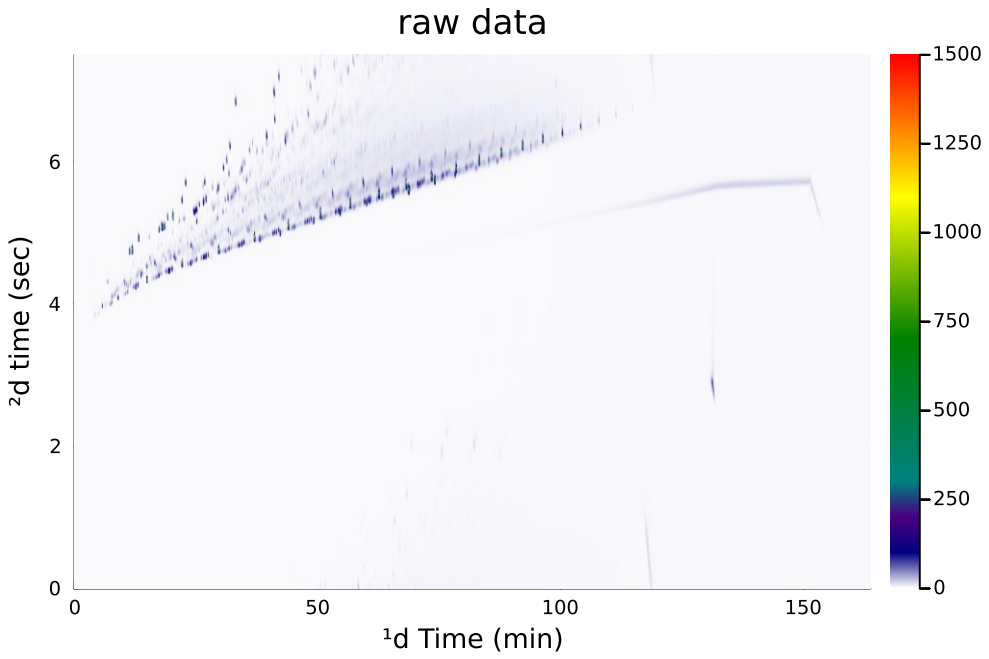
Unlocking Objective Numerical Evaluation of data analysis strategies:

A Novel Platform to Generate Highly Realistic LC×LC and GC×GC data

Nino Milani – University of Amsterdam

How to assess data analysis ?



Our goal:
create a way to perform objective numerical evolution without sacrificing data nuance

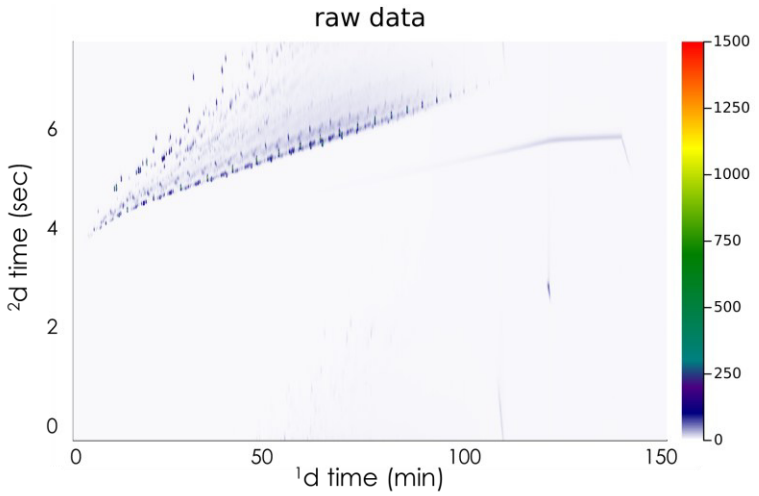
Currently two ways main:

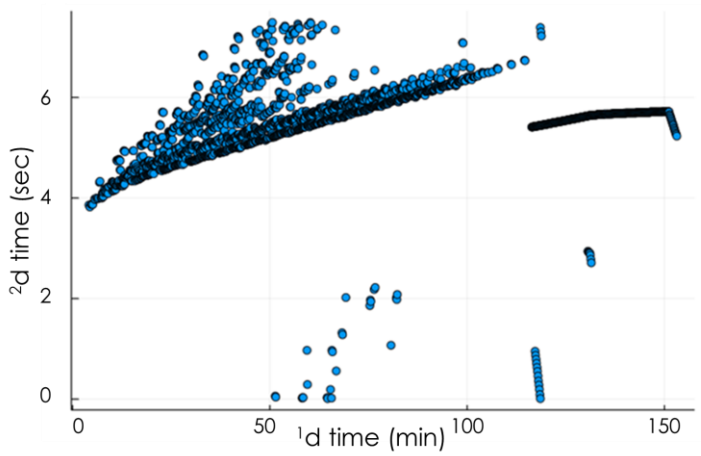
Experimental data:
Pro: Highly relevant
Con: No way to objectively compare results

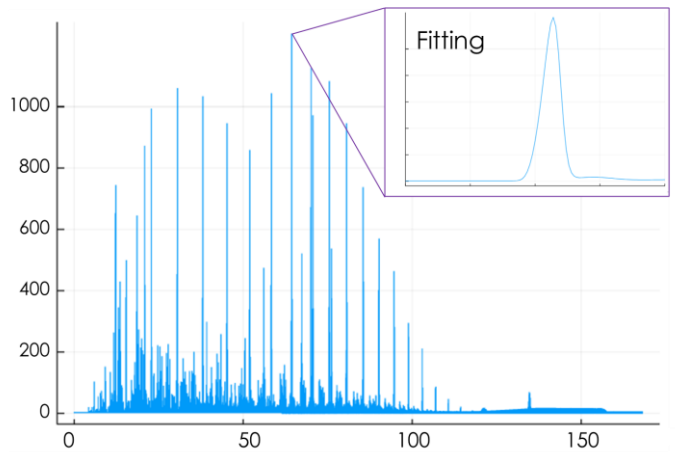
Simulated data:
Pro: Numerically defined
Con: Very general so no guaranty the experimental data behaves the same

Example of good simulated data: L. Niezen et al. Analytica Chimica Acta 1201 (2022) 339605 *https://doi.org/10.1016/j.aca.2022.339605*

Our approach





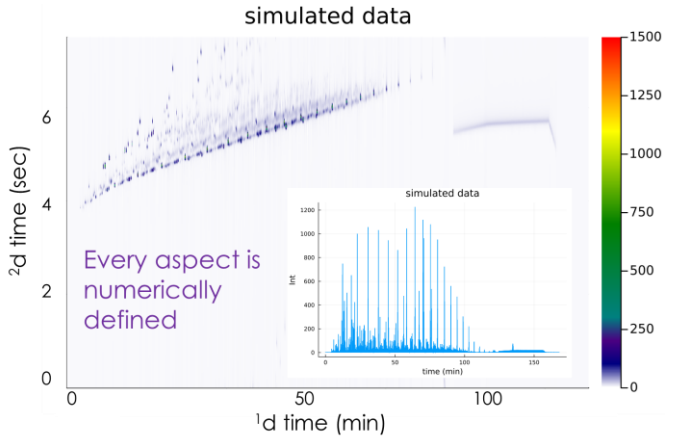


Model equation:

Skewed Lorentz-Normal

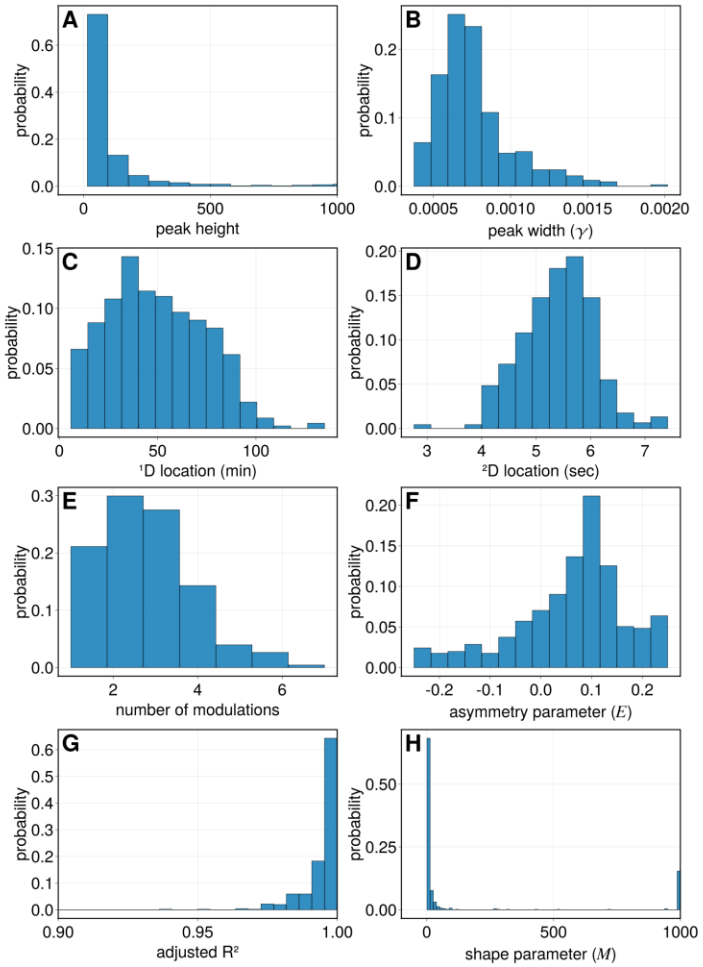
$$f(x) = \frac{h}{\pi\gamma} \times \left(1 + \frac{(x-\mu)^2}{M \times \left(\gamma + E \times (x-\mu)\right)^2}\right)^{-M}$$

Were:
h = peak height
γ = peak width
M = peak shape
E = peak skew
μ = peak position



Milani et al. Analytica Chimica Acta 1312 (2024) 342724 *https://doi.org/10.1016/j.aca.2024.342724*

The set up



Tested parameters:

- ^{1}D width
- ^{2}D width
- peak ratio
- modulation shifting
- peak shape
- peak asymmetry

All taken at several points of the distribution resulting in 70 total parameters

Tested algorithms:

Two-step

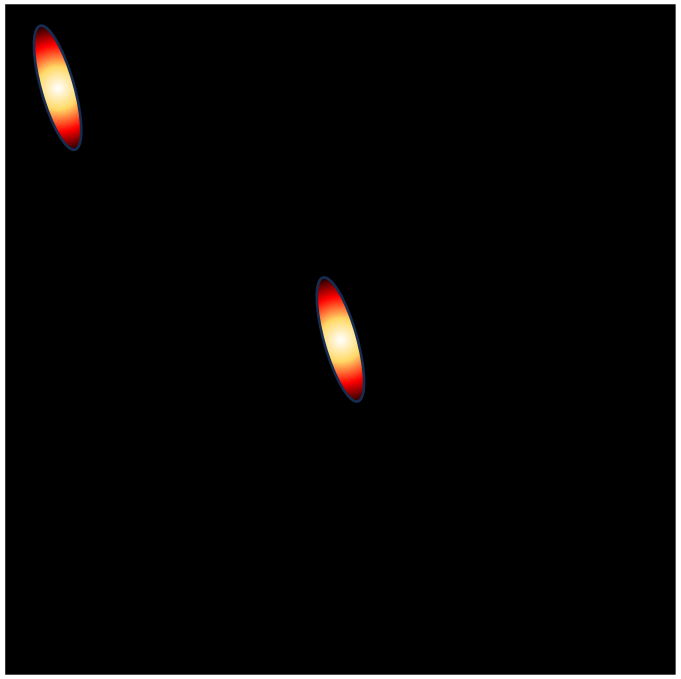
One dimensional peak detection combined with clustering to get the full 2D peak

