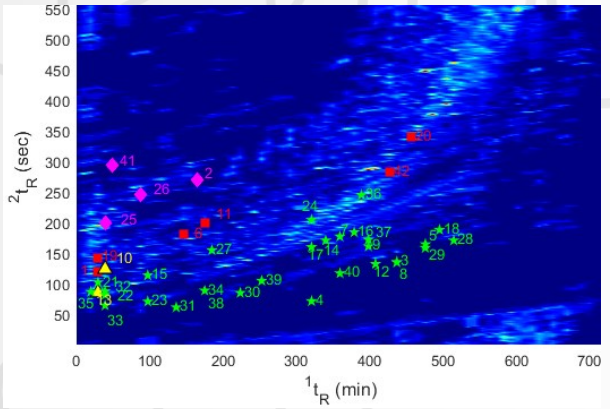
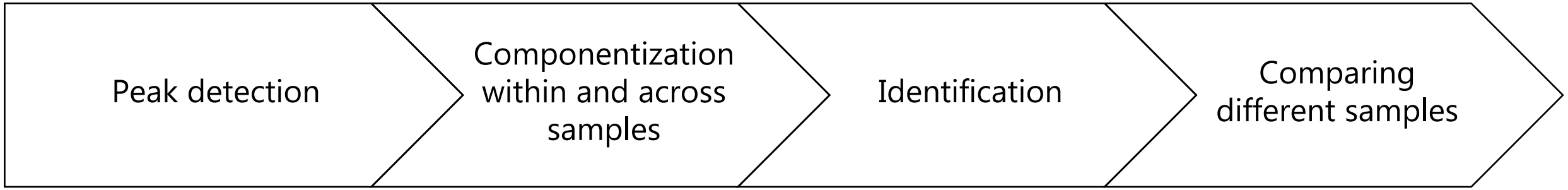
DATA PROCESSING WORKFLOWS FOR NON-TARGET SCREENING ON LC×LC-HRMS DATA: READY TO GO?

Oskar M. Kronik[1], Jan H. Christensen[1], Giorgio Tomasi[1], Paul-Albert Schneide[2,3], Selina Tisler[1], Nikoline J. Nielsen[1]

