One-shot tensor decomposition of full-scale GC×GC-VUV datasets for resolving petrochemical groups

Paul-Albert Schneide[1], Jesper L. Hinrich[2], Aleksandra Lelevic[3]

[1]Department of Food Science, University of Copenhagen
[2]Department of Computer Science, Technical University of Denmark
[3]IFP Energies Nouvelles (France)

UNIVERSITY OF COPENHAGEN

Acknowledgements

Aleksandra Lelević
Research Scientist - IFPEN
Verified email at ifpen.fr
Comprehensive two-dimen…

From: LELEVIC Aleksandra <aleksandra.lelevic@ifpen.fr>
Sent: 10. marts 2025 13:15
To: Rasmus Bro <rb@food.ku.dk>
Subject: Question regarding PARAFAC (for GCxGC-VUV data of gas-oils)

Dear prof Bro,

is important

I am writing to inquire regarding PARAFAC and the possibility of decomposing the spectral contributions in the GCxGC-VUV data, which seems to me as an interesting problem that should respond very well to PARAFAC. My question is purely scientific as I have this data collected during my PhD work (some years ago) and I am very curious to see what PARAFAC can do as I have a hint that it should be a very fitting approach (but I never managed to make it work well for my data). This is why I thought that that I could present to you the problem in case it will interest you (I will also share my data in case you want to go further https://ifpen-webfile.appcollaboratif.fr/?s=download&token=a5855454-95ff-44b6-82c8-78254be42c29

Acknowledgements

Aleksandra Lelević

Research Scientist - IFPEN
Verified email at ifpen.fr
Comprehensive two-dimen…

From: LELEVIC Aleksandra <aleksandra.lelevic@ifpen.fr>
Sent: 10. marts 2025 13:15
To: Rasmus Bro <rb@food.ku.dk>
Subject: Question regarding PARAFAC (for GCxGC-VUV data of gas-oils)

is important

Dear prof Bro,

I am writing to inquire regarding PARAFAC and the possibility of decomposing the spectral contributions in the GCxGC-VUV data, which seems to me as an interesting problem that should respond very well to PARAFAC. My question is purely scientific as I have this data collected during my PhD work (some years ago) and I am very curious to see what PARAFAC can do as I have a hint that it should be a very fitting approach (but I never managed to make it work well for my data). This is why I thought that that I could present to you the problem in case it will interest you (I will also share my data in case you want to go further https://ifpen-webfile.appcollaboratif.fr/?s=download&token=a5855454-95ff-44b6-82c8-78254be42c29

Rasmus Bro

Professor in chemometrics,
Bestätigte E-Mail-Adresse b
chemometrics tensor anal

On 10. Mar 2025, at 16:12, Rasmus Bro <rb@food.ku.dk> wrote:

Hi Aleksandra

Interesting problem. I think that Paul or Jesper can better answer your question as they have worked on similar data.

Best regards
Rasmus

Acknowledgements

Aleksandra Lelević

Research Scientist - IFPEN
Verified email at ifpen.fr
Comprehensive two-dimen…

From: LELEVIC Aleksandra <aleksandra.lelevic@ifpen.fr>
Sent: 10. marts 2025 13:15
To: Rasmus Bro <rb@food.ku.dk>
Subject: Question regarding PARAFAC (for GCxGC-VUV data of gas-oils)

is important

Dear prof Bro,

I am writing to inquire regarding PARAFAC and the possibility of decomposing the spectral contributions in the GCxGC-VUV data, which seems to me as an interesting problem that should respond very well to PARAFAC. My question is purely scientific as I have this data collected during my PhD work (some years ago) and I am very curious to see what PARAFAC can do as I have a hint that it should be a very fitting approach (but I never managed to make it work well for my data). This is why I thought that that I could present to you the problem in case it will interest you (I will also share my data in case you want to go further https://ifpen-webfile.appcollaboratif.fr/?s=download&token=a5855454-95ff-44b6-82c8-78254be42c29

Rasmus Bro

Professor in chemometrics,
Bestätigte E-Mail-Adresse b
chemometrics tensor anal

On 10. Mar 2025, at 16:12, Rasmus Bro <rb@food.ku.dk> wrote:

Hi Aleksandra

Interesting problem. I think that Paul or Jesper can better answer your question as they have worked on similar data.

Best regards
Rasmus

Paul-Albert Schneide

Jesper Løve Hinrich

Internal

Acknowledgements

Aleksandra Lelević

Research Scientist - IFPEN
Verified email at ifpen.fr
Comprehensive two-dimen…

From: LELEVIC Aleksandra <aleksandra.lelevic@ifpen.fr>
Sent: 10. marts 2025 13:15
To: Rasmus Bro <rb@food.ku.dk>
Subject: Question regarding PARAFAC (for GCxGC-VUV data of gas-oils)

Thank you!

is important

Dear prof Bro,

I am writing to inquire regarding PARAFAC and the possibility of decomposing the spectral contributions in the GCxGC-VUV data, which seems to me as an interesting problem that should respond very well to PARAFAC. My question is purely scientific as I have this data collected during my PhD work (some years ago) and I am very curious to see what PARAFAC can do as I have a hint that it should be a very fitting approach (but I never managed to make it work well for my data). This is why I thought that that I could present to you the problem in case it will interest you (I will also share my data in case you want to go further https://ifpen-webfile.appcollaboratif.fr/?s=download&token=a5855454-95ff-44b6-82c8-78254be42c29

Rasmus Bro

Professor in chemometrics,
Bestätigte E-Mail-Adresse b
chemometrics tensor anal

Paul-Albert Schneide Jesper Løve Hinrich

On 10. Mar 2025, at 16:12, Rasmus Bro <rb@food.ku.dk> wrote:

Hi Aleksandra

Interesting problem. I think that Paul or Jesper can better answer your question as they have worked on similar data.

Best regards
Rasmus

Outline of the presentation

All-samples-at-once shift-invariant tri-linear decomposition provides a simple and robust data analysis strategy to group-level analysis of petrochemicals, using GC×GC-VUV.

Outline of the presentation

All-samples-at-once shift-invariant tri-linear decomposition provides a simple and robust data analysis strategy to group-level analysis of petrochemicals, using GC×GC-VUV.

Simple implementation, small number of tunable parameters.	Accurate estimates of peak pattern, UV spectra and peak areas.

Outline of the presentation

All-samples-at-once shift-invariant tri-linear decomposition provides a simple and robust data analysis strategy to group-level analysis of petrochemicals, using GC×GC-VUV.

Simple implementation, small number of tunable parameters.

Accurate estimates of peak pattern, UV spectra and peak areas.

Data properties

Low-rank structure of GC×GC-VUV permits all-at-once decomposition for group-level data analysis.

Outline of the presentation

All-samples-at-once shift-invariant tri-linear decomposition provides a simple and robust data analysis strategy to group-level analysis of petrochemicals, using GC×GC-VUV.

Simple implementation, small number of tunable parameters.

Accurate estimates of peak pattern, UV spectra and peak areas.

Data properties

Low-rank structure of GC×GC-VUV permits all-at-once decomposition for group-level data analysis.

