Zentrum für Technomathematik

SCiLS

Visualizing MALDI-TOF data of FFPE Tissue for Quality Assessment and Comparison

Tobias Boskamp

Bioinformatics Group
Center for Industrial Mathematics
University of Bremen, Germany

SCiLS GmbH, Bremen, Germany

Universität Bremen*

Clinical MSI – Where's The Challenge?

Increasing attention in MSI research

What we CAN do …

Identify significant biomarkers, metabolic and proteomic pathways, prognostic factors, …

Improving diagnosis or therapy outcome in patients

Clinical MSI

There's a gap!

Improve
- **reproducibility**
- **robustness**
- **standardization**

Neurology

Cardiology

Oncology

Dermatology

Regenerative medicine

Transplantation

Gastro-enterology

What we CANNOT yet do …

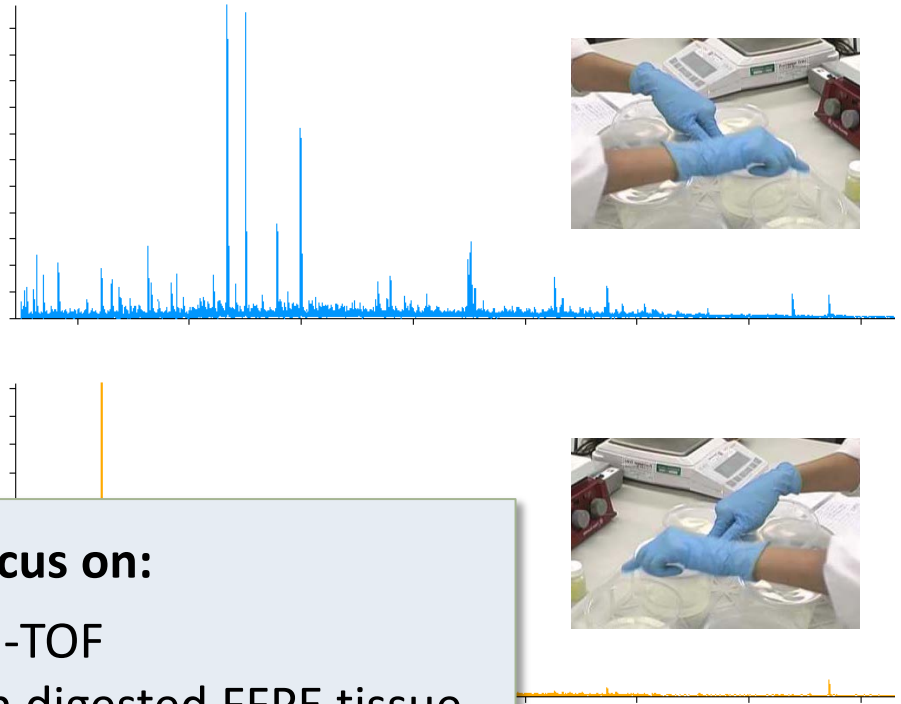
Reproducibly and routinely detect and quantify diagnostic and prognostic biomarkers

Universität Bremen

MALDI MSI Reproducibility

MALDI-TOF on FFPE tissue is a complex task!

- High sensitivity to process variations
- Differences between measurements often larger than between tissue types
- Data analysis and interpretation affected by
  - Delocalization
  - Noise
  - Intensity / sensitivity variations
  - Mass distortions
  - …



Here, focus on:

- MALDI-TOF
- Trypsin digested FFPE tissue
- Peptide signal features
- **Mass distortions**

Chemical Noise – Peptide Mass Rule

Idea: Use peptide background signal as an intrinsic reference

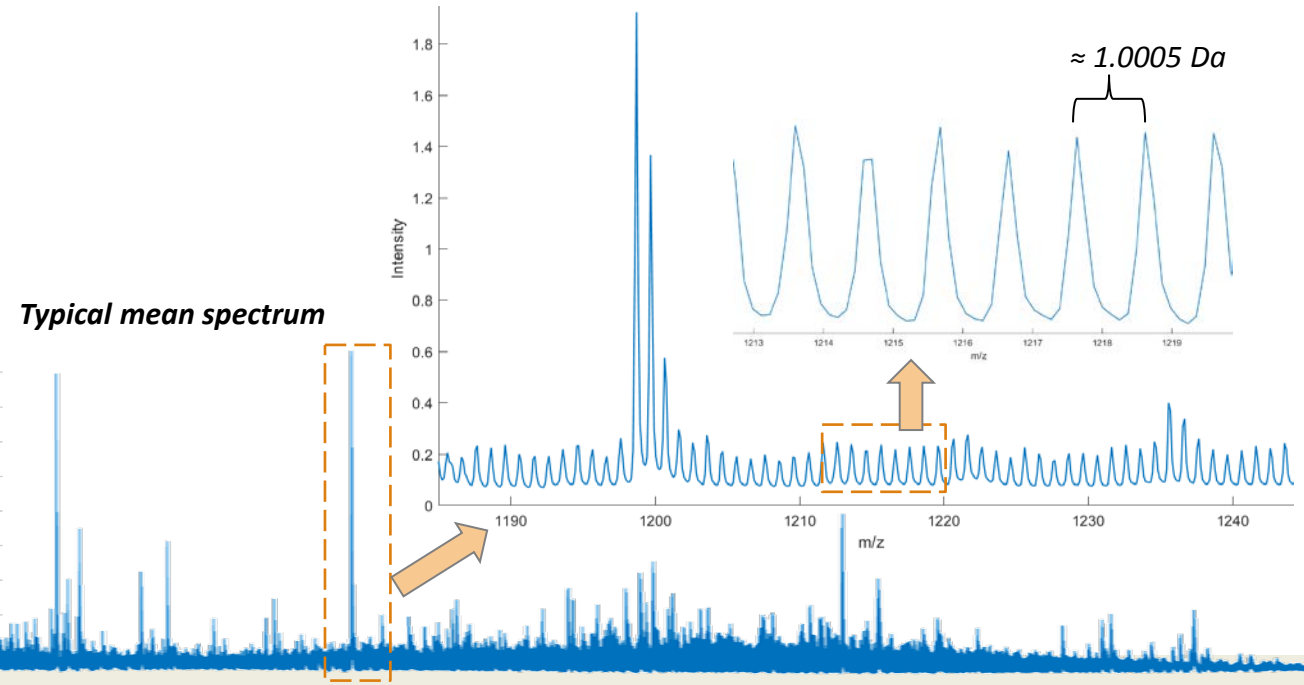
- Chemical noise largely dominated by peptides
- Characteristic wavelength = 1+δ Da
- Mass defect δ factor determined by peptide mass rule

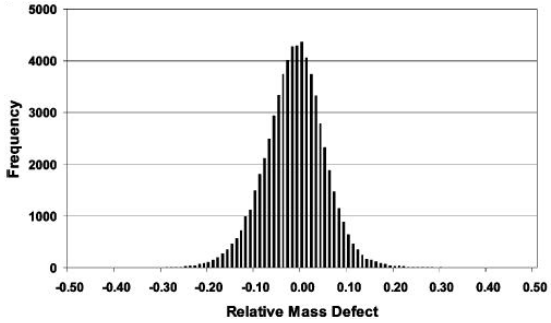
Peptide mass rule:

$m = (1+ \delta)\, m_N$

$\delta \approx 4.95e\text{-}4$

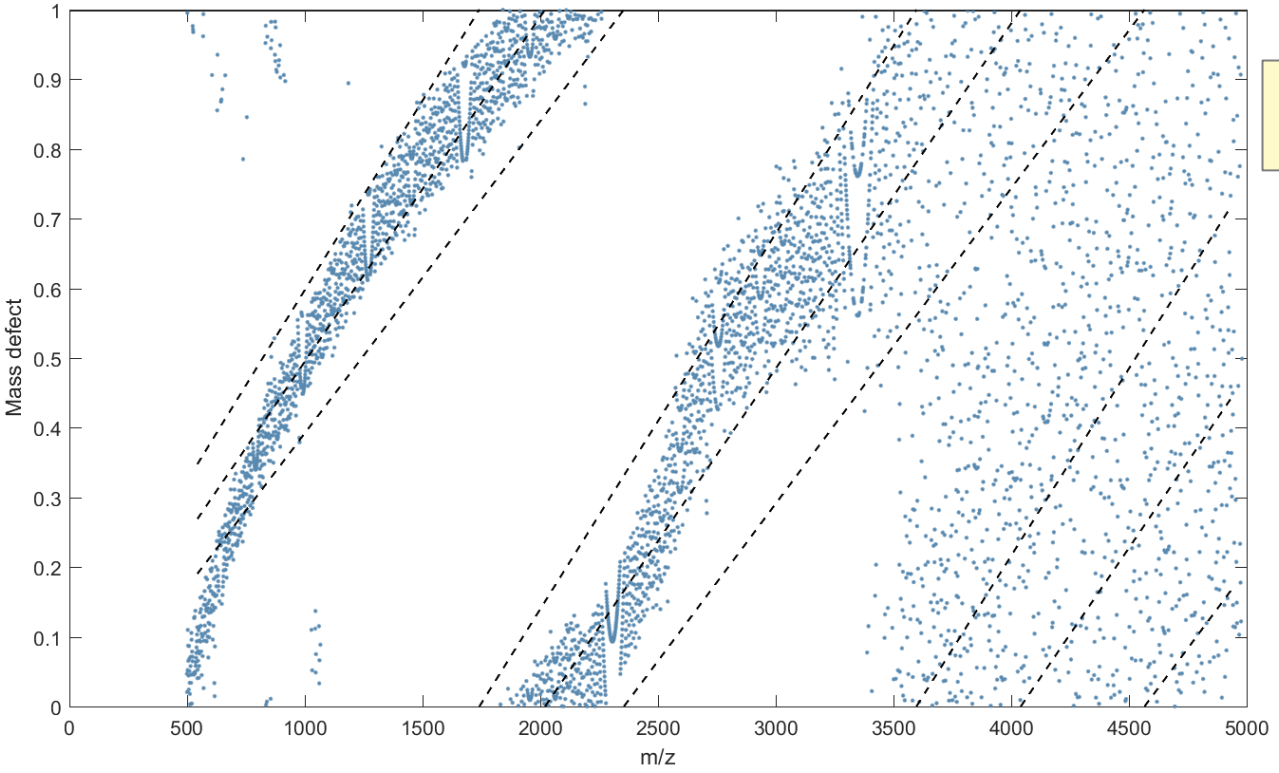
Element	Nominal mass m_N	Mass defect
H	1	0.0078
C	12	0.0000
N	14	0.0031
O	16	-0.0051
S	32	-0.0279





