



A multiway approach to overcome the peak resolution challenges in gas-chromatography coupled to ion mobility spectrometry

Samuele Pellacani^a, Lorenzo Strani^a, Daniele Tanzilli^{a,b,c}, Rebecca Barbieri^a, Raniero Marabottini^d, Francesca Martone^d, Angela Napolitano^d, Marina Cocchi^a, Caterina Durante^{a,*}

^a Department of Chemical and Geological Sciences, University of Modena and Reggio Emilia, via Campi 103, Modena, 41125, Italy

^b ITAP, Univ. Montpellier, INRAE, Institut Agro, Montpellier, 34196, France

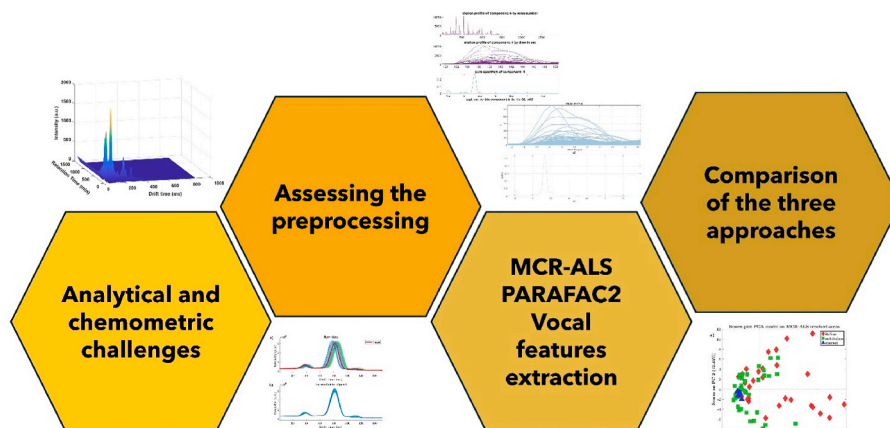
^c LabCom Aioly, Artificial Intelligence and Optics Laboratory, Montpellier, 34196, France

^d Ministero dell'agricoltura, della Sovranità Alimentare e delle Foreste, Dipartimento dell'Ispettorato Centrale della tutela della Qualità e Repressioni Frodi dei prodotti agro-alimentari, ICQRF Laboratory of Modena, Via Domenico Cucchiari 12, Modena, 41124, Italy

HIGHLIGHTS

- Thorough development and assessment of a GC-IMS method from data acquisition to data analysis.
- First application of PARAFAC2 for features extraction, and comparison with MCR.
- The proposed pipeline could allow reaching a high degree of automation in the analysis of GC-IMS data.

GRAPHICAL ABSTRACT



ARTICLE INFO

Handling Editor: Prof. L. Buydens

Keywords:

GC-IMS
Features extraction
MCR
PARAFAC2
Honey
Authenticity

ABSTRACT

Background: Gas-chromatography coupled to ion mobility spectrometry (GC-IMS) is a powerful platform for untargeted analysis of volatile compounds, combining high sensitivity and short analysis times. However, its practical implementation still presents several analytical and computational challenges. From the instrumental perspective, the high sensitivity of isothermal GC-IMS makes it prone to signal saturation and memory effects, which can compromise the reproducibility of measurements. On the data-analysis side, the nature of GC-IMS outputs produce large and collinear datasets, in which baseline drift, shift in ion mobility and retention time modes, and peak overlap hinder the reliable extraction of relevant chemical information.

* Corresponding author.

E-mail address: caterina.durante@unimore.it (C. Durante).

<https://doi.org/10.1016/j.aca.2026.346326>

Received 12 May 2026; Received in revised form 3 August 2026; Accepted 25 September 2026

Available online 25 September 2026

0003-2670/© 2026 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Results: This work proposes a comprehensive workflow from the analytical strategy to the data analysis using as benchmark a data set of honey samples of different origin. Analysing the GC-IMS data, obtained with an instrument operating in isothermal condition, a prominent and marked memory effect was observed; it could be contained only through the systematic use of frequent blanks. Notably, most papers in the literature do not adopt blanks at such frequency. Attention was then devoted to the preprocessing, to improve data quality prior to modelling while limiting the number of preprocessing steps. A multiway approach, i.e. Parallel Factor Analysis 2 (PARAFAC2) applied interval wise, was adopted, for the first time, to resolve overlapping peaks and extract chemically meaningful components from the GC-IMS landscape. The results obtained were also compared with Multivariate Curve Resolution–Alternating Least Squares (MCR–ALS) and recent improved PARAFAC2 algorithms.

Significance: A complete pipeline for the GC-IMS analysis was carefully developed and validated. PARAFAC2, was successfully applied for the deconvolution of GC-IMS data. This approach provided comparable results with previous methods and indicated that the same grade of automation reached in the analysis of GC-MS data with the same approach is feasible.

1. Introduction

Headspace Gas Chromatography-Ion Mobility Spectrometry (HS-GC-IMS) is a robust and sensitive benchtop analytical technique that combines gas chromatographic separation with ion mobility separation [1, 2]. The analytes are first separated by their retention in a capillary column and then by their drift time through a nitrogen buffer gas under an electric field. Ionization occurs when β particles generate ionized water clusters (Reactant Ion Peak, RIP), which subsequently ionize the analytes through chemical reactions. This dual separation provides high selectivity and sensitivity for detecting volatile and semi-volatile organic compounds [2].

GC-IMS has found extensive applications across various fields [3,4]. In food analysis, it is widely used for the authentication and discrimination of the geographical or botanical origins of products such as honey [5], olive oil [6], bacon [7], and saffron [8]. It has also been employed to detect adulteration in edible oils [9] and to assess the quality and spoilage of meats [10] and fermented products [11]. Beyond food analysis, its utility extends to environmental monitoring, where it is applied to identify pollutants and assess air quality [12], and to forensic investigations for analysing fire debris [13], explosives [14], and illicit substances [15]. GC-IMS is well-suited for untargeted analysis due to its minimal or even negligible sample preparation requirements. Moreover, since it operates at ambient pressure, it is a less cost- and resource-intensive alternative to other analytical techniques, such as GC-MS, which requires a high-vacuum system to operate [16]. Finally, it inherently functions in an untargeted mode, detecting all possible ions without the sensitivity loss commonly observed in full-scan modes of other detectors [2].

Despite its versatility, GC-IMS presents several challenges from both analytical and data analysis points of view. Regarding the analytical challenges, the high sensitivity of the instrument can lead to saturation phenomena [17]. Indeed, since ionization involves a reaction between gas-phase molecules and the RIP, excessively high concentrations of molecules in the gas phase can lead to signal saturation and a consequent loss of linearity [17]. This phenomenon may occur either in the presence of highly concentrated compounds, as it can be ethanol in many foodstuffs [18], or as a result of an excessive injection volume [19]. Therefore, this effect must be carefully considered during sample preparation. Another major issue is the memory effect, in which volatile organic compounds (VOCs), that are not fully eluted from the GC column contaminate subsequent samples, causing a decrease in the RIP signal and potentially producing "ghost peaks" [20]. This carry-over is particularly problematic in instruments operating at constant temperatures, where the absence of temperature ramping prevents thorough column cleaning [20]. To mitigate this effect, the following measures could be employed: syringe flushing with nitrogen; increasing the flow of carrier gas; and inserting blank runs between samples [20].

Data analysis remains the most time-consuming step in the GC-IMS workflow. First of all, like all hyphenated techniques, the instrumental

output has three dimensions, which are: samples, retention times, and drift times. However, unlike other hyphenated techniques such as GC-MS or LC-MS, the GC-IMS data suffer from shifts along both chromatographic and drift time dimensions [4]. In fact, during prolonged analysis, instrumental variations can cause misalignments in retention times (GC dimension) and drift times (IMS dimension) of detected compounds [4]. These shifts necessitate rigorous data preprocessing and alignment protocols, often using internal standards or intrinsic signals, such as the RIP, to ensure accurate identification and comparability of results, especially in untargeted analyses where unknown features are compared across samples. The sampling frequency of a GC-IMS is relatively high, resulting in a computationally intensive output [21]. Furthermore, in most untargeted studies, the number of samples is also large, meaning that the final dataset can reach several gigabytes in size [6]. It is therefore necessary to implement an effective data compression strategy to ensure that the dataset remains manageable on a standard computer. Finally, within the framework of untargeted analysis, GC-IMS signals typically exhibit a high degree of co-elution and overlap in both dimensions. For these reasons, the features extraction process, where the relevant information is separated from random noise, emerges as the bottleneck of the overall analysis. This also highlights the need for advanced chemometric techniques capable of resolving and quantifying the analytes.

Currently, the predominant approach employed in published studies involves the use of VOCal software (Gesellschaft für Analytische Sensoren mbH, Dortmund, Germany), which calculates the area under detected peaks following manual definition of integration intervals [7, 22,23]. Although this method has been shown to be reliable in many studies, it is labour-intensive and depends on meticulous manual selection of the GC-IMS regions that enclose the areas to be integrated (i.e. the peaks) and on the sample selected as reference to define the regions set (i.e. the grid map). This increases the risk of discarding peaks and consequently losing valuable information, particularly when analysing complex matrices where peaks strongly overlap. Obviously, the approach is highly operator sensitive.

Alternative approaches reported in the literature rely on peak-picking algorithms that first define regions of interest (ROIs), namely peaks with intensities above a predefined threshold, and subsequently merge these ROIs using clustering techniques [4,24]. Although effective, this strategy is inherently affected by peak-picking failures and signal shifts in both dimensions (retention time and drift time), thus requiring bidimensional alignment prior to analysis [4]. More recently, Multivariate Curve Resolution–Alternating Least Squares (MCR–ALS) has been applied to GC-IMS data [8] for its capability to resolve pure component profiles from complex mixtures. Notably, this approach requires alignment only along the drift time dimension, as, when the three-way GC–IMS dataset is unfolded by augmenting the retention-time profiles across samples, MCR–ALS can accommodate shifts in the retention time dimension [25,26] as well as baseline effects in the same dimension. On the other hand, MCR requires constraints to be applied to

cope with rotational ambiguity, in the previous work several constraints were used including component correspondence which require some previous knowledge about the matrix composition [8].

The aim of this study was to address the challenges outlined above and to provide a workflow from data acquisition to features extraction in the context of untargeted analysis, especially when data show strong chromatographic overlap. The focus has been on limiting the number of preprocessing steps and tuneable parameters taking advantage of the multivariate/multiway resolution methods to extract relevant chemical information. In particular, to the best of our knowledge, this study represents the first application of a multiway modelling approach, i.e. PARAFAC2 [27], to the analysis of GC-IMS data. PARAFAC2 does not suffer from rotational ambiguity and furnishes a unique solution under mild conditions [28], hence constraints are not needed to define the solution. On the other hand, it may suffer from algorithmic issues, such as local minima, convergence speed, stability, and constraints may be needed or helpful to overcome them. Moreover, in many cases, it is desired that the parameters are constrained to be nonnegative, i.e. in chromatography, because ideally, concentrations, elution profiles and spectra are nonnegative and imposing non-negativity allows gathering components profiles better conforming to pure spectra. To this aim PARAFAC2 algorithms consenting to constraints all modes to be non-negative have been recently developed [29–31]. It has also to be underlined that PARAFAC2 requires constant cross-products of the elution profiles, rather than constant peaks positions across samples, which translates in the condition that only linear shifts of the elution profiles that do not violate this constraint are allowed. While in general, it is not expected that this condition is always fulfilled, the PARAFAC2 model has been proven to be successful in modeling gas chromatography coupled with mass spectrometric detection (GC-MS) in many applications [32,33]. At the same time recent improvements of PARAFAC2 have been recently proposed to cope with cases violating the constant cross-products, e.g. presence of non-linear retention time shifts or occurrence of peak inversion. Specifically, the Shift-Invariant Tri-linearity (SIT) [34] and Shift-Invariant Soft Tri-linearity (SIST) constraints [35] methods which we also considered in this work.