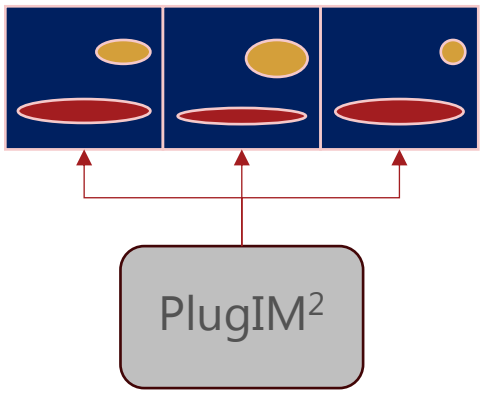
Model properties

Shift-invariant tri-linearity is an efficient method to deconvolute shifted, co-eluting peaks.

Overview: Data analysis workflow

Data analysis workflow

GC×GC-VUV measurements

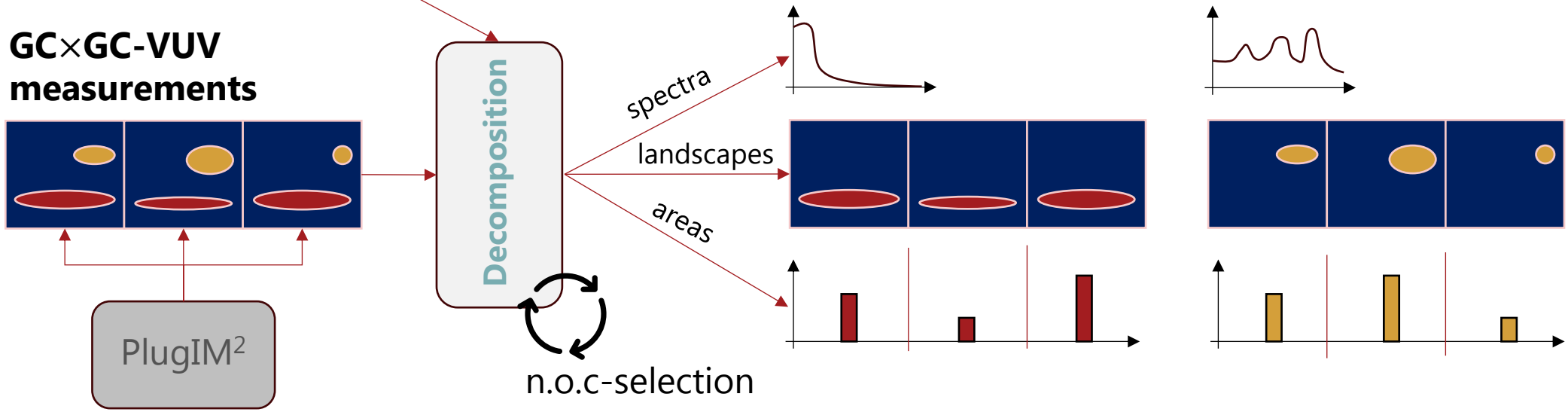


[1]Schneide P-A, Bro R, Gallagher NB. *Journal of Chemometrics*. 2023; 37(8):e3501
[2]Lelevic et al *Chemometrics and Intelligent Laboratory Systems,* Volume 231, 2022, 104708

Overview: Data analysis workflow

Data analysis workflow

All-samples-at-once shift-invariant tri-linear decomposition[1] provides a simple and robust data analysis strategy to group-level analysis of petrochemicals, using GC×GC-VUV.

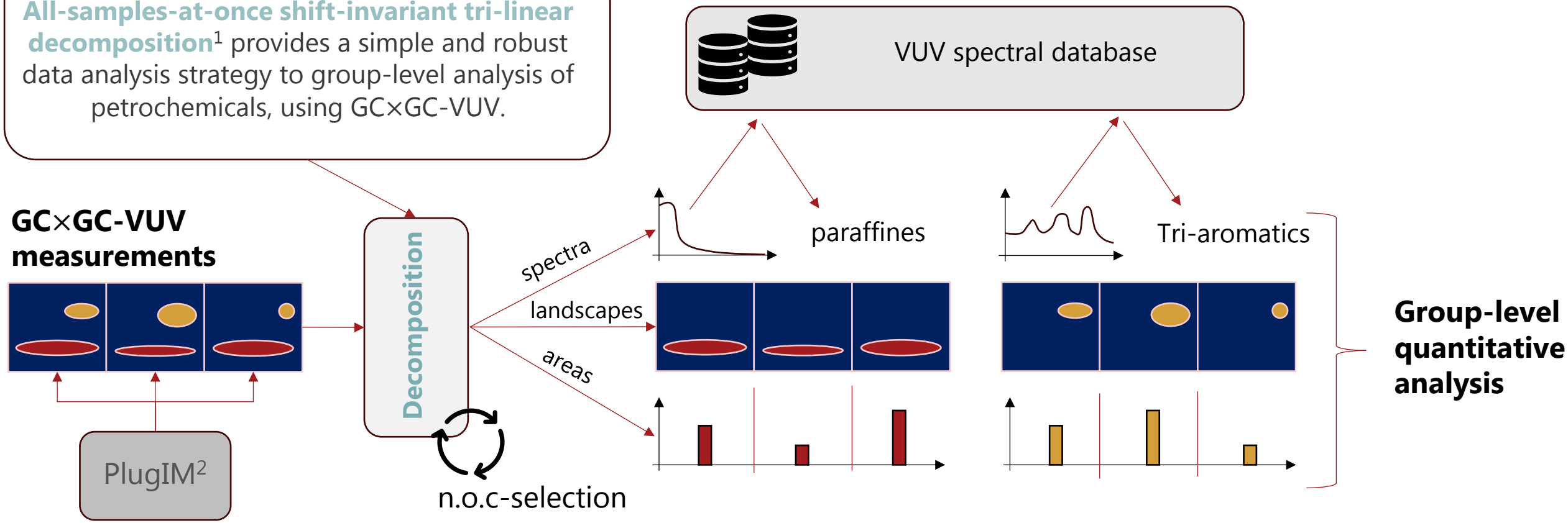


[1]Schneide P-A, Bro R, Gallagher NB. *Journal of Chemometrics*. 2023; 37(8):e3501
[2]Lelevic et al *Chemometrics and Intelligent Laboratory Systems,* Volume 231, 2022, 104708