(Hernandez et al, Anal Chem 2006)

Peptide Mass Defect Diagram

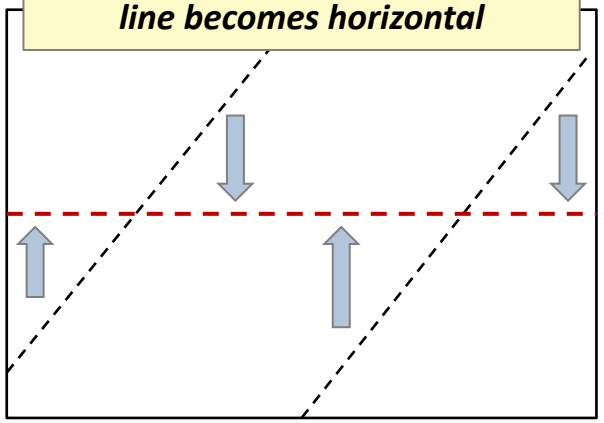


True peptide mass defects

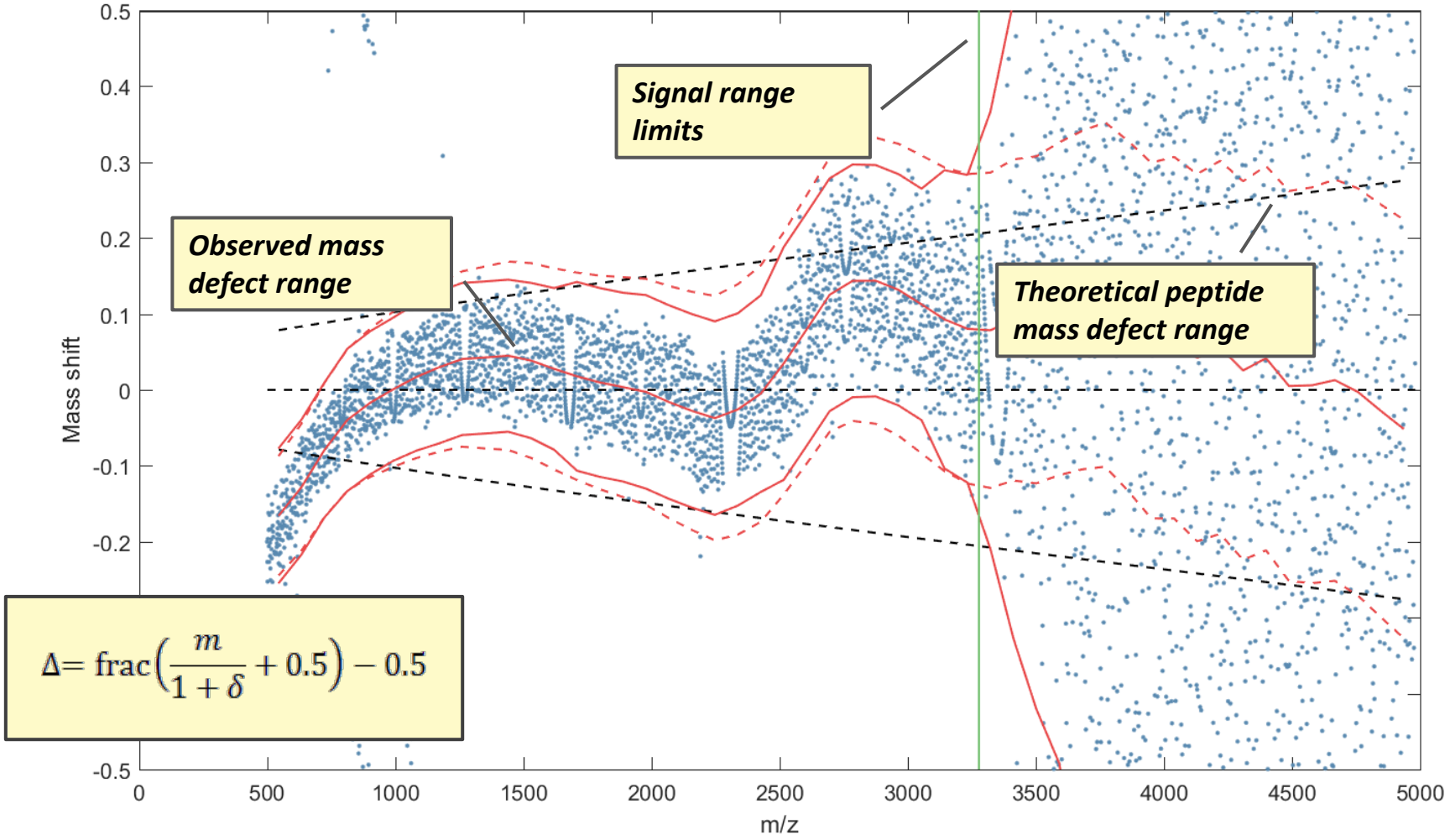
Peptide reference range

Mass defect vs. m/z of local maxima in mean spectrum

Transform diagram s.t. reference line becomes horizontal



Mass Shift Profile



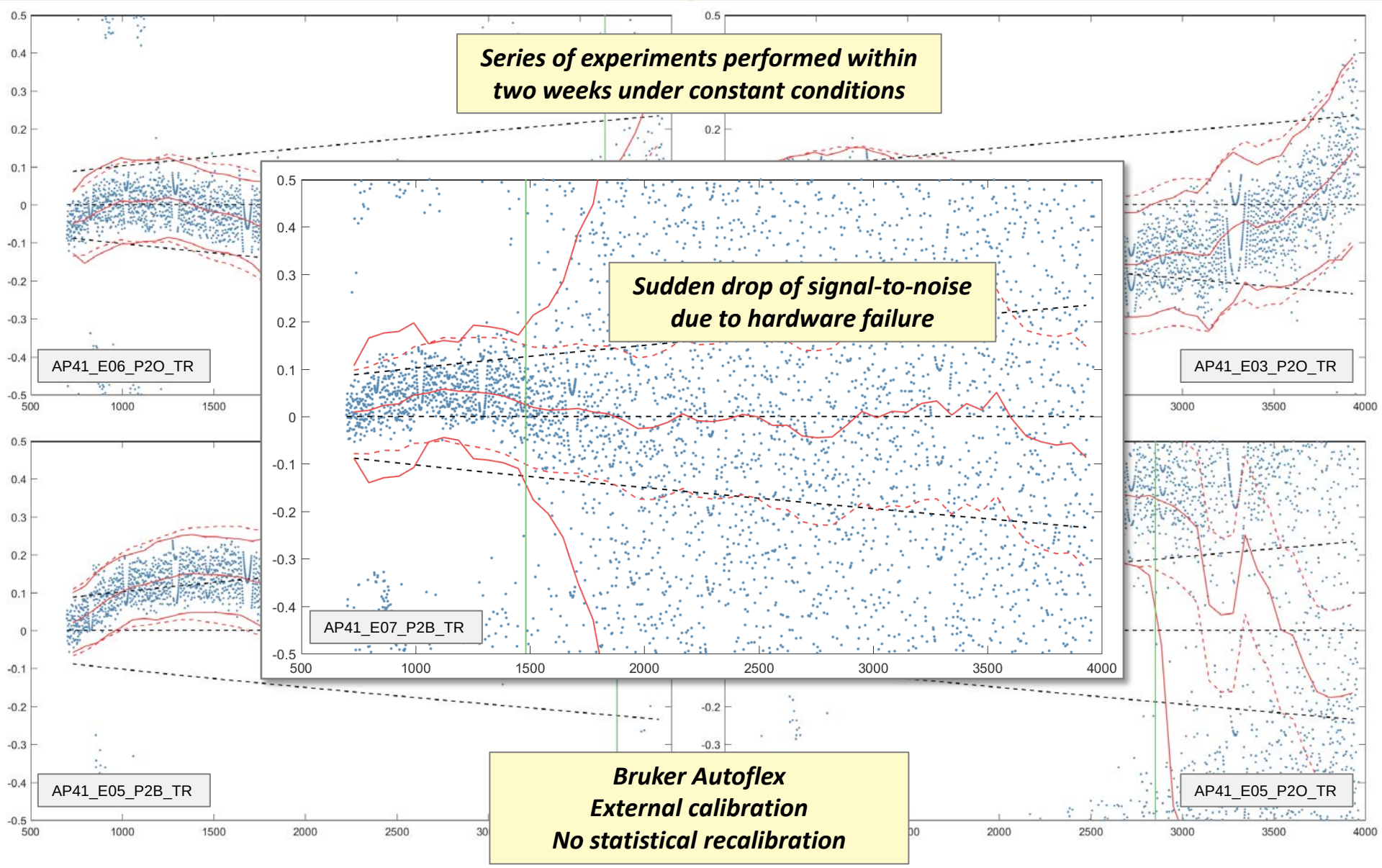
$$\Delta = \mathrm{frac}\left(\frac{m}{1+\delta} + 0.5\right) - 0.5$$