Due to the complexity and strong overlap in the chromatographic dimension, distinct PARAFAC2 models have been derived after defining intervals in the retention time direction. Intervals definition has been done manually but their definition is not critical as intervals can be allowed to overlap and in perspective can be automatised [28,29].

MCR-ALS has been applied in parallel adopting the same intervals definition, in fact it has been demonstrated that an interval strategy enhances MCR-ALS resolution capability [36,37]. MCR-ALS and PARAFAC2 were selected because of their demonstrated effectiveness in GC-MS applications and their capability to decompose complex multiway data, resolve overlapping signals, and extract meaningful chemical information without requiring strong prior assumptions about the system. As benchmark, we have considered a data set of honey samples of different geographical origin for authentication purpose. The resolved components were thoroughly examined, and exploratory data analysis carried out on the features matrix holding the peaks area. These results were compared with those obtained by using the peak areas provided by the VOCal software.

GC-IMS has been previously applied to analyse the volatile fraction of honey samples mainly to characterize the botanical origin [5,38]. Gerhardt et al. [5] proved that GC-IMS, combined with chemometric methods such as partial least squares – discriminant analysis (PLS-DA), could be a robust and fast tool for the classification of the botanical origin of honeys. Moreover, Arroy-Monzaranes et al. [39] applied GC-IMS to differentiate pure honeys from those adulterated with sugar cane or corn syrups by analysing 198 samples and selecting 130 characteristic markers from the topographic plots for multivariate analysis. Feng et al. [40] used HS-GC-IMS to investigate the volatile organic compound (VOC) profile of Acacia honey powders obtained by vacuum belt drying at different temperatures. In that study, 23 volatile

compounds belonging to several chemical classes were identified, highlighting the capability of this technique to characterize aroma-related compounds in processed honey products. To the best of our knowledge, this is the first time HS-GC-IMS was used for studying the geographical origin of honey with different botanical origins.

2. Experimental section

2.1. Honey samples benchmark

In this study, 67 honey samples were analysed using GC-IMS: 31 of Italian origin and 36 of not-Italian origin. These samples belong to three botanical varieties: polyfloral (36 samples), acacia (21 samples), and citrus (10 samples). The aim was to evaluate the possibility of recognizing honey of Italian origin notwithstanding the botanical origin. Detailed information is reported in Table 1.

The honey samples were provided by the Laboratory of the Central Inspectorate for Quality Protection and Fraud Prevention of Agri-Food Products in Modena (ICQRF). It is important to note that all samples had previously undergone rigorous verification procedures, as outlined in the annex to Legislative Decree 179/2004 (the year of analysis is reported in Table 1). These verification procedures included chemical-physical analyses, organoleptic examination, and melissopalynological analysis. Although the resulting values could not be disclosed, it was confirmed that all honeys were declared compliant in all tests.

2.2. GC-IMS analysis

The samples were analysed by HS-GC-IMS FlavourSpec G.A.S. (Gesellschaft für Analytische Sensorsysteme mbH, Dortmund, Germany) coupled with the PAL-RSL autosampler (CTC Analytics AG, Zwingen, Switzerland).

The analyses were carried out under isothermal conditions, meaning the temperature was kept constant, since the instrument could only operate at constant temperature, while the carrier gas flow was varied. The analysis conditions were selected based on preliminary tests and literature data [41,42]. The flow ramp was set as follows: from an initial

Table 1
Classification of honey samples by class of origin, country of origin, botanical origin, including number of samples for each typology.

Class of origin	Specific country of origin	Botanical variety	Number of samples
italian	Italy	Acacia	9
	Italy	Citrus	10
	Italy	polyfloral	12
not-italian	Argentina	polyfloral	2
	Brasil	polyfloral	1
	Bulgaria	polyfloral	1
	Bulgaria-Romania	polyfloral	2
	Bulgaria-Romania	polyfloral	1
	China	polyfloral	1
	China	polyfloral	10
	Cuba	polyfloral	1
	Hungary	polyfloral	1
	Turkey	polyfloral	3
	Italy - Hungary - Argentina - Moldova	polyfloral	1
	China	Acacia	1
	China	Acacia	1
	Hungary	Acacia	1
	Hungary	Acacia	1
	Hungary-Italy	Acacia	1
	Hungary-Romania	Acacia	1
	Hungary-Serbia	Acacia	1
	Romania	Acacia	1
	Serbia-Hungary	Acacia	1
	Ukraine	Acacia	2
	Spain	Citrus	1

value of 2 mL/min for 2 min, increasing linearly over 18 min until reaching 50 mL/min. Each run lasted 28 min, followed by a 5-min cleaning step, during which the syringe was flushed with a nitrogen flow of 2 mL/min to remove any contamination and minimize memory effects [20]. A summary of the GC-IMS setup is reported in Table S1, Supplementary Materials.

A mix of n-ketones (2-butanone, 2-pentanone, 2-hexanone, and 2-heptanone, named from now on as "ketomix") was analysed under the same chromatographic conditions as the samples. This ketomix was used for calculating the retention index (RI) in GC-IMS since the alkanes, used for instance in GC-MS, can't be ionized with the chemical ionization of the instrument [20]. The calculated RI and the drift times were matched with available libraries using VOCal software (GAS, Dortmund, Germany) for putative identification. Moreover, the ketomix standards were searched in the volatile profile of the samples for performing the comparison between MCR-ALS and PARAFAC2 and assessing the resolution of the ion mobility spectra (see Section 3.1).

2.3. Dataset

Preliminary tests showed a strong memory effect, which is typical of instruments operating at constant temperature as reported in the literature [20]. To overcome this issue, in this study, one empty vial bracketed each honey sample (i.e. two blanks per sample). Thus, blanks were analysed under the same experimental conditions as the honey samples. One of the honey samples provided by the ICQRF was selected as control sample and analysed 8 times during the measurement sequence to check for instrumental drift or batch effect.

Hence, a total of 67 honey samples, 8 replicates of the control sample, and 150 blanks were analysed. Due to a pronounced carry-over effect observed in the blanks measured immediately after the samples, only the blanks collected prior to the sample runs were considered for the following data evaluation. As a result, data analysis was performed on 150 measurements (75 samples and 75 blanks before the sample).

2.4. Preprocessing

The preprocessing strategy was selected based on the nature of the GC-IMS data, with respect to the preprocessing pipeline present in the literature [6,24] ours, as illustrated in Fig. 1, represent a simplified workflow; in fact baseline correction and alignment are only applied on the drift time dimension, since both PARAFAC2 and MCR can handle it.

At the end of the experimental analysis, a three-way data array $\mathbf{X}$ ($I \times J \times K$) was obtained, where I corresponds to the sample mode (blanks and honey samples), J represents the chromatographic dimension defined by retention times (RT), and K represents the ion mobility dimension described by drift times (DT). The initial array had the following dimensions: $150 \times 11432 \times 3150$. During import, retention times were downsampled by retaining one point every 8 (for details see section 3.1). This allowed us to reduce the array size to $150 \text{ samples} \times 1429 \text{ retention times} \times 3150 \text{ drift times}$. Ion mobility spectra were aligned by using the algorithm *iCoshift* as detailed in Section 3.1. Other alignment strategies are available in literature using, for example, added standards in sample as references and cubic spline interpolation as algorithm [24]. After alignment, useless regions in both drift and retention times were cropped. In particular, the drift time region (0–4.3 msec) enclosing the RIP signal, which can dominate the spectra, was removed. It should be noted, however, that in the presence of pre-RIP signals, the selection of the RIP interval becomes particularly critical, since peaks preceding the RIP are generally closely spaced in terms of drift time. Also, in retention time the last segment of the chromatograms (1200–1600 s), which showed no peaks, was removed. This allowed a further reduction of the data dimensions to $150 \times 921 \times 555$. Savitzky-Golay smoothing was applied to the chromatographic and ion mobility dimensions (window width = 7, polynomial order = 2) to reduce random noise while preserving peak shapes [43]. Finally, a

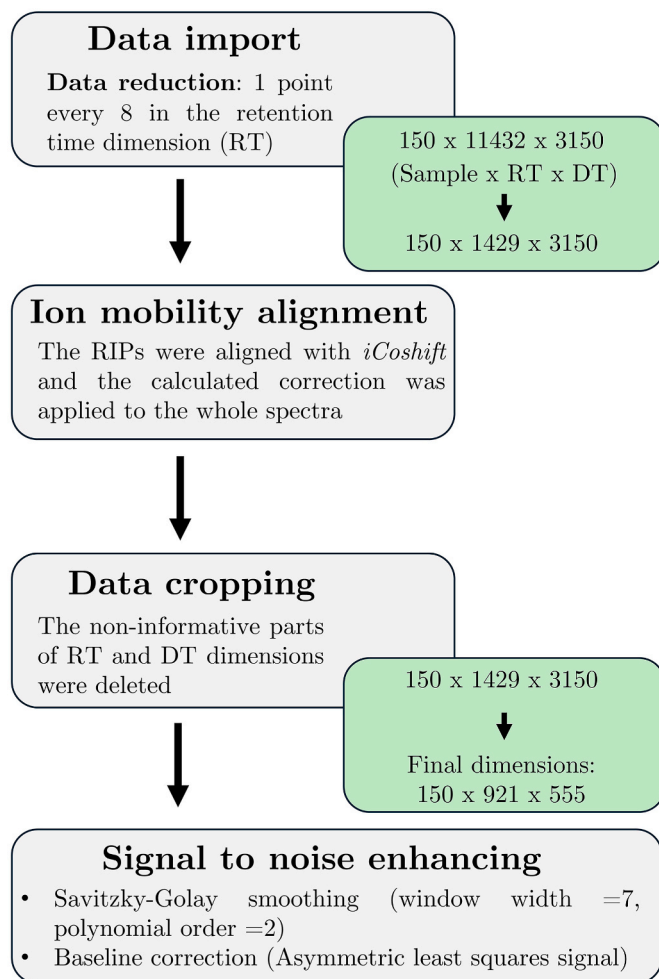


Fig. 1. Scheme of the proposed preprocessing of GC-IMS data. The green rectangles refer to data array dimensions. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

baseline correction was performed on the dimension of ion mobility using the asymmetric least squares (ALSs) [44] with default parameters ($\lambda = 100000$, $p = 10^{-3}$ and $d = 2$), ensuring a more accurate representation of the true signal by minimising baseline drift.

2.5. Peak resolution

GC-IMS signals contain a large amount of information embedded in a highly complex signal, which is strongly affected by chromatographic coelution. To address this issue, the pre-processed GC-IMS signals consisting of 75 samples and 75 blanks were analysed by PARAFAC2 as implemented in the PARADISE software [45]. For comparative purposes they were also analysed by MCR-ALS by using the MCR-GUI function [46]. For the methodological and algorithmic aspects of both techniques the reader is referred to the relevant literature [25,28,31,46–48].

Prior to both PARAFAC2 and MCR-ALS the chromatograms were divided into 38 intervals as shown in Fig. S1, Supplementary Materials, and MCR-ALS and PARAFAC2 models were calculated for each interval as shown in Fig. 2.