KØBENHAVNS UNIVERSITET



Overview of non-target screening workflow



Risk of loosing relevant information

Data size

Challenges

- Low intensity compounds
- Slow processing

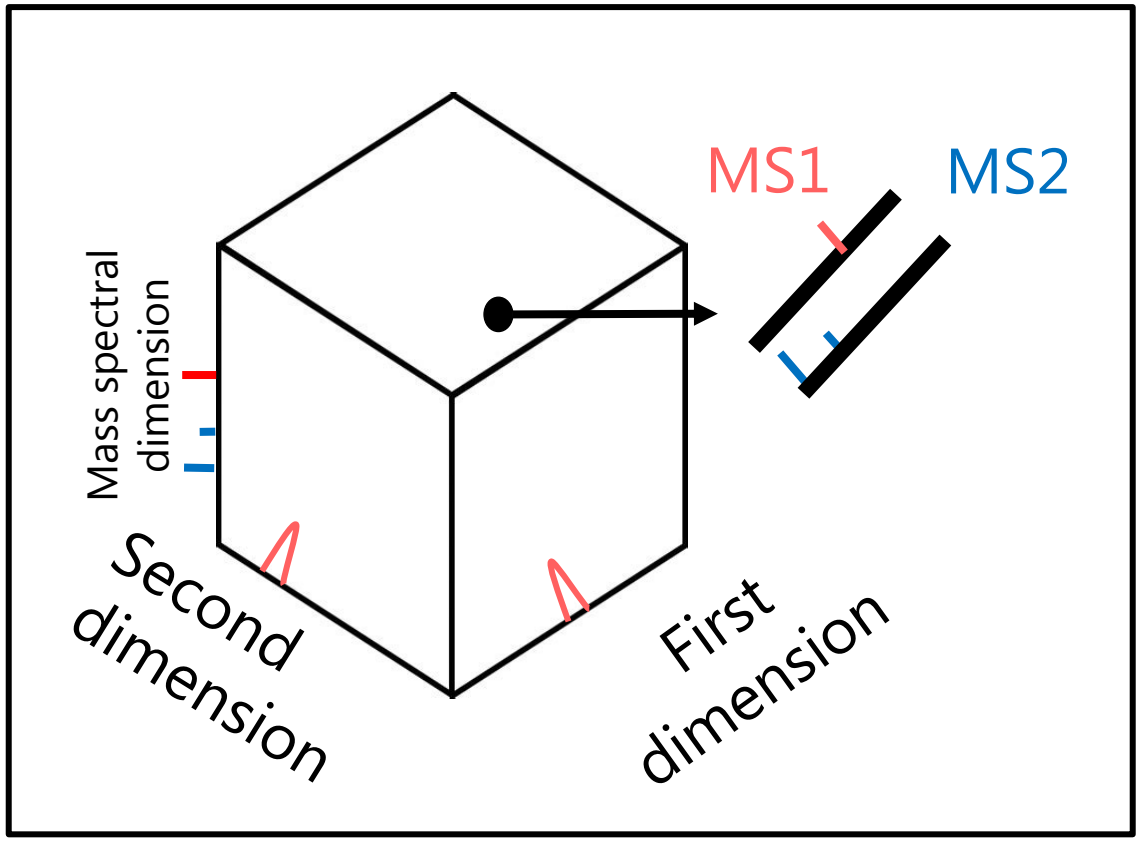
- Poor grouping of m/z

- Non-diagnostic ions in mass spectrum

Data structure of LC×LC-HRMS in data independent mode

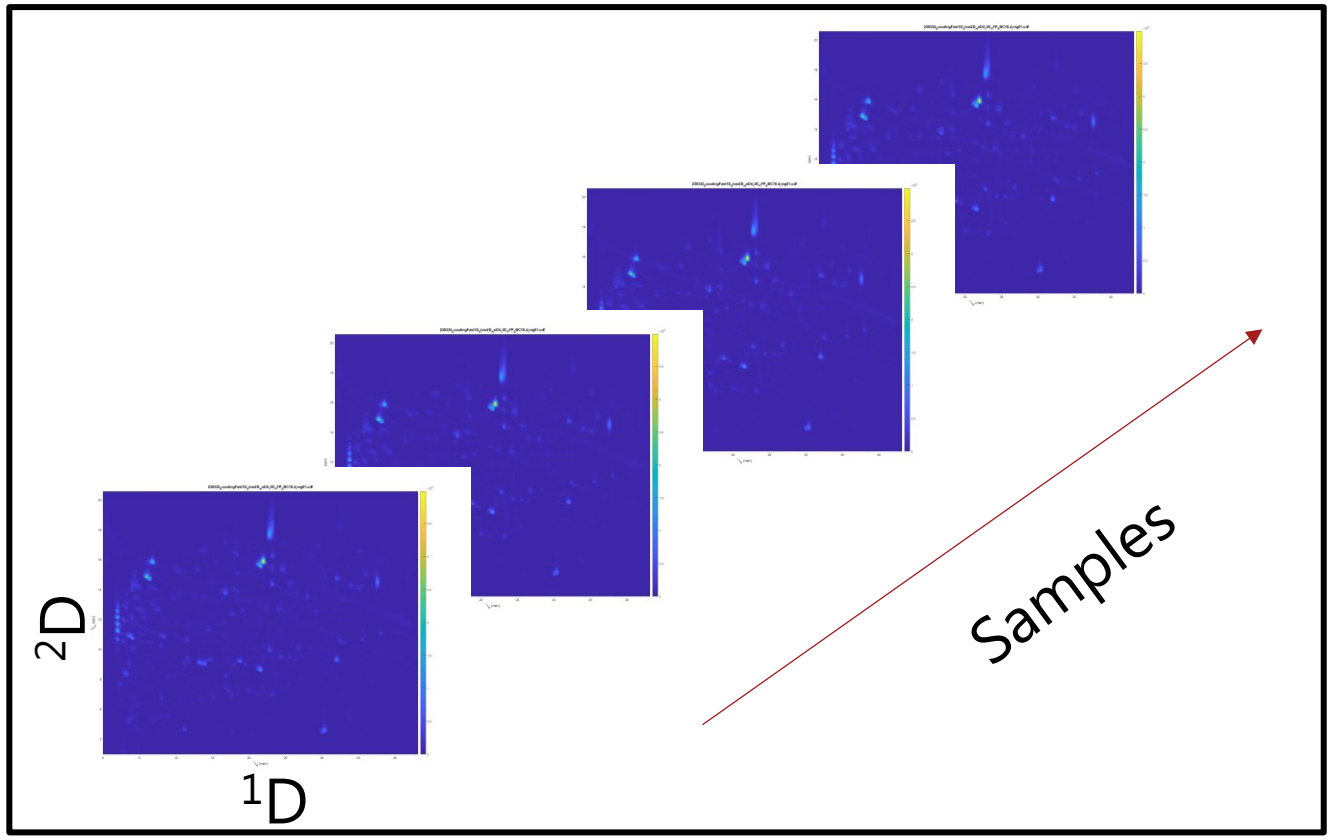
Single sample

Data size: 5-7 GB



Multiple samples

Data size: too large for normal computers!



