Finnee – A Matlab Toolbox for In-Depth Analysis of Separation Techniques Hyphenated to High-Resolution Mass Spectrometry

Guillaume L. Erny (**guillaume.erny@fe.up.pt**)

LEPABE - Laboratory for Process Engineering, Environment, Biotechnology and Energy, Faculty of Engineering, University of Porto, Rua Dr. Roberto Frias, 4200-465 Porto, Portugal

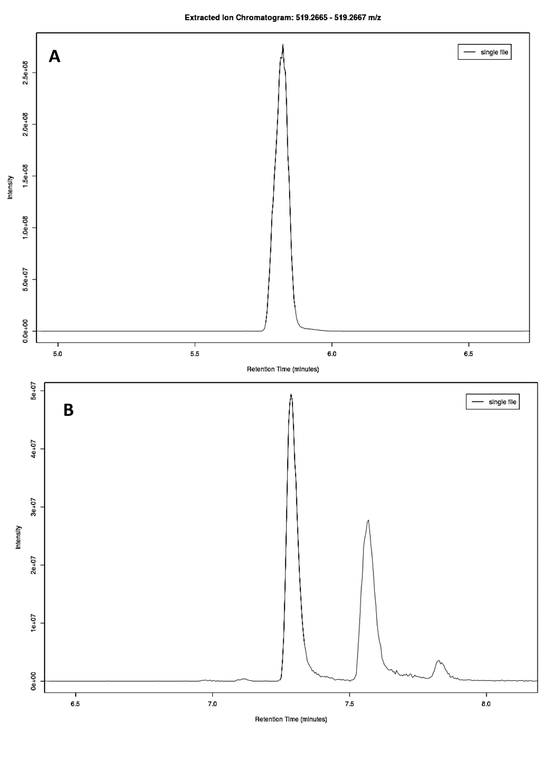
Introduction

Untargeted analyses in MS1 mode is a promising approach. This approach consists, using computer-assisted tools, in mining datasets to recognise and measure any chromatographic-like features in each dataset resulting from a single separation. With liquid chromatography hyphenated to high-resolution mass spectrometry and complex samples, tens of thousands of features are often extracted. It is not possible to identify each, and every feature and the analyst often uses differential analyses where the features extracted from multiple experiments are pooled together in a single large matrix. Each line corresponds to a single feature, with its measured intensity or area in each experiment in each row. With the help of chemometrics tools, such as clustering, principal components analysis or partial least squares, it is possible to obtain features that are significantly different, which could be specific markers of a population or biologic process. However, this approach is often erroneous due to the many errors arising when mining for peaks. *Finnee* is a Matlab toolbox that aims to decrease error associated with peak mining.

The "Classical" Approaches and Errors Associated

step	Comments	Errors associated
1. Centroid transformation and filtering	Peak picking to detect every peaks in a single MS spectrum and generate the usual centroid bar plot (position vs intensity). Low intensity peaks are removed.	Near Isobaric peaks may lead to one or two lines. Difficulties in selecting an "optimal" threshold value?
2. Features extraction	Generation of features that are a succession of centroid peaks whose accurate masses (position of the peak) do not differ by more that a set value	
3. Peak detection and deconvolution	Scanning each features to detect if and how many peaks are present. Peaks figure of merits are recorded on a peak table. Examples are shown in Figure 1.	Many features contain multiple peaks that are **not baseline resolved**! Features as in **C or D should be discarded.**
4. Peak alignment	Aligning peaks between multiple table	How to have reliable results when peaks may be split or missing?

Ideal Features



Real Features

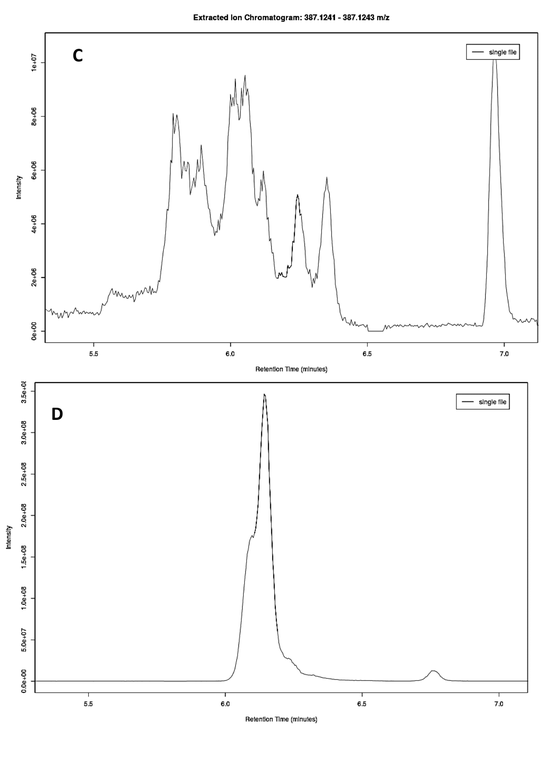


Figure 1. Example of features from metabolomics analysis of flour analysed by LC-Orbitrap/MS mined by xcms online. A and B: ideal features that will give reliable peaks figure of merits. C and D: problematic features.