2.5.1. Resolution by PARAFAC2

PARAFAC2 is a tensor decomposition method that can directly operate on the three-way data structure $\mathbf{X}$ (drift time x retention time x samples), differently from PARAFAC, the PARAFAC2 model allows every sample to have its own distinct set of loadings in one of the modes

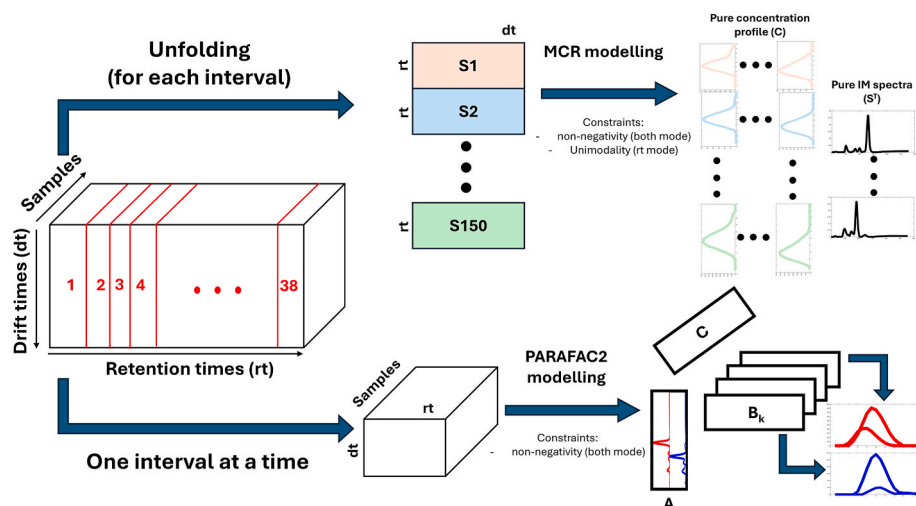


Fig. 2. Schematic representation of the workflow for the data analysis.

[28,49] in our case the retention time mode (J), whereas the other independent mode is restricted to be linear, in our case the drift time mode (I). In PARAFAC2, each sample (each slab of $\underline{X}(I \times J \times K)$ $\mathbf{X}_k(I \times J)$ is modelled according to the equation:

$$\mathbf{X}_k = \mathbf{A}\mathbf{D}_k(\mathbf{B}_k)^T + \mathbf{E}_k \quad \text{Eq. (1)}$$

where $\mathbf{A}(I \times F)$ holds the independent mode (i.e. drift time) loadings for each F factor (drift time profiles of “pure” component). $\mathbf{D}_k(F \times F)$ is a diagonal matrix that holds the k -th row of the sample mode loadings matrix $\mathbf{C}$ (see Fig. 2). The matrix $\mathbf{B}_k(J \times F)$ holds the individual sample loadings for the mode affected by the shift (i.e. retention time) for each F factor (i.e. elution profile for each resolved component). The matrix $\mathbf{E}_k$ represents the residuals (unmodeled part of the data).

The PARAFAC2 model assumes the cross-product terms $\mathbf{B}_k^T\mathbf{B}_k$ to be constant. This means that PARAFAC2 cannot accommodate non-linear retention time shifts, varying degree of shape changes, or occurrence of peak inversion. While this assumption is quite strong and it is not expected to hold true in general, reasonably good results were achieved in the present case, where PARAFAC2 is applied in narrow retention time intervals mostly holding a single or few peaks. Nonetheless, recent methods, such as SIT [34] and SIST [35] (specifically proposed to deal with varying peak shifts and shapes) were also applied to two selected intervals, namely intervals 16 and 22 (Table S2), spanning different degrees of overlap and shape changes among neighbouring peaks (the top plots of Fig. S10 and S10 show the selected intervals 16 and 22, respectively).

Non-negativity constraint was applied to all three modes (samples, retention times, and drift times) [29,45]. The number of components for each PARAFAC2 model was chosen according to the explained variance, the entity and shape of the residuals, the core consistency [49] and the number of well-resolved peaks identified by each model [31]. The drift and retention times profiles of the resolved components were carefully inspected to select only those relating to peaks and peak areas were subsequently calculated. In all the selected models the explained variance was above 95% and the residuals were randomly distributed in the chromatographic dimension.

2.5.2. Resolution by MCR-ALS

MCR-ALS operates on a multiset structure obtained by augmenting column-wise the single sample GC-IMS data (see Fig. 2), which is decomposed according to the equation:

$$\mathbf{D} = \mathbf{C}\mathbf{S}^T + \mathbf{E} \quad \text{Eq. (2)}$$

Where $\mathbf{D} \in \mathbb{R}^{I \times J}$ is a two-way dataset, $\mathbf{C}$ contains concentration profiles,

$\mathbf{S}^T$ contains estimation of the spectra of the pure components and $\mathbf{E}$ is the residual matrix. The constraints applied were the following: non-negativity and unimodality for the concentration mode, non-negativity for the spectra mode. The number of MCR components was chosen based on lack of fit assessment [46] and the number of well-resolved peaks. In both cases, the optimal model was chosen as having the higher number of resolved peaks with the lower number of components. In all the selected models the explained variance was above 95% and the residuals were randomly distributed in the chromatographic dimension. The details of the resolved components are reported in Table S2, Supplementary Materials. When several components are resolved from a given data matrix by MCR-ALS, the resulting factorization may be affected by rotational ambiguity, i.e. different sets of concentration and spectral profiles may describe the experimental data equally well while satisfying the imposed constraints [50]. In this study, the extent of this ambiguity was therefore evaluated using the sensor-wise N-BANDS method [51]. In particular, the analysis was applied to the MCR-ALS models obtained for the same two representative retention-time intervals (16 and 22, Table S2), used for comparison with SIT and SIST (see section 2.5.1). Because of the computational cost of the sensor-wise optimization, only components exhibiting distinct peak-shaped profiles were selected for graphical reporting. For each selected component, the point-wise lower and upper boundaries for both C-mode and S-mode profiles were estimated under the same constraints used in the corresponding MCR-ALS analysis (reported above). Detailed results are provided in the Supplementary Material (Figs. S2–S4).

2.5.3. Peak areas from VOCal

The VOCal software, representing the state-of-the-art for GC-IMS data analysis, was also employed. In this case, 265 rectangular regions (a grid) were manually defined on the drift time $\times$ retention time map, within which peak areas were calculated (Fig. S2, Supplementary Materials). Defining this grid is both challenging and time-consuming for several reasons. First, the volatiliomic pattern of honey is highly complex; consequently, the analysed map (Fig. S5) is rich in analytes and strongly affected by co-elution. This aspect leads to the well-known issue of manual integration, namely user-dependence. Just as manual integration of peaks along the chromatographic dimension alone, such as the total ion current in GC-MS or the signal from a GC-Flame Ionization Detector (GC-FID), is user-dependent, the definition of the integration grid in GC-IMS output is similarly influenced by the operator.

Furthermore, although the integration grid is ultimately defined once, manual analysis of GC-IMS signals still requires the user to inspect each sample individually to ensure that peaks present only in certain samples are not excluded from integration, which would result in a loss

of information. Finally, integration was also performed on the 77 blank samples, enabling the identification of regions that were also present in the blanks (21 out of the 265 initially calculated regions).

2.6. Principal Component Analysis

Principal Component Analysis (PCA) [52,53] was applied as an unsupervised exploratory method [54] to investigate the main sources of variation among the honey samples and to assess whether the different signal-processing approaches provided comparable sample distributions in the reduced multivariate space. Three independent data matrices were obtained using the peak areas coming from MCR-ALS, PARAFAC2, and VOCal, respectively. The resulting sample $\times$ area matrices were auto scaled before PCA.

2.7. Software

GC-IMS data were exported using VOCal software (GAS, Dortmund, Germany). Subsequent data preprocessing was conducted with custom routines implemented in MATLAB v. R2024 (The MathWorks, Inc., Natick, MA, US) using functions in the PLSToolbox v. 9.2 (Eigenvector Research, Inc., Manson, WA, US). Icoshift is available at the developer website (<https://ucphchemometrics.com/186-2/algorithms/>), while applyicoshift routine was developed by G. Tomasi (Department of Plant and Environmental Science, University of Copenhagen, Thorvaldsensvej 40, 1871 Frederiksberg C, Denmark) and F. Savorani (Department of Applied Science and Technology, Polytechnic of Turin, Corso Duca degli Abruzzi 24, Turin 10129, Italy) and kindly made available to us. MCR-ALS (2.0) software was downloaded from the official developers' website (<https://mcrals.wordpress.com/theory/mcr-als/>) and is implemented in the MATLAB environment. PLSToolbox v. 9.5 software was used for PCA analysis and SIST (by using the ALS_SIT function with shiftmap option). PARAFAC2 and SIT models were developed within

PARADISE software (Kromath, Denmark), downloaded from the official developers' website (<https://www.kromath.com/>).

3. Results and discussion

3.1. Assessing the preprocessing step

A critical aspect during preprocessing, particularly when multiple steps are involved, as in the present work, is the potential introduction of artifacts that can propagate through subsequent steps. Therefore, each preprocessing step was carefully selected and evaluated both for its ability to perform the intended data treatment and to ensure that no artifacts were introduced.

To make the GC-IMS data array, $\mathbf{X}$, manageable, an initial data reduction step is necessary [4,6,21,24]. This has been done in various ways, including binning in both drift time and retention time dimension [6], where the average of a user-defined number of points is calculated, by decimation (down sampling) of the retention time dimension, in which one point every N (defined by the user) points is retained [4,24], or by wavelet compression in drift time [21]. In this work, down sampling with a factor of 8 was applied only to the chromatographic dimension without altering the chromatographic profile. This is clearly shown in Fig. 3, which compares the chromatogram of one of the samples (blue curve), i.e. a control replicate, and the retained points after reduction (red stars). The down sampled chromatogram closely retraces the original one. At this point, the $\mathbf{X}$ has the following dimensions: 150 samples $\times$ 1429 retention times $\times$ 3150 drift times.