Zentrum für Technomathematik

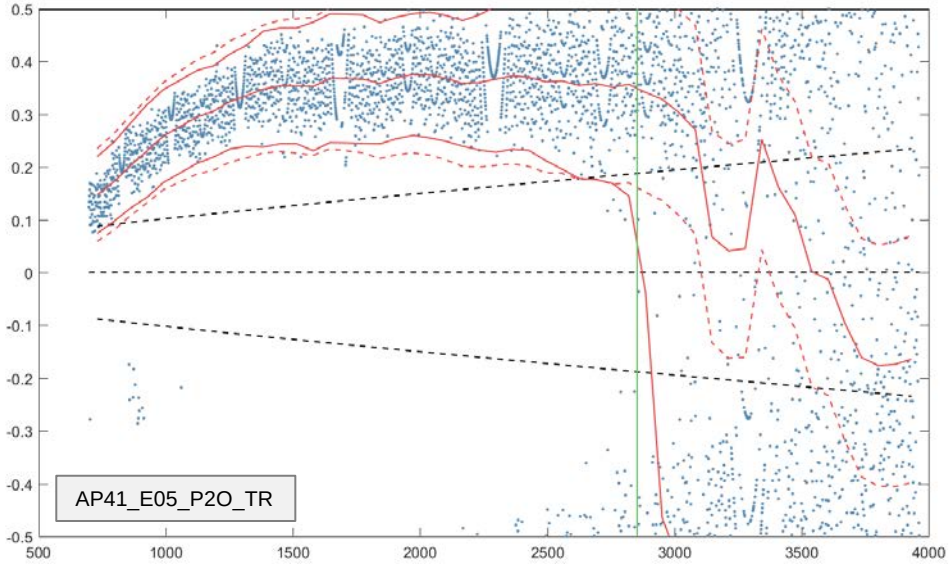
SCiLS

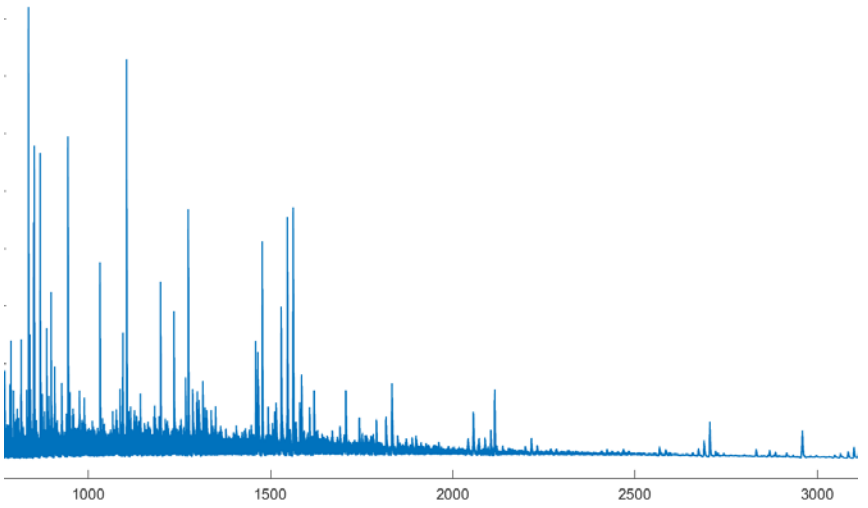
Universität Bremen*

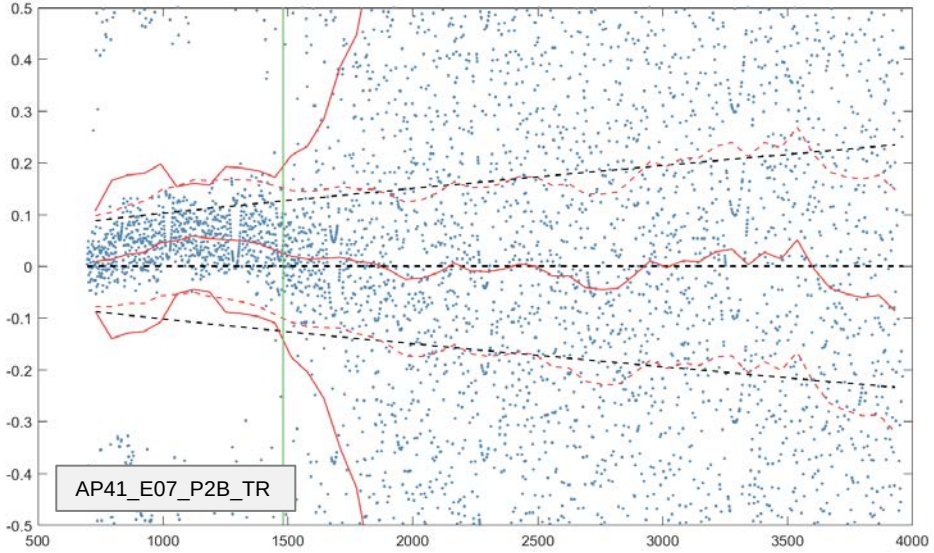
Variation in Mass Shift Profiles

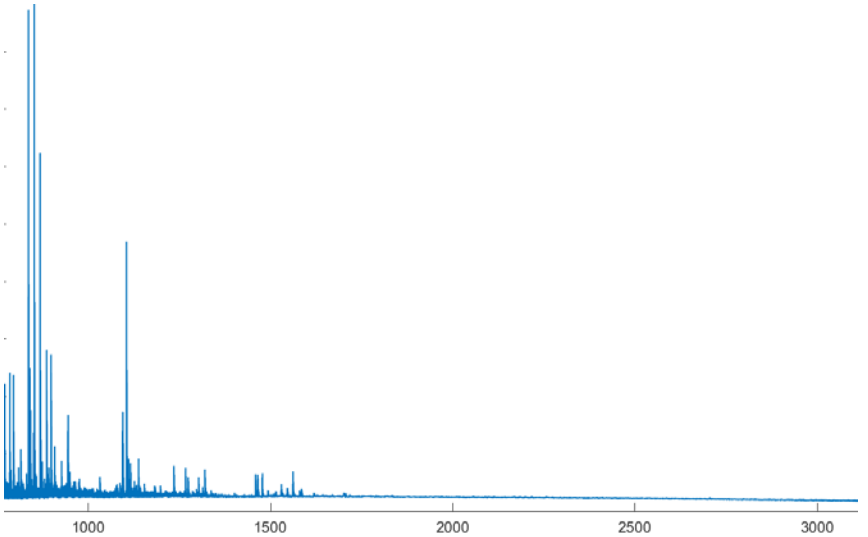


Mass Shift Profile and Mean Spectrum