watershed

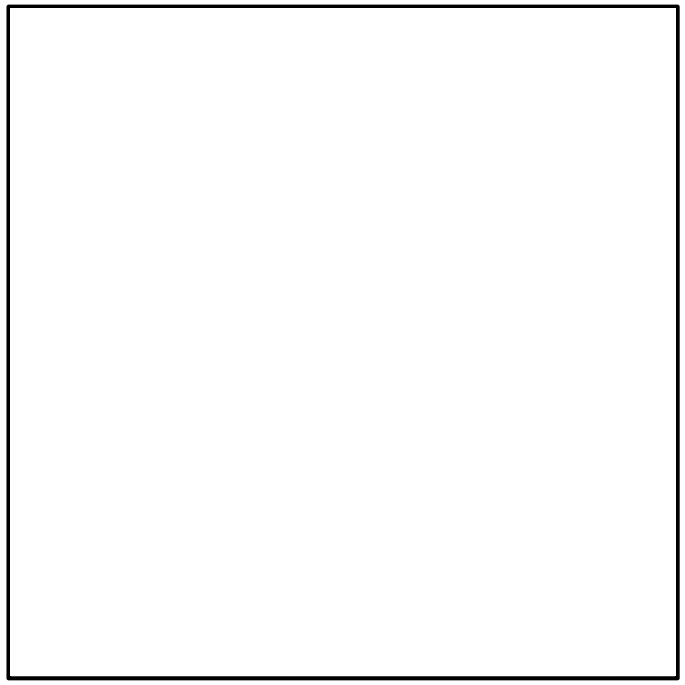
Image segmentation-based processing in only the 2D representation