Overview: Data analysis workflow

Data analysis workflow

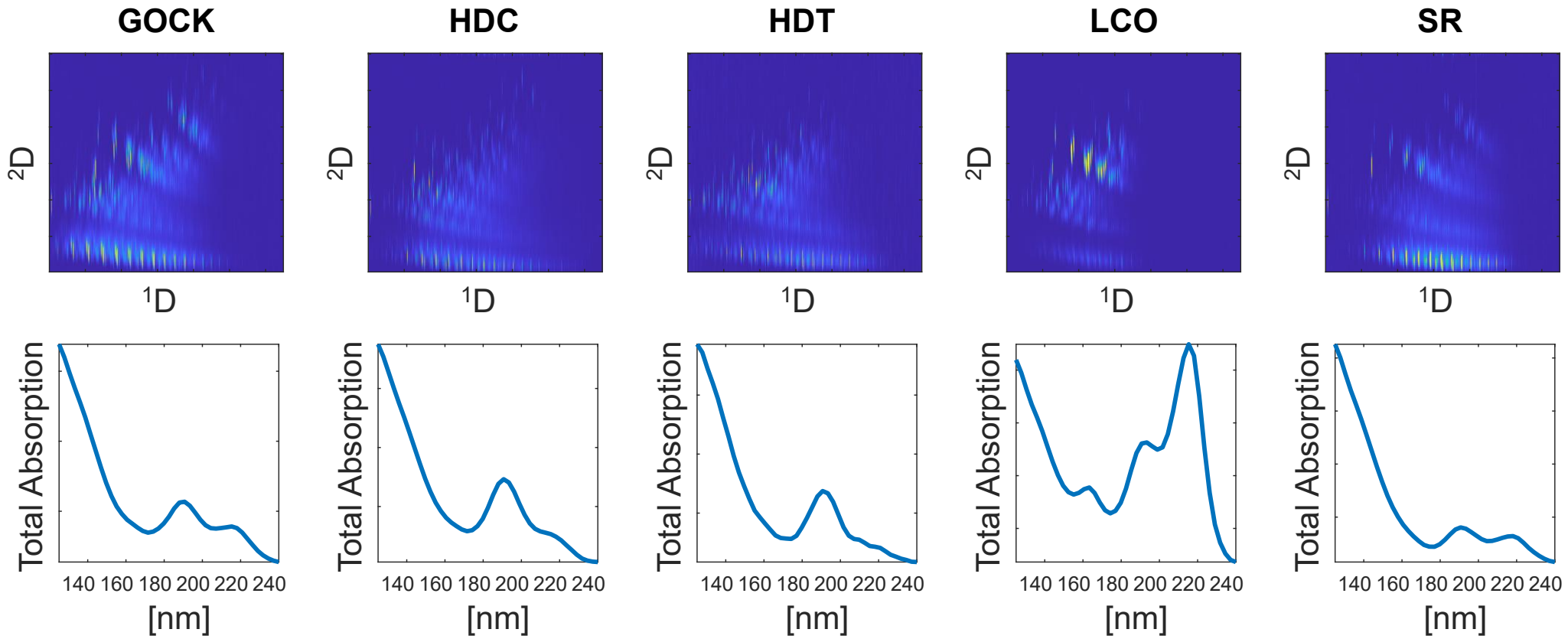
All-samples-at-once shift-invariant tri-linear decomposition[1] provides a simple and robust data analysis strategy to group-level analysis of petrochemicals, using GC×GC-VUV.



[1]Schneide P-A, Bro R, Gallagher NB. *Journal of Chemometrics*. 2023; 37(8):e3501
[2]Lelevic et al *Chemometrics and Intelligent Laboratory Systems,* Volume 231, 2022, 104708

Case study: Petrochemical samples

- GC×GC-VUV data: 14 gas oil samples, cocker gas oils (GOCK), hydroconverted (HDC), hydrotreated (HDT), light cycle oils (LCO), straight run gas oils (SR), 24 measurements[3]

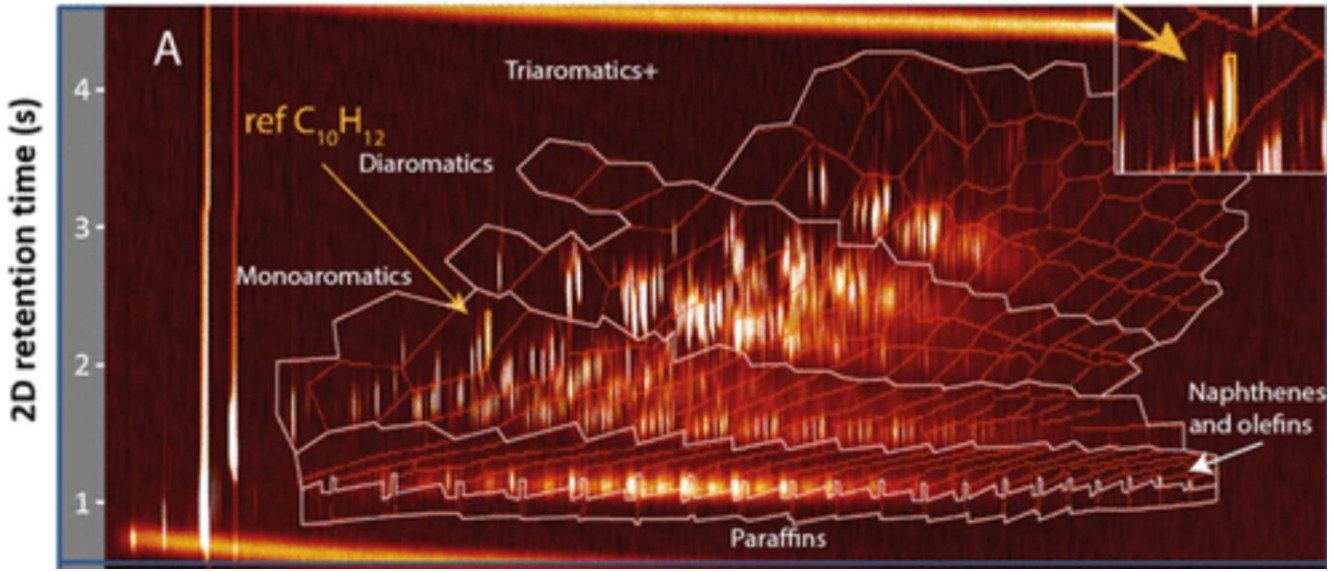


[3]Lelevic et al *Energy & Fuels* 2022 *36* (18), 10860-10869 DOI: 10.1021/acs.energyfuels.2c01960

Case study: Problem specification and benchmark

Characterizing and semi-quantifying chemical groups: *saturates, olefines, mono-aromatics, di-aromatics, tri-aromatics*.

Benchmark: *Template-based approach*[3,4], *pixel-based approach*[5]



Template-based approach has been compared to, and showed good agreement with[1]*:*

- *GC-MS derived from ASTM D2425.*
- *GC×GC-FID with prefractionation.*
- *Bromine Number (ASTM D1159).*

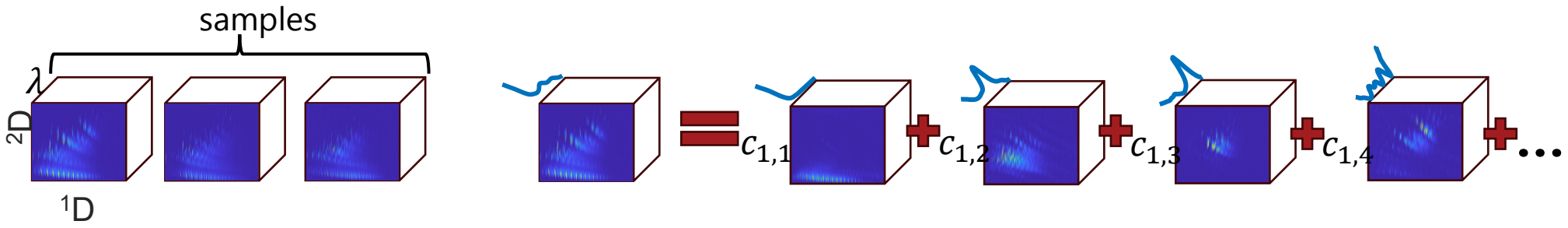
Limitations: sensitive to retention time shifts → manual adjustment of template required, resolving co-eluting signals requires *a priori* knowledge

[3]Lelevic et al *Energy & Fuels* 2022 *36* (18), 10860-10869 DOI: 10.1021/acs.energyfuels.2c01960
[4]Lelevic et al *Energy & Fuels 2021* 35 (17), 13766-13775 DOI: 10.1021/acs.energyfuels.1c01910
[5]Lelevic *Journal of Separation Science 2025* 48 (12), e70318. https://doi.org/10.1002/jssc.70318

