UNIVERSITÀ
DI TORINO

Dual parallel detection raw data fusion: Challenges and opportunities for accurate fingerprinting over large time frames

Andrea Caratti[1], Simone Squara[1], Angelica Fina[1], Fulvia Trapani[1], Erica Liberto[1], Stephen E. Reichenbach[2,3], Qingping Tao[2], Daniel Geschwender[2], Chiara Cordero[1]

[1] Dipartimento di Scienza e Tecnologia del Farmaco, Università di Torino, Via Pietro Giuria 9, Turin I-10125, Italy
[2] GC Image LLC, Lincoln, NE, USA
[3] Computer Science and Engineering Department, University of Nebraska –Lincoln, Lincoln, NE, USA

16TH Multidimensional Chromatography Workshop

UNIVERSITÀ
DI TORINO

Presentation outline

Background and strategies

- ✓ Advantages and disadvantages of multiple detectors in GC × GC
- ✓ Fingerprinting and profiling - strategies based on pattern recognition
- ✓ Data processing workflow

Challenges and current application

- ✓ Fusion of parallel signal channels

Applications

- ✓ Food volatilomics -> quantitative fingerprinting
- ✓ Raw fragrances volatile allergens quantitative profiling

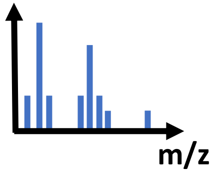
Conclusive remarks

Context

- Comprehensive GC×GC is fundamental for analyzing complex samples
- Profiling, fingerprinting, group-type analysis the most common approaches
- Different detectors can be coupled with the chromatographic system

- Most common are Mass Spectrometry (MS) and Flame Ionization detector (FID)

MS



✓

Specificity - **Molecular fingerprint**
High **sensitivity**
Accurate mass information

✗

Limited linearity range
Response stability over time could be an issue

FID H_2

✓

High **stability** over time
Broad **linearity** range
Use of **FID-predicted RRFs**

✗

No confirmatory attitudes
No molecular diagnosis

Context

Chromatographic fingerprinting based on GC × GC patterns enables efficient cross-comparative analyses for samples **discrimination** and **identitation**. With **MS detection** the process has **profiling attitude** informing about chemical composition. It is a **high-throughput process** capable of unravelling compositional **differences between samples**

Contents lists available at ScienceDirect

Trends in Analytical Chemistry

journal homepage: www.elsevier.com/locate/trac

Chromatographic fingerprinting by comprehensive two-dimensional chromatography: Fundamentals and tools

Federico Stilo [a], Carlo Bicchi [a], Ana M. Jimenez-Carvelo [b], Luis Cuadros-Rodriguez [b], Stephen E. Reichenbach [c, d, **], Chiara Cordero [a, *]

Untargeted/ Targeted fingerprinting

Step 1 - Untargeted Template Construction

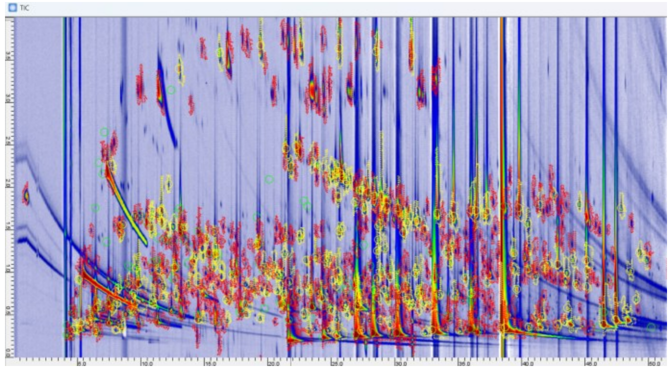
Beginning with several chromatograms randomly selected from all batches

- Pre-processing: file import, rasterization, colorization, baseline correction, **2D peaks detection** and integration
- **Comprehensive pair-wise** peak **matching**
- Determination of *reliable* registration peaks (most relaxed constraint)
- **Alignment of 2D chromatograms**
- Generation of a composite chromatogram
- Definition of pattern of peak-regions for all detected 2D peaks
- Building of ***feature* templates with *reliable* peaks and peak-regions**

Step 2 - Cross-Sample Analysis

Feature template built at Step 1 is matched on all sample images

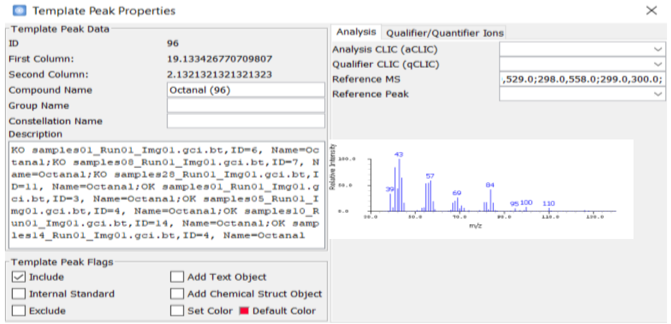
- Template matching for *reliable* peaks and peak-regions
- Alignment of peak-regions relative to matched peaks
- Save processed chromatograms and export re-aligned metadata



(IMAGE) PATTERN RECOGNITION

Specificity level 1 relative retention ${}^1t_R$ and ${}^2t_R$

Specificity level 2 EI fragmentation pattern similarity



Template Peak Properties

Template Peak Data

ID 96

First Column: 19.133426770709807

Second Column: 2.1321321321321323

Compound Name: Octanal (96)

Group Name

Constellation Name

Description

KO samples01_Run01_Img01.gci.bt,ID=6, Name=Octanal;KO samples08_Run01_Img01.gci.bt,ID=7, Name=Octanal;KO samples28_Run01_Img01.gci.bt,ID=11, Name=Octanal;OK samples01_Run01_Img01.gci.bt,ID=3, Name=Octanal;OK samples05_Run01_Img01.gci.bt,ID=4, Name=Octanal;OK samples10_Run01_Img01.gci.bt,ID=14, Name=Octanal;OK samples14_Run01_Img01.gci.bt,ID=4, Name=Octanal

