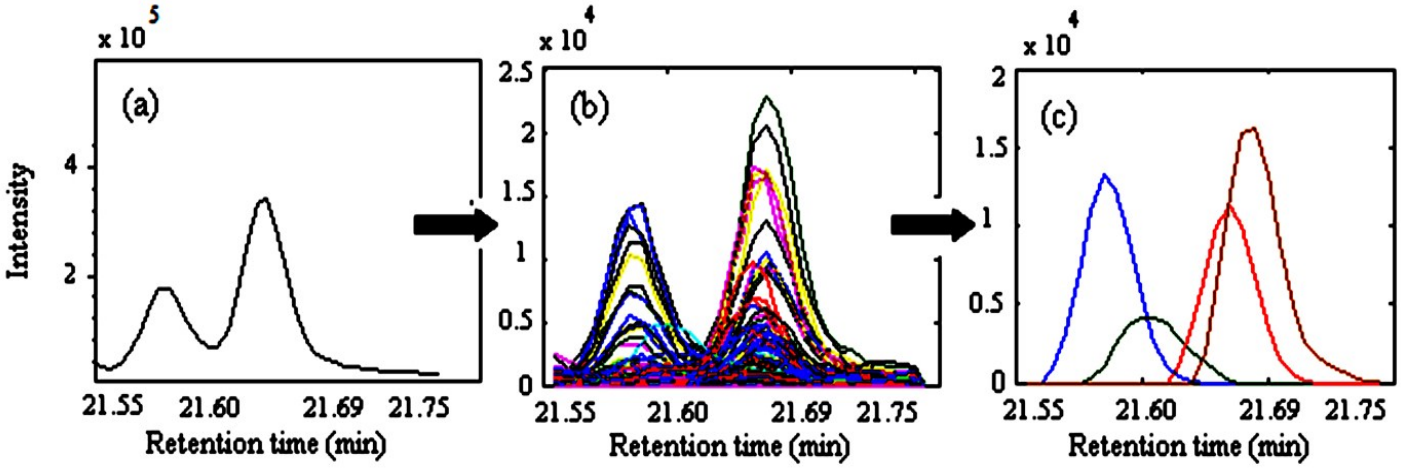

Fig. S9. Original Peak cluster C and pure chromatographic peaks after resolution by chemometric method MCR-ALS: (a) the TIC of peak cluster B from C. aurantium L. peel extract; (b) the corresponding two dimention plot; (c): Resolved MCR-ALS chromatographic profiles (1–4).