Due to small instrumental fluctuations, mainly arising from variations in pressure and potential differences within the ion mobility spectrometer, the ion mobility dimension also experiences shift between chromatographic runs. Since MCR and PARAFAC2 handle shifts in only one dimension (typically the chromatographic dimension), it was necessary to align the ion mobility dimension. For this reason, after data

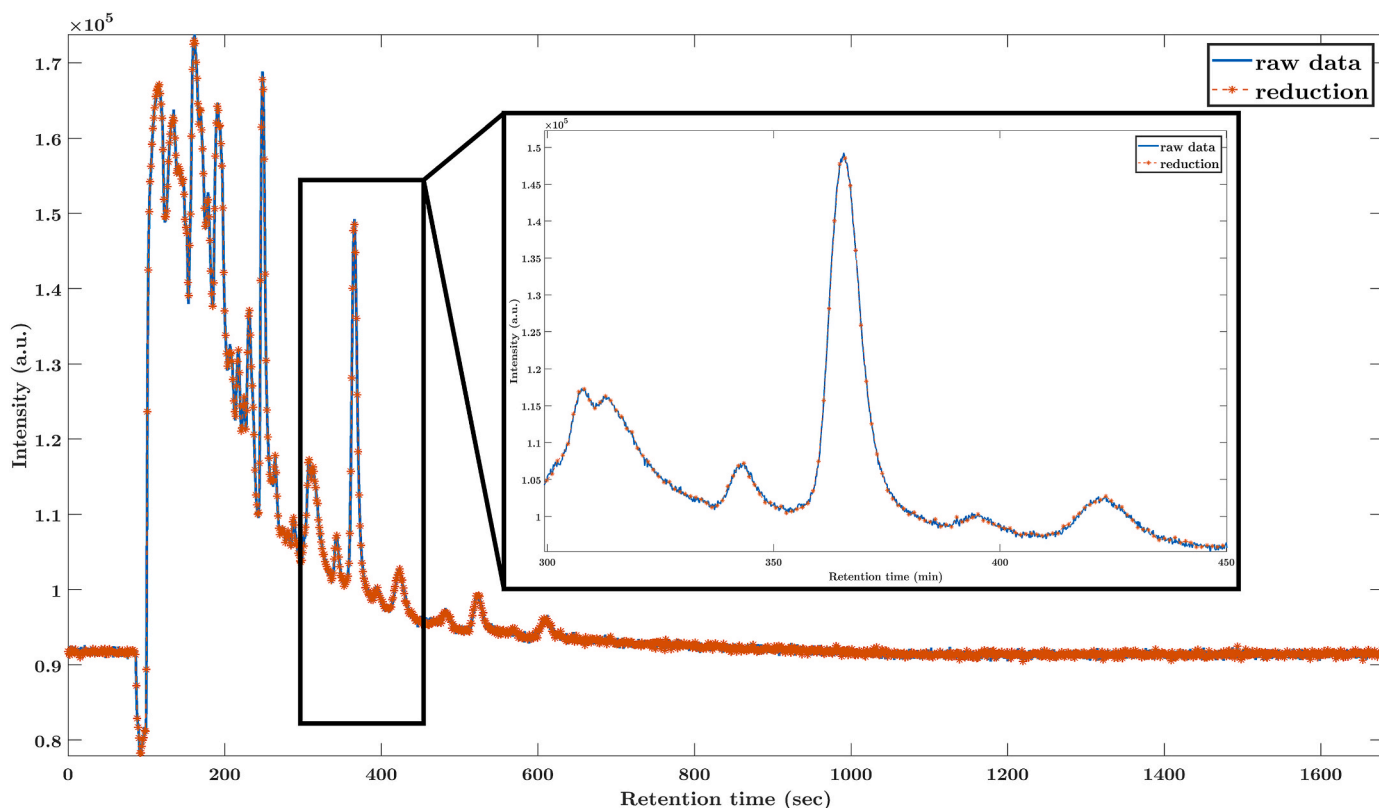


Fig. 3. Raw chromatogram (blue line) and retained point (red stars) of a control sample replicate. A zoomed-in view is also provided to highlight that no artifacts were introduced by the down sampling. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

reduction, the ion mobility dimension was aligned.

To achieve a good alignment, the chemical ionization occurring within the instrument was exploited. During ionization, the Reactant Ion Peak (RIP) forms due to the ionization of residual water by means of β particles. The RIP is a cluster of water molecules ionized, and it reacts with the volatilized molecules, A, producing ions $A(H_2O)_m^+$ and $A_2(H_2O)_m^+$, as function of analyte concentration, as reported in Fig. S3, Supplementary Materials. Although the absolute drift time of an analyte may vary between runs as a result of the fluctuations, the drift time difference between an analyte peak and the RIP remains constant because it depends solely on the mobility difference between the ions. This phenomenon is illustrated in Fig. S6, which shows ion mobility spectra of the same ionic species, A, (i.e., registered at the same retention times from three chromatograms of the control sample). Fig. S6 highlights the RIP and shows that the drift time difference between the RIP and the analyte signal is consistent across chromatographic runs.

Therefore, the RIPs were aligned using *iCoshift*, with the signal of a control sample (red line in Fig. 4) selected as the reference for alignment. The effect of this process is shown in Fig. 4 and in Fig. S7, Supplementary Materials. Initially, the shift parameter was determined by running *iCoshift* on the drift time signals obtained by summing over the retention times dimension and considering only the control samples. After that, the alignment parameter was applied to single drift time signal (i.e. to the whole spectrum) at each retention time and to all the samples by using a MATLAB routine (*applycoshift*, see Section 2.7).

Honey presents a particularly challenging case for alignment because its volatile profile is highly complex, containing many isomers that have nearly identical drift times. In addition, different honey samples may contain different volatile compounds, which makes it essential to ensure that the same ions are being compared when evaluating the quality of the alignment. Consequently, the success of the alignment can only be verified when it is certain that the spectra compared originates from the same analyte. To ensure this, ion mobility (IM) spectra recorded at identical retention times were compared across replicates of the control samples, confirming that the ion mobility dimension was accurately aligned.

After alignment the non-informative signal regions in drift and

retention times were cropped. Finally, smoothing and baseline correction were applied on the drift time dimension, with the methods explained in Section 2.4. The effect of the overall preprocessing is shown in Fig. 5, which shows the summed signal along the drift times and retention times, both before (a) and after (b) preprocessing.

3.2. Assessing the peak resolution step

Before applying MCR-ALS and PARAFAC2, the chromatographic signals were divided into 38 intervals (see Fig. S1, Supplementary Materials). The intervals were selected with the aim of including only one chromatographic peak per interval, in line with recommendations from the literature [55,56]. This interval-based segmentation was motivated by two main considerations. Given the high degree of coelution observed in the volatile organic compound profile of honey (see Fig. 6), dividing the chromatogram into smaller intervals greatly helps both algorithms in resolving overlapping peaks. From a computational standpoint, this segmentation also markedly reduces processing demands, making the calculations feasible on a standard computer.

In principle, MCR-ALS could achieve good resolution also when applied to intervals of larger size [8,57]. To explore this option, we initially divided the chromatogram into four large intervals (Fig. S8, Supplementary Materials) and computed MCR-ALS models for each of them. For sake of brevity, only the first interval is discussed. Several MCR models with different numbers of components were built. The model with 60 components turned out to be the optimal in terms of resolved peak and lack of fit. However, this approach presented a major limitation: many resolved components contained multiple chromatographic peaks (an example for the first interval is shown in Fig. 6a), meaning that a single component is going to describe several compounds and prevent full signal resolution. This result is consistent with the expected behaviour of MCR-ALS-based models, since compounds sharing identical or highly correlated profiles in at least one dimension may be grouped within the same component. Selectivity constraints would be needed, within the MCR framework, if further separation of these correlated peaks were required.

This behaviour differs from typical MCR results in LC-HRMS or GC-

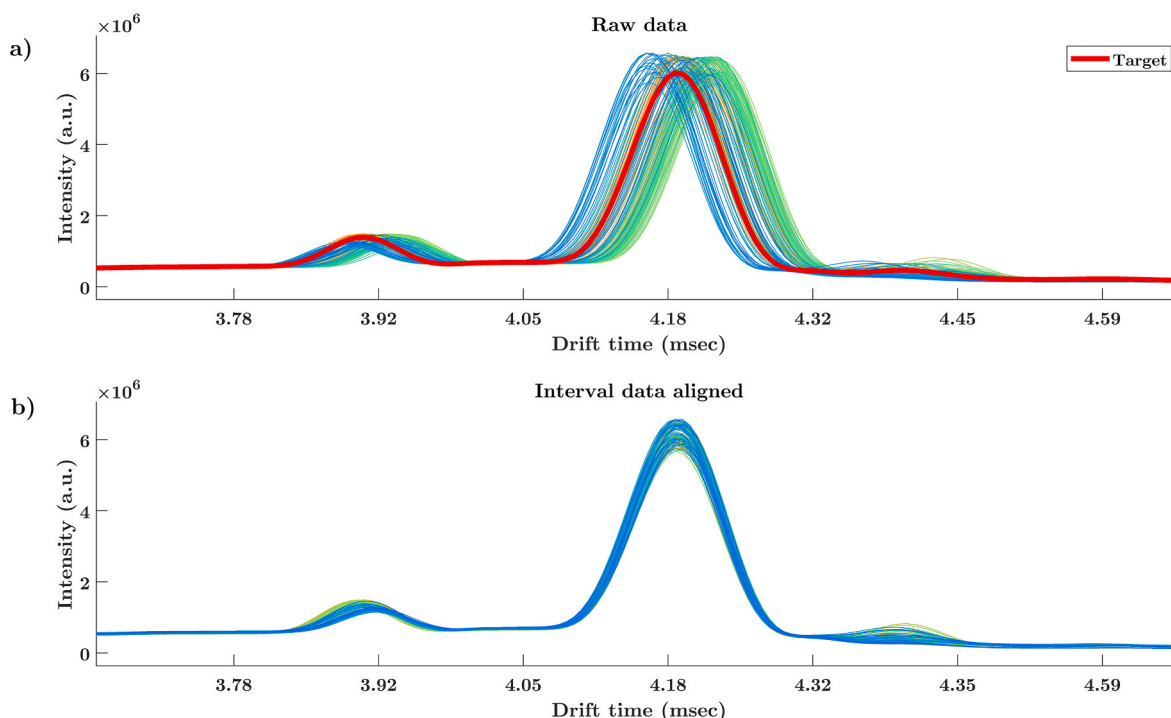


Fig. 4. Impact of alignment on the ion mobility (IM) spectral region corresponding to the RIP. Before *iCoshift* alignment (a) and afterwards (b).

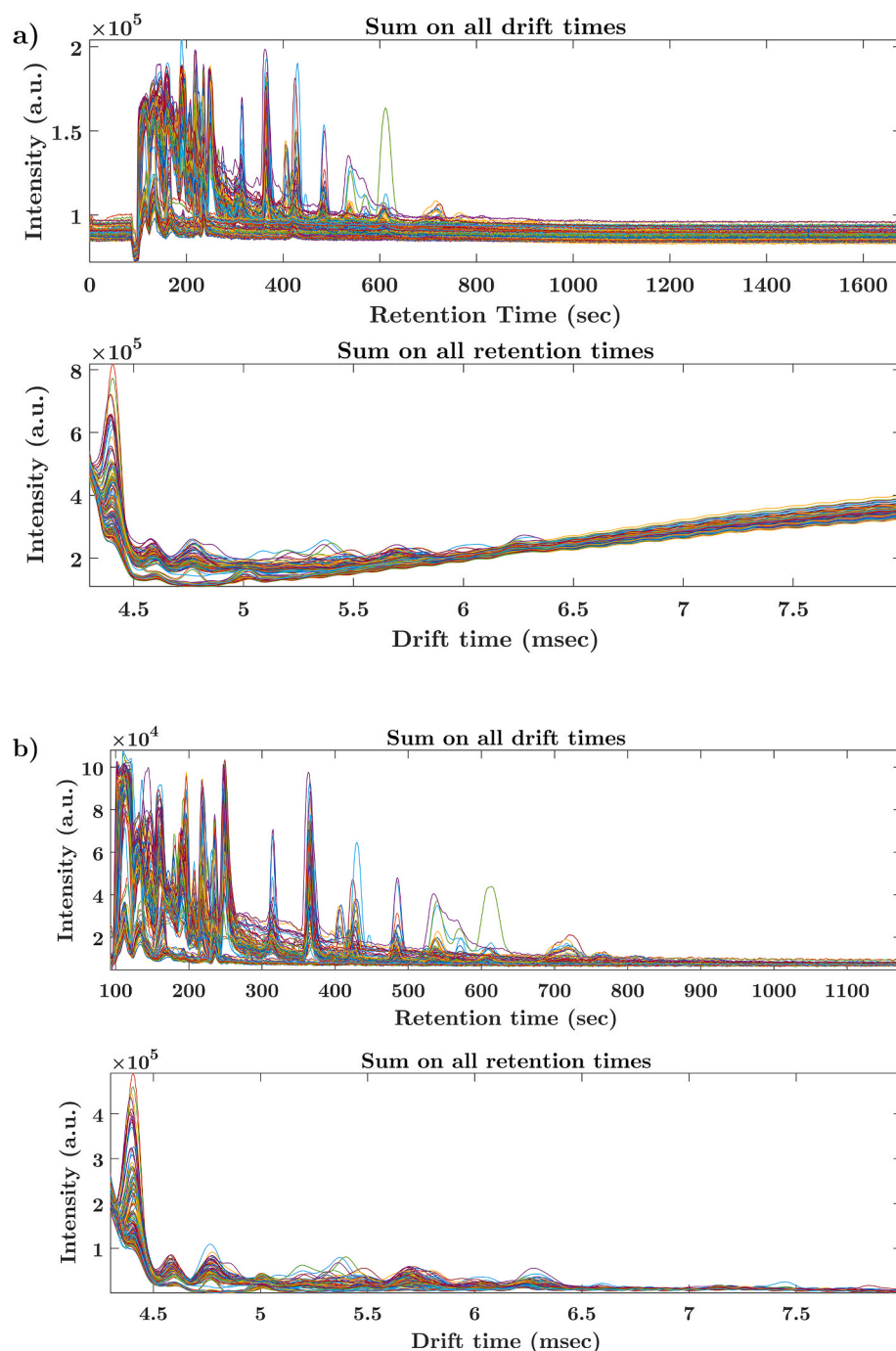


Fig. 5. GC-IMS data before (a) and after (b) the preprocessing.