Template Peak Flags

☑ Include ☐ Add Text Object

☐ Internal Standard ☐ Add Chemical Struct Object

☐ Exclude ☐ Set Color 🟥 Default Color

Analysis | Qualifier/Quantifier Ions

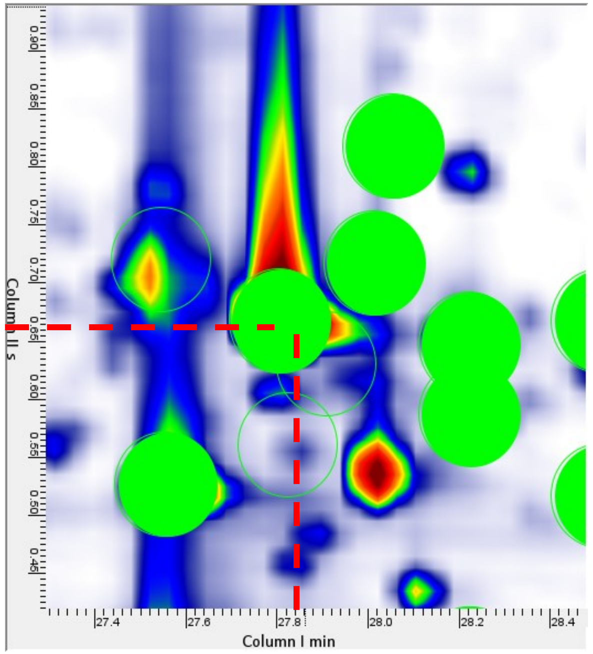
Analysis CLIC (aCLIC)

Qualifier CLIC (qCLIC)

Reference MS: ,529.0;298.0,558.0;299.0,300.0;

Reference Peak

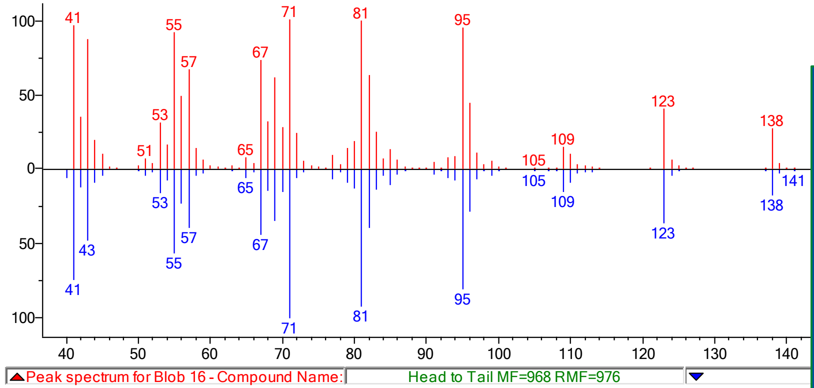
Context – Specificity in Pattern Recognition MS/FID



2D-peaks have unique time coordinates (retention times)

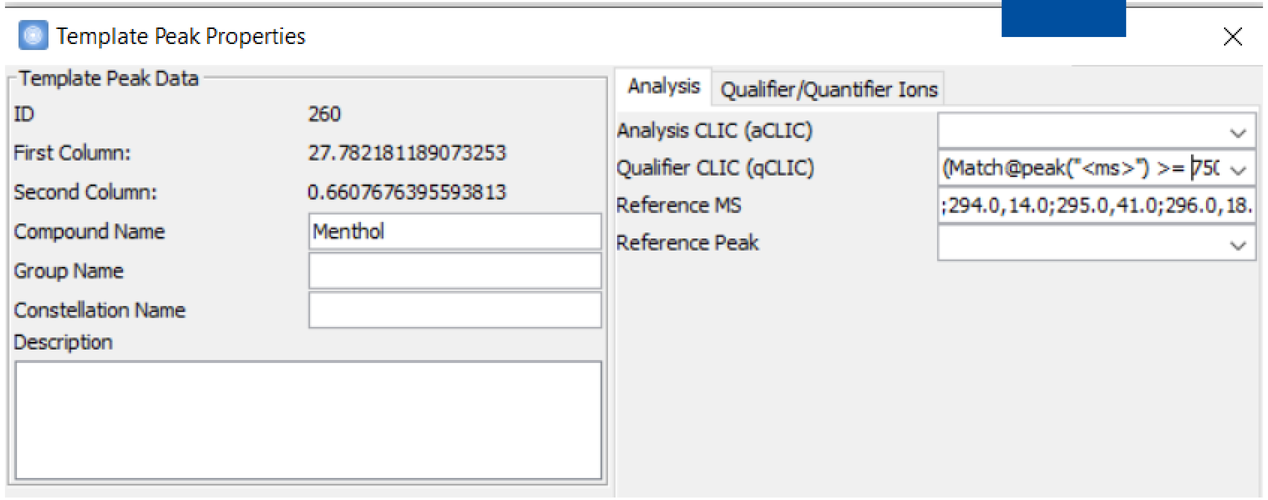
Additional metadata informative for identity and concentration

- ✓ MS spectrum
- ✓ Detector response
- ✓ Diagnostic ions/ratios
- ✓ LRI …



Matching specificity is verified by MS spectral similarity (NIST Match Factors) between reference and target peak MS spectra

Scripting:
(Match@peak("<ms>") >= 750.0) &
(RMatch@peak("<ms>") >= 750.0)



Template Peak Properties

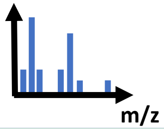
Template Peak Data

ID	260
First Column:	27.782181189073253
Second Column:	0.6607676395593813
Compound Name	Menthol
Group Name	
Constellation Name	
Description	

Analysis CLIC (aCLIC)	
Qualifier CLIC (qCLIC)	(Match@peak("<ms>") >= 75(
Reference MS	;294.0,14.0;295.0,41.0;296.0,18.
Reference Peak	

Context – common application

MS



✓

Specificity - **Molecular fingerprint**
High **sensitivity**
Accurate mass information

✗

Limited linearity range
Response stability over time could be an issue

(IMAGE) PATTERN RECOGNITION
Specificity level 1 relative retention ^1t_R and ^2t_R
Specificity level 2 EI fragmentation pattern similarity

FID

✓

High **stability** over time
Broad **linearity** range
Use of **FID-predicted RRFs**