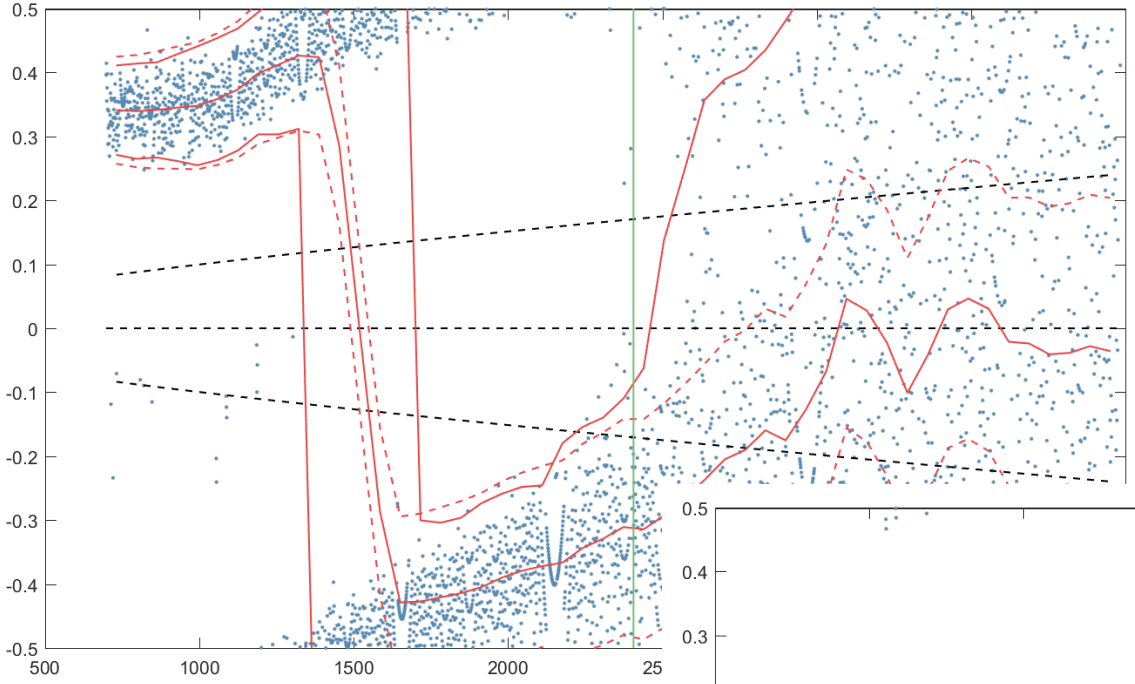




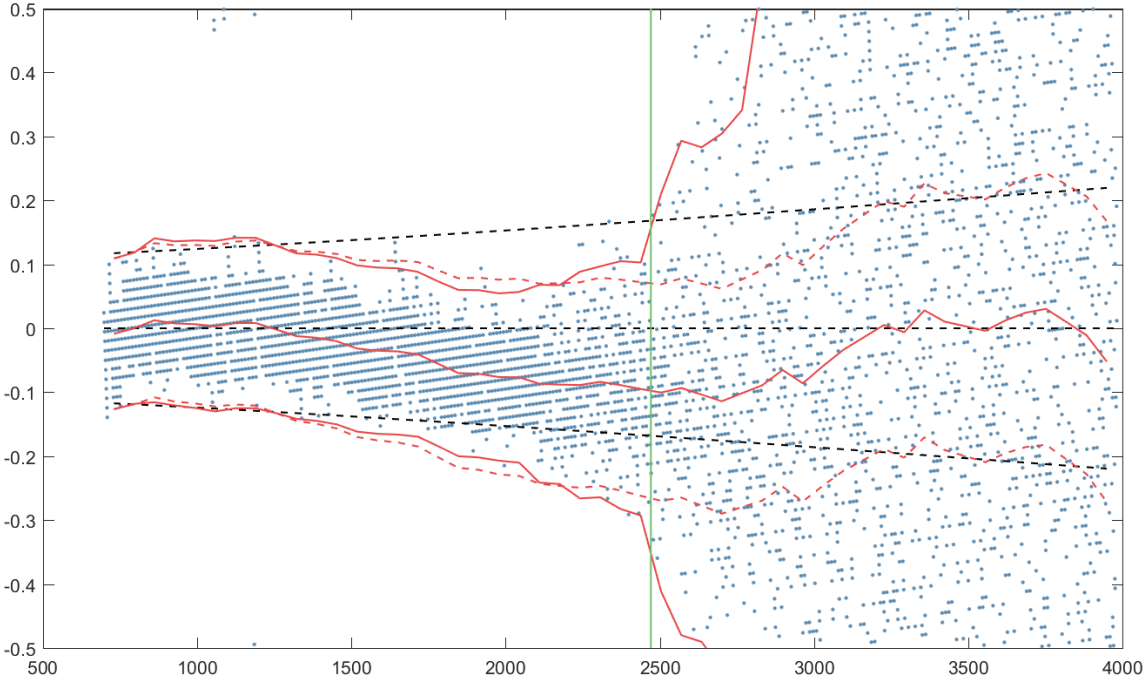




Effect of Mass Calibration

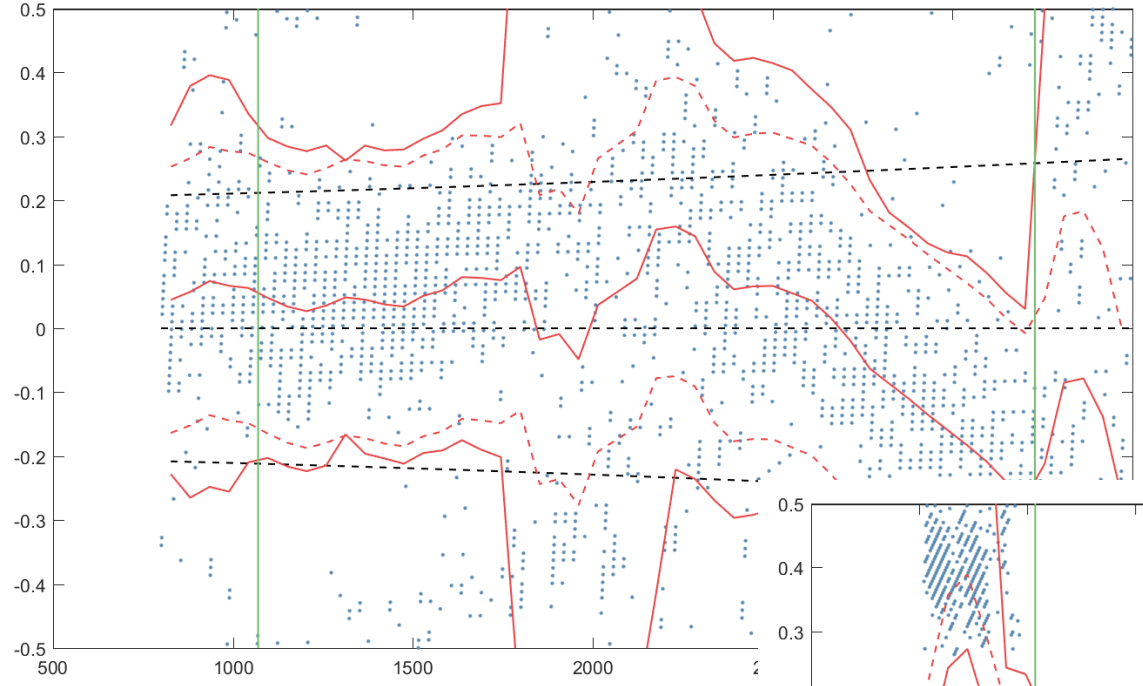


Artificial mass shift by deliberately incorrect external calibration

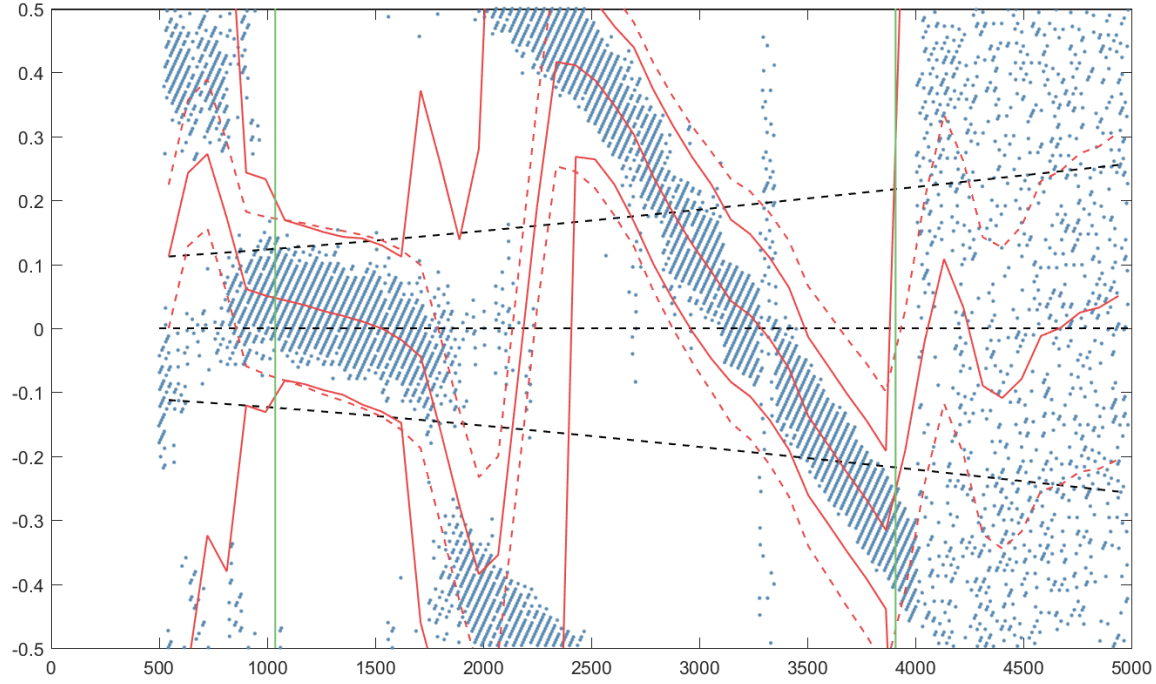


After statistical recalibration (Bruker software)

