Finnee 2024 – a Matlab toolbox for untargeted analysis of LC-HRMS datasets with MS1 scans acquired as continuous scans.

Guillaume Laurent Erny[a,b]

[a] Associate Laboratory i4HB - Institute for Health and Bioeconomy, University Institute of Health Sciences - CESPU, 4585-116 Gandra, Portugal.
[b] UCIBIO - Applied Molecular Biosciences Unit, Translational Toxicology Research Laboratory, University Institute of Health Sciences (1H-TOXRUN, IUCS-CESPU), 4585-116 Gandra, Portugal.

*Corresponding author: Dr. Guillaume Erny (guillaume.erny@outlook.com)

Metabolite identification and quantification in biological samples are crucial for understanding biochemical processes, disease mechanisms, and biomarker discovery. While targeted metabolomics focuses on specific compounds, untargeted metabolomics comprehensively assesses the entire metabolome. Liquid chromatography hyphenated with high-resolution mass spectrometry (LC-HRMS) has become a powerful tool for untargeted metabolomics due to its ability to detect various chemicals. LC-HRMS combines liquid chromatography's separation capabilities with mass spectrometry's high-resolution mass analysis. MS1 spectra, acquired in LC-HRMS, provide detailed mass-to-charge ratio information for detected ions, enabling simultaneous profiling of metabolites without prior knowledge of their structures or retention times. Pre-processing steps and feature detection algorithms are applied to raw MS1 spectra, facilitating downstream analysis, including statistical analysis, metabolite annotation, and pathway mapping. The choice between profile and centroid scans for MS1 data acquisition entails trade-offs in resolution, data complexity, and processing requirements. Several software tools have been developed to handle LC-HRMS data, providing functionalities for pre-processing, peak detection, metabolite identification, and statistical analysis. This article presents Finnee2024, a Matlab toolbox for analysing profile scan-based MS1 data. The toolbox allows good coverage, low false positives, and improved visualisation and control of extracted features. It offers enhancements over previous versions, facilitating faster calculations and leveraging the advantages of profile scans. The article highlights the changes in Finnee2024 using datasets from a comparative study of software packages, demonstrating its performance and usability.

Introduction

Identifying and quantifying metabolites in biological samples is essential for understanding underlying biochemical processes, elucidating disease mechanisms, and discovering potential biomarkers.[1,2] Targeted metabolomics approaches focus on analysing known or expected chemicals, limiting the scope to pre-selected compounds. On the other hand, untargeted metabolomics provides an unbiased and comprehensive assessment of the entire metabolome, offering the opportunity to discover novel biological pathways and explore the intricate interactions within biological systems.[3,4]

In recent years, liquid chromatography hyphenated high-resolution mass spectrometry (LC-HRMS) has emerged as a powerful tool for untargeted metabolomics due to its ability to separate and detect a wide range of chemicals in complex matrices.[5] LC-HRMS combines the chromatographic separation capabilities of liquid chromatography (LC) with the high-resolution mass analysis provided by mass spectrometry (HRMS). HRMS instruments offer superior mass accuracy, resolution, and sensitivity, enabling the detection of metabolites at trace levels.[6] One of the approaches to LC-HRMS-based untargeted metabolomics is the acquisition of MS1 (or full scan) spectra, which provide detailed mass-to-charge ratio (m/z) information for all ions detected within a given mass range. Collecting MS1 spectra makes it possible to simultaneously profile a wide range of metabolites without prior knowledge of their structures or retention times. MS1 spectra can be processed using sophisticated data analysis tools to extract valuable information about the detected metabolites' presence, abundance and potential identity. Compared to MSn acquisition modes, MS1 offers increased coverage, sensitivity, and simplified data interpretation and analysis.[7] However, MSn provides better structural information and specific patterns, allowing for better identification.

Analysis of such data is performed using computerised tools. After data acquisition, the raw MS1 spectra are subjected to pre-processing steps, including noise removal, baseline correction and peak alignment, to improve data quality and comparability.[8] Feature detection algorithms are then applied to identify metabolite peaks in the spectra and generate a list of detected features represented by their respective m/z values, retention times and intensities. This feature list forms the basis for downstream data analysis, including statistical analysis, metabolite annotation and pathway mapping.[9,10] While MS1 scans are the starting objects, these can be either profile scans or centroid scans. Each approach has advantages and considerations that are important to understand for practical untargeted metabolomics analysis. Profile scans, also known as continuum scans, record the entire mass spectrum, capturing detailed information about the intensity distributions of the multiple ions separated at a given time in the separation by the mass analyser (ion peak) across the *m/z* range. In profile scans, each ion peak is described by multiple points, providing a high-resolution representation of the separation in the mass analyser

at any time. The advantages of profile scans for untargeted analysis include high mass accuracy, fine resolution, and improved quantification.[11] However, the scan files are larger than centroid scans and more complex to analyse.