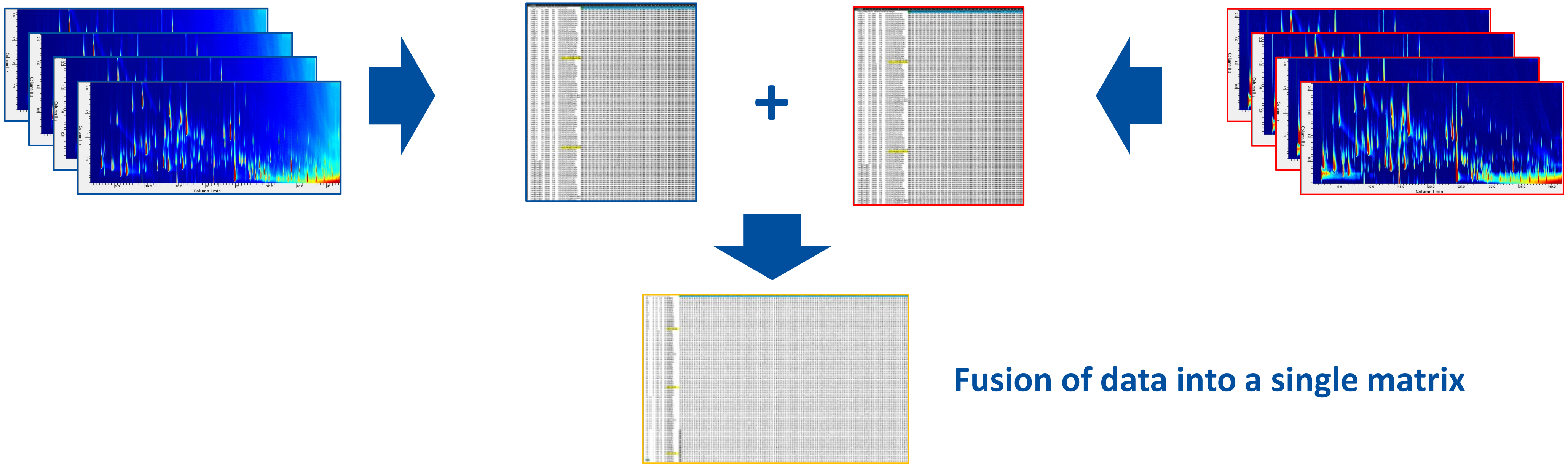
✗

No confirmatory
No molecular diagnosis

(IMAGE) PATTERN RECOGNITION
Specificity level 1 relative retention ^1t_R and ^2t_R
~~Specificity level 2 EI fragmentation pattern similarity~~

Channel 1 data processing

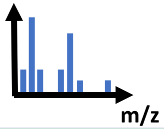
Channel 2 data processing



Fusion of data into a single matrix

Current application

MS



Specificity - **Molecular fingerprint**
High **sensitivity**
Accurate mass information

Limited linearity range
Response stability over time could be an issue

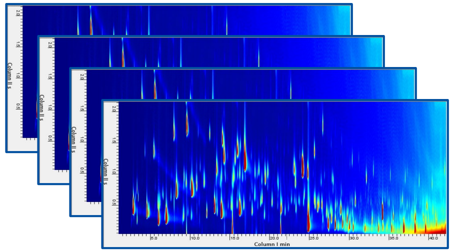
(IMAGE) PATTERN RECOGNITION
Specificity level 1 relative retention ^1t_R and ^2t_R
Specificity level 2 EI fragmentation pattern similarity

FID

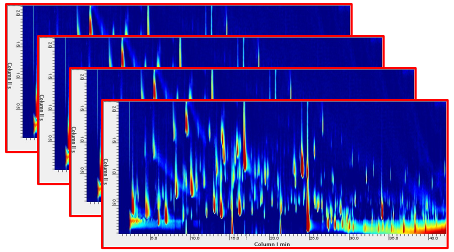
High **stability** over time
Broad **linearity** range
Use of **FID-predicted RRFs**

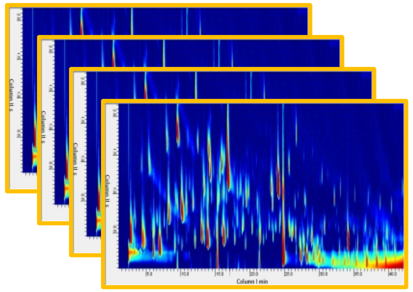
No confirmatory
No molecular diagnosis

(IMAGE) PATTERN RECOGNITION
Specificity level 1 relative retention ^1t_R and ^2t_R
~~Specificity level 2 EI fragmentation pattern similarity~~



+





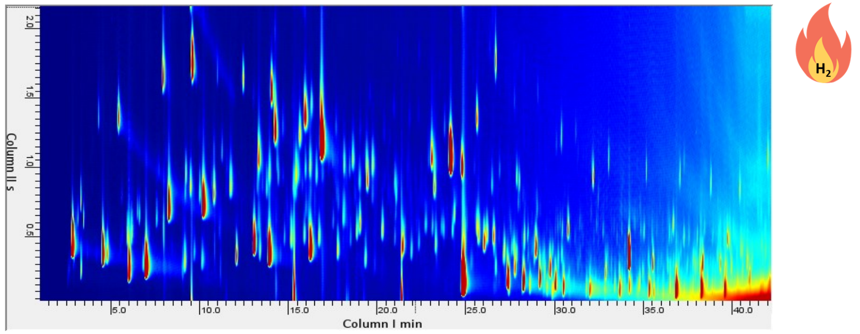
- Fusion of chromatograms
- Data processing for fused chromatograms

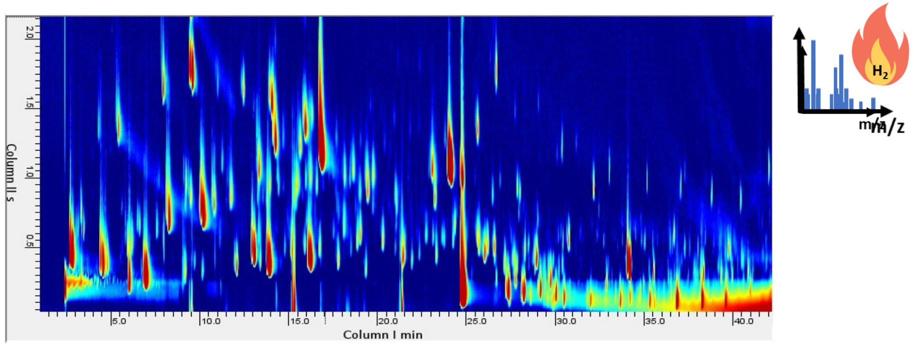
Single data matrix

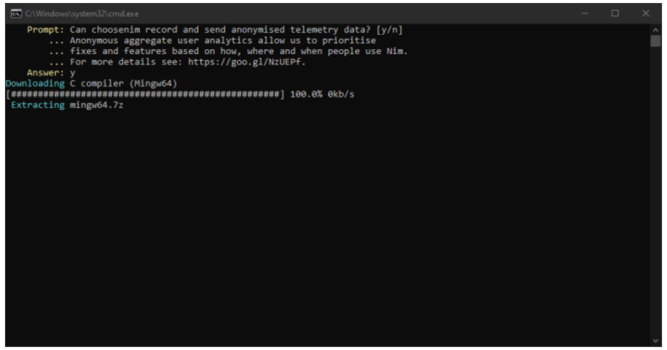
- FID responses
- TIC responses
- Quant responses