We need to:

1. Compress the data without loosing information
   Methods: Region of Interest

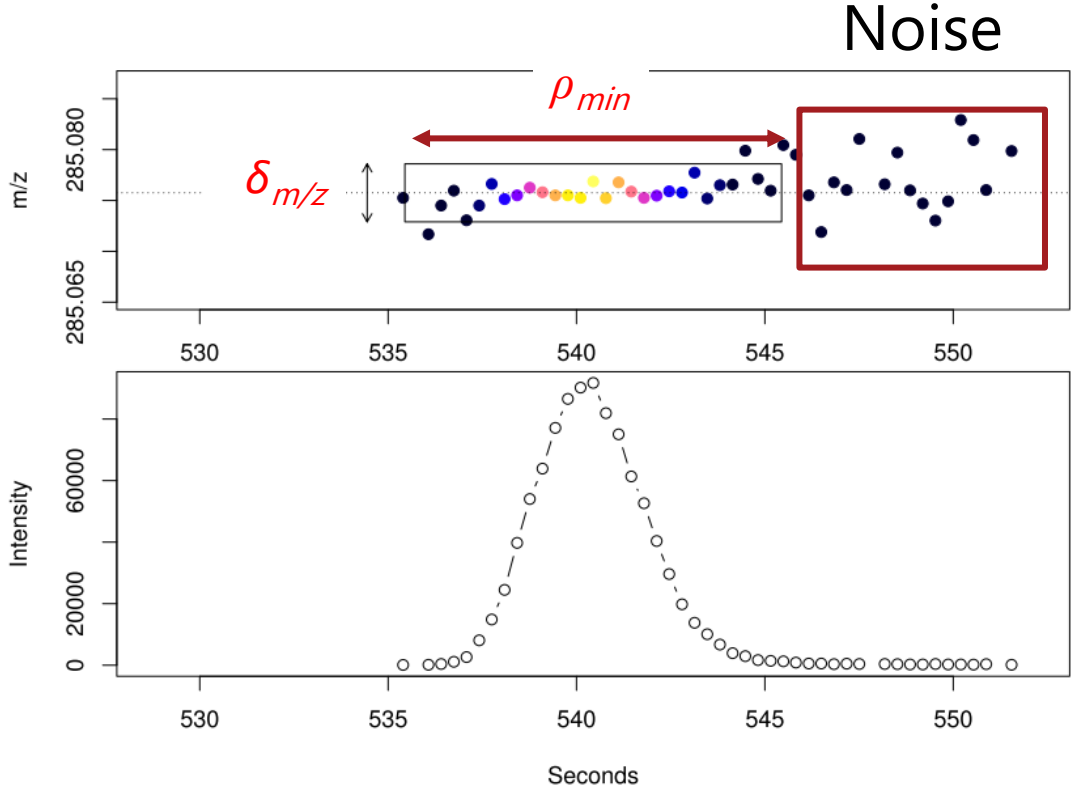
2. Extract high-quality mass spectra of trace-level compounds
   Methods: mass filtering, multivariate curve-resolution

3. Group compounds across samples

Focus of today

Region of interest yields high mass accuracy and high data compression

Input parameters
- m/z deviation ($\delta_{m/z}$)
- Minimum peak width (ρ_{min})