Centroid scans, also known as centroided or peak-picked scans, convert each ion peak in the MS1 scans into discrete data points representing each peak's centroid (centre). Centroid scans record only the m/z positions and intensity information of each detected MS peak, drastically reducing the file size and allowing faster data processing. However, there is potentially a substantial loss of information, with peak overlap becoming undetectable.[12] Most instrument software can export MS1 data as either centroid or profile scans. Freeware can also convert proprietary files into profile or centroid scans in the open mzML format.[13,14]

Several freeware tools have been developed to facilitate untargeted analysis of LC-HRMS data with MS1 or MSn scans. XCMS,[15] MZmine,[16] OpenMS,[17] MSdial[18] and MetaboAnalyst[19] are among the most widely used software frameworks providing comprehensive functionalities for pre-processing, peak detection, metabolite identification and statistical analysis of untargeted metabolomics data. These tools have proven invaluable resources for researchers to unravel complex metabolic networks and identify potential biomarkers. It should be emphasised that while centroid and profile scans can be used, with most software, the first step will be to convert every profile scan to centroid.

This article aims to present the latest version of Finnee, a Matlab toolbox for analysing MS1 scans acquired by LC-HRMS.[20] While Finnee offers similar functionalities to other freeware, it specifically aims to work with MS1 scans acquired as profile scans, providing better coverage and fewer false positives while allowing better control of the quality of the extracted features.[21,22] The critical step that allowed efficient working with profile scans was to use a master m/z axis, estimated from the data, and to interpolate all MS scans to this master axis. The new version of Finnee, Finnee2024, presented in this article, is a complete rewrite of the previous toolbox, allowing faster computations and taking full advantage of profile scans. The changes are highlighted in the text and illustrated with the data provided by Li and coworkers.[23] They use 1100 compounds, with 130 of them at varying concentrations, separated by LC-HRMS (Orbitrap - Thermo Q exactives). Each mixture was analysed five times. The original data files are freely available for download. Their manuscript compared the performance of five software packages. Those data will validate and compare the Finnee version with other existing approaches.

Finnee2023 Objectives and Data Structures

Core objectives. Features mining and alignment that aims to extract features consistent with chromatographic peaks from HRMS datasets and align them across multiple datasets is an essential step in metabolomics. The results of this step are used to detect isotopic and adduct patterns and merge all features originating from the same compound into a single group, identify and annotate putative molecular formulae, test for significant differences between two or more populations and identify important features between these observed differences. Not only is the correct identification and alignment of key features essential, but precision and accuracy in measuring key chromatographic figures of merit, such as areas, accurate masses, and migration time, are also necessary. However, error rates can be high in untargeted analysis, where thousands of peaks per dataset within, ideally, hundreds of datasets. In addition, classical chromatographic optimisation to maximise precision (e.g. resolution greater than 1.5) is not possible.[24] To address this limitation, Finnee2024 has been designed with the following objectives:

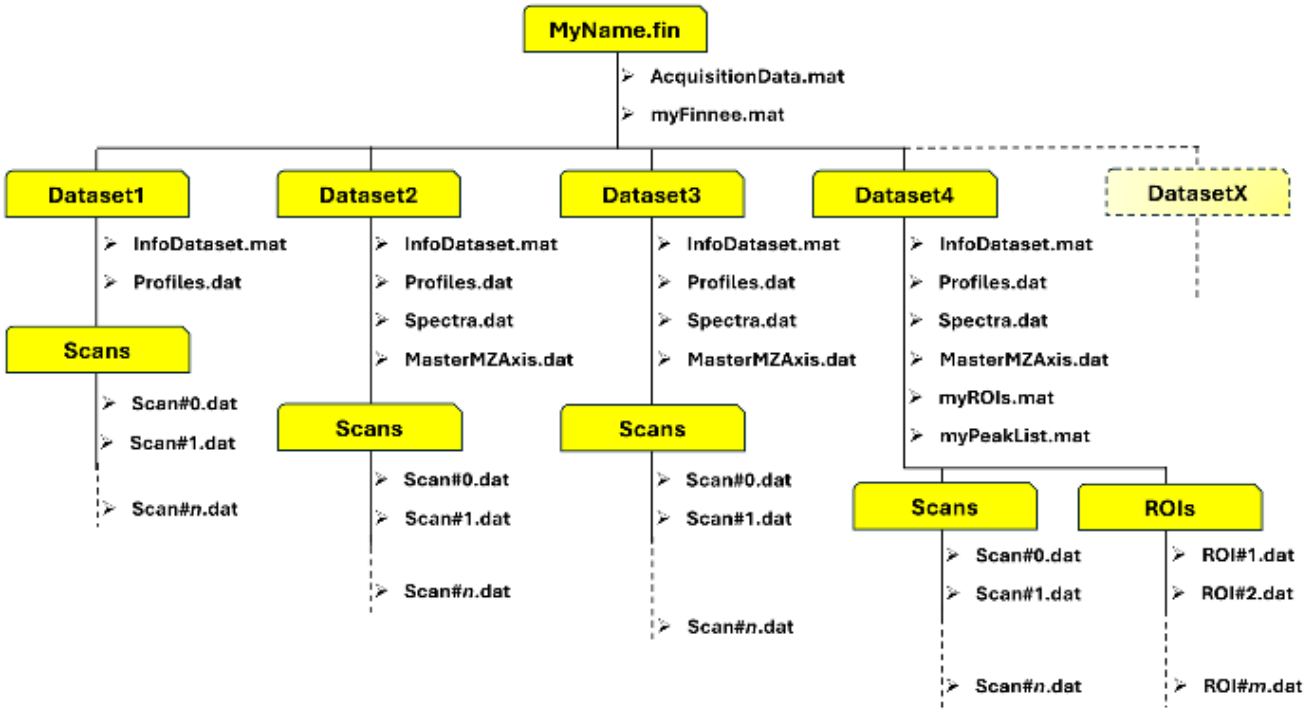
1. Transparency. Each transformation is recorded in detail, allowing for the visualisation of data integrity and data loss.
2. Error analysis. Errors are inherent in large-scale metabolomics. Finnee is equipped with algorithms that detect and visualise potential errors.
3. Flexibility to re-analyse individual features as required.

Finnee2023 is not a closed pipeline but a series of individual functions that can be tailored to the specifics of a given study. However, the toolbox and all associated functions have been designed solely to process MS1 data acquired as profile scans. The input file should be in mzML format.

Data structure. The data structure reflects these core objectives. The data structure has been redesigned from the previous version to facilitate parallel computing and allow profile scans to be processed even with large-scale metabolomics studies that may involve processing thousands of datasets.

With Finnee2024, each file part of a large study can be processed independently, allowing easy parallelism. For each mzML file obtained for the original instrumental data, a folder with a user-defined name and a "fin" extension is created. This folder will contain two Matlab files; "AquisitionData.mat" will include all the metadata associated with the original mzML file, and "myFinnee.mat" is a Matlab object that will summarise any transformation done on the original data as well as archiving any used parameters. A subfolder within this folder is created with the generic name <Dataset1>. This folder contains two files; the one called "infoDataset.mat" has information about the dataset, such as the date of creation, options for creations, labels and units for the m/z, time and intensity axis. It also contains information allowing loading the binary file "Profiles.dat". This file contains the name of each MS scan file recorded as a profile scan and its corresponding time of acquisition together with simple descriptors such as the total intensity and maximum intensity in each MS scan, allowing to obtain the total ion profile (**TIP**) and base peak profile (**BPP**). Per IUPAC, we have used the term profile rather than chromatogram to include electropherogram or any traces obtained from a separation device.[25] Additional two-dimensional representations of the date can easily be added. Individual MS scans are stored in binary files in an additional sub-folder <Scans> with the generic name "Scan#X.dat". The

size of each binary file is stored in "Profiles.dat" for quick retrieval of information. A schematic of the Finnee2024 data chart is given in Scheme 1. Such structure allows to rapidly calculate extracted ion profiles or obtained targed scans as only the necessary information has to be retrieved. However, the primary rationale for such a data structure was to allow for recording any data transformation. Those are recorded in new folders named <Dataset2>, <Dataset3>,…, <DatasetX> with the same structure as in <Dataset1>, allowing to compare any MS scans or profiles between each transformation and assess data loss. Scheme 1 illustrates the four datasets used in this work, with <Dataset1> corresponding to the MS scans as acquired by the instruments, <Dataset2> the MS scans and associated profiles after alignment to the master mz axis, <Dataset3> the MS scans after baseline correction and the <Dataset4> the MS scans after noise removal. Additional transformation can easily be added. The ones used in this work are detailed below.