Development of MS/FID signal fusion workflow

After preprocessing the MS and FID channels separately, a new image with both traces is created through the use of a script

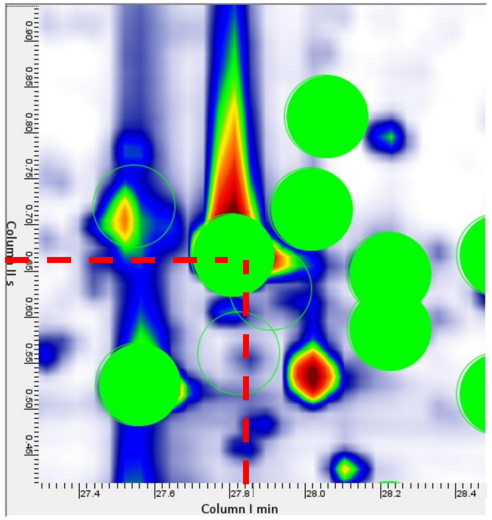






```
    Prompt: Can choosenim record and send anonymised telemetry data? [y/n]
        ... Anonymous aggregate user analytics allow us to prioritise
        ... fixes and features based on how, where and when people use Nim.
        ... For more details see: https://goo.gl/NzUEPf.
    Answer: y
Downloading C compiler (Mingw64)
[###################################################] 100.0% 0kb/s
 Extracting mingw64.7z
```

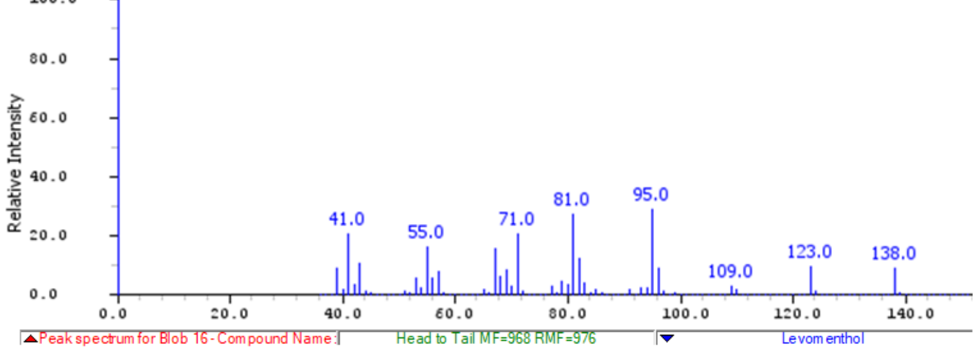
GC Image
Software for Multidimensional Chromatography



2D-peaks have unique time coordinates (retention times)

Additional metadata informative for **identity** and concentration
- ✓ **MS spectrum**
- ✓ FID response
- ✓ Diagnostic ions/ratios
- ✓ LRI …

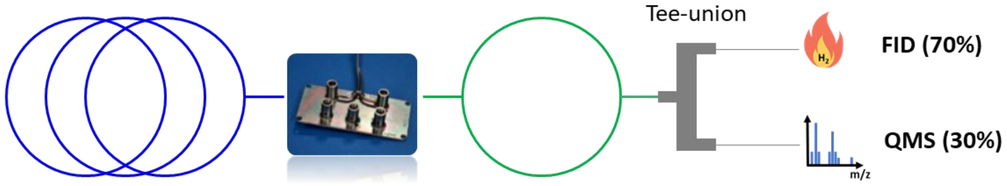
FID channel coded in 0 *m/z* fragment



Applications

Food volatilomics fingerprinting

Corylus avellana L.



^{1}D - Polar Heavy Wax (J&W)
20 m × 0.18 mm × 0.18 µm
He carrier @ **0.4 mL/min**

^{2}D - Medium polarity OV17
1.8 m × 0.18 mm × 0.18 µm
He carrier @ **10 mL/min**

Contents lists available at ScienceDirect
Journal of Chromatography A
journal homepage: www.elsevier.com/locate/chroma

Artificial Intelligence decision-making tools based on comprehensive two-dimensional gas chromatography data: the challenge of quantitative volatilomics in food quality assessment

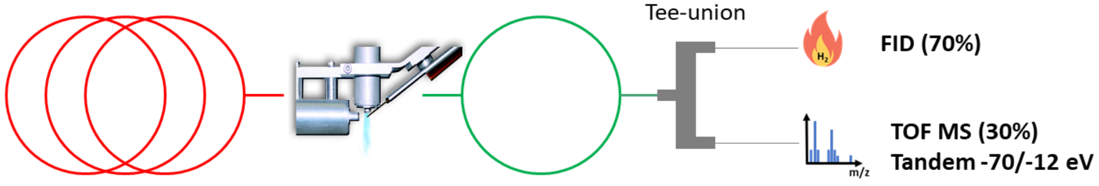
Simone Squara[a], Andrea Caratti[a], Angelica Fina[a], Erica Liberto[a], Nicola Spigolon[b], Giuseppe Genova[b], Giuseppe Castello[b], Irene Cincera[b], Carlo Bicchi[a], Chiara Cordero[a,*]

Samples – Quantitative profiling

- Cultivars
- Harvest regions (Italy and Turkey)
- Storage times (0-6-12 months: T0-1-2)

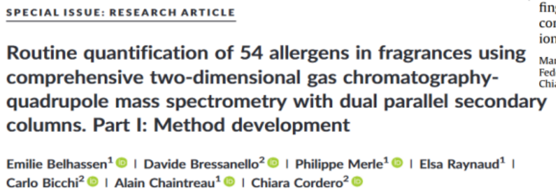
Raw fragrances profiling

Allergenes mixture



^{1}D - Apolar DB1 (J&W)
60 m × 0.25 mm × 0.25 µm
He carrier @ **2.0 mL/min**

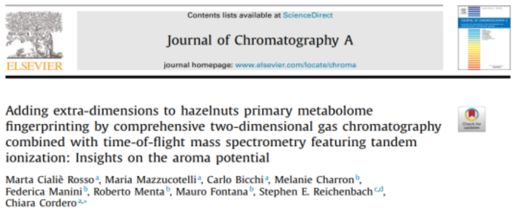
^{2}D - Medium polarity OV1701
1.8 m × 0.18 mm × 0.18 µm
He carrier @ **2.0 mL/min**



SPECIAL ISSUE: RESEARCH ARTICLE

Routine quantification of 54 allergens in fragrances using comprehensive two-dimensional gas chromatography-quadrupole mass spectrometry with dual parallel secondary columns. Part I: Method development

Emilie Belhassen[1] | Davide Bressanello[2] | Philippe Merle[1] | Elsa Raynaud[1] | Carlo Bicchi[2] | Alain Chaintreau[1] | Chiara Cordero[2]