MS data and can be attributed to the relatively low selectivity of ion mobility spectrometry [58]. Moreover, while mass spectra are rich in information, featuring base peaks, isotopic distributions, and fragmentation patterns, ion mobility spectra typically consist of only one or two signals, corresponding to $A(H_2O)^+$ and $A_2(H_2O)^+$ (Fig. S6, Supplementary Materials), with additional signals being rare. As a result, the probability that two chemical species share the same ion mobility spectrum is significantly higher in a low-resolution IMS system like the one used here, compared to mass spectrometry, where such overlap typically occurs only among isomers. For these reasons, 38 intervals were also used in MCR analysis.

It was not always feasible to keep only a single peak per interval, especially during early retention times when extensive peak overlap was

observed. Nevertheless, the results obtained by the 38 defined intervals were satisfactory in terms of resolved components. In Fig. 6 an illustrative example is reported for two of the defined intervals (Fig. 6a), showing that it was possible to resolve the different individual compounds (Fig. 6b and c) for each interval that indeed exhibited highly similar ion mobility spectra.

Having confirmed the appropriateness of the interval-based separation, we proceeded with a detailed evaluation of the components resolved by both MCR-ALS and PARAFAC2.

Our evaluation was therefore conducted as follows. First, we considered only those components (as reported in Table S2, Supplementary Materials) whose elution profiles corresponded to a well-defined peak, thereby excluding components arising from background

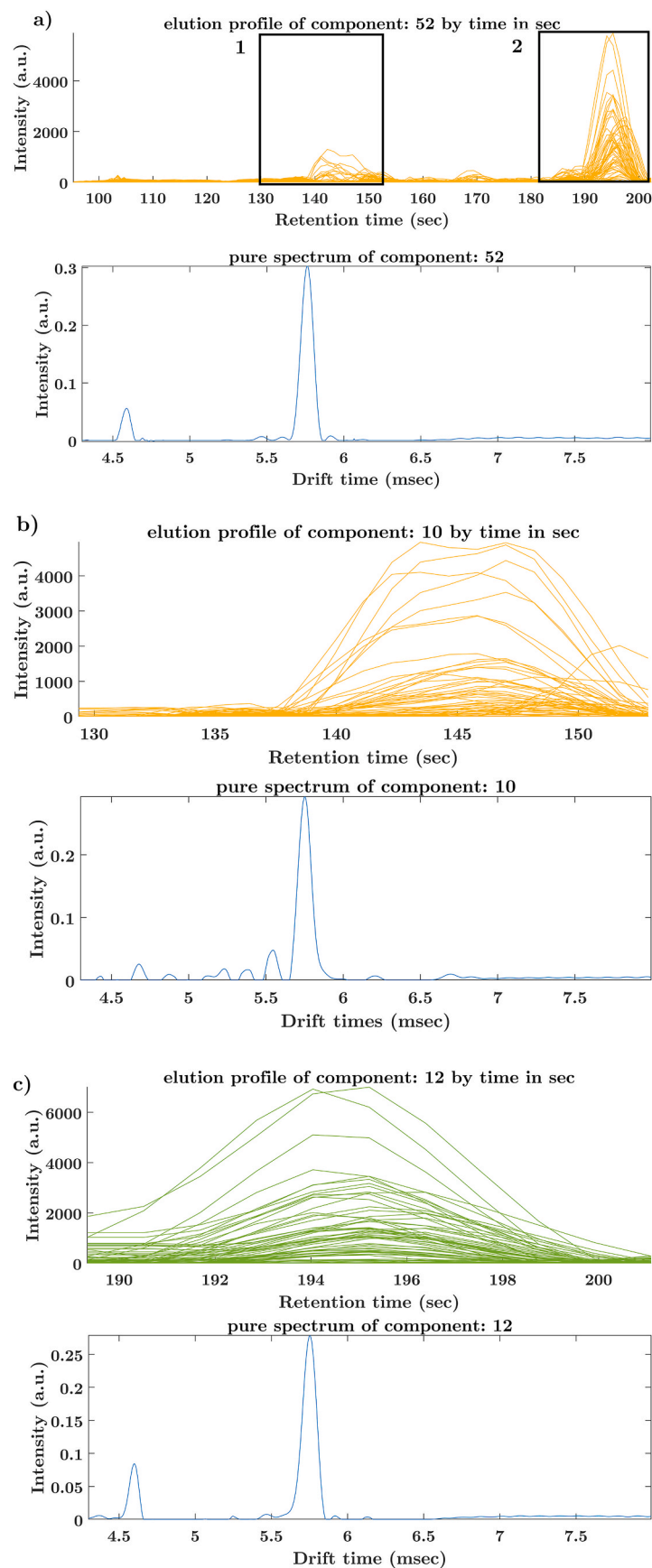


Fig. 6. Resolved component (a) using the intervals defined according to Fig. S8, Supplementary Materials. Subsequently, 38 intervals were defined, and two of them are reported as black rectangles. The components resolved for those two intervals are also reported (b-c).

signals or peaks split across multiple intervals (for example, the reader is referred to Fig. S9, Supplementary Materials).

PARAFAC2 resolved 114 peak areas, while MCR-ALS resolved 133 (for details regarding the models developed and the components describing a chromatographic peak the reader is referred to Table S2, Supplementary Materials). Secondly, in all three approaches (i.e., MCR-ALS, PARAFAC2 and VOCal), peaks were excluded when the ratio between the largest area observed in the samples and the average area measured in the blanks was below five [59]. This threshold was selected because such a low ratio indicates that the signal in the samples is not sufficiently distinct from the background detected in the blanks, suggesting that these components likely represent background artifacts rather than true analytical contributions. In all three cases, approximately 5% of the features were eliminated, resulting in the matrices of the areas, with 125 areas retained for MCR-ALS, 108 for PARAFAC2, and 244 for VOCal. It should be noted that the number of areas retained by MCR and PARAFAC2 corresponds to resolved compounds, characterized by both a chromatographic peak and an ion mobility spectrum. In contrast, for VOCal, the areas are defined as signals within the retention time versus ion mobility plot (see Fig. S5). As a single compound may produce multiple signals, the count of areas does not necessarily equate to the number of detected compounds.

For comparative purposes, we also estimated SIT and SIST models for two selected intervals (interval number 16 and 22 of Table S2) which are representative of different complexity (in terms of number of overlapped peaks) and varying degree of peaks retention times shift and shape change across samples. Also, for the same intervals, the extent of rotational ambiguity in the MCR-ALS models has been evaluated.

As far as MCR-ALS results are concerned, sensor-wise N-BANDS analysis showed that the MCR-ALS profiles were consistently enclosed within the corresponding lower and upper feasible boundaries for all the investigated components. Very narrow feasible bands were obtained for both the C and S profiles in both investigated intervals, indicating limited rotational ambiguity. Detailed results are reported in the Supplementary Material (Figs. S2–S4).

The comparison of MCR-ALS, PARAFAC2, SIT and SIST models showed substantial agreement of the resolved spectral profiles for the selected components in the respective intervals. In fact, the profiles' similarity (Table S3) is high, as can also be appreciated in Figs. S10–S11, reporting both resolved peaks (Figs. S10b–S10e and S11b–S11e) and drift times (S10f–S10g and S11f–S11g) obtained by the different models (for interval 22 the component n°4 was taken as an example).

These results support the effectiveness of the PARAFAC2 approach also dealing with the GC-IMS data.

Next, a targeted evaluation of the resolved components was performed. Specifically, we searched for all the ketones present in the analysed ketomix and successfully identified two, 2-butanone and 2-hexanone, based on their retention times and ion mobility spectra (reported in Fig. 7 and in Fig. S12, Supplementary Materials, respectively). Fig. 7 illustrates the results for 2-butanone: it shows the components resolved by PARAFAC2 and MCR-ALS (Fig. 7a and b, respectively) with their corresponding elution profiles (upper part of the figure) and ion mobility spectra (lower part of the figure), alongside the spectrum of the pure standard recorded in the ketomix run (Fig. 7c). The comparison demonstrates excellent agreement between the components resolved by the two techniques and between the resolved ion mobility spectra and the spectrum of the pure compound.

3.3. Exploratory PCA on the features matrix (comparison of the three approaches)

For an initial qualitative characterization of the aroma profile of the investigated samples, 15 analytes were putatively identified (see Section 2.2 for identification details). These compounds comprise five aldehydes, five alcohols, three ketones, one ester, and one organic acid (Table S4, Supplementary Materials). Identification of the MCR and

PARAFAC2 components was achieved by matching retention and drift times using VOCal. PCA models were then built by using the refined area matrices obtained using MCR-ALS, PARAFAC2, and VOCal, respectively. In all cases, up to 3 components were inspected. Fig. 8 displays the scores of the first two principal components for all three models, with samples colored according to their geographical origin, while the loadings plots are reported in the Supplementary Materials (Fig. S13).

Some observations can be made from these results. First, the control samples (blue triangles in Fig. 8a) showed good reproducibility, clustering with similar PC1 and PC2 scores. To further assess reproducibility across the control replicates, the dispersion of six areas obtained with the three approaches spanning different orders of magnitude (in terms of peaks area values) is shown as a box plot in the Supplementary Materials (Fig. S14), covering both intense and very low-intensity signals. This finding also underlines the importance of introducing two blanks between each sample in the measuring sequence, no drift was observed during the sequence, meaning that memory effects were effectively minimized. To further inspect the magnitude of the effect, a box plot of the sum of the areas (obtained with VOCal) for the first (blanks immediately following the samples) and second series of blanks (blanks following the first) is reported in Fig. S15, Supplementary Materials. The distribution of the blanks immediately following the samples (first one) is considerably broader and more heterogeneous, with a large number of outliers, with respect to the second blank. This indicates that the first set of blanks is strongly influenced by the preceding sample, with samples particularly rich in volatile compounds producing a pronounced memory effect. Moreover, in Fig. S16 the difference plot with VOCAL is reported where it is possible to see the higher magnitude of signals in the blank right after the sample with respect to the second one. In contrast, the second set of blanks exhibits a much narrower distribution, indicating substantially greater reproducibility. This is a key observation because the purpose of including blank runs between samples is not to completely eliminate the signals detected in the blanks, which are expected to persist due to unavoidable sources of background contamination (e.g., the vial cap septum or other system-related contributions), but rather to ensure consistent analytical conditions for every sample. Therefore, these results demonstrate that a single blank is insufficient to fully remove the memory effect from highly concentrated samples, whereas the second blank provides a stable and reproducible baseline. This allows us to conclude that the observed differences among samples reflect their intrinsic chemical composition rather than residual contamination from previously analysed samples.