Extreme Examples



Coarse resampling at ≈0.3 Da

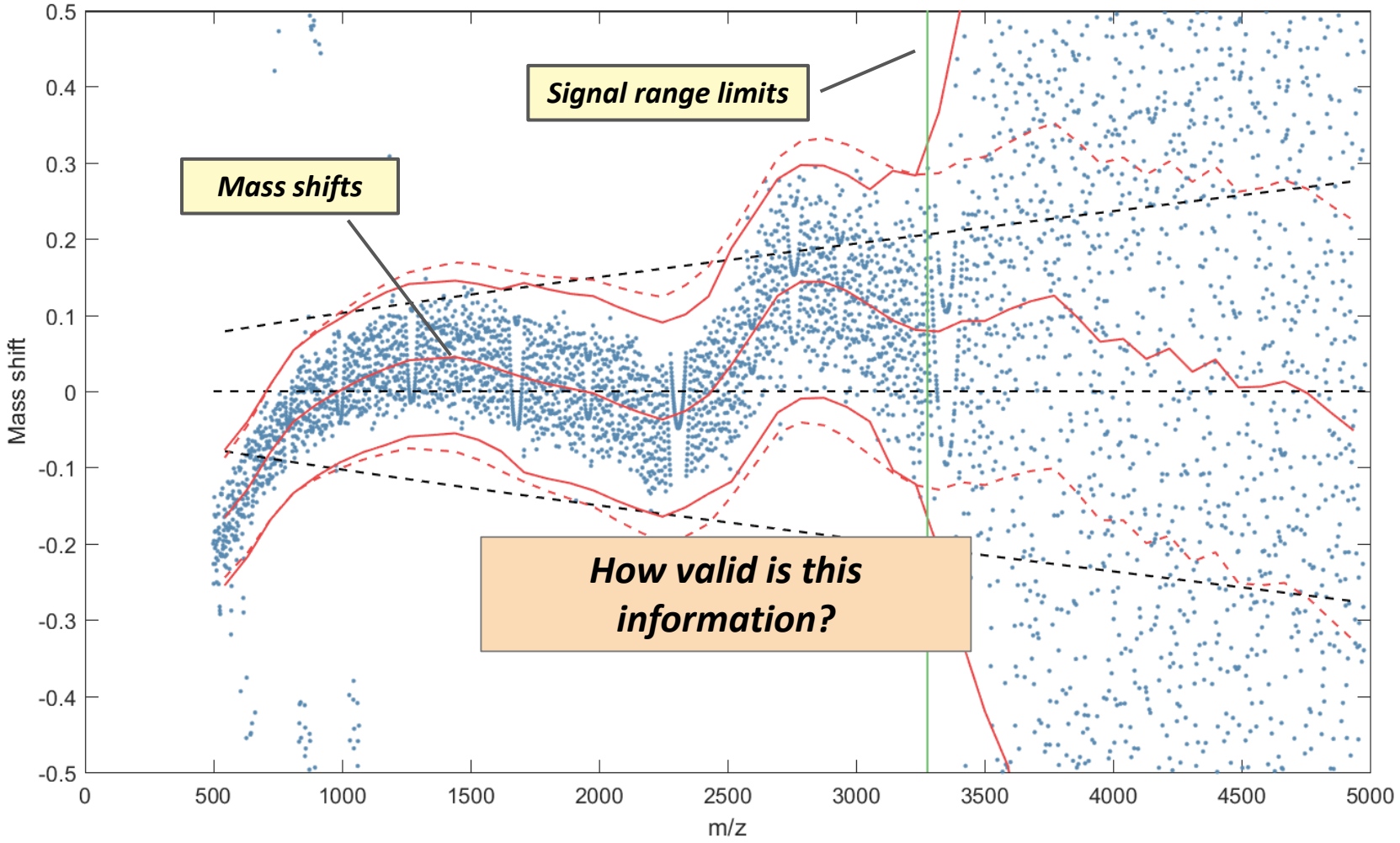


Severe mass distortion –
Caused by software failure?

Zentrum für Technomathematik

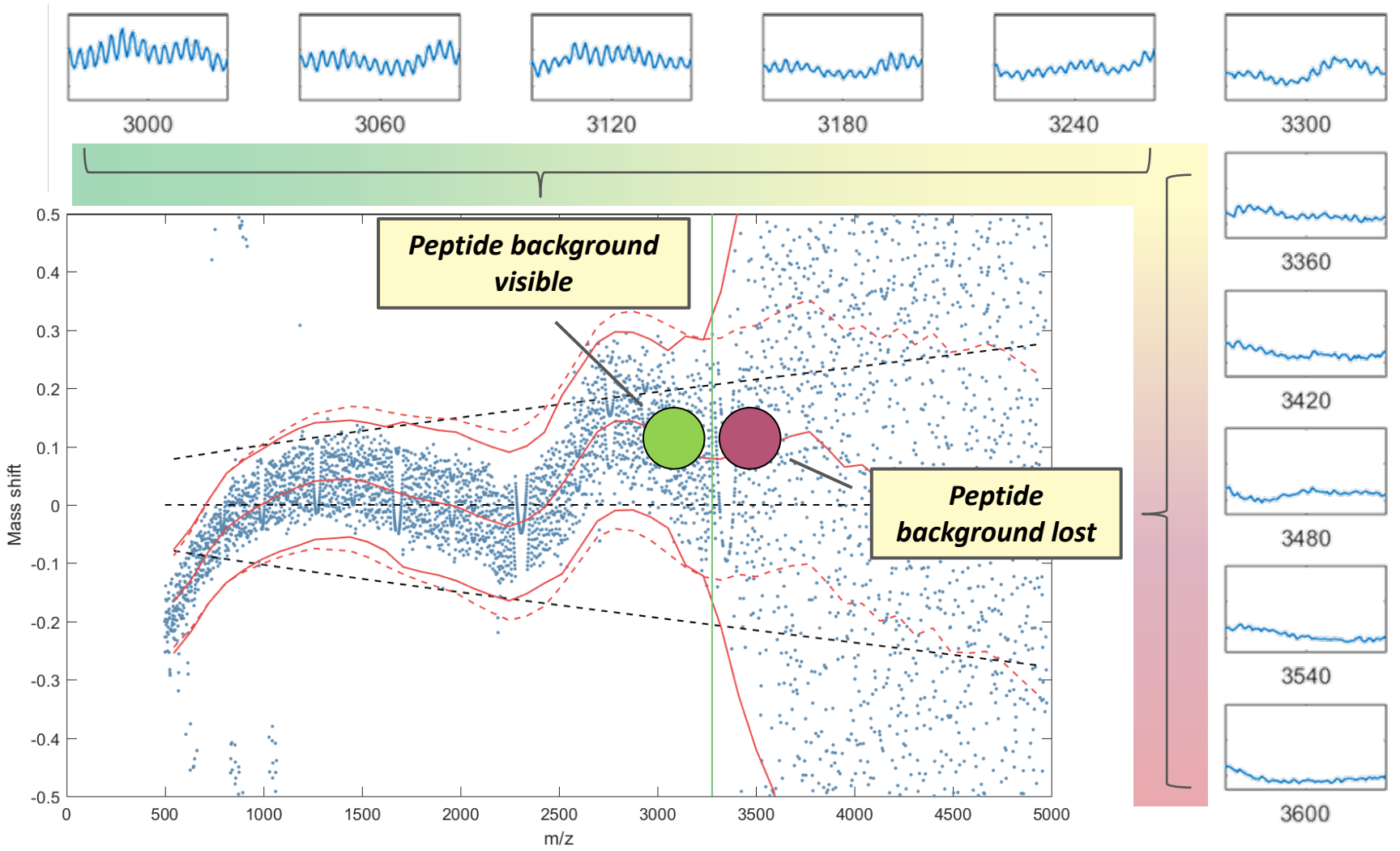
Reality Check

SCiLS



Universität Bremen*

Reality Check – Signal Range Limits



Zentrum für Technomathematik

SCiLS

Universität Bremen*

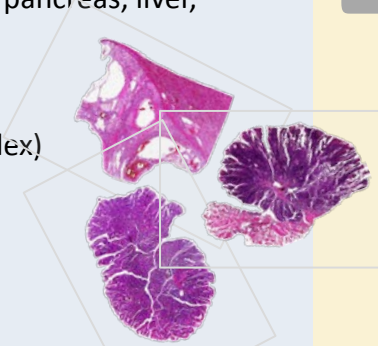
Reality Check – Mass Shifts

Zentrum für Technomathematik

SCiLS

1 MALDI MSI data from human tissue samples

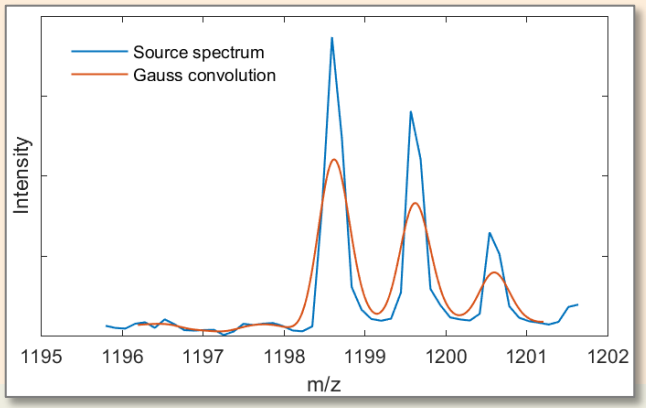
- 40 datasets (19 TMAs, 21 single sections)
- Taken from breast, ovary, colon, lung, pancreas, liver, lymph nodes, …
- Acquired on 3 sites
  (2 x Bruker Autoflex, 1 x Bruker Ultraflex)