Contents lists available at ScienceDirect
Journal of Chromatography A
journal homepage: www.elsevier.com/locate/chroma

Adding extra-dimensions to hazelnuts primary metabolome fingerprinting by comprehensive two-dimensional gas chromatography combined with time-of-flight mass spectrometry featuring tandem ionization: Insights on the aroma potential

Marta Cialiè Rosso[a], Maria Mazzucotelli[a], Carlo Bicchi[a], Melanie Charron[b], Federica Manini[b], Roberto Menta[b], Mauro Fontana[b], Stephen E. Reichenbach[c,d], Chiara Cordero[a,*]

Samples – Profiling

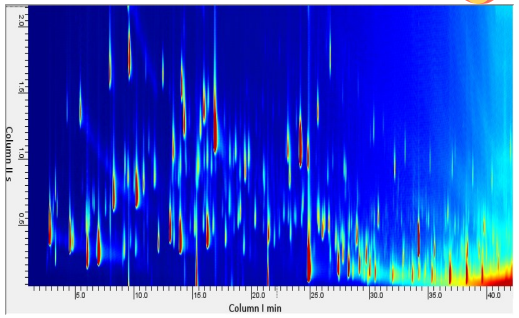
- Allergens mixtures
- Quantification of regulated substances (≈ 60 allergens)[1,2]
- Qualification of different raw fragrances

---

[1] Regulation (EC) No 1223/2009 on cosmetic products, November 2009 [2] Commission Regulation (EU) 2023/1545, July 2023

Pattern Recognition Results

FID

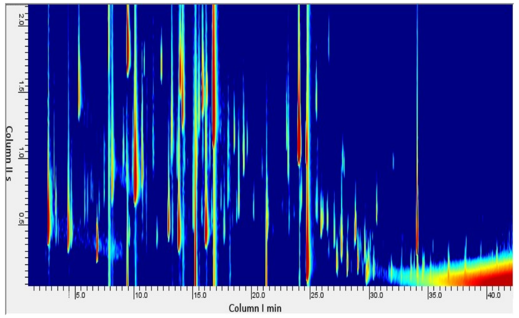


Processing: 30 FID images
Reliable peaks: 96
Detected peaks: 445
Peak matching by position

Reliable Peaks: 2D peaks matching in 50% of the analyzed images

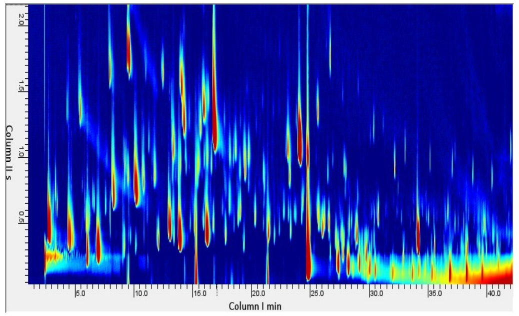
30 samples

MS



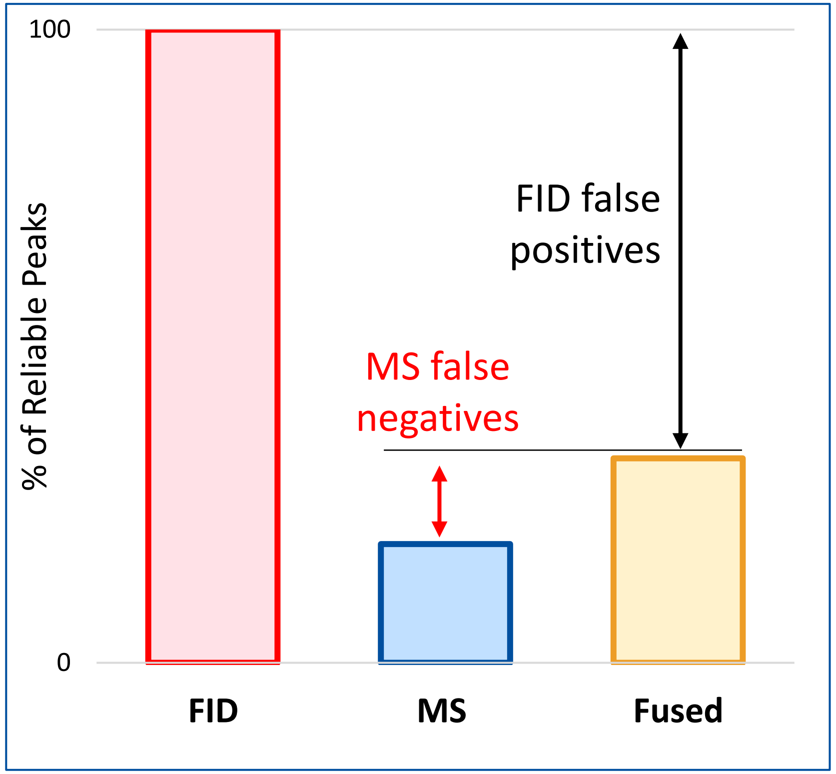
Processing: 30 MS images
Reliable peaks: 20
Detected peaks: 210
Peak matching by position and MS

Fused



Processing: 30 fused images
Peak matching by position and MS
Responses: FID (m/z 0), TIC, single Ion (Qualifiers/quantifiers)
Quantification: FID predicted RRF

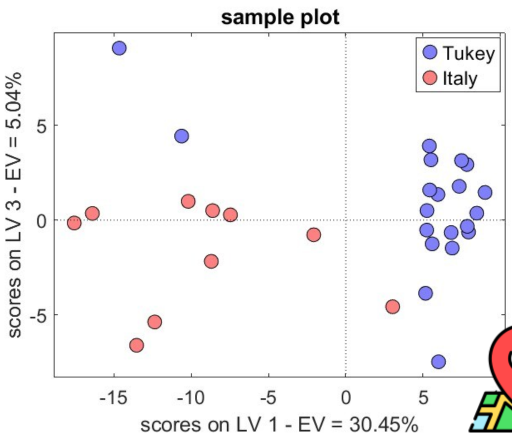
Due to the lack of FID specificity



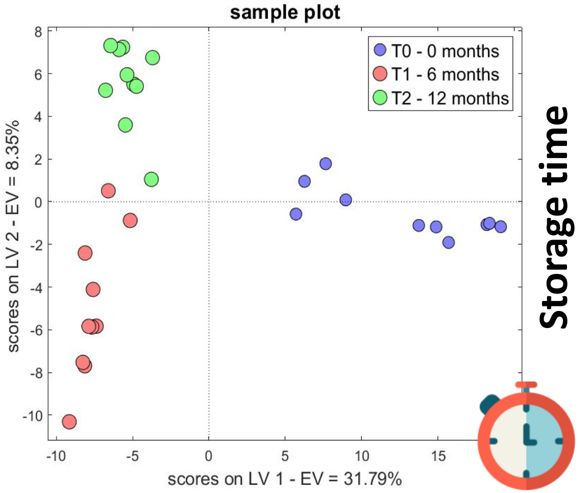
Due to the lack of pattern recognition efficiency