Scheme 1. Schematic representation of the data structure of a Finnee file.

Main transformations and associated datasets

The datasets obtained in Scheme 1 correspond to the data after various global transformations. While this can easily be tailored to one needs, the following ones were used. The first dataset corresponds to the data as recorded by the MS instrument.

Dataset2: Linear Interpolation to Master m/z Axis. This first transformation is the main originality of the Finnee approach. It consists of two steps: first, a master m/z axis is generated using the experimental data, and then MS1 scans are interpolated to this master m/z axis, thus providing the same axis for all spectra within the same dataset.

In the first step, the master m/z axis is created from the scan with the most information. The m/z profile scan axis cannot readily be used as portions with null intensities are missing. To build the master axis, the value of the differences of m/z between two successive m/z with non-null intensities is extracted. A polynomial is fitted to this data to predict the m/z increment as a function of the m/z value. More than 10,000 data points can be obtained from one MS spectrum in an orbitrap file to build the master m/z axis. Results are shown in Figure 1.

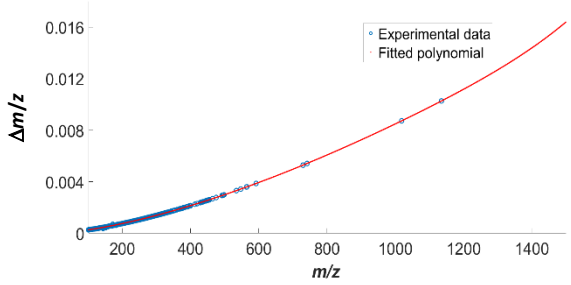


Figure 1. $\Delta m/z$ as a function of m/z calculated from one MS scan**.** The circles correspond to the experimental data (n = 10241) and the line to the fitted polynomial.

The polynomial is then used to obtain the completed m/z axis. This axis is a vector with 552,377 points in the orbitrap dataset recorded in <Dataset2> as "MasterMZAxis.dat"

Each MS1 scan in the original dataset is then interpolated using a linear spline, and the new scan is recorded in the new dataset, <Dataset2>. Following the intensity variation at a specific m/z over time is now possible with such a transformation. Emphasis should be placed on this approach, which does not involve a bucketing approach where all intensities with a set m/z interval are summed in a single bucket. Here, we maintain the spatial resolution where multiple points describe each ion peak's m/z positions and intensities. By analogies, we named the profile obtained at a specific m/*z*, single mz profile (**SmzP**). The folder <Dataset2> contains similar information as <Dataset1>, but a new file is now used: "Spectra.dat". While "Profile.dat" contains information about profiles, that is, variation of information as a function of time (total ion profile, base peak profile,…), "Spectra.dat" contains information about spectra, which is a variation of information as a function m/z that can now be calculated. For example, the Total Intensity Spectra (**TIS**) is the sum of all intensities recorded at any time at a specific m/z, and the Base Intensity Spectra (**BIS**) is the maximum intensity recorded at any time. Those two spectra allow for pinpointing relevant SmzP. However, an essential spectrum is the Maximum Continuous Profile (**MCP**), which records for each m/z the most extended sequence of non-null intensity. This representation allows the differentiation of SmzP, where only spikes are present, from SmzP with probable chromatographic peaks and SmzP with background ions. For example with SA1.mzml, one of the files from the orbitrap dataset, out of the 490,031 SmzP that have non-null intensities, 293,282 of them do not have a continuous sequence of points with non-null intensities higher than 3, 192,347 of them have at least one sequence of non-null intensity between 4 and 500, and 4,402 of them have, at least, one sequence of points of non-null intensities higher than 500. 500 has been chosen as a threshold because it is roughly 10 % of the total time axis length (5824). Such information allows us to remove and calculate the MS base noise (**bNoise**) due to random spikes in the absence of any ions entering the mass analyser and select the SmzP that may need to be corrected for baseline drift. This is illustrated in Figure 2, with one representative file from the orbitrap dataset.

