[1] Research and Innovation Center, Fondazione Edmund Mach, San Michele aA (TN) Italy,
[2] Center Agriculture Food Environment, University of Trento, S. Michele aA (TN), Italy

Mazzucotelli M.[1,2], Franceschi P. [1]

Automated untargeted peak detection for GC-IMS data

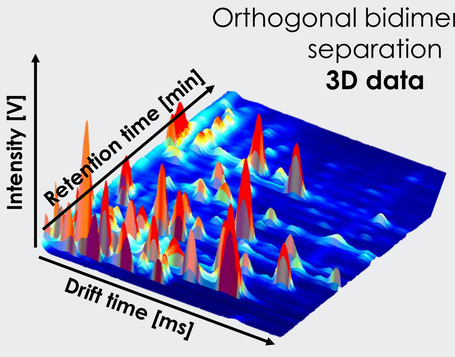
• Introduction

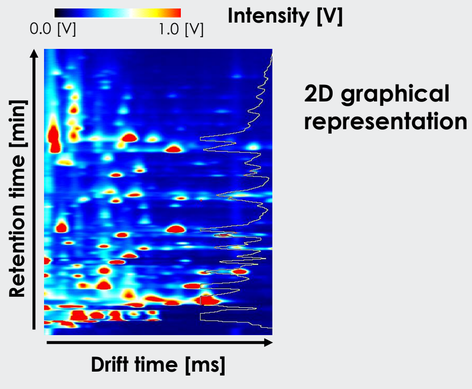
Gas Chromatography (GC) coupled with Ion Mobility Spectrometry (IMS) is an analytical technique that has been rapidly gaining popularity in the field of food volatilomics[1].

The formation of multiple ionized species from a single analyte significantly complicates the interpretability of the analytical output, and the mobility coefficient of the detected ions (even if combined with the chromatographic retention index) is not sufficient for a reliable peak identification, which requires the comparison with pure standards or using an additional MS detector. For these reasons, **GC-IMS is mostly used as a tool for untargeted fingerprinting analysis.**

GC-IMS data

The hyphenation of GC and IMS results in an **orthogonal bidimensional separation of analytes**. The analytical output of GC-IMS analysis is a tridimensional data, and its common graphic representation is as 2D plots with the signal intensity visualized as color scale.





The most common approach for **peak detection** is based on the **manual selection of the 2D peaks** by creating a set of retention time and drift time coordinates, which are subsequently used to extract peak intensities and generate a peak table for the entire sample set.

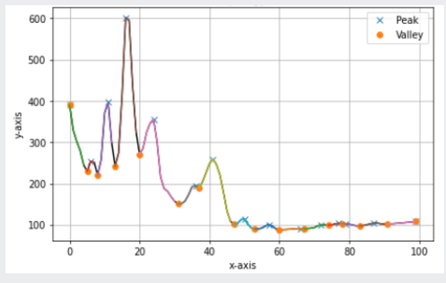
This is suitable as part of a targeted strategy when a set of analytes of interest are monitored. On the opposite, in case of untargeted fingerprinting studies this approach become **time-consuming**, and it is likely to be **influenced by the operator's personal judgment.**

Advancements in GC-IMS data processing

The development of automated workflows for GC-IMS data processing is a research subject of increasing interest, and different strategies have been proposed.

These strategies can be classified into two approaches:
- processing of the entire spectral fingerprint obtained for each sample [2,3]
- peak detection to obtain a peak table for each sample[4]

For food matrices with a complex volatile profile, the approach based on peak detection is a promising solution. Recently, Oller-Moreno, S. *et al.*[4] proposed the application of wavelet transform for GC-IMS peak detection. Methods based on image processing and topological data analysis represent viable alternatives, that are worth evaluating.



Persistent homology[5,6]

Persistent homology is a tool for topological data analysis and found application in many fields.

It is a method to identify relative maxima and to quantify their significance, which is expressed as persistence.

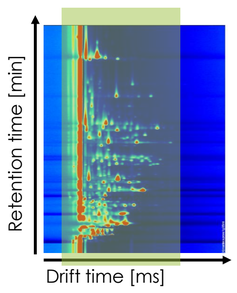
This nonparametric approach enables the detection of peaks without requiring assumptions about their shape or width. Persistent homology can be applied to both 1D and 2D peaks.

• Aim

Development of an automated peak detection workflow for GC-IMS data based on persistent homology and its application to roasted hazelnut paste samples.

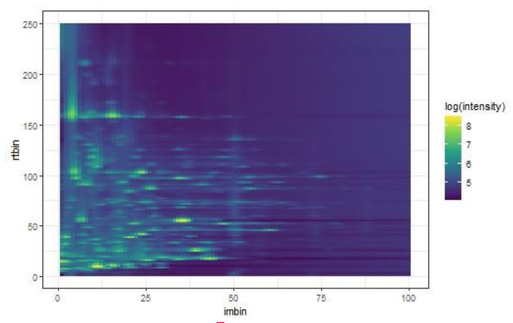
• Experimental

1 Automated untargeted data processing



1) Selection of the region with peaks

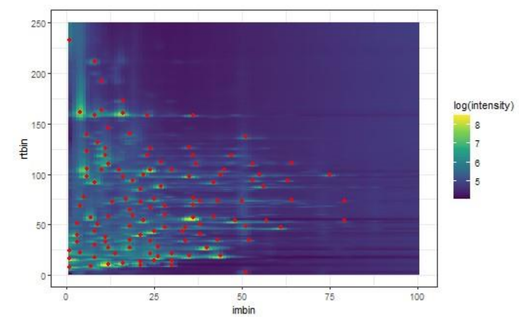
2) Binning of the bidimensional matrix



Python package *findpeaks* [7]

3) Denoising (median filter)

4) Peak detection based on persistent homology



The R package *reticulate*[8] was used to integrate Python code and R code

110-120 peaks/samples

The intensities of the automatically detected peaks corresponding to the target peaks were extracted for each sample.