Classification and Markers Quantification Results

Responses



Harvest region

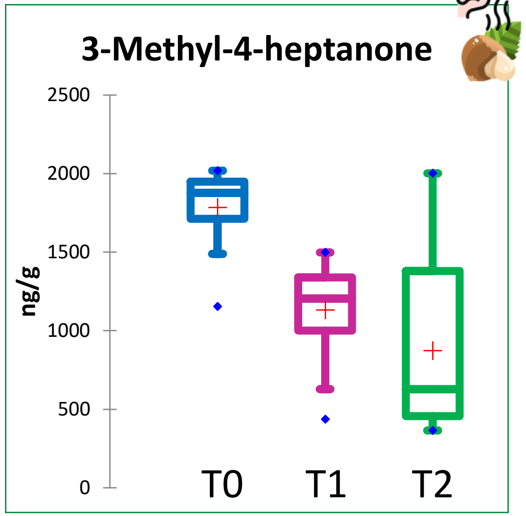


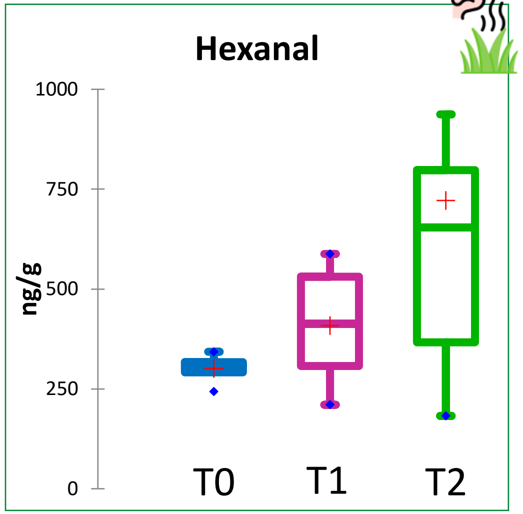
Storage time

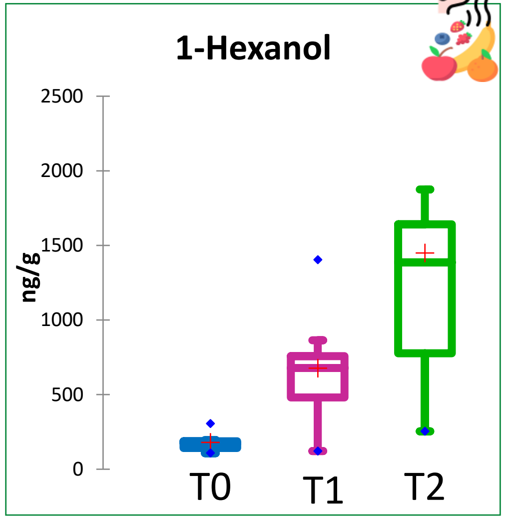
Amounts

MHS-SPME enables accurate quantitation of **45 markers** (ESTD and response factors) including

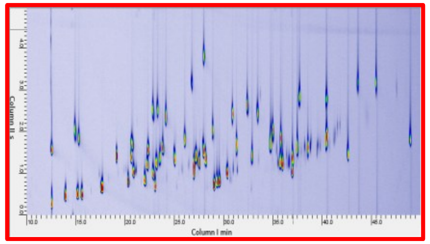
secondary products of lipid oxidation (hexanal, heptanal, octanal, nonanal, (E)-2-octenal, (E)-2-nonenal)
key-aroma compounds (3-methylbutanal, ethyl 2-methylbutanoate, (E)-β-damascenone, 2-nonanone, heptanoic acid etc);
markers of defect (nonanoic acid, butyric acid, 4-heptanol, 1-pentanol, propanoic acid, 2-heptanol, pentanoic acid etc)





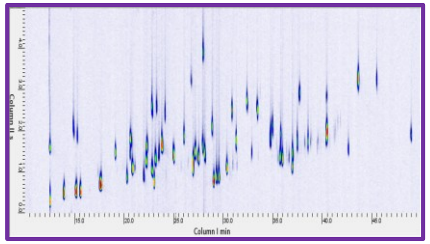


Profiling Results



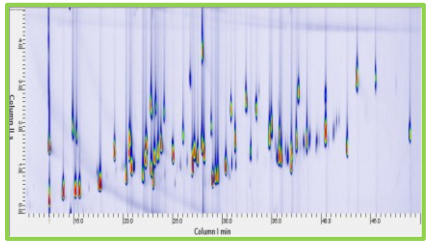
FID

Processing: FID images
Peak matching by position
Quantification by external calibration or predicted RRFs[1]



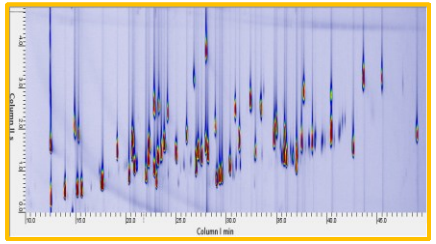
MS
70 eV

Processing: MS (70eV) images
Peak matching by position and MS spectrum
Identification with standard database



MS
12eV

Processing: MS (12eV) images
Peak matching by position and MS spectrum
Limited fragmentation, method flexibility and specificity



Fused

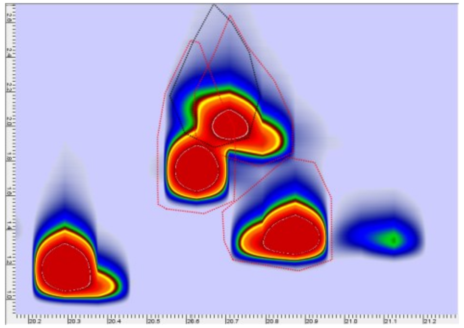
Processing: Fused images with 3 channels
Peak matching by position and MS spectrum
Responses: FID (m/z 0), TIC, single Ions (Qual/Quant)

---

[1] De Saint Laumer, J. Y., Cicchetti, E., Merle, P., Egger, J., & Chaintreau, A., Analytical Chemistry, (2010), 82(15), 6457–6462.