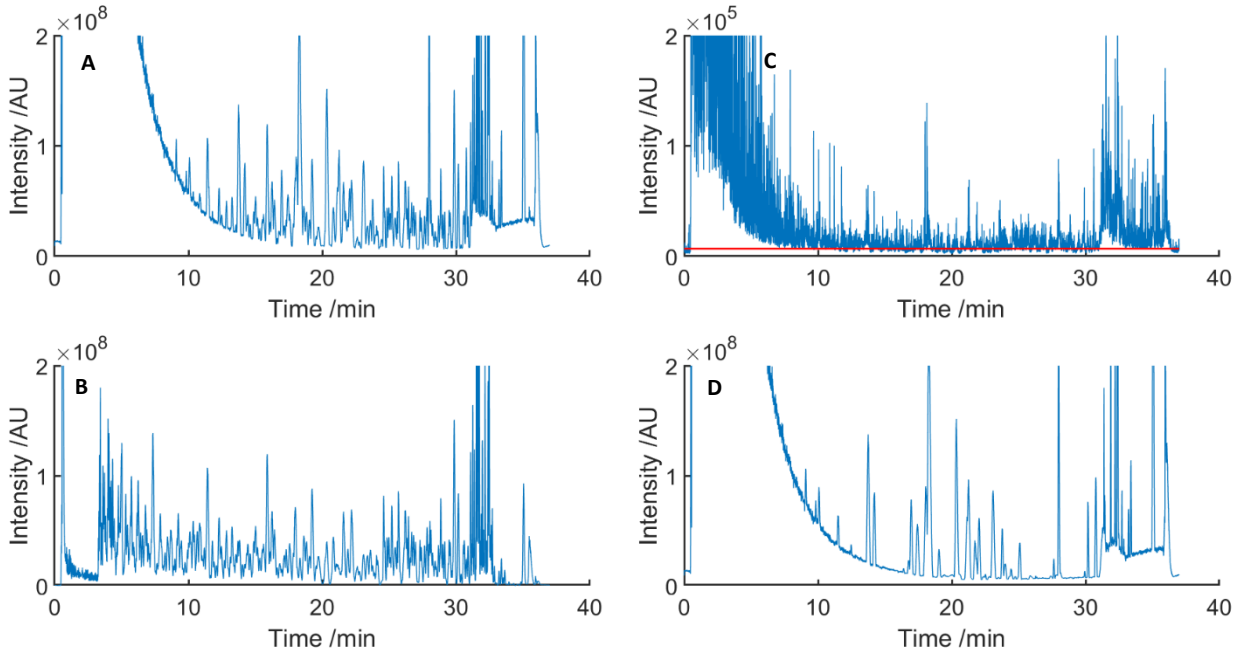


Figure 2. Base peak profiles were obtained with (A) all data, (B) using SmzP with continuous non-null intensity between 3 and 500, (C) using SmzP with continuous non-null intensity below 3, and (D) using SmzP with continuous non-null intensity higher than 500. In (C), the continuous line represents the bNoise.

Dataset3: Baseline Drift Correction. The following dataset is obtained after correction of the selected SmzP for baseline drift. This part also allows us to estimate the background noise at every *m/z*. For SmzP corrected for baseline drift, the noise is calculated as twice the standard deviation of the point detected as baseline intensity. For the other, it is the value of the minimal noise, calculated as twice the standard deviation of all non-null intensities in SmzP with MCP < 3. This approach limits the number of profiles to be corrected for baseline drift. While many algorithms have been developed,[26] for the moment, only the symmetrically reweighted penalised least squares (arPLS) has been implemented. [27]

Dataset4: Noise removal. The noise removal step used in the previous version of Finnee, Finnee2016, has been optimised, allowing an x2 time gain. As in the earlier version, the intensity at every m/z in every spectrum is scanned and set to zeros if its intensity and the intensities of neighbouring points (in the m/z and time dimension) are below ten times the noise, allowing a decrease in the quantity of data drastically. Here, for example, while the size of the

scan folder is 997 MB in <Dataset1>, it is reduced by 3 to 314 MB in <Dataset4> without losing important information. This is illustrated in Figure 3, within A the original TIF, B the TIF after baseline correction and noise removal, and C the difference between the two profiles.

