Exploring data quality of multivariate HRMS data in food authenticity research

Marios Kostakis, Reza Aalizadeh, Marilena Dasenaki, Nikolaos Thomaidis

Laboratory of Analytical Chemistry, Department of Chemistry, National and Kapodistrian University of Athens, Panepistimioupolis Zografou, 157 71 Athens, Greece; emails – makostak@chem.uoa.gr; ntho@chem.uoa.gr

Abstract

The rapid development of HRMS-based metabolomics boosted the research and the development of new approaches in food authenticity studies. The large amount of data produced with HRMS techniques can be used to determine the molecular fingerprint of food and to detect food fraud, using both targeted and untargeted metabolomics approaches. However, this technique has not been widely used so far by official authorities in EU for food adulteration control, with the main drawback being the difficulty of data treatment and integrity assessment by the end user. Moreover, most of the times the results of multivariate statistical processing are binary, and there is not always a clear compliance with parameters characterizing an adulterated product. The aim of this study is to investigate the development of validation protocols to assess the data quality of multivariate statistical processing. Juice-to juice adulteration of pomegranate juice with apple juice was used as a case study in order to produce and validate a statistical model that can reliably assess food adulteration.

Adulteration

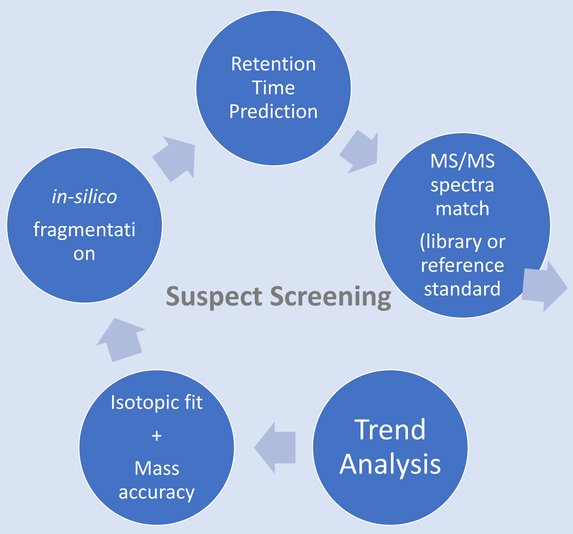
Experimental Part

LC-ESI-qTOF
Bruker, MaXis Impact
Ultra High Resolution
Time-of-Flight Mass Spectrometer
Column: Acclaim RSLC 120 C18
2.2 µm, 2.1 × 100 mm
Pre-column: VanGuard (Waters):
Acquity UPLC BEH C18 1.7 µm, 2.1 × 5 mm
Negative Ionizaton (-)

Samples: pomegranate brix 15°
apple brix 11.2°
Procedure: filtration with RC filters 0.22 µm
Injection in LC-qTOF-MS

Replicates: 3
Percentage of adulteration: pure apple, pure pomegranate, 1-2-3-5-10-20% adulteration of pomegranate with apple

Building the Model

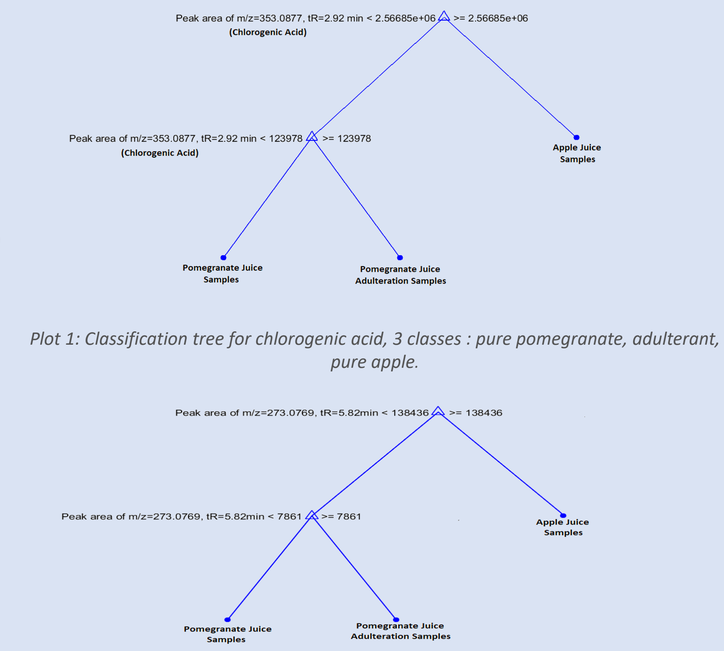


26 potential suspects

Table 1: 2 of 26 suspect masses that Identified.