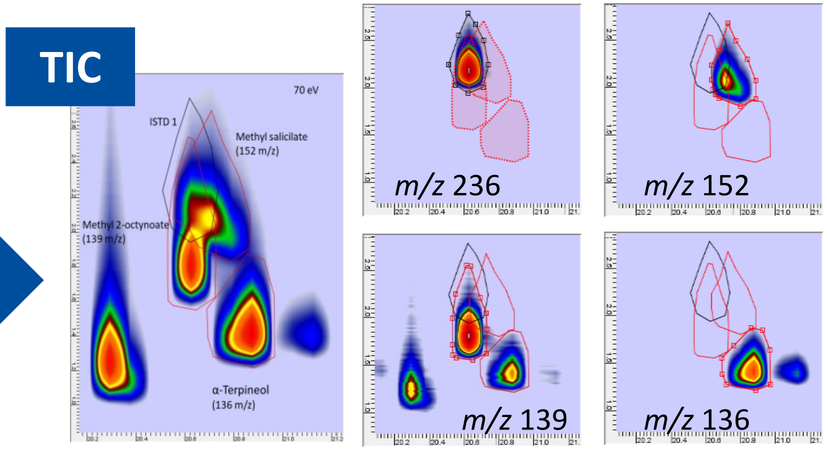
Profiling Results

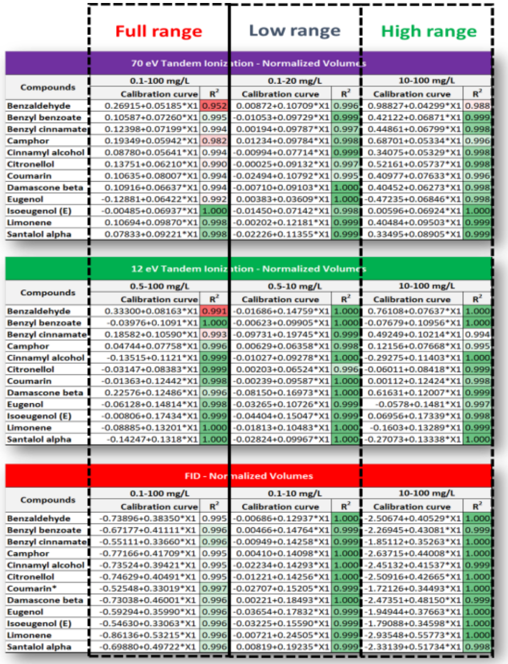
FID data stream



Possible issue: FID co-elution cannot be solved and quantitation results inaccurate

Solved





Full range | Low range | High range

70 eV Tandem Ionization - Normalized Volumes

Compounds	0.1-100 mg/L		0.1-20 mg/L		10-100 mg/L	
	Calibration curve	R^2	Calibration curve	R^2	Calibration curve	R^2
Benzaldehyde	0.26915+0.05185*X1	0.952	0.00872+0.10709*X1	0.996	0.98827+0.04299*X1	0.988
Benzyl benzoate	0.10587+0.07260*X1	0.995	-0.01053+0.09729*X1	0.999	0.42122+0.06871*X1	0.999
Benzyl cinnamate	0.12398+0.07199*X1	0.994	0.00194+0.09787*X1	0.997	0.44861+0.06799*X1	0.998
Camphor	0.19349+0.05942*X1	0.982	0.01234+0.09784*X1	0.998	0.68701+0.05334*X1	0.996
Cinnamyl alcohol	0.08780+0.05641*X1	0.994	-0.00994+0.07714*X1	0.999	0.34075+0.05329*X1	0.998
Citronellol	0.13751+0.06210*X1	0.990	-0.00025+0.09132*X1	0.997	0.52161+0.05737*X1	0.998
Coumarin	0.10635+0.08007*X1	0.994	-0.02494+0.10792*X1	0.995	0.40977+0.07633*X1	0.996
Damascone beta	0.10916+0.06637*X1	0.994	-0.00710+0.09103*X1	1.000	0.40452+0.06273*X1	0.998
Eugenol	-0.12881+0.06422*X1	0.992	0.00383+0.03609*X1	1.000	-0.47235+0.06846*X1	0.998
Isoeugenol (E)	-0.00485+0.06937*X1	1.000	-0.01450+0.07142*X1	0.998	0.00596+0.06924*X1	1.000
Limonene	0.10694+0.09870*X1	0.998	-0.00202+0.12181*X1	0.999	0.40484+0.09503*X1	0.999
Santalol alpha	0.07833+0.09221*X1	0.998	-0.02226+0.11355*X1	0.999	0.33495+0.08905*X1	0.999

12 eV Tandem Ionization - Normalized Volumes

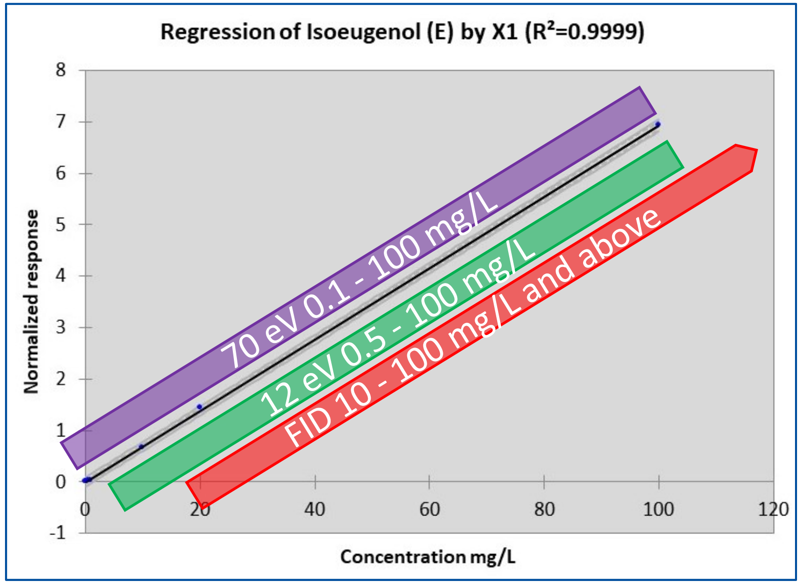
Compounds	0.5-100 mg/L		0.5-10 mg/L		10-100 mg/L	
	Calibration curve	R^2	Calibration curve	R^2	Calibration curve	R^2
Benzaldehyde	0.33300+0.08163*X1	0.991	-0.01686+0.14759*X1	1.000	0.76108+0.07637*X1	1.000
Benzyl benzoate	-0.03976+0.1091*X1	1.000	-0.00623+0.09905*X1	1.000	-0.07679+0.10956*X1	1.000
Benzyl cinnamate	0.18582+0.10590*X1	0.993	-0.09731+0.19745*X1	0.999	0.49249+0.10214*X1	0.994
Camphor	0.04744+0.07758*X1	0.996	0.00629+0.06358*X1	0.998	0.12156+0.07668*X1	0.995
Cinnamyl alcohol	-0.13515+0.1121*X1	0.999	-0.01027+0.09278*X1	1.000	-0.29275+0.11403*X1	1.000
Citronellol	-0.03147+0.08383*X1	0.999	0.00203+0.06524*X1	0.996	-0.06011+0.08418*X1	0.999
Coumarin	-0.01363+0.12442*X1	0.998	-0.00239+0.09587*X1	1.000	0.00112+0.12424*X1	0.998
Damascone beta	0.22576+0.12486*X1	0.996	-0.08150+0.16973*X1	1.000	0.61631+0.12007*X1	0.999
Eugenol	-0.06128+0.14814*X1	0.998	-0.03265+0.10726*X1	0.999	-0.0578+0.1481*X1	0.997
Isoeugenol (E)	-0.00806+0.17434*X1	0.999	-0.04404+0.15047*X1	0.999	0.06956+0.17339*X1	0.998
Limonene	-0.08885+0.13201*X1	1.000	-0.01813+0.10483*X1	1.000	-0.1603+0.13289*X1	0.999
Santalol alpha	-0.14247+0.1318*X1	1.000	-0.02824+0.09967*X1	1.000	-0.27073+0.13338*X1	1.000