Simulations

100 x 100 grid

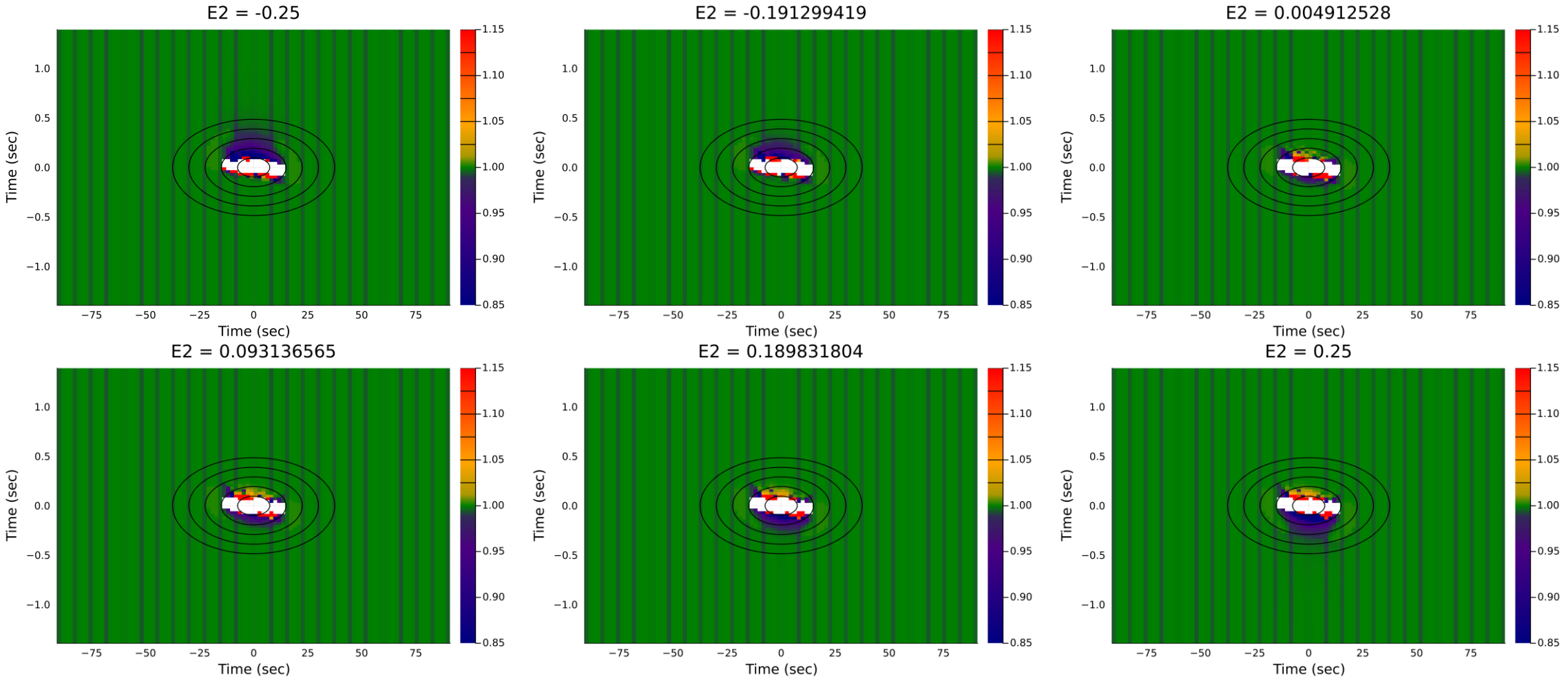


100 x 100 grid

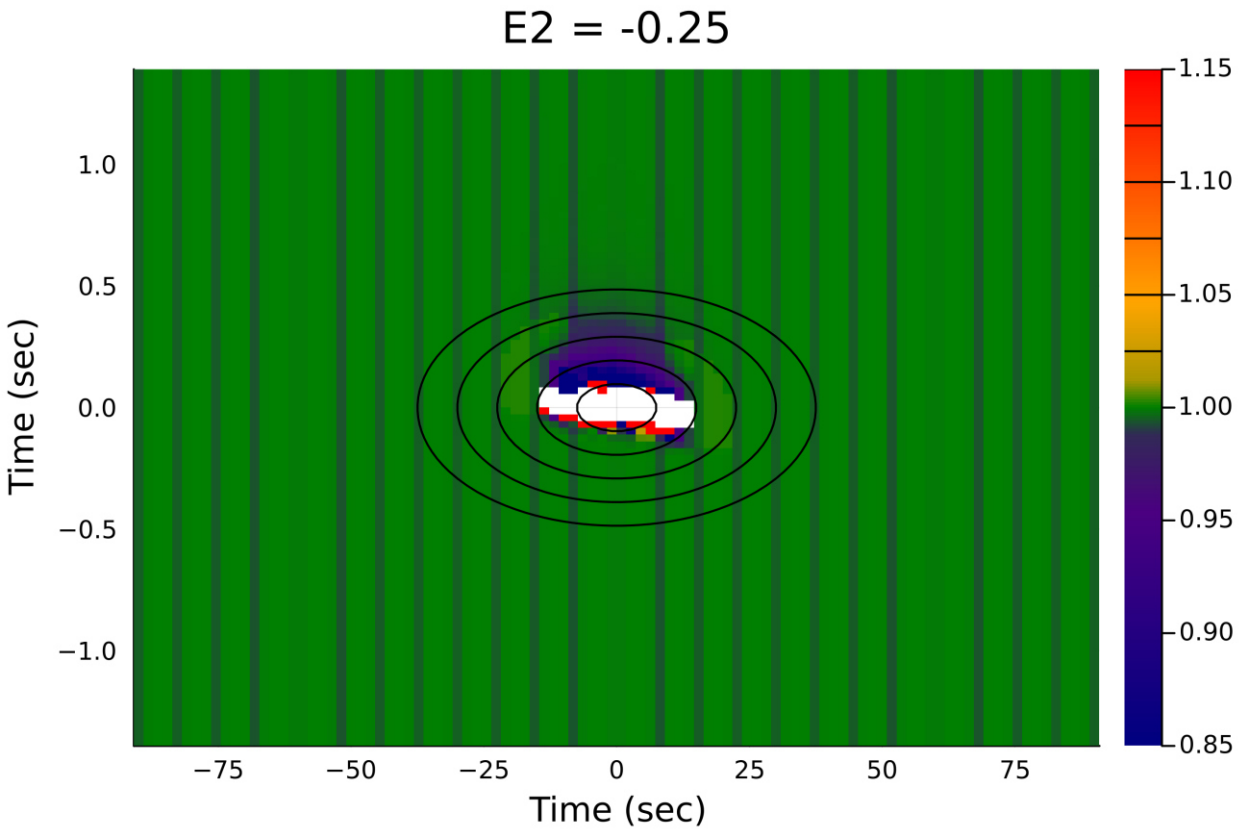


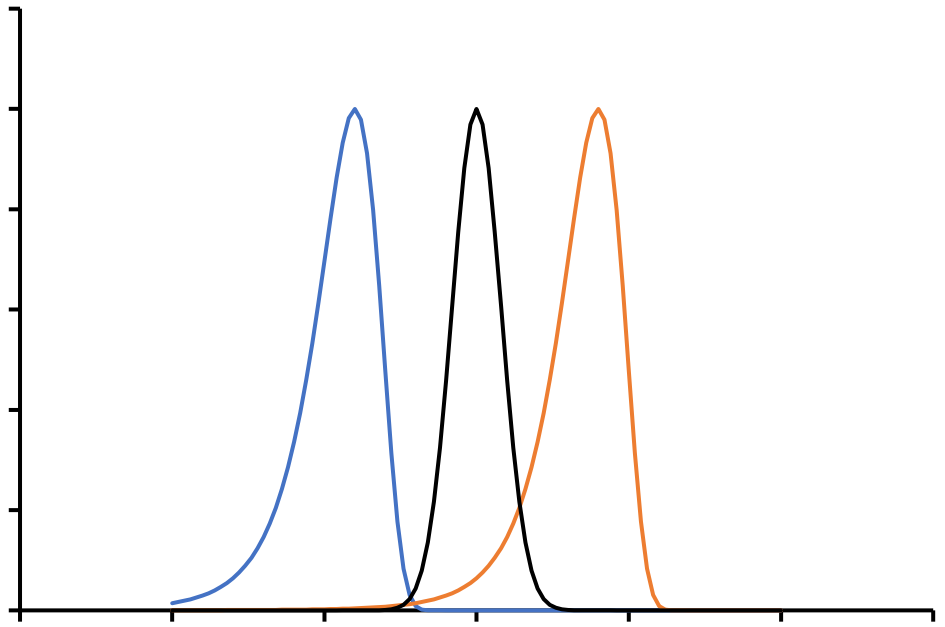
- The peak in the middle stays constant and is there to provide interference with the moving peak.
- The moving peak has a known area, and therefore the recover can be calculated.
- Green means ~100% recovery
- Warm colours are >101%
- Cool colours are <99%