Important masses (m/z)	Identification
353.0877	Chlorogenic Acid
273.0769	Phloretin

Classification tree



Plot 1: Classification tree for chlorogenic acid, 3 classes : pure pomegranate, adulterant, pure apple.

Plot 2: Classification tree Phloretin, 3 classes : pure pomegranate, adulterant, pure apple.

Validation of the Model

Experiments

Replicates: 5 x 2 injections (n=10)

Percentage of adulteration: 0-1-3-5 %

Estimation of the uncertainty of identification (reliability) with two approaches:
contingency tables & Bayesians

Setting thresholds for identification of adulteration according to the results of classification tree

Table 2: Thresholds of peak areas from classification tree for every compound.

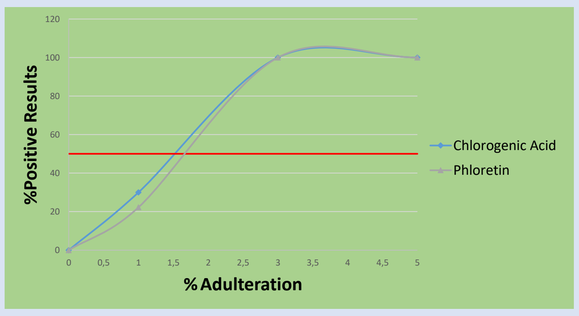
Peak Area	Pure pomegranate	Adulteration of pomegranate with apple
Chlorogenic acid	<123978	>123978
Phloretin	<7861	>7861

Calculation of TP,FP,TN,FN ratios

Table 3: The results of TP,FP,TN,FN ratios for chlorogenic acid and phloretin

	%Adulteration	%TP	%FP	%TN	%FN
Chlorogenic Acid	0	0	0	100	0
	1	30.0	0	0	70.0
	3	100	0	0	0
	5	100	0	0	0
phloretin	%Adulteration	%TP	%FP	%TN	%FN
	0	0	0	100	0
	1	22.2	0	0	77.8
	3	100	0	0	0
	5	100	0	0	0

Performance Curve



Plot 3: Performance curve plot: Percentage of Positive results versus the percentage of adulteration .

Conclusions & Perspectives

- Chlorogenic acid and Phloretin were detected as potential markers for detection the adulteration of pomegranate with apple.
- Classification tree was the statistical technique that it was applied to data in order to distinguish pure juices (pomegranate and apple) and juices with adulteration.
- 0-1-3-5% of adulteration was used for the validation of model.
- It was estimated the uncertainty of identification in order to validate the model.
- For both compounds, uncertainty of identification is the same.
- Identification of adulteration is reliable above the 3% of adulteration, for both compounds.
- A perspective is to include to model more compounds in order to improve the reliability (uncertainty) of the model and to identify the adulteration in lower levels.
- A next step is to compare different multivariate statistical techniques (e.g. PLC-DA) for building the same model.

Table 4: Results of uncertainty for chlorogenic acid and phloretin with both approaches.

	contingency table approach		Bayes approach	
	PPV	NPV	P(A/A)	P(nA/nA)
Chlorogenic Acid	100	68,2	1,0	0,811
Phloretin	100	68,2	1,0	0,814

Estimation of uncertainty

Contingency Table approach
PPV: Positive Predictive Value, True positive results with respect to total positive results[1,2]
NPV: Negative Predictive Value, true negative results with respect to total negative results[1,2]
Bayes approach
P(A|A): Conditional probability of true positive results[2,3]
P(nA|nA):Conditional probability of true negative results[2,3]

TP: True Positive
FP: False Positive
TN: True Negative
FN: False Negative

Literature.

1. L. Cuadros-Rodriguez et al, Trends in Analytical Chemistry, 80 (2016), 612-624.
2. A. Pulido, I. Ruisanchez, R. Boque, F.X. Rius, Trends in Analytical Chemistry, 22 (2003), 647-654.