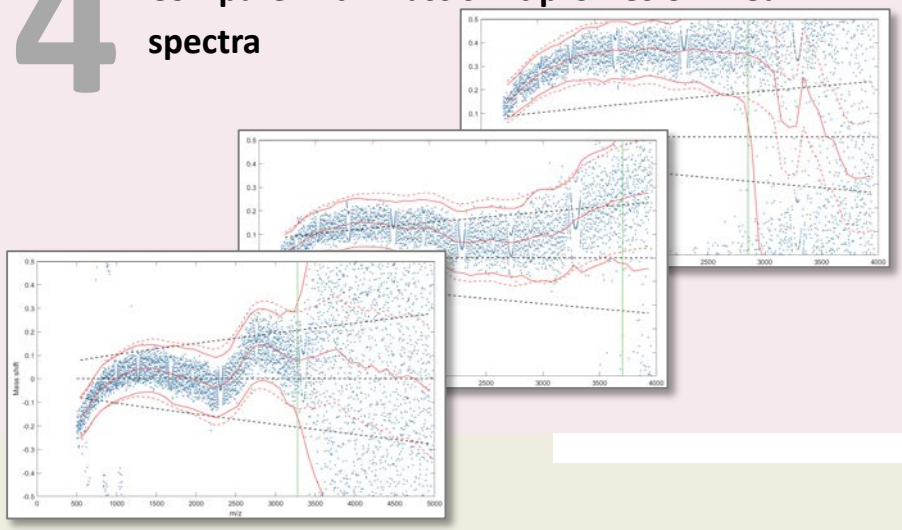
2 Analyze signals from 8 ubiquitous proteins (21 reference peptide peaks)

Actin, cytoplasmic 1	ACTB
Serum albumin	ALB
Collagen alpha-1(I) chain	COL1A1
Hemoglobin subunit alpha	HBA1
Hemoglobin subunit beta	HBB
Histone H4	HIST1H4A
78 kDa glucose-regulated protein	HSPA5
Vimentin	VIM

3 Identify peak intensities and shifts in single spectra by Gauss convolution within search interval

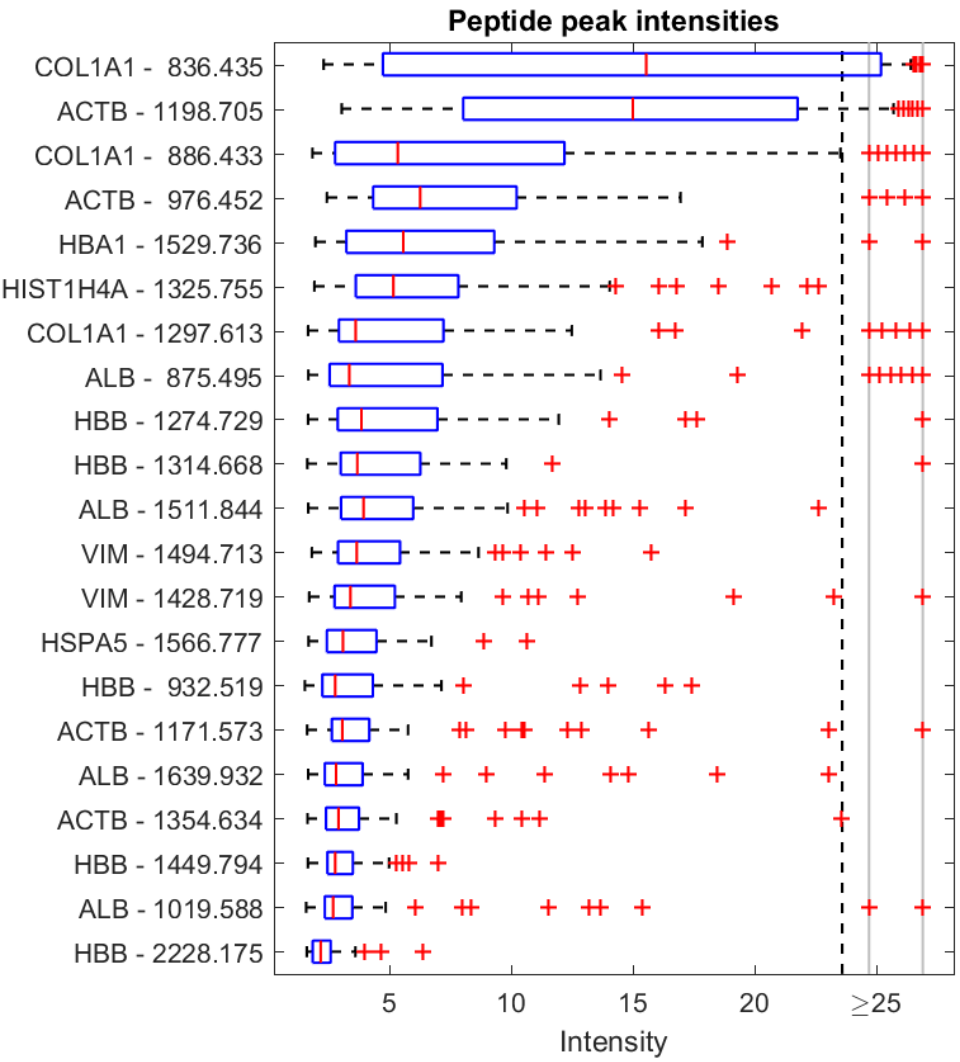


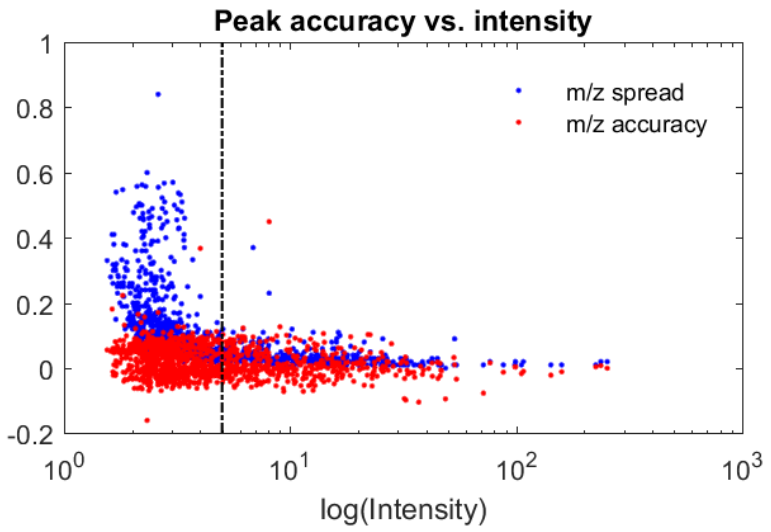
4 Compare with mass shift profiles on mean spectra