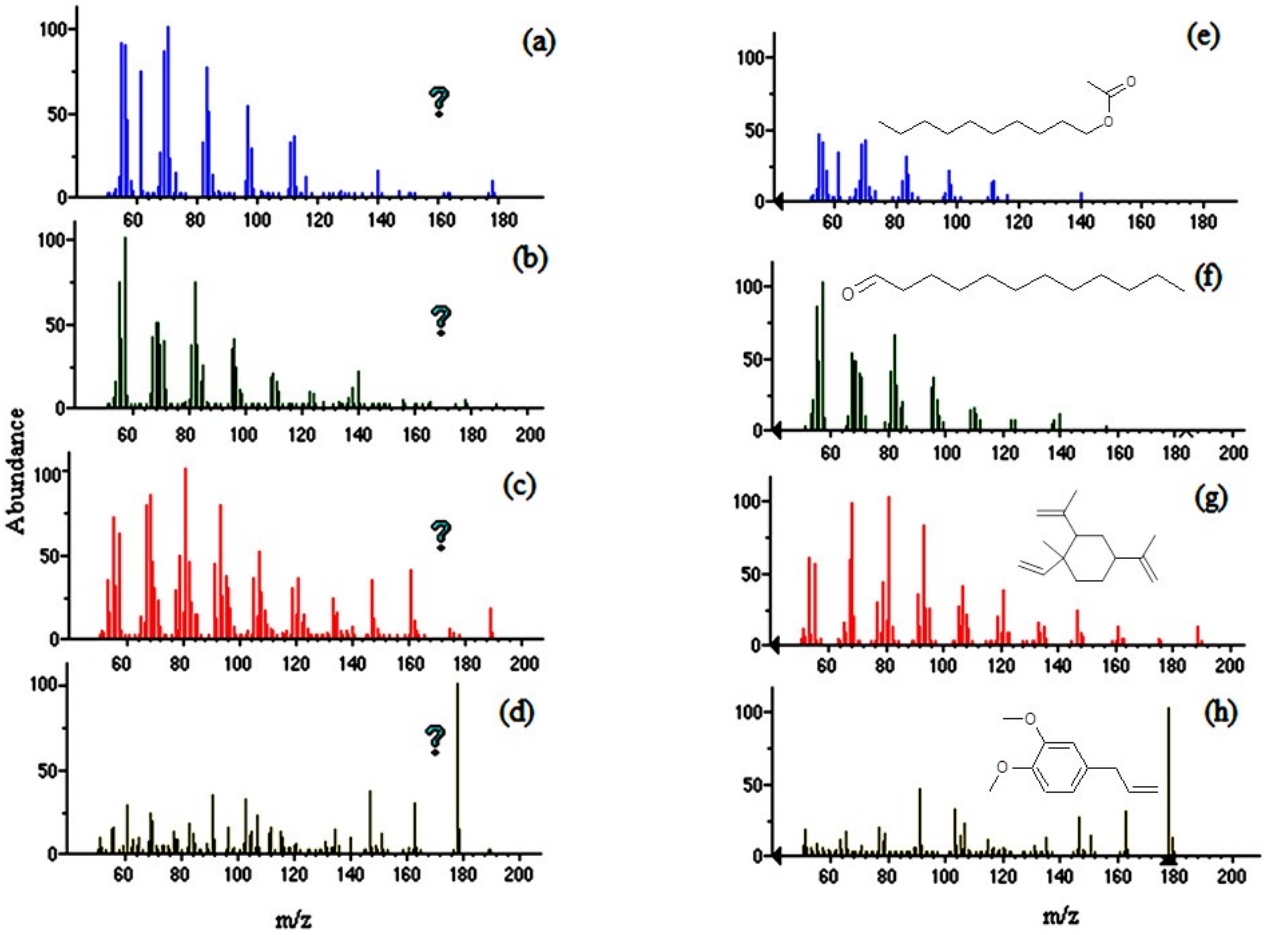

Fig. S10. Resolved mass spectra and their corresponding standard mass spectra for peak cluster C. Resolved (a) and standard (e) mass spectra of Acetic acid, decyl ester; resolved (b) and standard (f) mass spectra of Dodecanal; resolved (c) and standard (g) mass spectra of Cyclohexane, 1-ethenyl-1-methyl-2,4-bis(1-methylethenyl)-; resolved (d) and standard (h) of Benzene, 1,2-dimethoxy-4-(2-propenyl)-.

Section (4.2)

Table S1. Relative peak area related to volatile constituents of C peak cluster before and after applying chemometric methods

Constituents	Acetic acid, decyl ester (%)	Cyclohexane, 1-ethenyl-1-methyl-2,4-bis(1-methylethenyl)- (%)	Benzene, 1,2-dimethoxy-4-(2-propenyl)- (%)	Dodecanal (%)
Vertical cutting	0.022	-	-	0.046
Overall volume integration	0.017	0.011	0.010	0.024

References

1. S. Pack, *Journal of Chemometrics*, 1991, **5**, 545-545.
2. T. Azzouz and R. Tauler, *Talanta*, 2008, **74**, 1201-1210.
3. R. Tauler, *Chemometrics and Intelligent Laboratory Systems*, 1995, **30**, 133-146.

4. C. Tistaert, H. P. Bailey, R. C. Allen, Y. V. Heyden and S. C. Rutan, *Journal of Chemometrics*, 2012, **26**, 474-486.
5. A. de Juan, J. Jaumot and R. Tauler, *Analytical Methods*, 2014, **6**, 4964-4976.
6. M. Antunes, J. Simao, A. Duarte, M. Esteban and R. Tauler, *Analytica Chimica Acta*, 2002, **459**, 291-304.
7. J. Jaumot, R. Gargallo, A. de Juan and R. Tauler, *Chemometrics and Intelligent Laboratory Systems*, 2005, **76**, 101-110.