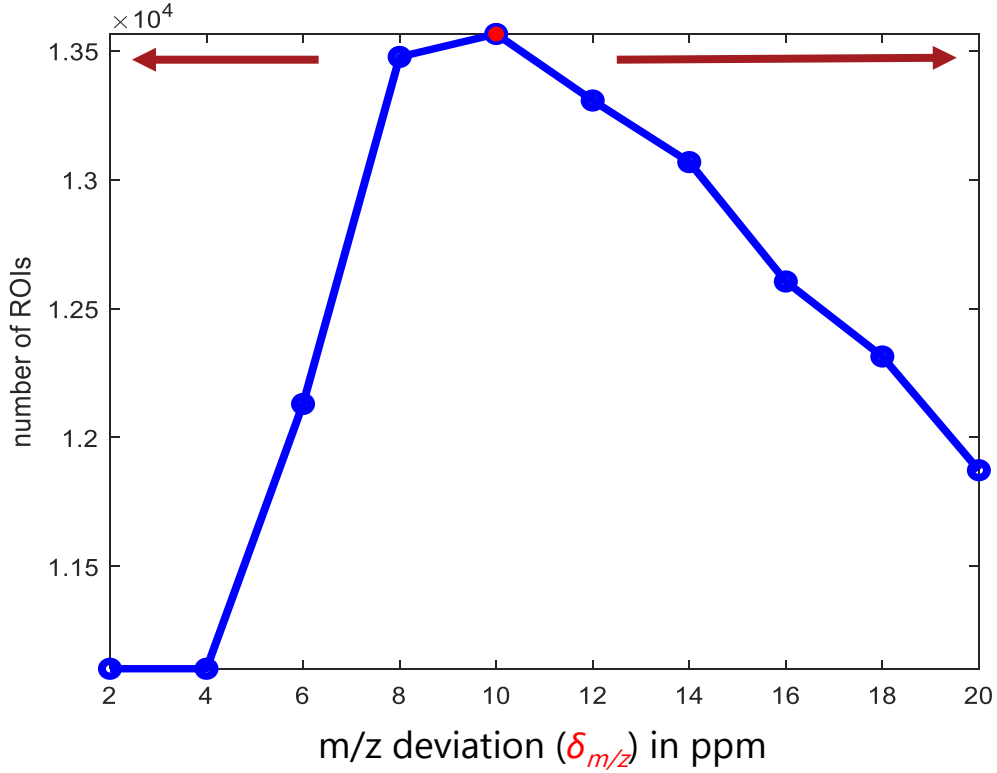


Tautenhahn et al., BMC Bioinformatics 2008, 9:504 (2008)

Optimization
m/z deviation ($\delta_{m/z}$)