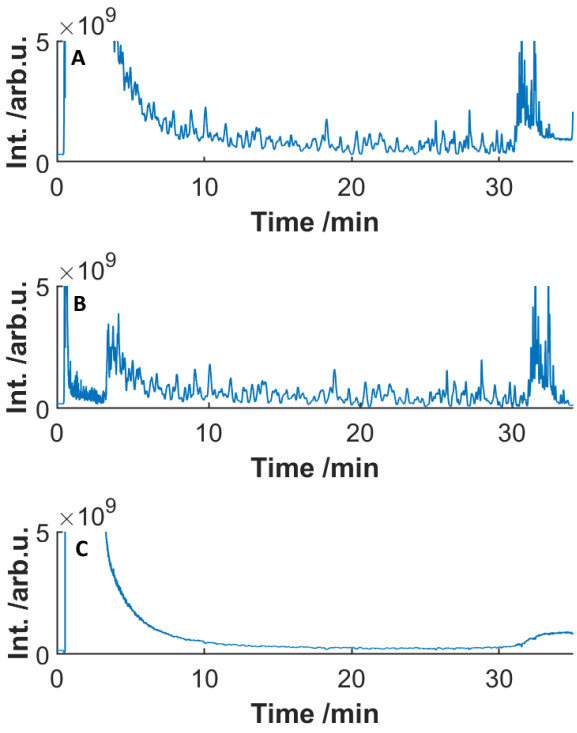


Figure 3. Comparison of the total ion profiles (TIP), (A) after alignment to the master mz axis and (B) after baseline drift correction and noise reduction. The profile displayed in (C) corresponds to the difference between both profiles.

Region of Interests and Feature List. The feature extraction takes advantage of the thorough data cleaning performed previously. Features are obtained in two steps. First, regions of interest (ROI) are extracted from the noise-corrected dataset. ROIs are defined as areas within the time and m/z space with non-null intensities surrounded by null intensities. Each ROI is recorded as an *m*x*n* matrix and saved as binary files in a new folder. A set of descriptors is recorded in the file myROIs for each ROI.mat (see Scheme 1, allowing assessing the relevance of each ROI. ROI may contain noise and one or more chromatographic features. ROI can be independently visualised. A random forest ML approach has been developed to allow a prefiltration of the ROI based on those descriptors. Users are first asked to label a random selection of more than 150 ROIs to discard or to keep. Examples of ROIs are displayed in Figure 4. It should be emphasised that some may be difficult to classify (Figure 4D) and will depend on the user's choice. It should be emphasised that ROIs are extracted as surfaces surrounded by null intensity points. As such, ROIs may contain more than one peak if they are not baseline-separated (Figure 4C). This is normal, and ROIs will be further processed in the next step.

The labelling stops when each class contains at least 75 ROI (this value can be changed). Label ROI are then separated into training (66% of the data) set and test (33% of the data). A random forest model is optimised with the training set and validated with the test set. In this example, an accuracy of 82.0% was obtained (false negative: 6%, false positive 12%, n = 50), allowing filtering of 50 % of the ROI (from 41,568 ROI to 20,861). Individual features that should be single chromatographic peaks are extracted from the ROIs by determining the limits of the peak's surfaces (chromatographic peak x ms peak). Different figures of merits are calculated from those data, including the peak volume, the centroid position in the m/z axis, allowing the calculation of the accurate masses, the centroid position in the time axis, allowing the calculation of the retentions, as well as the peaks' variances in both dimensions.

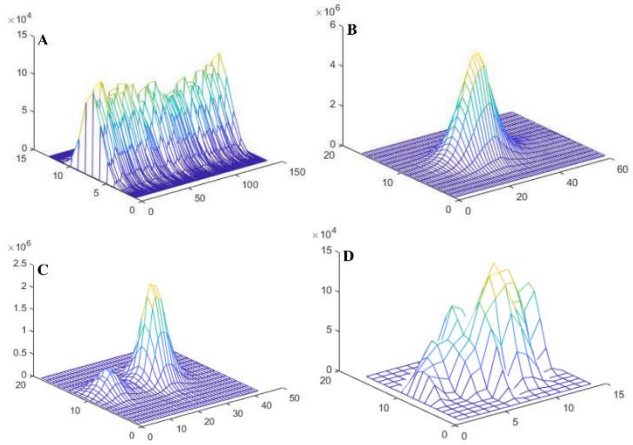


Figure 4. Representative examples of ROI, with (A) only noise or (B) and (C) one or multiple chromatographic peaks. (D) is an example of hard-to-label ROI, which may depend on the user's decision.