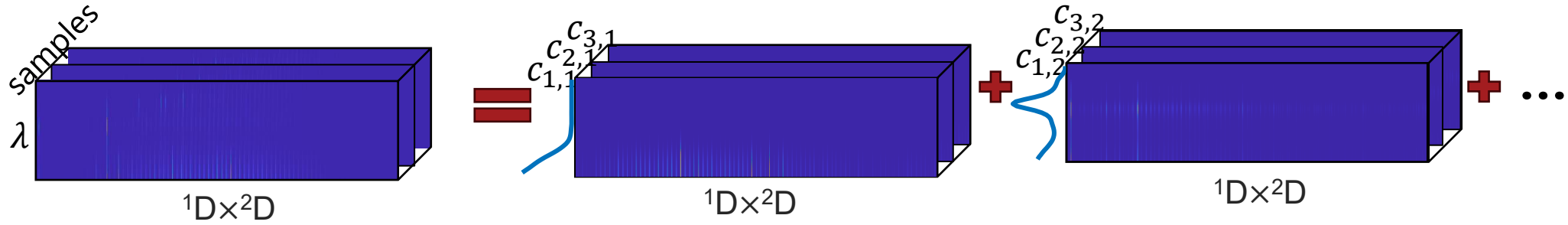
The proposed method: Overview

Shift-invariant tri-linear decomposition → **Exploiting Lambert-Beers law**

Strategy 1:



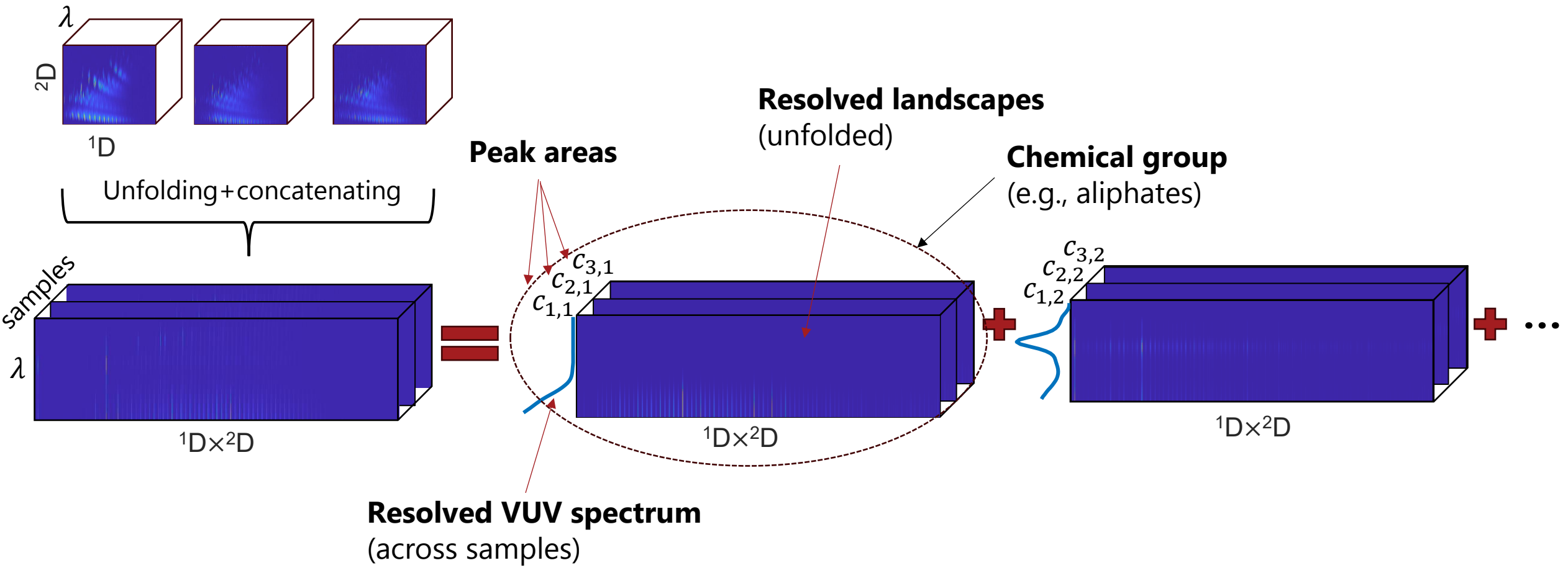
Strategy 2:



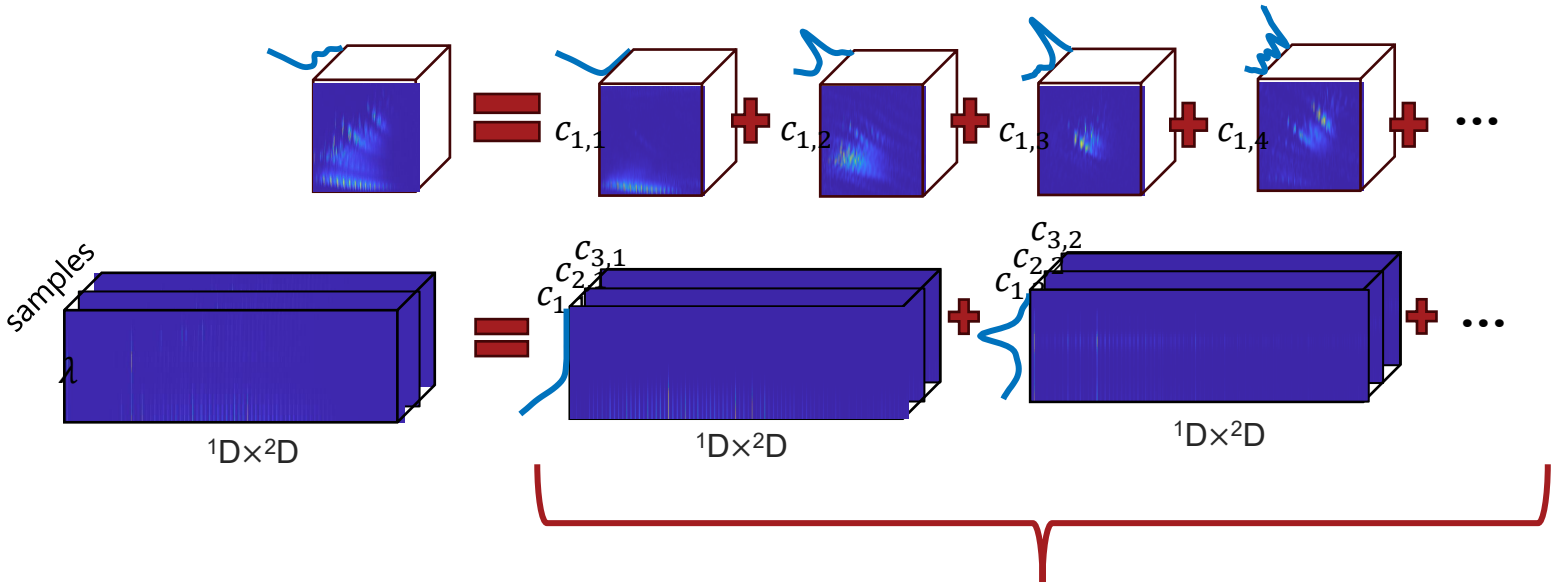
Conceptually: Doing Parallel Factor Analysis (PARAFAC)[6] / Multivariate Curve Resolution (MCR)[7] on the whole GC×GC-VUV dataset instead of decomposing selected Regions of Interest (ROIs).