Low $\delta_{m/z}$:
Split ROIs

High $\delta_{m/z}$
Loss of resolution

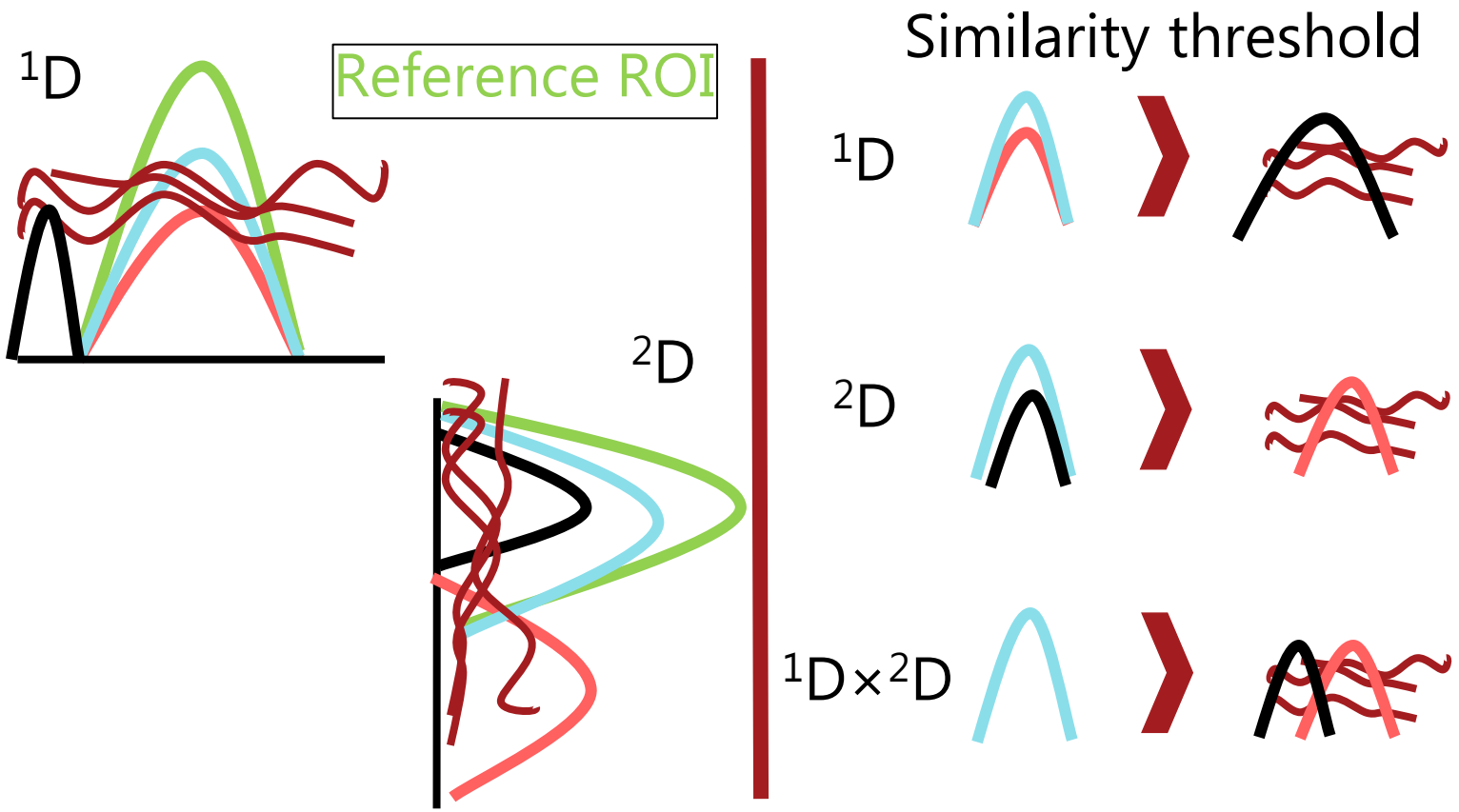


Kronik et al., Analytical and Bioanalytical Chemistry (2025)

Leveraging the two-dimensional data through mass filtering reduces data complexity

Mass filtering:
Similarity between ROIs in the ^1D and ^2D

Relevant ROIs after mass filtering



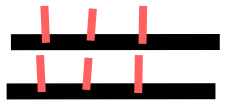
Schneide and Kronik. Analytical Chemistry (2025)

The mass spectral quality improves with mass filtering and multivariate curve-resolution (MF+MCR)

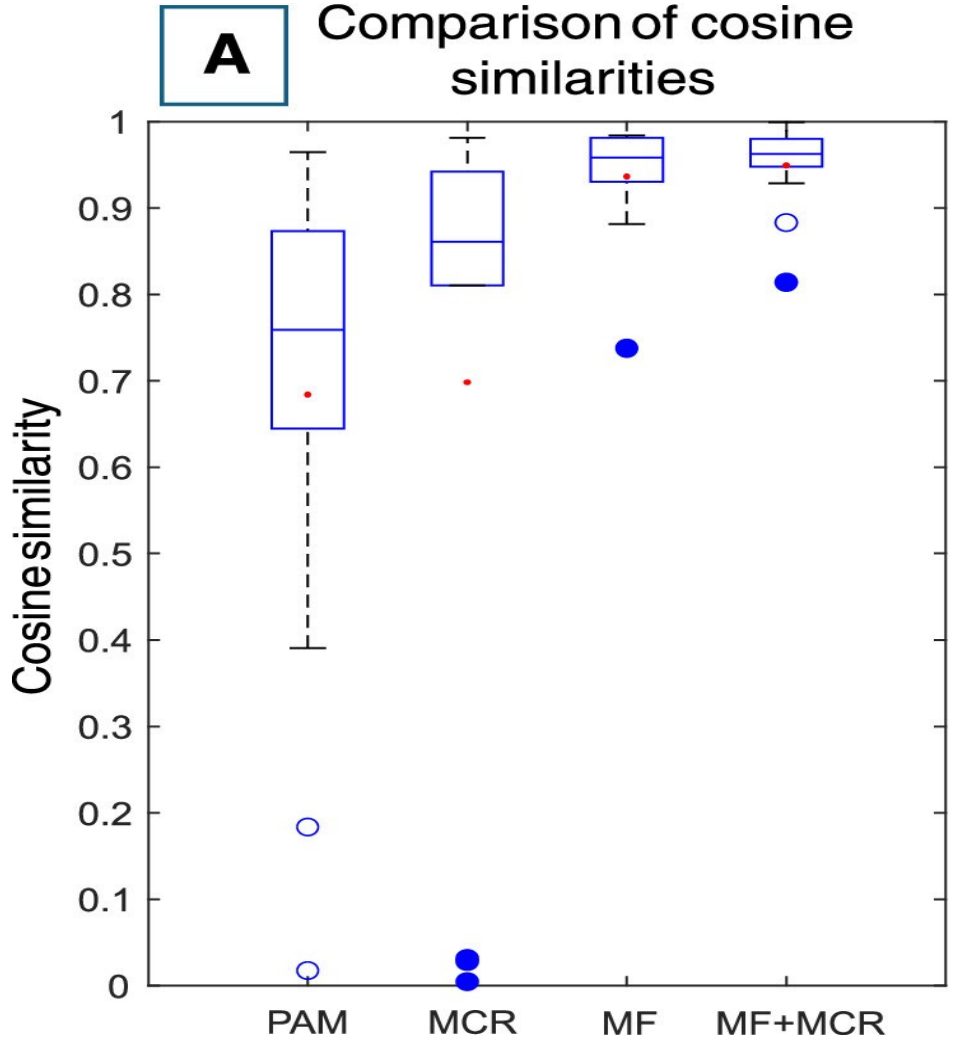
- The **cosine similarity** is a measure of similarity between two vectors