Effect of asymmetry on peak area recovery in two-step peak detection

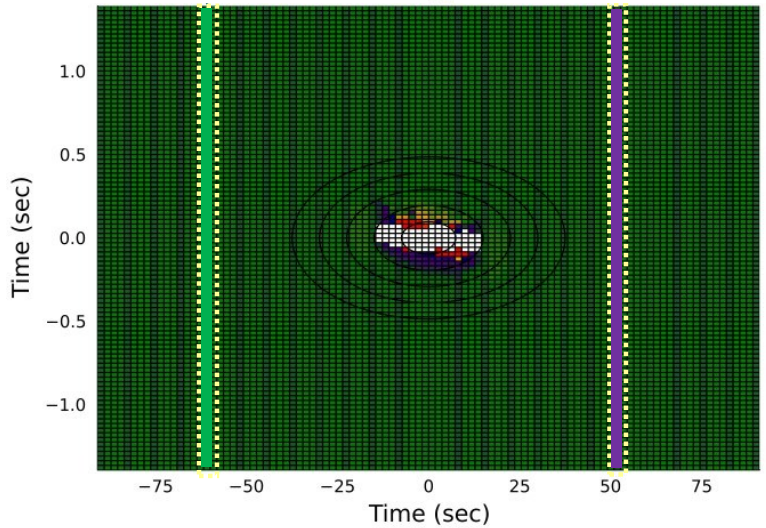


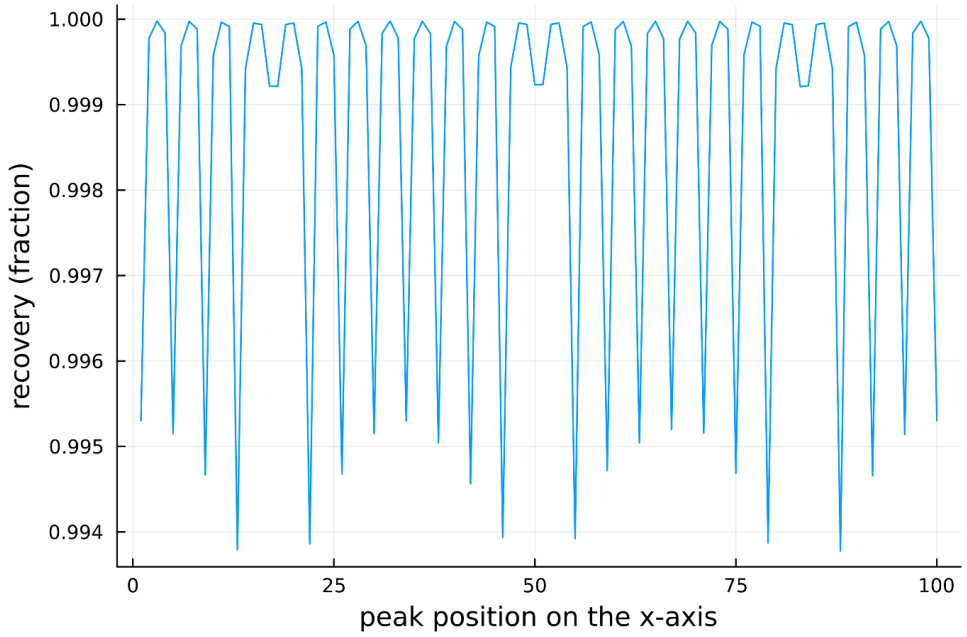
Effect of asymmetry on peak area



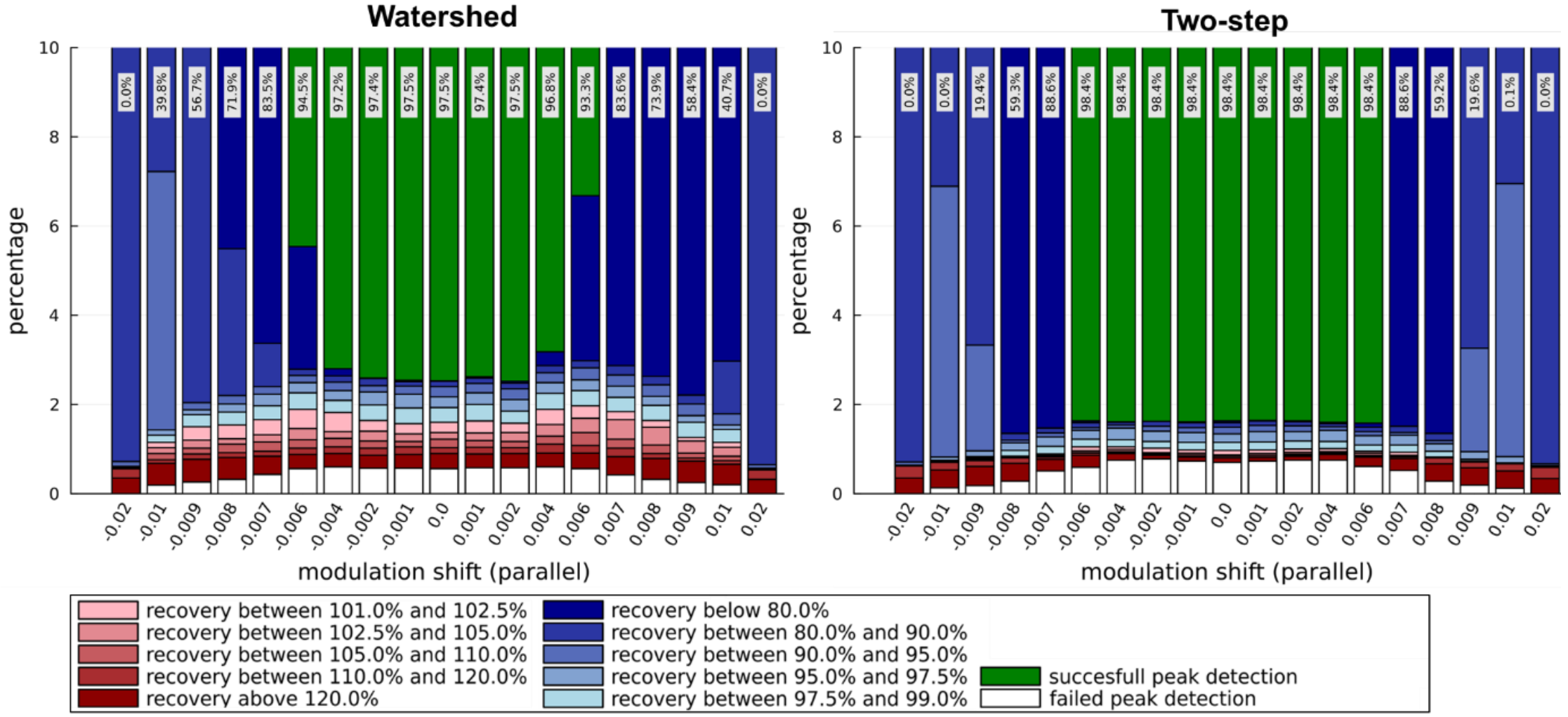


Vertical lines





Modulation shifting



Conclusion and outlook

- Using simulated data based on experimental parameters can aid in data analysis method selection
- A detailed and systematic look at peak detection's common pitfalls
- There is no "good" or "bad" peak detection method
- Just be aware of the characteristics of your data and how to minimize the impact on area recovery.
- Sanne Boot is working on adding MS to this concept (Poster P3-6)