[6]Hoggard et al. *Analytical Chemistry* 2007 *79* (4), 1611-1619
[7]Parastar et al. *Analytical Chemistry* 2014 *86* (1), 286-297

How does it work: Strategy 2 (all-samples-at-once)

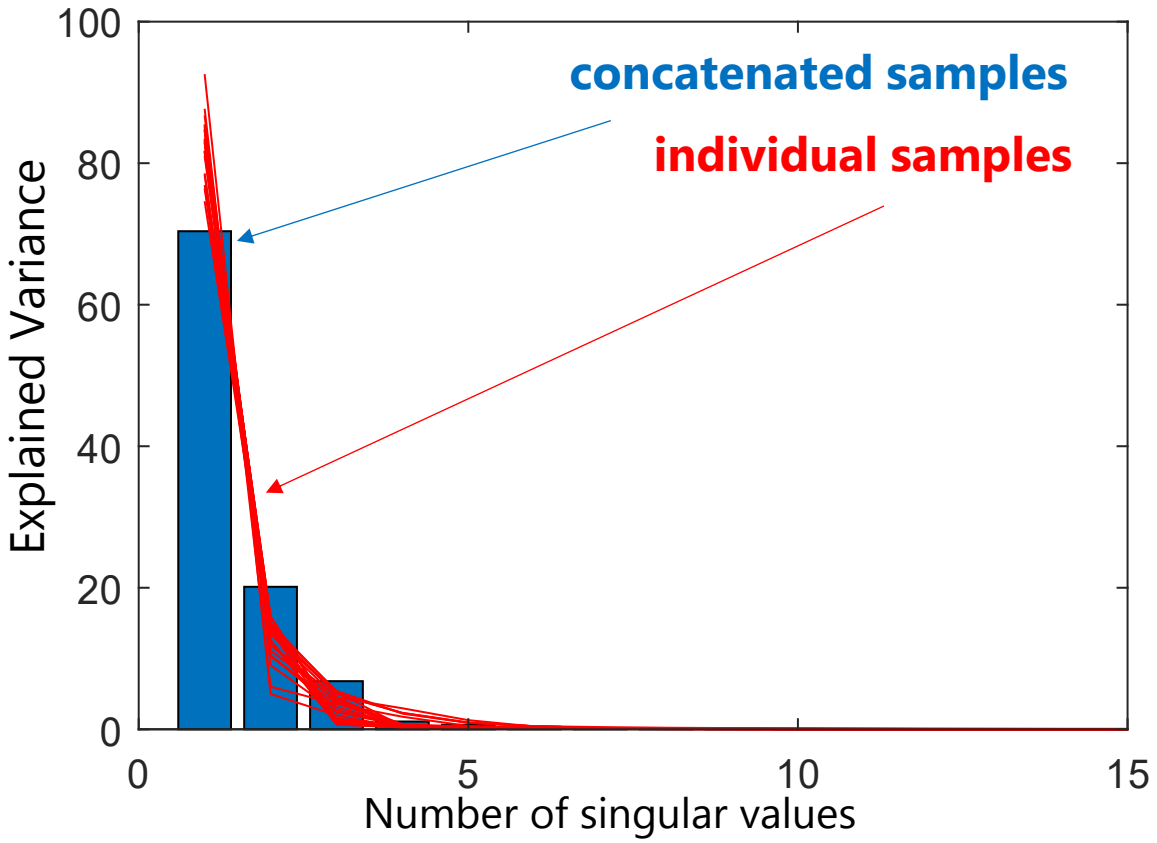


Data Properties: Low-rank-structure of GC×GC-VUV data



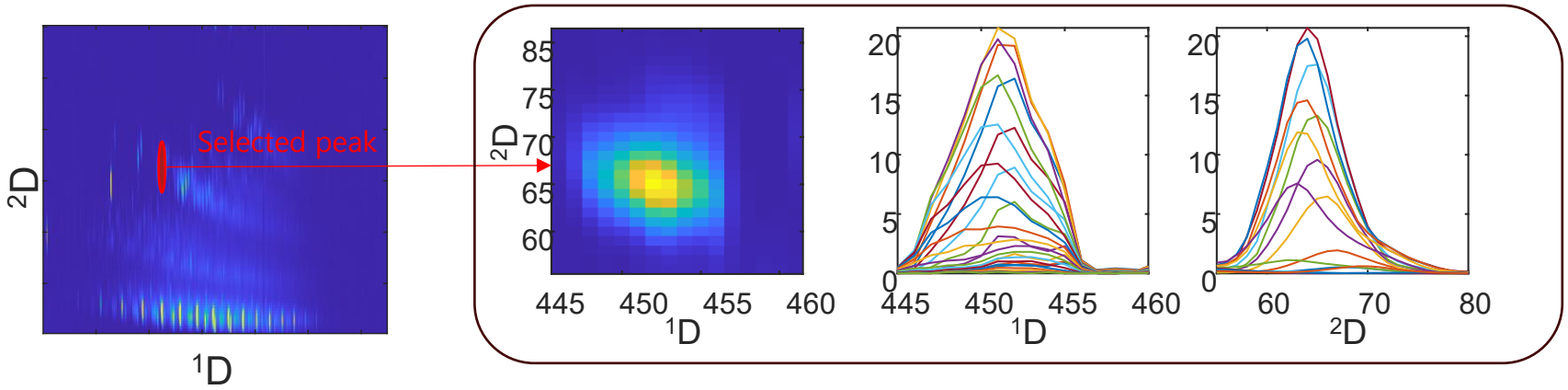
We need to get a "good" approximation of the raw data for a "small" number of terms.