  0 = no similarity =

  1 = identical =

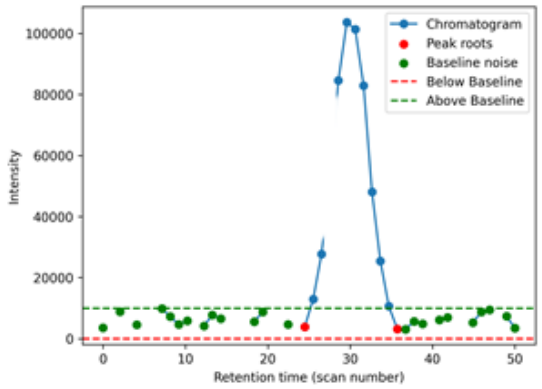


- The cosine similarity was highest for the MF+MCR workflow when comparing towards a reference spectrum

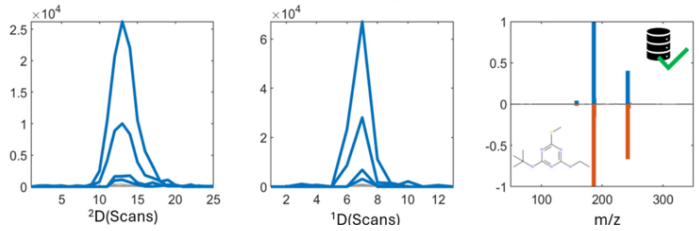


Conclusion

- High data compression rates achieved using ROI
- Allowing for processing of multiple samples at the same time
- Fewer challenges with exceeding the computer's RAM



- Leveraging the data structure improved mass spectral quality for HRMS data obtained in data independent acquisition. Therefore, the number of identifiable compounds increased.



- Future work includes:
  - Validating workflows across diverse datasets to ensure robustness
  - Comparing across samples

Acknowledgements

- The organizing committee
- The VANDALF team involved in acquiring the samples
- The work was funded by:
  - **Innovation Fund Denmark** project VANDALF, Grant Number: 9067-00032A,
  - **Novo Nordisk Foundation** project The Matrix (Grant number: NNF19SA0059348).