Acknowledgements

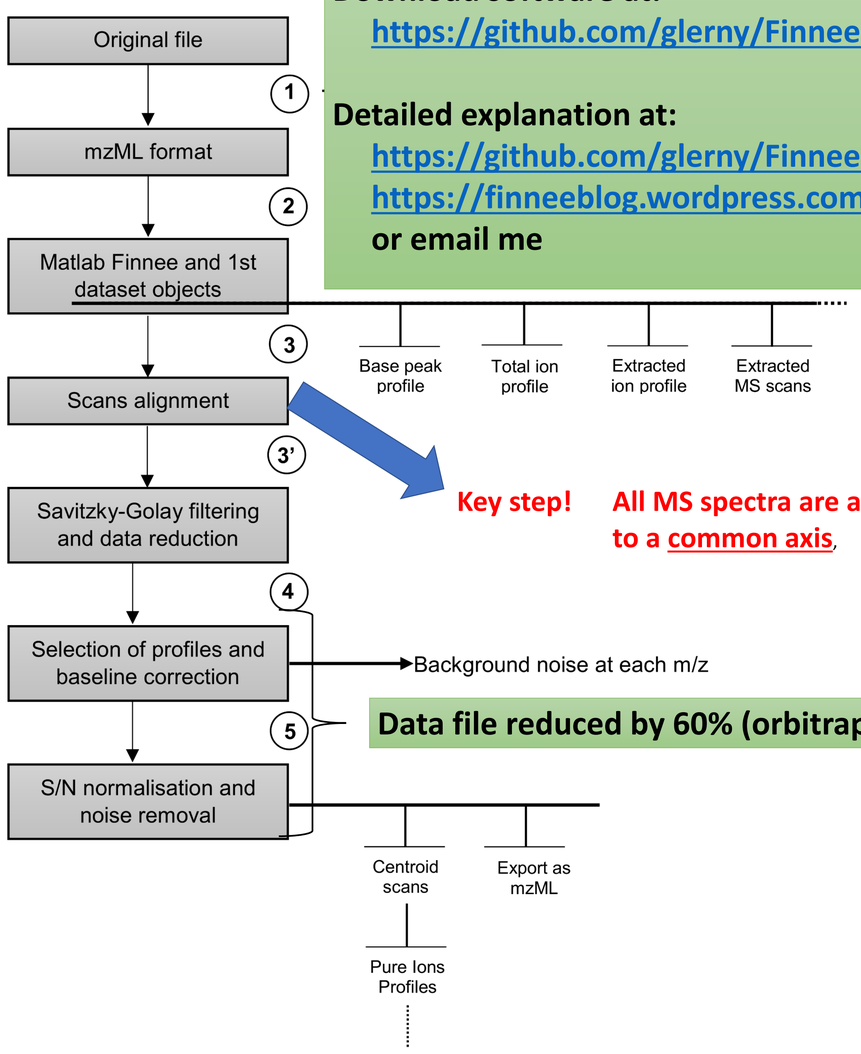
This work was financially supported by the projects:

(i) project UID/EQU/00511/2019 - Laboratory for Process Engineering, Environment, Biotechnology and Energy – LEPABE funded by national funds through FCT/MCTES (PIDDAC);

(ii) Project POCI-01-0145-FEDER-029702, funded by FEDER funds through COMPETE2020 – Programa Operacional Competitividade e Internacionalização (POCI) and by national funds (PIDDAC) through FCT/MCTES;

(iii) Project "LEPABE-2-ECO-INNOVATION" – NORTE-01-0145-FEDER-000005, funded by Norte Portugal Regional Operational Programme (NORTE 2020), under PORTUGAL 2020 Partnership Agreement, through the European Regional Development Fund (ERDF).