We are in the process of publishing this work,

follow along at **projectparadise.nl**

Acknowledgements

Supervisors:

Arian van Asten

Bob Pirok

Tijmen Bos

Colleagues:

Analytical Chemistry group

Joshka Verduin

Rick van den Hurk

Collaborators:

Alan Garcia Cicourel

Rob Edam

Jan Blomberg

Gabriel Vivo Truyols

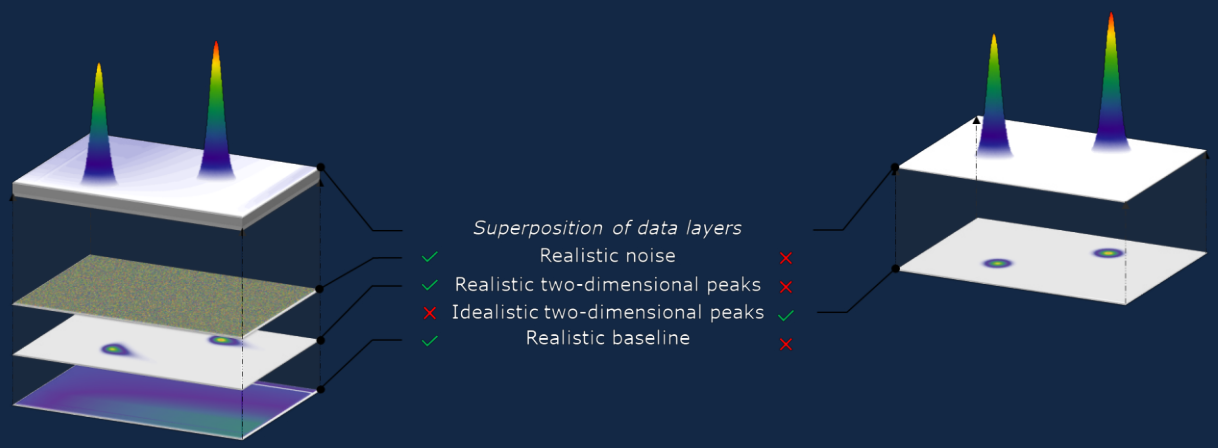
Students:

Tesse Ravenshorst

Nard Schellekens

Thank you for your attention!
Are there any questions?

	Experimental data	New approach to simulated data	Traditional simulated data
Realism	++	+	--
Known statistical moments	--	++	++
Universally applicable	--	++	-



Poster from our team:

P2: Rick van den Hurk

projectparadise.nl