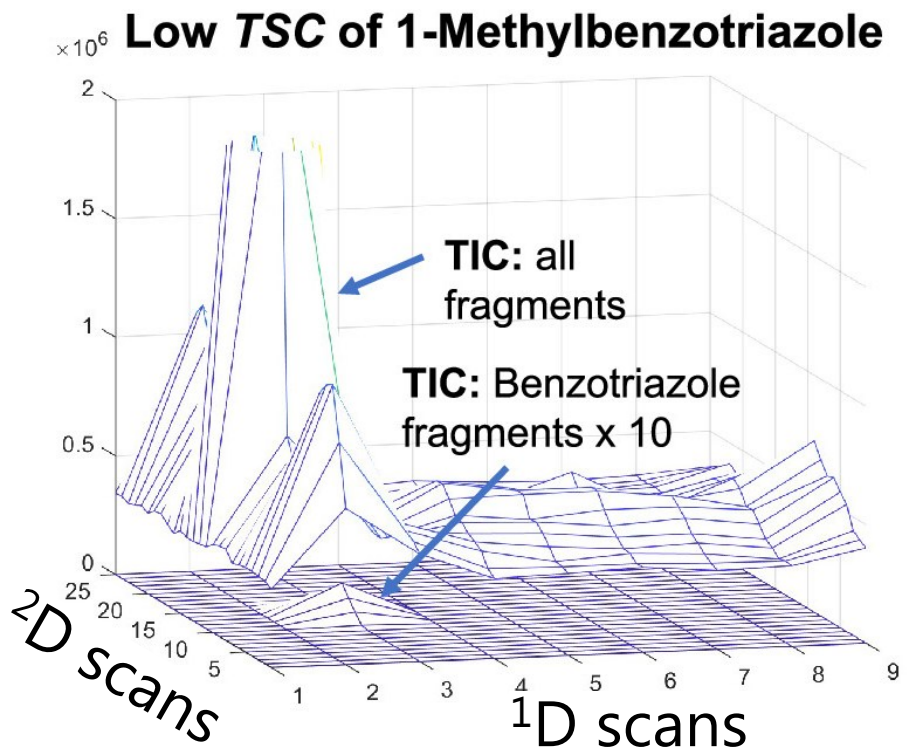
Code availability:
Region of interest: https://github.com/OskarMunkKronik/regionofinterest
Mass filtering: https://github.com/PaulAlbertAnselm/MassFiltering

Check out
Wednesday 2:00 – 2:10: FL22 Nadine Gawlitta - Ambient ultrafine particles: classification, chemical characterization, and quantification of ubiquitous PAHs via DTD-GC×GC-TOFMS

Extra slides

The MF+MCR enables identification of trace-level compounds in municipality wastewater