The score plots showed a comparable data structure between the three approaches, consistently revealing three primary clusters: i) samples lying at positive PC1 and negative PC2 values in MCR (Fig. 8a) and PARAFAC2 (Fig. 8b) and positive PC2 values in VOCal (Fig. 8c). This group consists predominantly of Italian samples and it seems characterized by ethyl acetate and 2-hexanone (Fig. S13); ii) a second group also mainly composed of Italian samples at positive PC1 and PC2 values (for MCR and PARAFAC2, negative PC2 scores for VOCal) characterized by most of the analytes detected and identifies (Fig. S13); and a mixed cluster where the samples of the two-origin overlap appear at negative scores on PC1 rich in furfural and pentanal (Fig. S13).

In the case of PCA on PARAFAC2 and MCR extracted features, it is also possible to observe the prevalence of not-italian samples at negative PC1 and PC2 values, which is not observable in PCA on VOCal features. Accordingly, by looking at the loadings (Fig. S13) it can be noticed that features distribution is almost coincident for PARAFAC2 and MCR and there are group of features located in each of the corresponding position of the clusters observed in scores space, while in the case of VOCal loadings at negative PC1 values there are very few features. In general, the Italian samples seem richer in volatile compounds as compared to not-italian.

Overall, these findings suggest that the three approaches yielded essentially the same samples distribution in PCA with slightly better separation when areas are obtained after resolution.

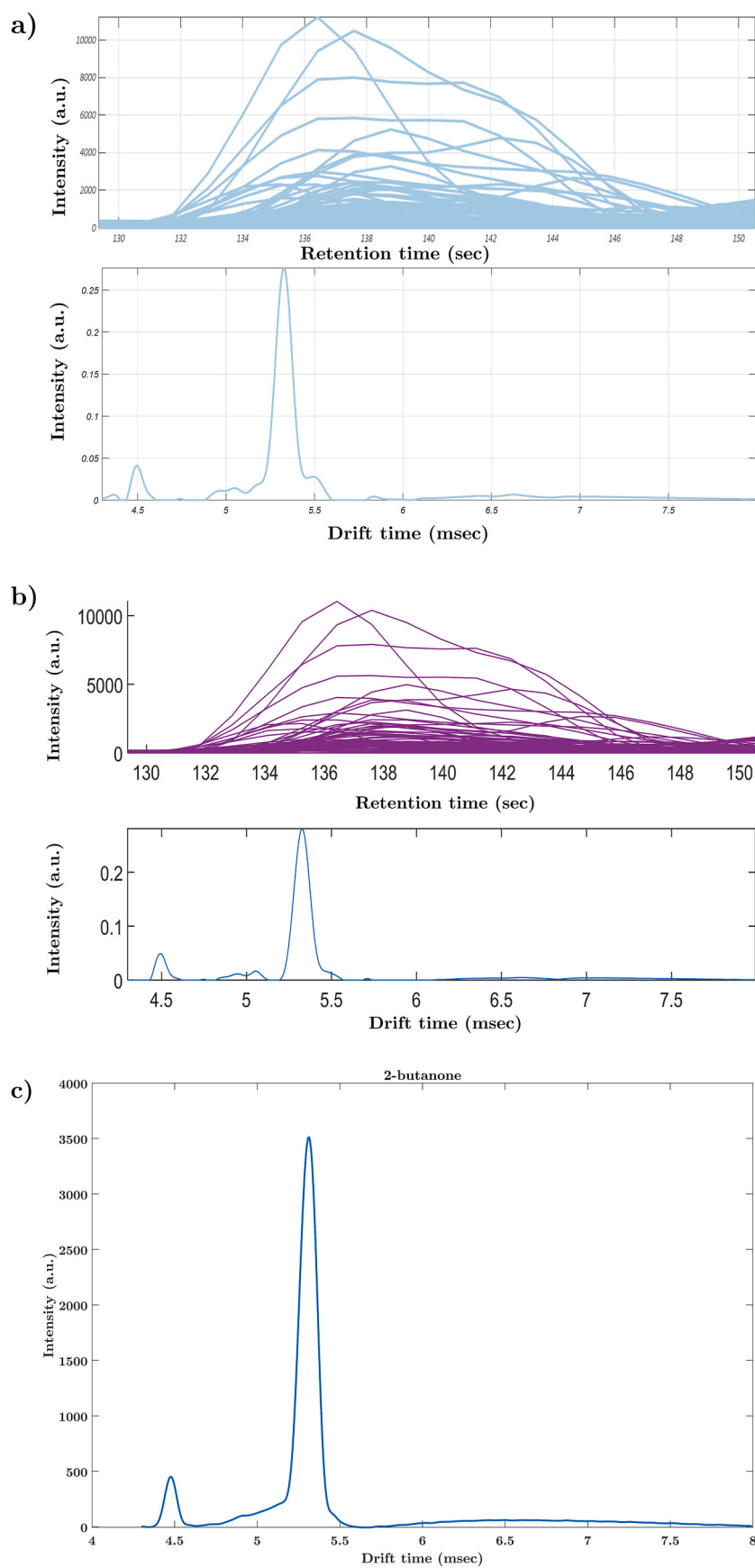


Fig. 7. Correct resolution of 2-butanone by PARAFAC2 (a) and MCR-ALS (b). The pure spectrum is also reported (c).

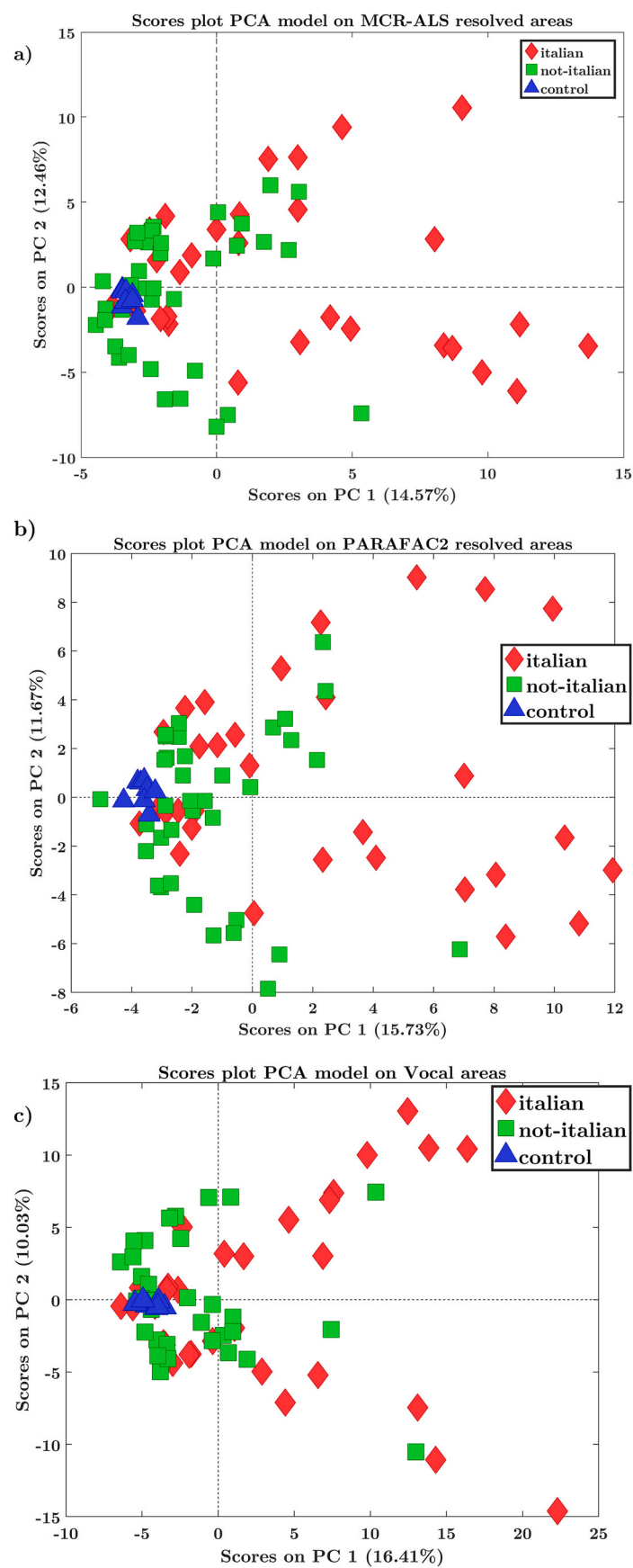


Fig. 8. Scores plot of the first two components of the PCA calculated on the areas resolved by MCR (a), PARAFAC2 (b) and VOCal (c).

To explore more in depth the samples' structure, the score plots of the PCA on PARAFAC2-resolved areas are presented and discussed in the Supplementary Materials, colored both by geographical and botanical origin (Fig. S17, Supplementary Materials). The scores plot indicates that the botanical origin is the predominant factor that accounts for the differences among the samples, in line with observations from the literature [60]. When considering subclasses, there is some evidence of differentiation between Italian and non-Italian polyfloral honeys, although these differences are relatively subtle and will require further validation in future work. The third component was also analysed, although no meaningful patterns or information emerged (Fig. S18, Supplementary Materials). Moreover, further components were examined but they were not informative (not reported). Taken together, the results imply that GC-IMS is a promising technique for assessing the botanical origin of honey, but additional studies with larger sample sets and more botanical/geographical diversity would help to confirm robustness and generalizability to geographical origin.

3.4. Summary considerations

Both PARAFAC2 and MCR-ALS proved particularly effective in detecting very low-intensity signals that might be missed with manual inspection or less sensitive methods. This capability is clearly illustrated in Fig. S19, Supplementary Materials, which demonstrates that even very weak peaks can be reliably resolved and potentially quantified using these approaches. As a result, the effective limit of detection (LOD) for trace or less abundant species could be reduced, similar to the improvements typically achieved with GC-MS.

In contrast, VOCal, although widely adopted, requires an extremely meticulous and time-consuming procedure. In theory, the procedure using the VOCal software could be semi-automated, meaning that a grid is constructed for one sample and then applied to all the others. This approach could significantly reduce processing time. However, it also presents some drawbacks, particularly when, as in the present study, samples of the same nature are analysed but may exhibit markedly different VOCs profiles, for example due to variations in botanical origin. For this reason, each sample must be visually inspected within the software interface, and the user is responsible for manually defining a grid of integration intervals that simultaneously accommodates all peaks in all samples. This manual step introduces operator dependence and represents a potential bottleneck when dealing with large datasets.

An additional outcome of this work is confirmation that blank samples are essential for obtaining meaningful results [20], regardless of the approach employed. The comparison of sample signals with those recorded in blanks allowed the exclusion of background contributions and ensured that only genuine analytical features were considered, thus avoiding the risk of interpreting artifacts as sample-specific peaks.