Data Properties: Low-rank-structure of GC×GC-VUV data

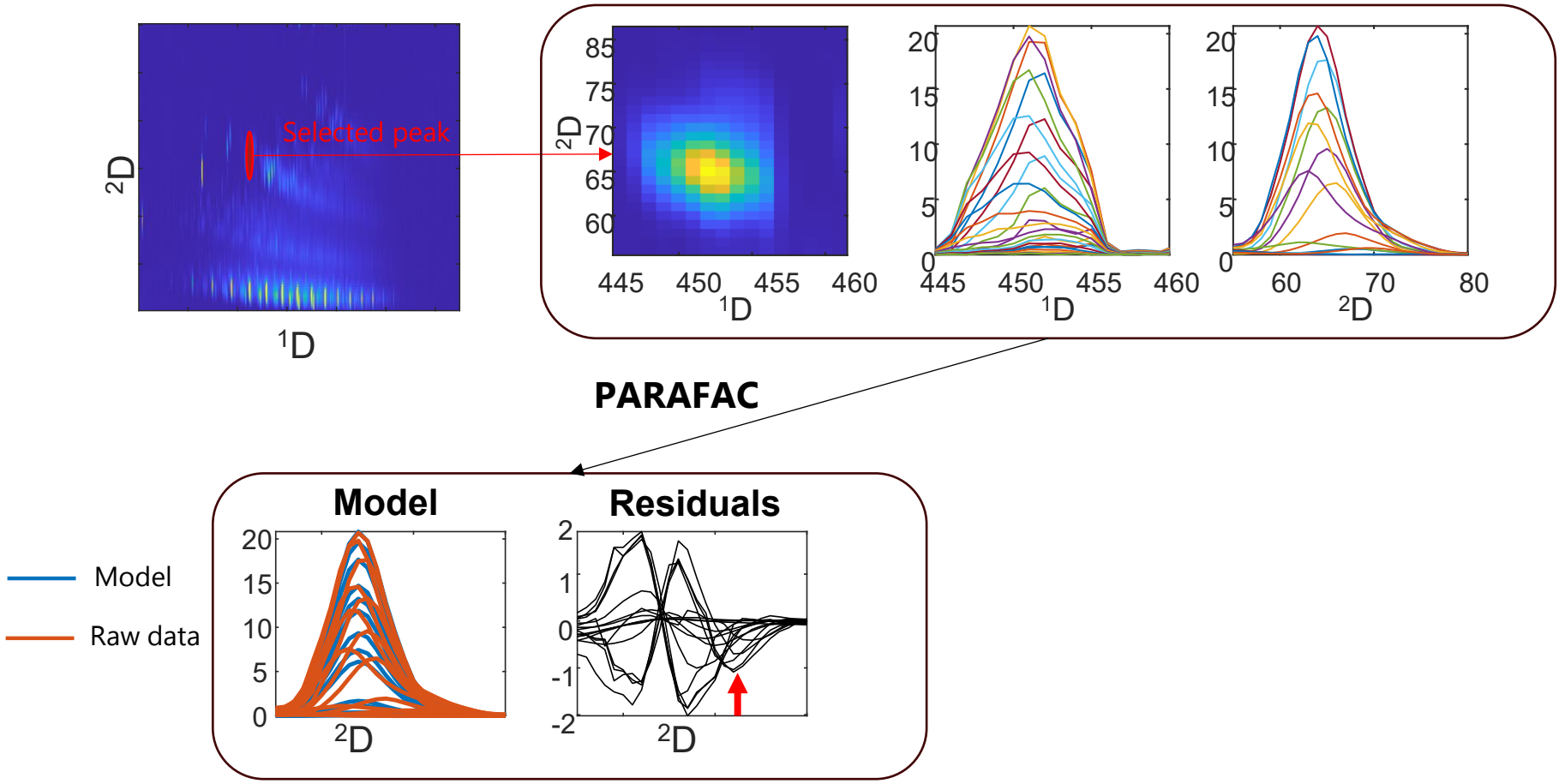


Singular value decay curves show that **< 10 components explain > 99.9 % of the variance** for all individual samples.

Model Properties: shift-invariant, tri-linear decomposition

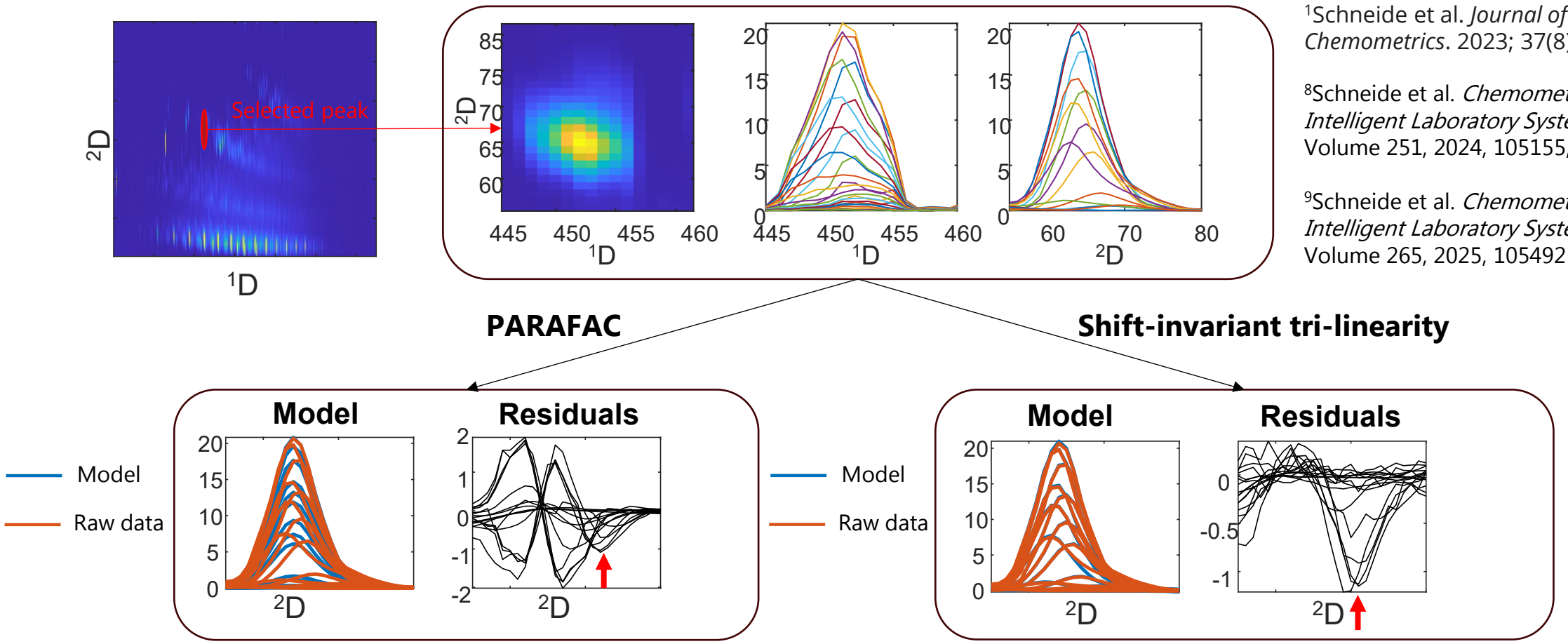


Model Properties: shift-invariant, tri-linear decomposition



Problem: Residuals from unmodelled shift overshadow residuals from an unresolved peak

Model Properties: shift-invariant, tri-linear decomposition



[1]Schneide et al. *Journal of Chemometrics*. 2023; 37(8):e3501.

[8]Schneide et al. *Chemometrics and Intelligent Laboratory Systems*, Volume 251, 2024, 105155,

[9]Schneide et al. *Chemometrics and Intelligent Laboratory Systems*, Volume 265, 2025, 105492

Improvement: Agnostic towards retention time shift, facilitating extraction of chemically interpretable information.[1,8,9]

Model Properties: Math

- We use FFT to model shift.
- It's faster than PARAFAC2

Matricizing $\underline{\mathbf{X}}$:

$\underline{\mathbf{X}} \in \mathbb{R}^{I \times J \times K} \rightarrow \mathbf{X}_{(1)} \in \mathbb{R}^{I \times JK}$
$\underline{\mathbf{X}} \in \mathbb{R}^{I \times J \times K} \rightarrow \mathbf{X}_{(2)} \in \mathbb{R}^{IK \times J}$
$\underline{\mathbf{X}} \in \mathbb{R}^{I \times J \times K} \rightarrow \mathbf{X}_{(3)} \in \mathbb{R}^{IJ \times K}$

Tri-linearity constrained NNMF (MCR):

$\mathbf{X}_{(2)} \approx \mathbf{C}'\mathbf{S}^{\mathrm{T}}, s.t. \mathbf{C}'_r = \mathbf{u}_{1,r}\boldsymbol{\sigma}_{1,r}\mathbf{v}_{1,r}^{\mathrm{T}}$
$\mathbf{u}_1 = \mathbf{c}_r$
$\left(\boldsymbol{\sigma}_1\mathbf{v}_1^{\mathrm{T}}\right)^{\mathrm{T}} = \mathbf{f}_r$

$\mathbf{C}' \in +\mathbb{R}^{IK \times R}$
$\mathbf{S} \in +\mathbb{R}^{J \times R}$
$\mathbf{C}'_r \in +\mathbb{R}^{I \times K}$
$r \in \{1, \ldots, R\}$
$\mathbf{u}_{1,r}\boldsymbol{\sigma}_{1,r}\mathbf{v}_{1,r}^{\mathrm{T}}$ is a rank-one SVD decomposition.