2 Manual targeted data processing

To validate the automated peak detection, a set of **identified target peaks** were manually selected in the GC-IMS profile, and their signal intensities were extracted with the **quantitative module** of the commercial software VOCal (G.A.S., Dortmund, Germany).

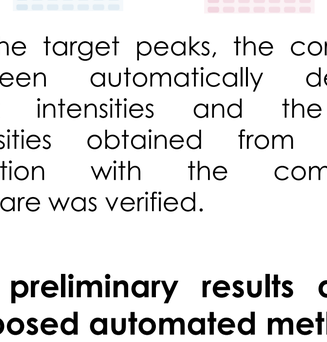
Peak height (above area min) was selected as the intensity measure.

The intensities of the manually detected peaks were extracted for each sample.

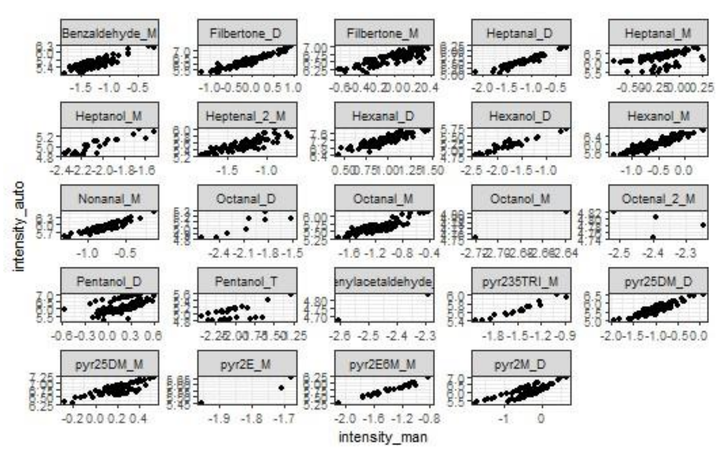
3 Intensity correlation assessment

Target peaks

peak intensities manual vs **peak intensities auto**



For the target peaks, the correlation between automatically detected peak intensities and the peak intensities obtained from manual selection with the commercial software was verified.



The preliminary results demonstrated the effectiveness and reliability of the proposed automated method.

• Conclusion

An automated peak detection workflow for GC-IMS data was developed using persistent homology. It was evaluated by processing the GC-IMS profiles of roasted hazelnut paste samples from different geographical and botanical origins.
The combination of a rapid analytical technique with an automated data processing strategy removes the bottleneck of manual peak selection and provides a promising analytical solution for food volatilomics studies.

Instrumental set-up

- GC-IMS (FlavourSpec GC-IMS system (3H-IMS) (G.A.S., Dortmund, Germany))
MXT-Wax column 30 m, 0.53 mm dc, 0.5 µm df (Restek Corporation, Bellefonte, US). Injector and transfer line (inj-oven): 80°C; GC and transfer line (oven-IMS): 60°C, Carrier gas: nitrogen. GC column flow program: 2 ml/min for 6 min, from then increased up to 12 ml/min at 16 min, up to 50 ml/min at 19.5 min, up to 75 ml/min at 22.5 min, up to 124 ml/min at 27 min, up to 150 ml/min at 27 min, ending with 3 min at 150 ml/min. The total GC runtime was 30 min.
IMS 45°C. Drift gas: nitrogen. IMS drift flow: 150 mL/min. Positive ionization mode. Averaging 6.

- SHS (HT2000H headspace autosampler (HTA, Brescia, Italy)).
Sample incubation 60°C - 20 min under constant agitation. HS syringe temperature 80°C. Injection volume 0.5 mL.

Instrument control and data acquisition were performed with the Sequence Designer software (G.A.S., Dortmund, Germany).

Samples

111 roasted hazelnut paste samples were obtained by processing kernels from different geographical and botanical origins.

Acknowledgements

SOREMARTEC

Bibliography

1. Gu, S., Zhang, J., Wang, J., Wang, X. & Du, D. Recent development of HS-GC-IMS technology in rapid and non-destructive detection of quality and contamination in agri-food products. *TrAC* vol. 144 (2021).
2. Freire, R. *et al.* Full workflows for the analysis of gas chromatography—ion mobility spectrometry in foodomics: Application to the analysis of iberian ham aroma. *Sensors* **21**, (2021).
3. Christmann, J., Rohn, S. & Weller, P. gc-ims-tools – A new Python package for chemometric analysis of GC–IMS data. *Food Chem.* **394**, 133476 (2022).
4. Oller-Moreno, S. *et al.* GCIMS: An R package for untargeted gas chromatography – Ion mobility spectrometry data processing. *Chemom. Intell. Lab. Syst.* **241**, (2023).
5. Huber, S. Persistent Homology in Data Science. *Data Sci. – Anal. Appl.* 81–88 (2021)
6. Fugacci, U., Scaramuccia, S., Iuricich, F. & de Floriani, L. Persistent homology: A step-by-step introduction for newcomers. *Ital. Chapter Conf. 2016 - Smart Tools Apps Comput. Graph. STAG 2016* 1–10 (2016)
7. Taskesen, Erdogan. (2023). findpeaks is for the detection of peaks and valleys in a 1D vector and 2D array (image). (2.5.1). Zenodo. (https://doi.org/10.5281/zenodo.8145101)
8. Ushey K, Allaire J, Tang Y (2023). reticulate: Interface to 'Python'. https://rstudio.github.io/reticulate/, https://github.com/rstudio/reticulate.