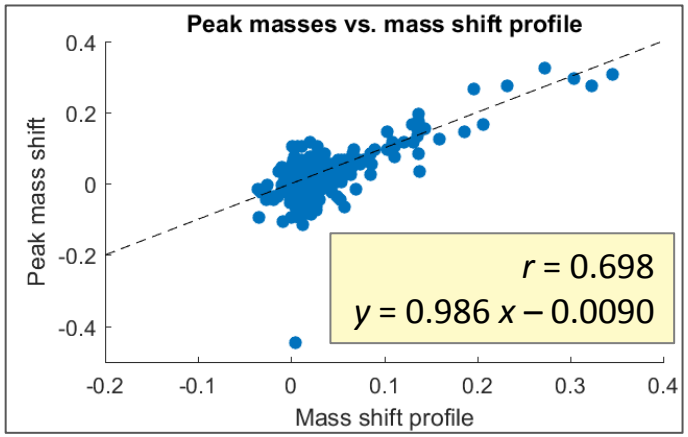


Universität Bremen*

Reality Check – Mass Shifts





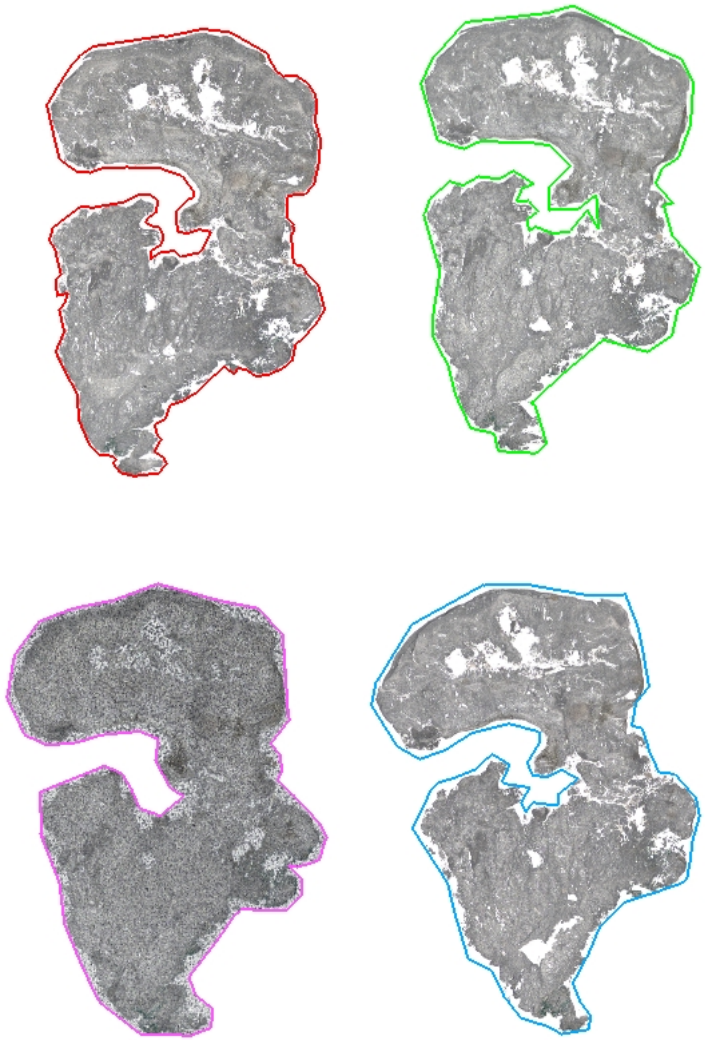


Zentrum für Technomathematik

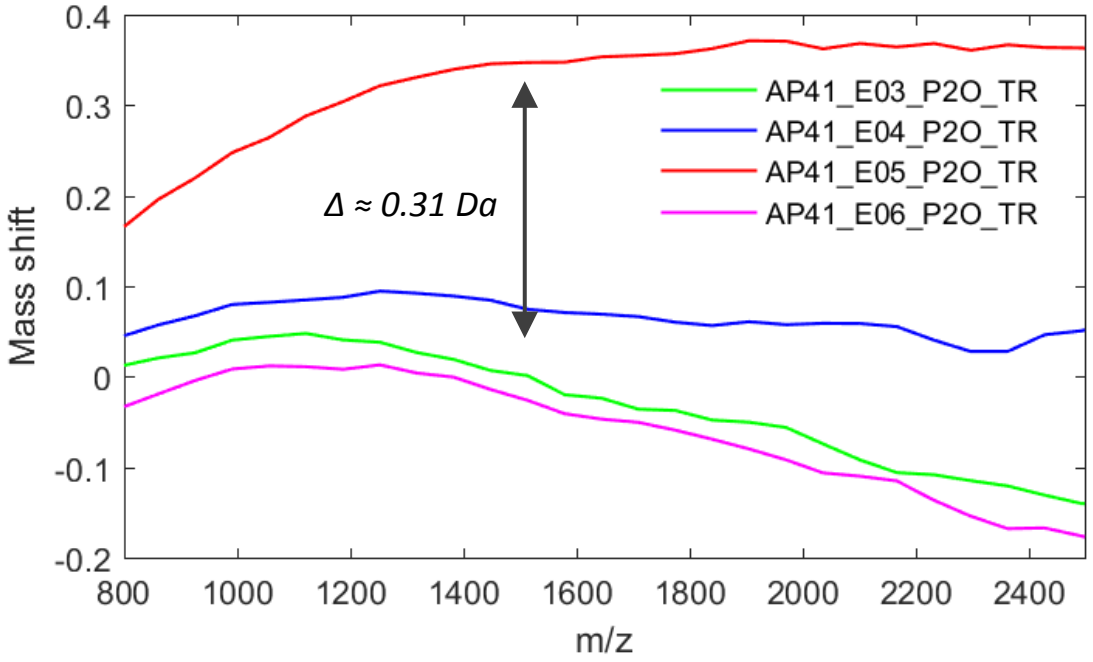
SCiLS

Universität Bremen*

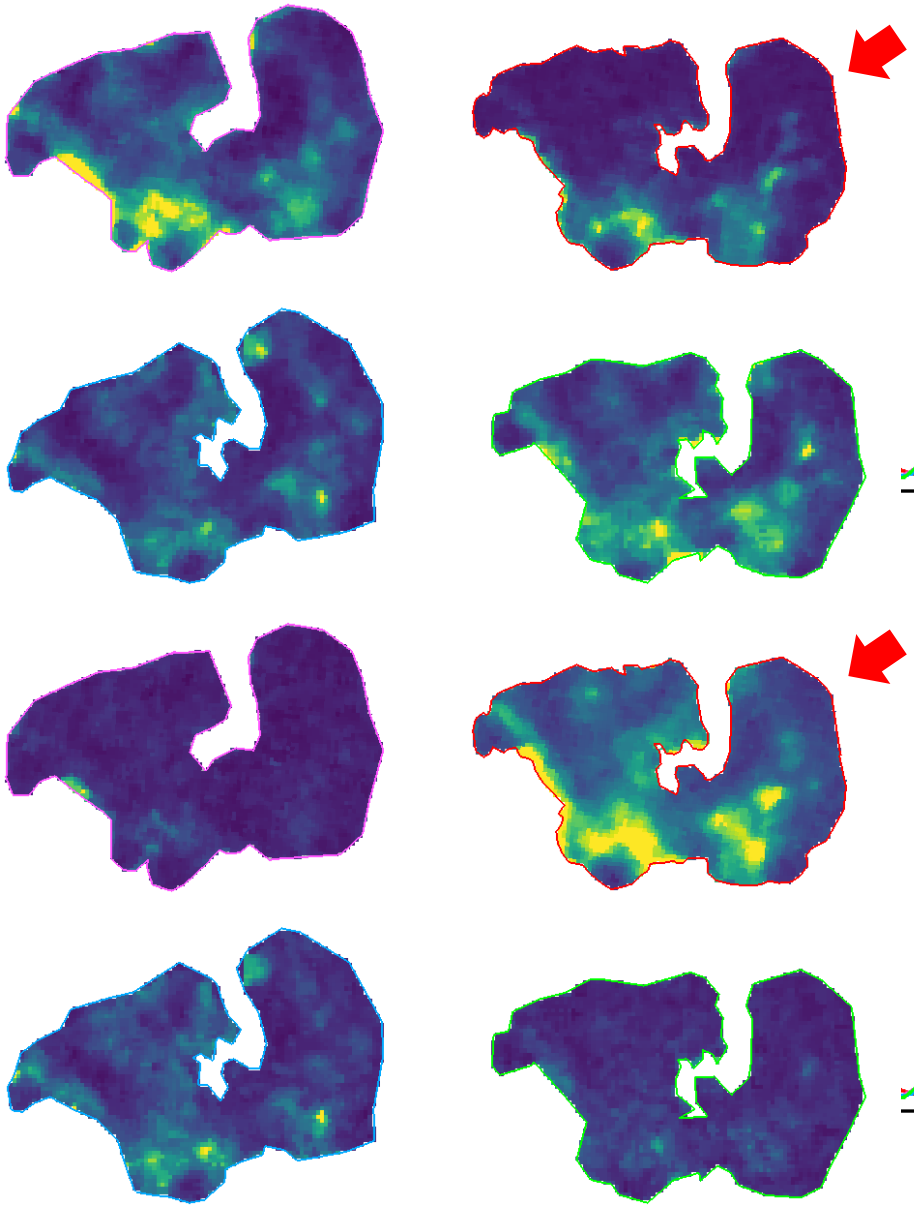
Real World Relevance: Ion Images



- Serial sections of an ovary cancer
- Experiments performed within two weeks under constant conditions
- Strong mass shift in one experiment
  > 0.3 Da at 1500 m/z



Real World Relevance: Ion Images



Real World Relevance: Classification

MCP MOLECULAR & CELLULAR PROTEOMICS

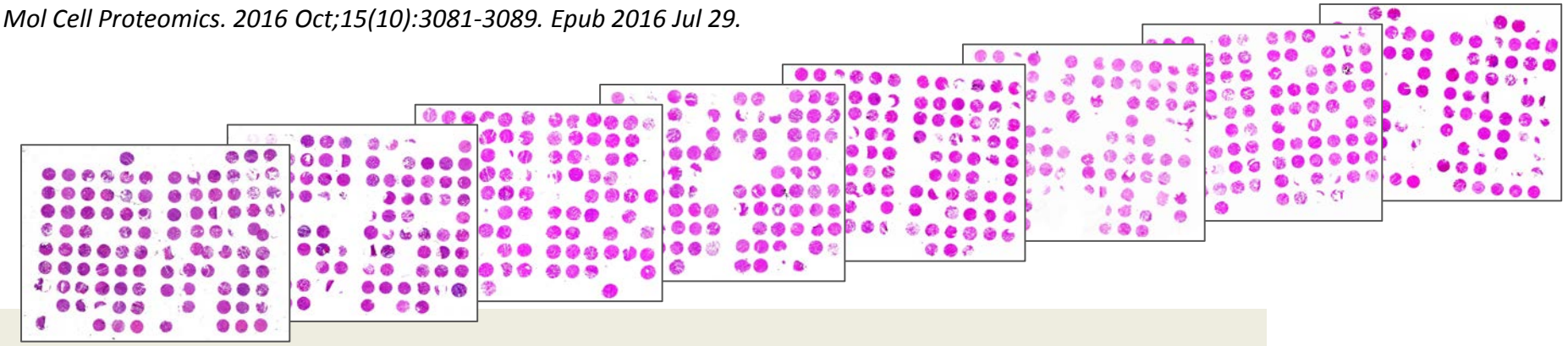
Reliable Entity Subtyping in Non-small Cell Lung Cancer by Matrix-assisted Laser Desorption/Ionization Imaging Mass Spectrometry on Formalin-fixed Paraffin-embedded Tissue Specimens*