Project and Validation

Master Finnee, Project, and merged Feature Lists. While it is necessary to optimise all steps described previously using the output of an LCMS experiment representative of the complete study, those steps can then be performed to all output of all experiments belonging to a study, as long as the conditions in each experiment are similar. This allows for leveraging the parallel computing toolbox and speeds up the whole analysis. The finnee file, with a structure similar to scheme 1, will be used as a template and is named the Master Finnee. All Finnee files belonging to a specific study are stored in an umbrella folder and linked to an object that will record basic information such as the name and location of all the original mzML files and the location of all output of the Finnee file. Features lists from all finnee files are aligned to obtain quantitative variation in each experiment. Average accurate mass, migration time, peak variances, and their standard deviation are calculated.

Validation with orbitrap files. The performance and accuracy of the Finnee2024 toolbox were measured using data provided by Li and coworkers. Out of the 1100 compounds used to build the test mixture, they demonstrated that 836 can be accurately detected and quantified using a targeted approach. These are present in 10 experiments grouped into two series (SA1-SA5 and SB1 and SB5). The concentrations of some of the compounds are varied between the

two series. The authors also provided an Excel file with, for each of the features, the theoretical accurate mass, the retention time and the theoretical concentration ratio between the SB and SA series, and the experimental ratio calculated using a targeted approach, thus providing two different validation tools, first a qualitative validation based on the number of features that were accurately detected with a mass difference lower than 10ppm and a retention time difference lower than 0.3 min between the theoretical values and the ones obtained using an untargeted approach. Features were assumed to be accurately quantified if the SB/SA ratio difference was lower than 20% between the values obtained using the targeted approach and the untargeted pipeline. The optimised parameters for the Finnee2024 approach are supplementary information in the short tutorial. With Finnee2024, *27947* common features were extracted from the ten files. Of those, 817 matched the expected features, and 711 (*27947*, **817**/711) of those matched the expected concentration ratio. This compared favourably with leading software such as XCMS (*35215*, **820**/731), MZMine 2 (*20021*, **769**/761), MS-Dial (*21545*, 799/654) and compounds discoverer (*10525,* 748/482). The relatively lower quantitative performance of Finnee2024 has been linked to (1) the absence of missing value implementation and (2) the various transformations performed along the successive dataset to extensively correct from baseline drift and background noise. This was verified by performing a semi-targeted approach, where the 27947 features were re-analysed using a targeted approach with **Dataset2,** as the only transformation is the alignment to the common mz axis. The targeted approach also allows us to accurately measure the noise for each peak, allowing us to exclude peaks with a signal-to-noise ratio lower than 10. After the target approach, 21008 common features remain; of those, 816 matched with the expected compounds, and 784 matched the expected concentration ratio.

Conclusions

Unlike existing pipelines, the Finnee toolbox aims to process MS scans recorded as profile spectra, thus avoiding the loss of information inherent to the transformation from profile to centroid scans. The Finnee2024 new version is faster and fully uses the profile spectra by analysing the 3-dimensional chromatographic objects recorded as a region of interest (ROI). With the test data provided by Li and coworkers, results were on par with the best software as assessed by Li and coworkers.[23]