Relevant pollutants in wastewater effluent can have low intensity

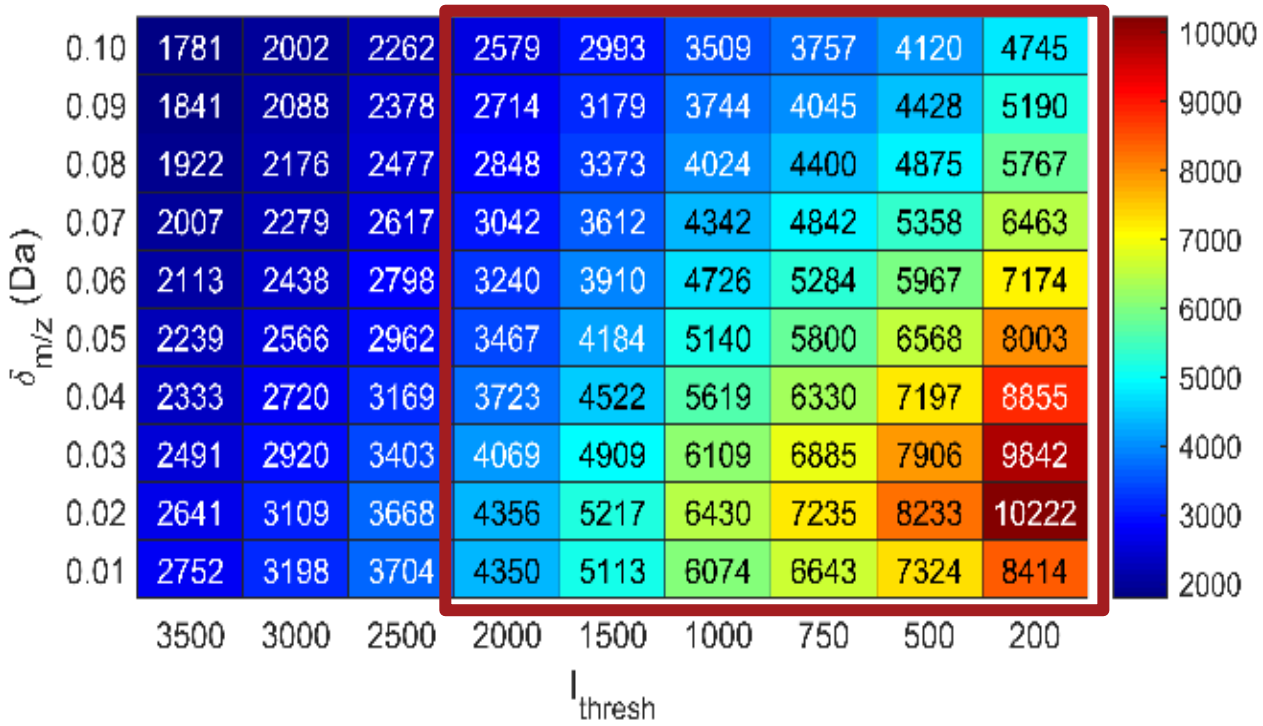


Peak apex method **(PAM)**,

Compound name	TSC	Detected suspects
		PAM
Fexofenadine	0.0608	1
Lamotrigine	0.0441	1
Losartan carboxylic acid	0.0190	1
Citalopram	0.0165	1
Metoprolol	0.0124	1
2-[4-(Diethylamino)-2-hydroxybenzoyl]benzoic acid	0.0116	1
Galaxolidone	0.0096	0
Amisulpride	0.0073	0
DEET	0.0068	1
2-Ethylidene-1,5-dimethyl-3,3-diphenylpyrrolidine	0.0061	0
Ensulizole	0.0061	1
Methadone	0.0048	0
Amitriptyline	0.0047	0
Cetirizine	0.0030	0
Diclofenac	0.0022	0
Losartan carboxaldehyde	0.0014	0
Carbamazepine	0.0013	1
Gabapentin lactam	0.0011	1
Terbutryn	0.0011	0
Clopidogrel carboxylic acid	0.0010	0
Verapamil metabolite	0.0007	0
Venlafaxine	0.0005	0
Climbazole	0.0004	0
1-Methylbenzotriazole	0.0004	0
Ritalinic acid	0.0004	0

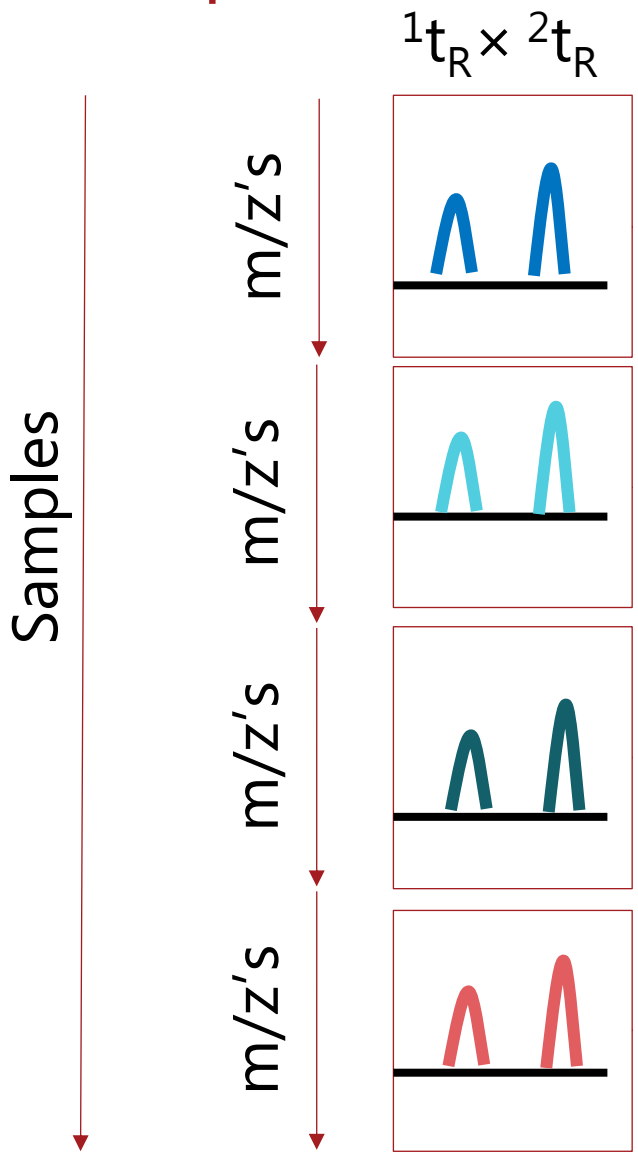
Optimization scheme to preserve the mass spectral resolution

Optimization



Kronik et al., Analytical and Bioanalytical Chemistry (2025)

The compounds detected in each sample can be grouped across samples



Group if

$\Delta m/z < \delta_{m/z}$