4. Conclusion

This study presents a comprehensive strategy for the untargeted analysis of GC-IMS signals, combining rigorous analytical design with advanced chemometrics tools. To ensure the reliability of the analytical data, multiple replicates of the control sample were acquired to verify the absence of instrumental drift. Since the instrument configuration used can operate only under isothermal conditions it was necessary to include two blank analyses between consecutive samples effectively minimising the memory effect commonly associated with these instruments.

A carefully validated preprocessing workflow was then applied, with particular attention to preserving data integrity throughout alignment, smoothing and baseline correction, notably considering the multiway/multiset approaches used for features extraction these preprocessing was applied only to the drift time dimension. This step proved essential for generating consistent and reproducible inputs for peak resolution. The preprocessed datasets were analysed with PARAFAC2, and MCR-

ALS adopting an intervals strategy to enhance resolution capability. The results showed that these methodologies enabled efficient deconvolution of over-lapping peaks and extraction of very low-intensity signals. The resulting area matrices were straightforward to handle, demonstrating the power of these algorithms in managing complex GC-IMS data.

Overall, this work shows that, when an adequate preprocessing pipeline is implemented, GC-IMS data can be processed with a degree of automation comparable to that achieved in GC-MS workflows by using PARADISE. Based on these results chemometric approaches such as MCR-ALS and PARAFAC2 seem to offer a robust alternative to manual methods like VOCal, which are inherently labor-intensive and operator-dependent. Nevertheless, key steps such as interval definition, component selection and validation still require expert supervision to ensure meaningful and chemically sound outcomes. However, the PARADISE software already provides sound suggestions to individuate components corresponding to peaks (AI based [61]) and some proposal towards automatic interval selection is in progress [31]. Finally, the dataset of this study is limited and unbalanced with respect to both botanical and geographical origin. In order to provide a more comprehensive comparison of the three approaches by means of classification models and metrics, it will be necessary to increase the number of samples.

Future developments will focus on standardizing pre-processing and peak resolution procedures and integrating them into user-friendly, fully automated pipelines. Furthermore, the proposed workflow will be further validated through the analysis of authentic reference standards, allowing direct comparison between pure compound spectra and the spectra resolved by PARAFAC2 and MCR, thus providing an additional level of confidence in compound identification.

CRedit authorship contribution statement

Samuele Pellacani: Writing – original draft, Methodology, Formal analysis, Data curation. **Lorenzo Strani:** Validation, Software, Formal analysis. **Daniele Tanzilli:** Formal analysis, Data curation. **Rebecca Barbieri:** Visualization, Formal analysis. **Raniero Marabottini:** Visualization, Resources. **Francesca Martone:** Validation, Methodology, Formal analysis. **Angela Napolitano:** Validation, Resources. **Marina Cocchi:** Writing – original draft, Validation, Supervision, Data curation, Conceptualization. **Caterina Durante:** Writing – original draft, Validation, Supervision, Project administration, Funding acquisition, Conceptualization.

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.

Acknowledgements

The authors would like to acknowledge Noemi Coppola for helping with the GC-IMS analysis of the honey samples used in this work. This research was supported by University of Modena and Reggio Emilia - Modena Foundation through FAR Mission Oriented 2024 funds.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.aca.2026.346326>.

Data availability

Data will be made available on request.

References

- [1] H. Parastar, P. Weller, Benchtop volatilities supercharged: how machine learning based design of experiment helps optimizing untargeted GC-IMS gas phase metabolomics, *Talanta* 272 (2024) 125788, <https://doi.org/10.1016/j.talanta.2024.125788>.
- [2] H. Parastar, P. Weller, Towards greener volatilities: is GC-IMS the new Swiss army knife of gas phase analysis? *TrAC, Trends Anal. Chem.* 170 (2024) 117438 <https://doi.org/10.1016/j.trac.2023.117438>.
- [3] M. del M. Contreras, N. Jurado-Campos, C. Sánchez-Carnerero Callado, N. Arroyo-Manzanares, L. Fernández, S. Casano, S. Marco, L. Arce, C. Ferreira-Vera, Thermal desorption-ion mobility spectrometry: a rapid sensor for the detection of cannabinoids and discrimination of *Cannabis sativa* L. chemotypes, *Sensor. Actuatur. B Chem.* 273 (2018) 1413–1424, <https://doi.org/10.1016/j.snb.2018.07.031>.
- [4] R. Freire, L. Fernandez, C. Mallafré-Muro, A. Martín-Gómez, F. Madrid-Gambin, L. Oliveira, A. Pardo, L. Arce, S. Marco, Full workflows for the analysis of gas chromatography–ion mobility spectrometry in foodomics: application to the analysis of Iberian ham Aroma, *Sensors* 21 (2021) 6156, <https://doi.org/10.3390/s21186156>.
- [5] N. Gerhardt, M. Birkenmeier, S. Schwolow, S. Rohn, P. Weller, Volatile-compound fingerprinting by headspace-gas-chromatography ion-mobility spectrometry (HS-GC-IMS) as a benchtop alternative to 1H NMR profiling for assessment of the authenticity of honey, *Anal. Chem.* 90 (2018) 1777–1785, <https://doi.org/10.1021/acs.analchem.7b03748>.
- [6] J. Christmann, S. Rohn, P. Weller, gc-ims-tools – a new python package for chemometric analysis of GC-IMS data, *Food Chem.* 394 (2022) 133476, <https://doi.org/10.1016/j.foodchem.2022.133476>.
- [7] Y. Yufei, L. Pan, C. Jia, W. Guofei, R. Hongqiang, W. Ling, L. Hongzhao, Z. Wenxian, Z. Tao, Using untargeted metabolomics and GC-IMS to analysis the influence of fat distribution on the flavor formation of bacon, *Food Biosci.* 59 (2024) 103986, <https://doi.org/10.1016/j.fbio.2024.103986>.
- [8] H. Parastar, H. Yazdanpanah, P. Weller, Non-targeted volatilities for the authentication of saffron by gas chromatography-ion mobility spectrometry and multivariate curve resolution, *Food Chem.* 465 (2025) 142074, <https://doi.org/10.1016/j.foodchem.2024.142074>.
- [9] K. Marcinkevicius, G. Gennari, V. Salomón, N. Vera, L. Maldonado, Detection of adulterations in native stingless bees honey from Argentina using UV–Vis spectroscopy coupled with chemometrics, *J. Food Meas. Char.* 18 (2024) 7283–7294, <https://doi.org/10.1007/s11694-024-02653-0>.
- [10] W. Yao, Y. Cai, D. Liu, Z. Zhao, Z. Zhang, S. Ma, M. Zhang, H. Zhang, Comparative analysis of characteristic volatile compounds in Chinese traditional smoked chicken (specialty poultry products) from different regions by headspace–gas chromatography–ion mobility spectrometry, *Poult. Sci.* 99 (2020) 7192–7201, <https://doi.org/10.1016/j.psj.2020.09.011>.
- [11] K. Lu, L. Liu, J. Zi, L. Song, W. Xie, New insights from flavoromics on different heating methods of traditional fermented shrimp paste: the volatile components and metabolic pathways, *Lebensm. Wiss. Technol.* 168 (2022) 113880, <https://doi.org/10.1016/j.lwt.2022.113880>.
- [12] X. Yang, R. Jiao, X. Zhu, S. Zhao, G. Liao, J. Yu, D. Wang, Profiling and characterization of odorous volatile compounds from the industrial fermentation of erythromycin, *Environ. Pollut.* 255 (2019) 113130, <https://doi.org/10.1016/j.envpol.2019.113130>.
- [13] M.J. Aliano-González, M. Ferreira-González, G.F. Barbero, M. Palma, C.G. Barroso, Application of headspace gas chromatography-ion mobility spectrometry for the determination of ignitable liquids from fire debris, *Separations* 5 (2018) 41, <https://doi.org/10.3390/separations5030041>.
- [14] G.W. Cook, P.T. LaPuma, G.L. Hook, B.A. Eckenrode, Using gas chromatography with ion mobility spectrometry to resolve explosive compounds in the presence of interferents, *J. Forensic Sci.* 55 (2010) 1582–1591, <https://doi.org/10.1111/j.1556-4029.2010.01522.x>.
- [15] C. Li, S. Tian, J. You, J. Liu, E. Li, C. Wang, Q. Wang, Z. Zhu, D. Fan, R. Tian, Qualitative determination of volatile substances in different flavored cigarette paper by using headspace-gas chromatography-ion mobility spectrometry (HS-GC-IMS) combined with chemometrics, *Heliyon* 9 (2023) e12146, <https://doi.org/10.1016/j.heliyon.2022.e12146>.
- [16] S. Wang, H. Chen, B. Sun, Recent progress in food flavor analysis using gas chromatography–ion mobility spectrometry (GC-IMS), *Food Chem.* 315 (2020) 126158, <https://doi.org/10.1016/j.foodchem.2019.126158>.
- [17] C. Mallafré-Muro, M. Cruz, A.B.I. Borrego, L.F. Romero, A.P. Martínez, S.M. Colás, Study of quality controls for stability check of the ROIs of a ketones mixture in different GC-IMS measurement campaigns, in: 2022 IEEE International Symposium on Olfaction and Electronic Nose (ISOEN), 2022, pp. 1–4, <https://doi.org/10.1109/ISOEN54820.2022.9789683>.
- [18] R. Brendel, S. Schwolow, S. Rohn, P. Weller, Volatilomic profiling of citrus juices by dual-detection HS-GC-MS-IMS and machine learning—an alternative authentication approach, *J. Agric. Food Chem.* 69 (2021) 1727–1738, <https://doi.org/10.1021/acs.jafc.0c07447>.
- [19] D. Geng, X. Chen, J. Jin, D. Lu, Y. Huo, B. Chen, Optimization of GC-IMS parameters and determination of volatile fingerprint of flavors and fragrances using GC-IMS, *CyTA - J. Food* 22 (2024) 2386422, <https://doi.org/10.1080/19476337.2024.2386422>.
- [20] N. Jurado-Campos, A. Martín-Gómez, D. Saavedra, L. Arce, Usage considerations for headspace-gas chromatography-ion mobility spectrometry as a suitable technique for qualitative analysis in a routine lab, *J. Chromatogr. A* 1640 (2021), <https://doi.org/10.1016/j.chroma.2021.461937>.
- [21] E. Szymańska, A.N. Davies, L.M.C. Buydens, Chemometrics for ion mobility spectrometry data: recent advances and future prospects, *Analyst* 141 (2016) 5689–5708, <https://doi.org/10.1039/C6AN01008C>.
- [22] Y. Wang, L. Jiang, D. An, Y. Zhao, Y. He, F. Wu, H. Zhao, X. Lu, H. Lu, Analysis of volatile flavor changes of sea urchin gonads and phospholipids after heating based on GC-IMS, *J. Food Process. Preserv.* 46 (2022) e17107, <https://doi.org/10.1111/jfpp.17107>.
- [23] L. Peng, X. Wang, M. He, X. Sha, Z. Dou, L. Xiao, W. Li, Discrimination and screening of volatile metabolites in *atractylodis rhizoma* from different varieties using headspace solid-phase microextraction-gas chromatography-mass spectrometry and headspace gas chromatography-ion mobility spectrometry, and ultra-fast gas chromatography electronic nose, *J. Chromatogr. A* 1725 (2024) 464931, <https://doi.org/10.1016/j.chroma.2024.464931>.
- [24] S. Oller-Moreno, C. Mallafré-Muro, L. Fernandez, E. Caballero, A. Blanco, J. Gumà, S. Marco, A. Pardo, GCIMS: an R package for untargeted gas chromatography – ion mobility spectrometry data processing, *Chemometr. Intell. Lab. Syst.* 241 (2023) 104938, <https://doi.org/10.1016/j.chemolab.2023.104938>.
- [25] A. de Juan, J. Jaumot, R. Tauler, Multivariate curve Resolution (MCR). Solving the mixture analysis problem, *Anal. Methods* 6 (2014) 4964–4976, <https://doi.org/10.1039/C4AY00571F>.
- [26] A. Smolinska, S. Pellacani, M. Skawinski, C. Durante, On the road to automation: a comparative review on chemometric strategies for GC-MS data analysis, *TrAC, Trends Anal. Chem.* 190 (2025) 118286, <https://doi.org/10.1016/j.trac.2025.118286>.
- [27] J.M. Amigo, T. Skov, R. Bro, J. Coello, S. Maspocho, Solving GC-MS problems with PARAFAC2, *TrAC, Trends Anal. Chem.* 27 (2008) 714–725, <https://doi.org/10.1016/j.trac.2008.05.011>.
- [28] R.A. Harshman, PARAFAC2: Mathematical and Technical Notes, UCLA Working Papers in Phonetics, vol. 22, University Microfilms, Ann Arbor, 1972, pp. 30–44. No. 10,085, <https://psychology.uwo.ca/faculty/harshman/wppp2.pdf>.
- [29] J.E. Cohen, R. Bro, Nonnegative PARAFAC2: a flexible coupling approach. In: Deville Y., Gannot S., Mason R., Plumley M., Ward D. (eds) *Latent Variable Analysis and Signal Separation. LVA/ICA 2018. Lecture Notes in Computer Science*, vol. 10891. Springer.
- [30] Marie Roald, Carla Schenker, Vince D. Calhoun, Tülay Adalı, Rasmus Bro, Jeremy E. Cohen, Evrim Acar, et al., An AO-ADMM approach to constraining PARAFAC2 on all modes, *SIAM Journal on Mathematics of Data Science* 4 (3) (2022) 1191–1222, <https://doi.org/10.48550/arXiv.2110.01278>.
- [31] G. Baccolo, B. Quintanilla-Casas, S. Vichi, D. Augustijn, R. Bro, From untargeted chemical profiling to peak tables – a fully automated AI driven approach to untargeted GC-MS, *TrAC, Trends Anal. Chem.* 145 (2021) 116451, <https://doi.org/10.1016/j.trac.2021.116451>.
- [32] G. Baccolo, H. Yu, C. Valsecchi, D. Ballabio, R. Bro, Comparison of machine learning approaches for the classification of elution profiles, *Chemometr. Intell. Lab. Syst.* 243 (2023) 105002, <https://doi.org/10.1016/j.chemolab.2023.105002>.
- [33] T. Skov, R. Bro, Solving fundamental problems in chromatographic analysis, *Anal. Bioanal. Chem.* 390 (1) (2008) 281–285, <https://doi.org/10.1007/S00216-007-1618-z>.
- [34] P.A. Schneide, N.B. Gallagher, R. Bro, Shift-invariant tri-linearity—A new model for resolving untargeted gas chromatography coupled mass spectrometry data, *J. Chemometr.* 37 (2023) e3501, <https://doi.org/10.1002/cem.3501>.
- [35] P.A. Schneide, N.B. Gallagher, R. Bro, Modelling shifts and shape changes in gas-chromatography coupled mass spectrometry, *Chemometr. Intell. Lab. Syst.* 251 (2024) 105155, <https://doi.org/10.1016/j.chemolab.2024.105155>.
- [36] N. Cavallini, F. Savorani, R. Bro, M. Cocchi, Beer's linguistics and chemistry: an investigation opening new research perspectives, *Chemometr. Intell. Lab. Syst.* 267 (2025) 105521, <https://doi.org/10.1016/j.chemolab.2025.105521>.
- [37] N. Cavallini, L. Strani, P.P. Becchi, V. Pizzamiglio, S. Michellini, F. Savorani, M. Cocchi, C. Durante, Tracing the identity of Parmigiano Reggiano “Prodotto di Montagna - progetto Territorio” cheese using NMR spectroscopy and multivariate data analysis, *Anal. Chim. Acta* 1278 (2023) 341761, <https://doi.org/10.1016/j.aca.2023.341761>.
- [38] X. Shen, H. Wang, L. Yao, M. Sun, C. Yu, T. Feng, W. Kang, Volatile compounds analysis and sensory profiling of Xinjiang ili different floral source honey via GC-IMS, GC-IMS and GC × GC-TOF MS, *J. Food Sci.* 90 (2025) e70750, <https://doi.org/10.1111/1750-3841.70750>.
- [39] N. Arroyo-Manzanares, M. García-Nicolás, A. Castell, N. Campillo, P. Viñas, I. López-García, M. Hernández-Córdoba, Untargeted headspace gas chromatography – ion mobility spectrometry analysis for detection of adulterated honey, *Talanta* 205 (2019) 120123, <https://doi.org/10.1016/j.talanta.2019.120123>.
- [40] D. Feng, J. Wang, Y. He, X. Ji, H. Tang, Y. Dong, W. Yan, HS-GC-IMS detection of volatile organic compounds in Acacia honey powders under vacuum belt drying at different temperatures, *Food Sci. Nutr.* 9 (2021) 4085–4093, <https://doi.org/10.1002/fsn3.2364>.
- [41] X. Wang, S. Yang, J. He, L. Chen, J. Zhang, Y. Jin, J. Zhou, Y. Zhang, A green triple-locked strategy based on volatile-compound imaging, chemometrics, and markers to discriminate winter honey and sapium honey using headspace gas chromatography-ion mobility spectrometry, *Food Res. Int.* 119 (2019) 960–967, <https://doi.org/10.1016/j.foodres.2019.01.004>.
- [42] P. Leng, H.-W. Hu, A.-H. Cui, H.-J. Tang, Y.-G. Liu, HS-GC-IMS with PCA to analyze volatile flavor compounds of honey peach packaged with different preservation methods during storage, *Lebensm. Wiss. Technol.* 149 (2021) 111963, <https://doi.org/10.1016/j.lwt.2021.111963>.