The Finnee Approach



Download software at:
https://github.com/glerny/Finnee2016

Detailed explanation at:
https://github.com/glerny/Finnee2016/wiki
https://finneeblog.wordpress.com/
or email me

Key step! **All MS spectra are aligned to a common axis,**

Data file reduced by 60% (orbitrap) to 99% (TOF)

Baseline Correction and Noise Removal – a Closer Look

All MS scans are profiles (original) scans and are aligned to a common *m/z* axis.
Definition: ***MZP – m/z* Profile*:*** Profile corresponding to the intensity as a function of scans/time at a specific *m/z* increment in the common axis

4.1. Selecting the MZP to correct from baseline drift by measuring the frequency of non-null values in each MZP

MZP with baseline drift (coelution with background ions): low amount of non-null values

MZP with one of few "separated" ions : high amount of non-null values

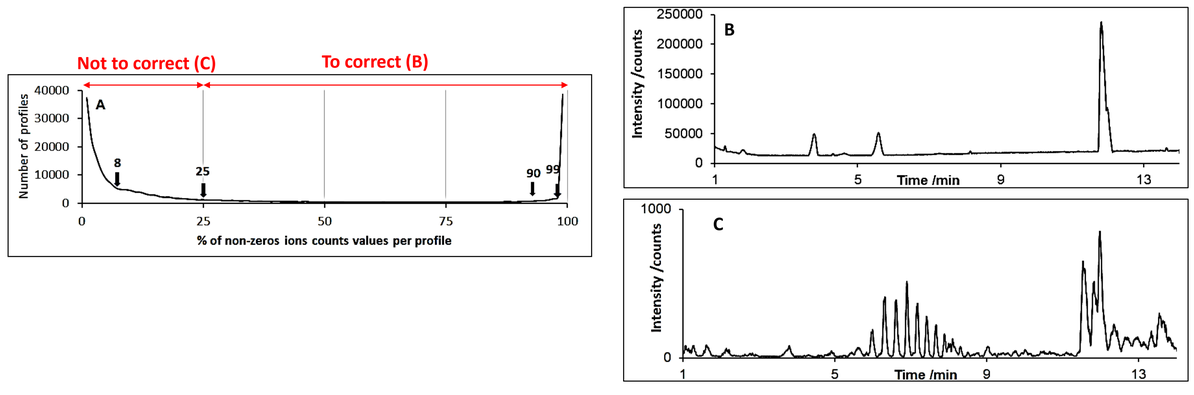


Figure 2. (A) frequency of MZP as a function of non-null values , (B) sum of MZP over the threshold, (C) sum of MZP under the threshold

4.2. Baseline correction and noise estimation using **a recursive fit**;

! Easy if and only if the MZP to correct contains a significant amount of baseline points.

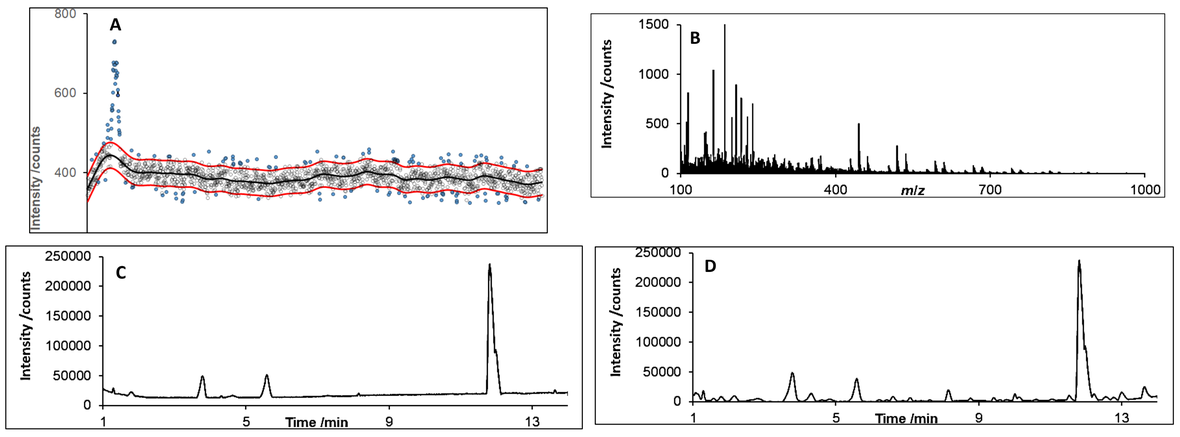


Figure 3. (A) Baseline correction and noise estimation, (B) noise estimated at each m/z, (C and D) total ion profile of a sample of exalted breath condensate separated by LC-tripleTOF/MS before and after correction respectively

After correction for baselines drift and noise removal, corrected scans can be converted from MS profile to MS centroid and a feature extraction step used, **BUT** there is no need for peak detection and deconvolution.