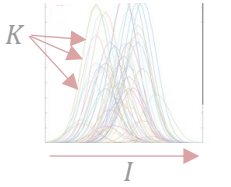
$\mathbf{X}_{(2)} \approx \mathbf{C}'\mathbf{S}^{\mathrm{T}}, s.t. \left|\tilde{\mathbf{C}}'_r\right| = \mathbf{u}_{1,r}\boldsymbol{\sigma}_{1,r}\mathbf{v}_{1,r}^{\mathrm{T}}$

$\mathbf{C}' \in +\mathbb{R}^{IK \times R}$
$\mathbf{S} \in +\mathbb{R}^{J \times R}$
$\tilde{\mathbf{C}}'_r \in \mathbb{C}^{I \times K}, \tilde{\mathbf{c}}'_{r,k} = \mathcal{F}(\mathbf{c}'_{r,k})$
$\left|\tilde{\mathbf{C}}'_r\right| \in +\mathbb{R}^{I \times K}$
$r \in \{1, \ldots, R\}$
$\mathbf{u}_{1,r}\boldsymbol{\sigma}_{1,r}\mathbf{v}_{1,r}^{\mathrm{T}}$ is a rank-one SVD decomposition.

The general idea is that the amplitude spectrum $\mathbf{a}_{r,k}$ of a time-domain signal is shift-invariant!

How to make shift invariant tri-linearity [32]

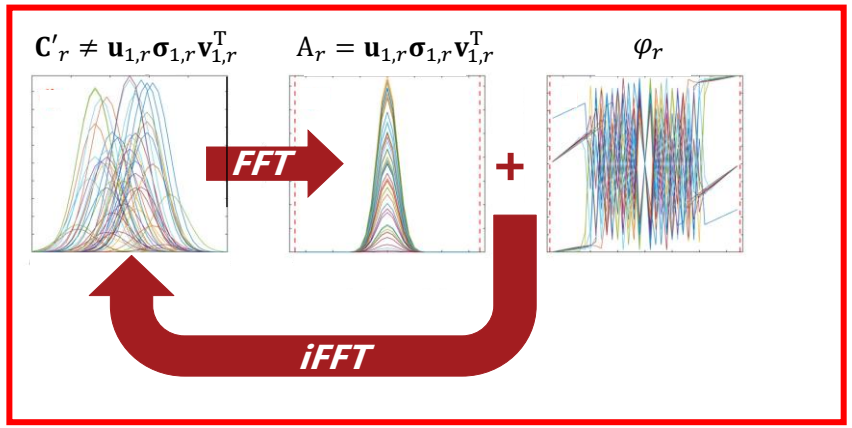
$\mathbf{C}'_r \neq \mathbf{u}_{1,r}\boldsymbol{\sigma}_{1,r}\mathbf{v}_{1,r}^{\mathrm{T}}$



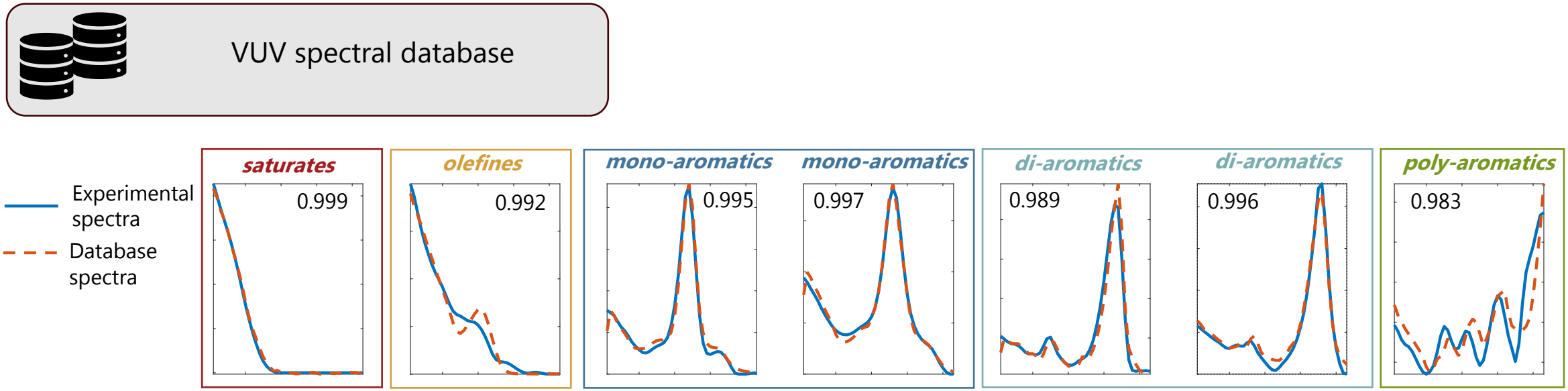
Looking at the shifted signal in the frequency domain:

$\tilde{\mathbf{c}}'_{r,k} = \mathcal{F}(\mathbf{c}'_{r,k})$
$\tilde{\mathbf{c}}'_{r,k}[i] = \mathbf{a}_{r,k}[i]e^{j\varphi_{r,k}[i]}$
$\mathbf{a}_{r,k} = \sqrt{Re\left(\tilde{\mathbf{c}}'_{r,k}\right)^2 + Im\left(\tilde{\mathbf{c}}'_{r,k}\right)^2} = \left|\tilde{\mathbf{c}}'_{r,k}\right|$
$\varphi_{r,k} = \arctan\left(\frac{Im(\tilde{\mathbf{c}}'_{r,k})}{Re(\tilde{\mathbf{c}}'_{r,k})}\right)$

[32] P.-A. Schneide, N. Gallagher, and R. Bro, Shift-invariant tri-linearity—A new model for resolving untargeted gas chromatography coupled mass spectrometry data , J. Chemometrics, 37, (2023)