- [43] Abraham Savitzky, M.J.E. Golay, Smoothing and differentiation of data by simplified least squares procedures, *Anal. Chem.* 36 (1964) 1627–1639, <https://doi.org/10.1021/ac60214a047>.
- [44] P. Eilers, H. Boelens, Baseline Correction with Asymmetric Least Squares Smoothing, Report Leiden University Medical Centre, The Netherlands, 2005. Available online: <https://prod-dcd-datasets-public-files-eu-west-1.s3.eu-west-1.amazonaws.com/dd7c1919-302c-4ba0-8f88-8aa61e86bb9d>. (Accessed 3 August 2026) (last accessed on).
- [45] B. Quintanilla-Casas, R. Bro, J.L. Hinrich, C.L. Davie-Martin, Tutorial on Paradise: Parafac2-Based Deconvolution and Identification System for Processing GC–MS Data, *PROTOCOL*, 2023, <https://doi.org/10.21203/rs.3.pex-2143/v1>.
- [46] 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>.
- [47] A. de Juan, R. Tauler, Multivariate curve resolution: 50 years addressing the mixture analysis problem – a review, *Anal. Chim. Acta* 1145 (2021) 59–78, <https://doi.org/10.1016/j.aca.2020.10.051>.
- [48] R. Bro, C.A. Andersson, H.A.L. Kiers, PARAFAC2—Part II. Modeling chromatographic data with retention time shifts, *J. Chemometr.* 13 (1999) 295–309, [https://doi.org/10.1002/\(SICI\)1099-128X\(199905/08\)13:3/4%3C295::AID-CEM547%3E3.0.CO;2-Y](https://doi.org/10.1002/(SICI)1099-128X(199905/08)13:3/4%3C295::AID-CEM547%3E3.0.CO;2-Y).
- [49] M.H. Kamstrup-Nielsen, L.G. Johnsen, R. Bro, Core consistency diagnostic in PARAFAC2, *J. Chemometr.* 27 (2013) 99–105, <https://doi.org/10.1002/cem.2497>.
- [50] A.C. Olivieri, Hazards of using multivariate curve resolution for processing first-order spectral data. A rotational ambiguity analysis, *Anal. Chim. Acta* 1367 (2025) 344304, <https://doi.org/10.1016/j.aca.2025.344304>.
- [51] A.C. Olivieri, Estimating the boundaries of the feasible profiles in the bilinear decomposition of multi-component data matrices, *Chemometr. Intell. Lab. Syst.* 216 (2021) 104387, <https://doi.org/10.1016/j.chemolab.2021.104387>.
- [52] H. Hotelling, Analysis of a complex of statistical variables into principal components, *J. Educ. Psychol.* 24 (1933) 417–441, <https://doi.org/10.1037/h0071325>.
- [53] R. Bro, A.K. Smilde, Principal component analysis, *Anal. Methods* 6 (9) (2014) 2812–2831, <https://doi.org/10.1039/c3ay41907j>.
- [54] M. Li Vigni, C. Durante, M. Cocchi, Chapter 3 - exploratory data analysis, in: F. Marini (Ed.), *Data Handling in Science and Technology*, vol. 28, Elsevier, 2013, pp. 55–126, <https://doi.org/10.1016/B978-0-444-59528-7.00003-X>.
- [55] S. Pellacani, M. Cocchi, C. Durante, L. Strani, Exploring the effect of different storage conditions on the aroma profile of bread by using Arrow-SPME GC-MS and chemometrics, *Molecules* 28 (2023), <https://doi.org/10.3390/molecules28083587>.
- [56] S. Pellacani, C. Durante, S. Celli, M. Mariani, A. Marchetti, M. Cocchi, L. Strani, Optimization of an analytical method based on SPME-Arrow and chemometrics for the characterization of the aroma profile of commercial bread, *Chemometr. Intell. Lab. Syst.* 241 (2023) 104940, <https://doi.org/10.1016/j.chemolab.2023.104940>.
- [57] L. Bodenbender, S. Rohn, H. Parastar, K. Sinderhauf-Gacioch, P. Weller, Towards “greener” strategies in quality control: rapid volatilomics of cocoa based on HS-GC-IMS and machine learning, *Anal. Bioanal. Chem.* (2026), <https://doi.org/10.1007/s00216-026-06415-3>.
- [58] C. Capitain, P. Weller, Non-Targeted screening approaches for profiling of volatile organic compounds based on gas chromatography-ion mobility spectroscopy (GC-IMS) and machine learning, *Molecules* 26 (2021) 5457, <https://doi.org/10.3390/molecules26185457>.
- [59] Ö.C. Zeki, C.C. Eylem, E. Nemutlu, Optimization of GC-MS run time for untargeted metabolomics: trade-offs between speed, coverage, and repeatability, *J. Pharmaceut. Biomed. Anal.* 266 (2025) 117068, <https://doi.org/10.1016/j.jpba.2025.117068>.
- [60] S. Schwolow, N. Gerhardt, S. Rohn, P. Weller, Data fusion of GC-IMS data and FT-MIR spectra for the authentication of olive oils and honeys—is it worth to go the extra mile? *Anal. Bioanal. Chem.* 411 (2019) 6005–6019, <https://doi.org/10.1007/s00216-019-01978-w>.
- [61] A.B. Risum, R. Bro, Using deep learning to evaluate peaks in chromatographic data, *Talanta* 204 (2019) 255–260, <https://doi.org/10.1016/j.talanta.2019.05.053>.