References

1. Fiehn, O. Metabolomics - The link between genotypes and phenotypes. *Plant Mol Biol* **48**, 155–171 (2002).
2. Wishart, D. S. Quantitative metabolomics using NMR. *TrAC - Trends in Analytical Chemistry* **27**, 228–237 (2008).
3. Dettmer, K., Aronov, P. A. & Hammock, B. D. Mass spectrometry-based metabolomics. *Mass Spectrom Rev* **26**, 51–78 (2007).
4. Dunn, W. B., Wilson, I. D., Nicholls, A. W. & Broadhurst, D. The importance of experimental design and QC samples in large-scale and MS-driven untargeted metabolomic studies of humans. *Bioanalysis* **4**, 2249–2264 (2012).
5. Kuehnbaum, N. L. & Britz-McKibbin, P. New advances in separation science for metabolomics resolving. *Chem Rev* **113**, 2437–2468 (2013).
6. Brown, M. *et al.* Mass spectrometry tools and metabolite-specific databases for molecular identification in metabolomics. *Analyst* **134**, 1322 (2009).
7. Guo, J. & Huan, T. Comparison of Full-Scan, Data-Dependent, and Data-Independent Acquisition Modes in Liquid Chromatography–Mass Spectrometry Based Untargeted Metabolomics. *Anal Chem* **92**, 8072–8080 (2020).
8. Macintyre, L. *et al.* Metabolomic tools for secondary metabolite discovery from marine microbial symbionts. *Mar Drugs* **12**, 3416–3448 (2014).
9. Boccard, J. & Rudaz, S. Harnessing the complexity of metabolomic data with chemometrics. *J Chemom* **28**, 1–9 (2014).
10. Schrimpe-Rutledge, A. C., Codreanu, S. G., Sherrod, S. D. & McLean, J. A. Untargeted Metabolomics Strategies—Challenges and Emerging Directions. *J Am Soc Mass Spectrom* **27**, 1897–1905 (2016).
11. Reuschenbach, M., Hohrenk-Danzouma, L. L., Schmidt, T. C. & Renner, G. Development of a scoring parameter to characterise data quality of centroids in high-resolution mass spectra. *Anal Bioanal Chem* **414**, 6635–6645 (2022).
12. Samanipour, S. *et al.* From Centroided to Profile Mode: Machine Learning for Prediction of Peak Width in HRMS Data. *Anal Chem* **93**, 16562–16570 (2021).
13. Martens, L. *et al.* mzML - A community standard for mass spectrometry data. *Molecular and Cellular Proteomics* vol. 10 Preprint at https://doi.org/10.1074/mcp.R110.000133 (2011).
14. Adusumilli, R. & Mallick, P. Data Conversion with ProteoWizard msConvert. in 339–368 (2017). doi:10.1007/978-1-4939-6747-6_23.
15. Forsberg, E. M. *et al.* Data processing, multi-omic pathway mapping, and metabolite activity analysis using XCMS Online. *Nat Protoc* **13**, 633–651 (2018).
16. Pluskal, T., Castillo, S., Villar-Briones, A. & Orešič, M. MZmine 2: Modular framework for processing, visualising, and analysing mass spectrometry-based molecular profile data. *BMC Bioinformatics* **11**, (2010).
17. Röst, H. L. *et al.* OpenMS: a flexible open-source software platform for mass spectrometry data analysis. *Nat Methods* **13**, 741–748 (2016).
18. Tsugawa, H. *et al.* MS-DIAL: data-independent MS/MS deconvolution for comprehensive metabolome analysis. *Nat Methods* **12**, 523–526 (2015).
19. Pang, Z. *et al.* Using MetaboAnalyst 5.0 for LC–HRMS spectra processing, multi-omics integration and covariate adjustment of global metabolomics data. *Nat Protoc* **17**, 1735–1761 (2022).
20. Erny, G. L., Acunha, T., Simó, C., Cifuentes, A. & Alves, A. Finnee — A Matlab toolbox for separation techniques hyphenated high resolution mass spectrometry dataset. *Chemometrics and Intelligent Laboratory Systems* **155**, 138–144 (2016).
21. Erny, G. L. *et al.* Mining for Peaks in LC-HRMS Datasets Using Finnee – A Case Study with Exhaled Breath Condensates from Healthy, Asthmatic, and COPD Patients. *ACS Omega* **5**, 16089–16098 (2020).
22. Mecha, E. *et al.* Metabolomics profile responses to changing environments in a common bean

(Phaseolus vulgaris L.) germplasm collection. *Food Chem* **370**, 131003 (2022).
23. Li, Z. *et al.* Comprehensive evaluation of untargeted metabolomics data processing software in feature detection, quantification and discriminating marker selection. *Anal Chim Acta* **1029**, 50–57 (2018).
24. Dyson, N. *Chromatographic Integration Methods*. (Royal Society of Chemistry, 1998).
25. Murray, K. K. *et al.* Definitions of terms relating to mass spectrometry (IUPAC Recommendations 2013). *Pure and Applied Chemistry* **85**, 1515–1609 (2013).
26. Niezen, L. E., Schoenmakers, P. J. & Pirok, B. W. J. Critical comparison of background correction algorithms used in chromatography. *Anal Chim Acta* **1201**, 339605 (2022).
27. Baek, S.-J., Park, A., Ahn, Y.-J. & Choo, J. Baseline correction using asymmetrically reweighted penalised least squares smoothing. *Analyst* **140**, 250–257 (2015).