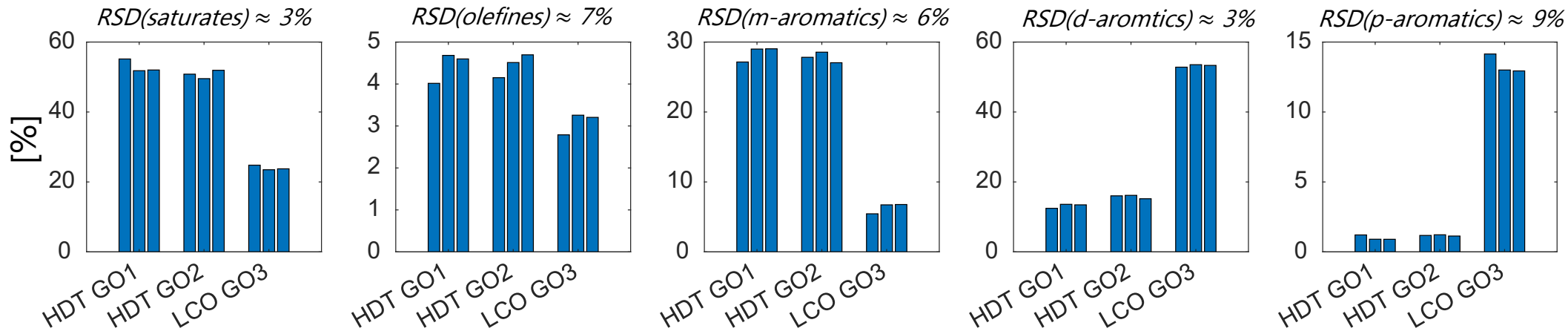


Results: Spectral similarity

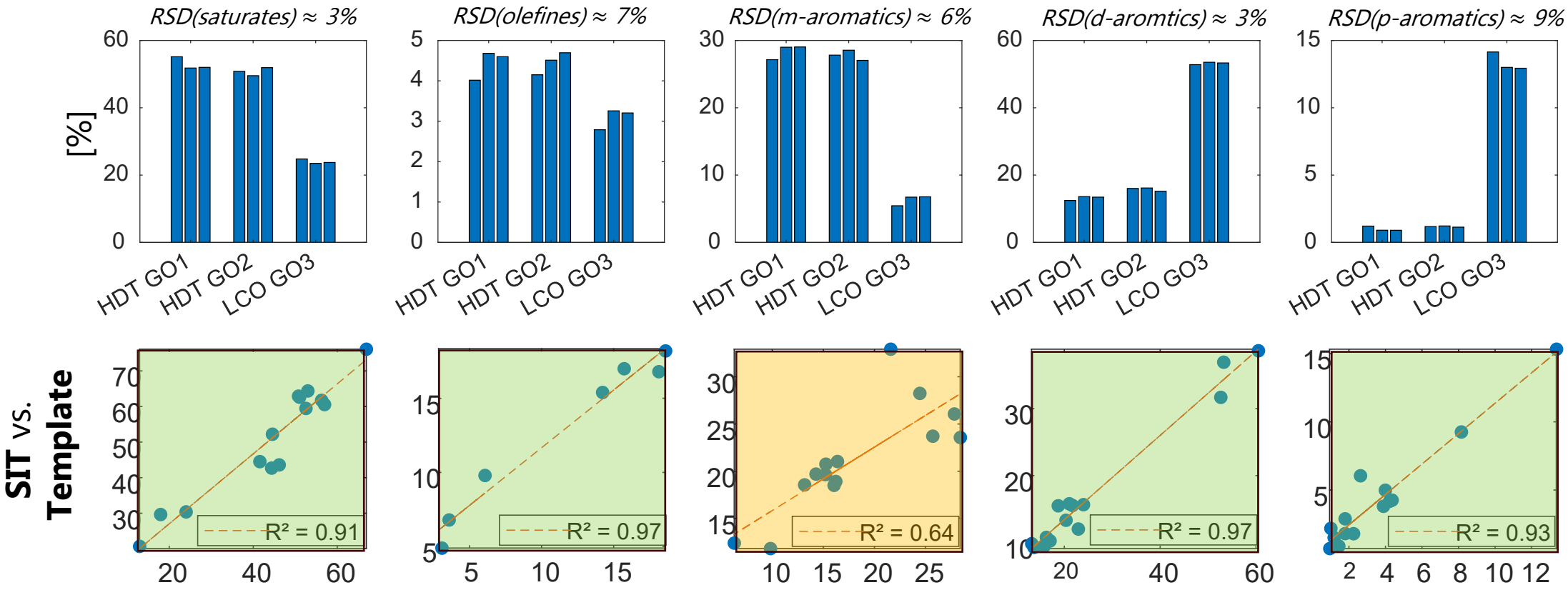


Extracted VUV spectra show high similarity to reference spectra → assigning components to chemical groups.

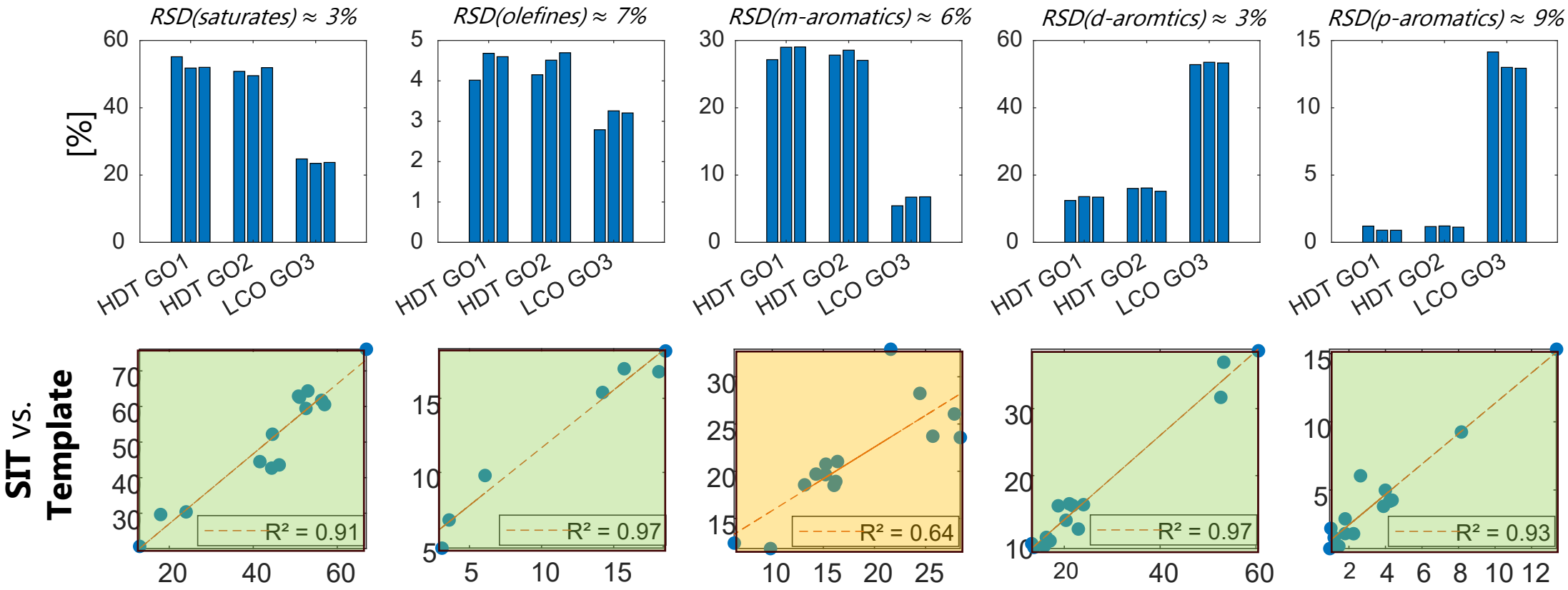
Results: Quantitative performance



Results: Quantitative performance



Results: Quantitative performance



Quantitative results obtained from a 7-component model show **high repeatability for replicates** and **high correlation with template-based reference method**.

Summary

Next steps

- ❑ Running models on individual samples
- ❑ Automating number of component selection in pipeline (e.g., split-half, cross validation)
- ❑ Understand where deviations for mono-aromatics come from
- ❑ Maybe integrate pipeline in existing project, e.g. GCDuo (?) ☺☺

Thank you very much for your attention!

I am happy to connect with you:

University of Copenhagen:
paul.schneide@food.ku.dk

BASF SE:
paul.schneide@basf.com