

A NEW DATA-DRIVEN ALGORITHM FOR UNTARGETED METABOLOMICS ANALYSIS

Sandia Machado^{1,*}, Luisa Barreiros^{1,2}, Ricardo N. M. J. Páscoa¹, Marcela A. Segundo¹, João A. Lopes³

¹LAQV, REQUIMTE, Departamento de Ciências Químicas, Faculdade de Farmácia da Universidade do Porto, Porto, Portugal.
sandia_machado@hotmail.com

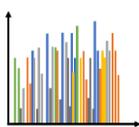
²Escola Superior de Saúde do Instituto Politécnico do Porto, Porto, Portugal

³iMed.ULisboa, Instituto de Investigação do Medicamento, Faculdade de Farmácia da Universidade de Lisboa, Lisboa, Portugal

INTRODUCTION

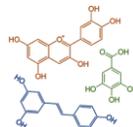
UNTARGETED ANALYSIS

Generates complex data hard to interpret due to its ability to access the entire metabolome.



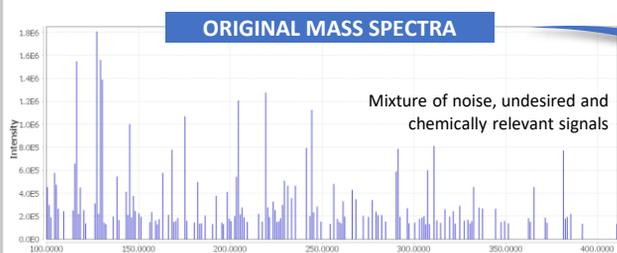
DATA-DRIVEN APPROACHES

Extract valuable information from chemical systems data.



There is no ultimate data-driven strategy to process data from untargeted mass spectrometry analysis

DEVELOPMENT OF AN ALGORITHM TO PROCESS UNTARGETED MASS SPECTROMETRY DATA



PEAK DETECTION (signals without characteristics of chromatographic peaks are excluded)

BLANK SUBTRACTION (signals present in blanks are excluded)



MATRIX WITH CHEMICALLY RELEVANT SIGNALS (possibility to compare several samples)



Algorithm was tested in Matlab with **24 grape samples** analyzed by liquid chromatography coupled to mass spectrometry using an untargeted method.

GRAPE FORMATION

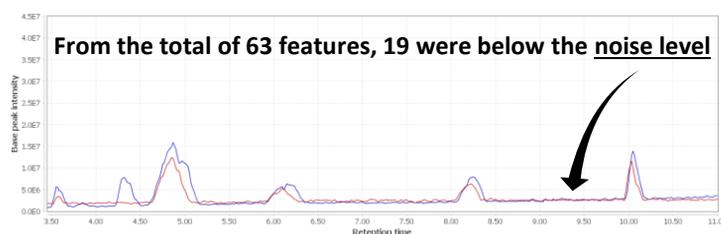
GRAPE RIPENING



Algorithm efficiency was evaluated by comparison with the results obtained in MZmine software.

RESULTS

Number of features	Formation	Ripening
18	✓	
12		✓
33	✓	✓



Total = 63 from a original matrix with more than 1100 signals

CONCLUSIONS

The proposed algorithm was able to select the chemically relevant features, creating a simpler and more understandable data matrix for further studies. It proved effective in discriminating the features typical of each developmental stage of grape and accessing the features present below the noise level, since it does not exclude any signal based on intensity.

ACKNOWLEDGEMENTS

This work received financial support from the European Union (FEDER funds) through COMPETE POCI-01-0145-FEDER-016735 and National Funds (FCT/MCTES, Fundação para a Ciência e a Tecnologia and Ministério da Ciência, Tecnologia e Ensino Superior) through project PTDC/AGR-PRO/6817/2014 and also under UIDB/50006/2020. S. Machado thanks FCT and POCH (Programa Operacional Capital Humano) for her PhD grant (SFRH/BD/122730/2016). L. Barreiros and R. N. M. J. Páscoa acknowledge funding from FCT through program DL 57/2016 – Norma transitória.