FID - Normalized Volumes

Compounds	0.1-100 mg/L		0.1-10 mg/L		10-100 mg/L	
	Calibration curve	R^2	Calibration curve	R^2	Calibration curve	R^2
Benzaldehyde	-0.73896+0.38350*X1	0.995	-0.00686+0.12937*X1	1.000	-2.50674+0.40529*X1	1.000
Benzyl benzoate	-0.67177+0.41111*X1	0.996	0.00466+0.14764*X1	0.999	-2.26945+0.43081*X1	0.999
Benzyl cinnamate	-0.55111+0.33660*X1	0.996	-0.00949+0.14258*X1	0.999	-1.85112+0.35263*X1	1.000
Camphor	-0.77166+0.41709*X1	0.995	0.00410+0.14098*X1	1.000	-2.63715+0.44008*X1	1.000
Cinnamyl alcohol	-0.73524+0.39421*X1	0.995	-0.02234+0.14293*X1	1.000	-2.45132+0.41537*X1	0.999
Citronellol	-0.74629+0.40491*X1	0.995	-0.01221+0.14256*X1	1.000	-2.50916+0.42665*X1	1.000
Coumarin*	-0.52548+0.33019*X1	0.997	-0.02707+0.15205*X1	0.999	-1.72126+0.34493*X1	1.000
Damascone beta	-0.73038+0.46001*X1	0.996	0.00221+0.18493*X1	1.000	-2.47351+0.48150*X1	0.999
Eugenol	-0.59294+0.35990*X1	0.996	-0.03654+0.17832*X1	0.999	-1.94944+0.37663*X1	1.000
Isoeugenol (E)	-0.54630+0.33063*X1	0.996	-0.03225+0.15590*X1	0.999	-1.79088+0.34598*X1	1.000
Limonene	-0.86136+0.53215*X1	0.996	-0.00721+0.24505*X1	0.999	-2.93548+0.55773*X1	1.000
Santalol alpha	-0.69880+0.49722*X1	0.996	0.00819+0.19235*X1	0.999	-2.33139+0.51734*X1	0.998

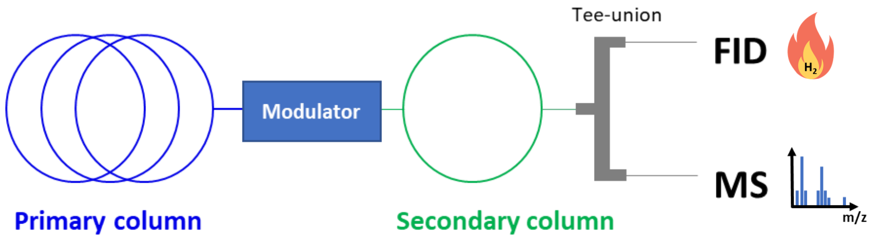
Possible issue: Wide dynamic range of linearity to be exploited with the processing of three data stream with three different workflow

Solved



UNIVERSITÀ
DI TORINO

Conclusion



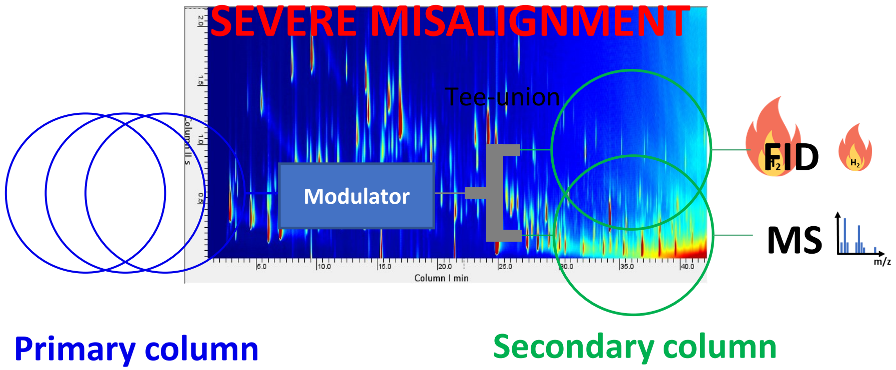
Challenges

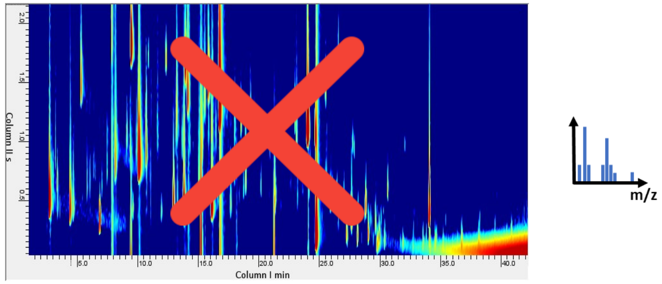
- Pattern complexity
- Accurate fingerprinting
- Different data streams to be processed

Opportunities

- Fusion of parallel raw data
- Chromatographic resolution
- Sensitivity
- Specificity – Exploiting multiple channels

Future perspectives





Pattern recognition can solve the problem by the **allignment** and **registration** of images

Agilent

GC Image
Software for Multidimensional Chromatography

University of Nebraska Lincoln

1404

UNIVERSITÀ
DI TORINO

Thank you for your attention