Mark Kriegsmann‡,[a], Rita Casadonte§, Jörg Kriegsmann§,¶, Hendrik Dienemann‖, Peter Schirmacher‡, Jan Hendrik Kobarg**, Kristina Schwamborn‡‡, Albrecht Stenzinger‡,§§, Arne Warth‡,¶¶ and Wilko Weichert‡,‡‡,§§,‖‖

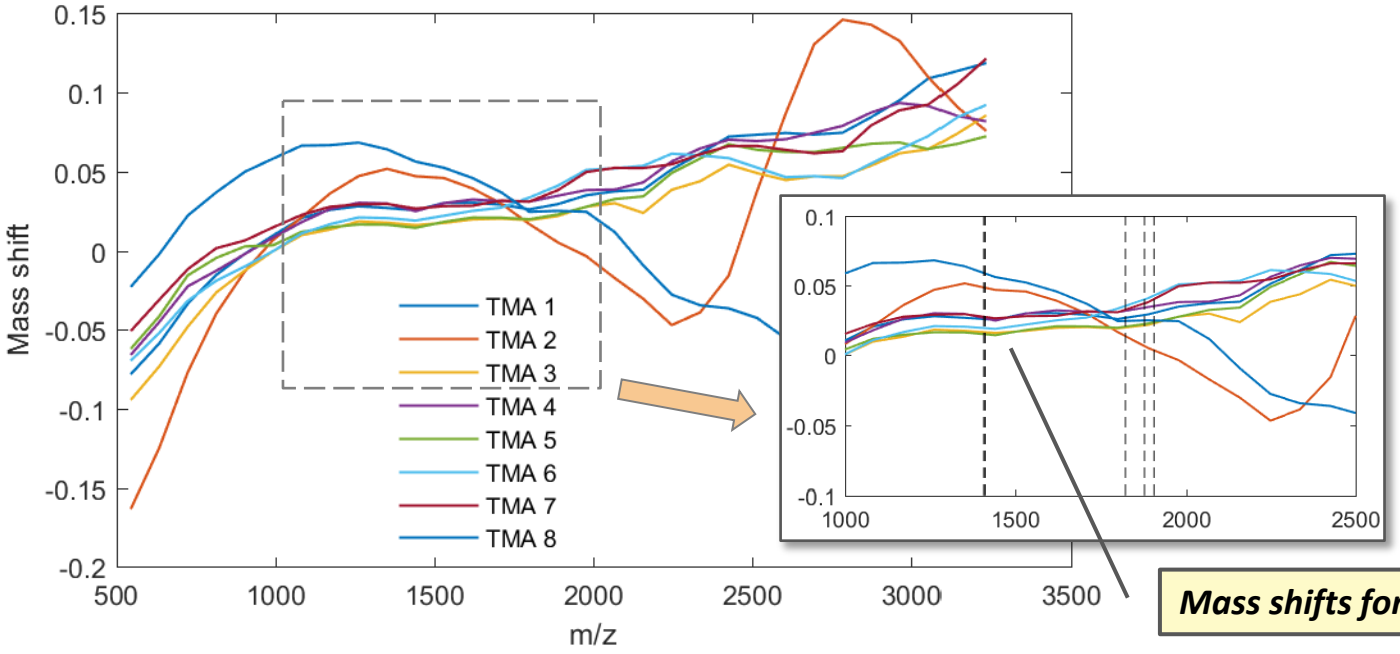
Mol Cell Proteomics. 2016 Oct;15(10):3081-3089. Epub 2016 Jul 29.

Classification model for lung adeno- vs. squamous cell carcinoma

- 8 TMAs, 326 patients, 168 ADC, 158 SqCC
- MALDI-TOF, Autoflex Speed (Bruker)
- LDA classification model based on 339 m/z values
- **Cross validation accuracy 99.1%**
- Subset of discriminating markers identified, including **CK5, CK7, CK15, HSP27**



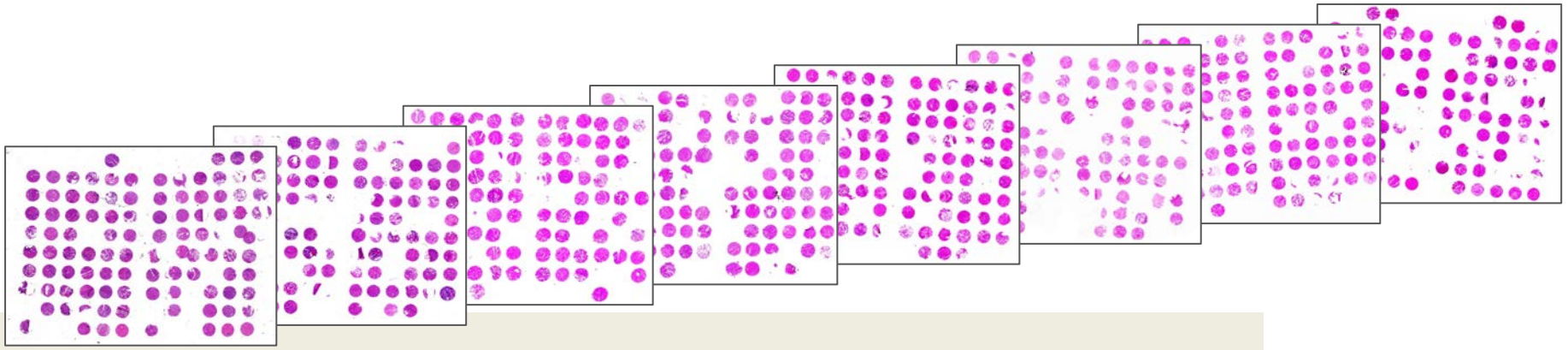
Real World Relevance: Classification



Peptide	m/z
CK5	1410.700
CK7	1406.600
CK15_1	1821.840
CK15_2	1877.850
HSP27	1905.940

Original m/z values from paper

Mass shifts for TMAs 1, 2 ≈ 0.05 Da

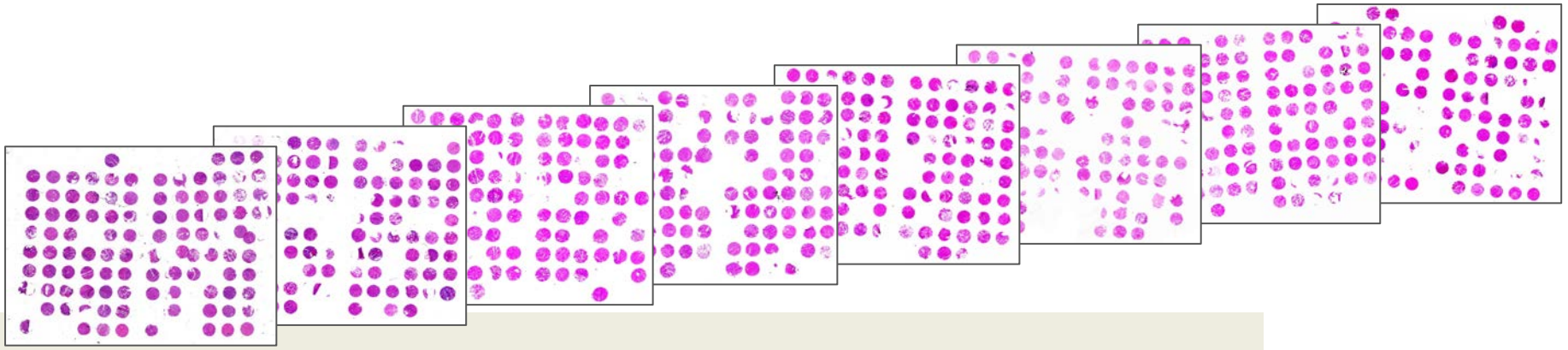


Universität Bremen*

Real World Relevance: Classification

- Mini classification model on 5 peptides
- Train on one TMA, validate on another
  - 8 models, 8 x 7 = 56 test combinations
- Two variants
  - Variant A: Original m/z values, same on all TMAs
  - Variant B: Per TMA mass shift adjusted m/z values
- Performance metric: Average sensitivity
- Compare performances in Variant A vs. B

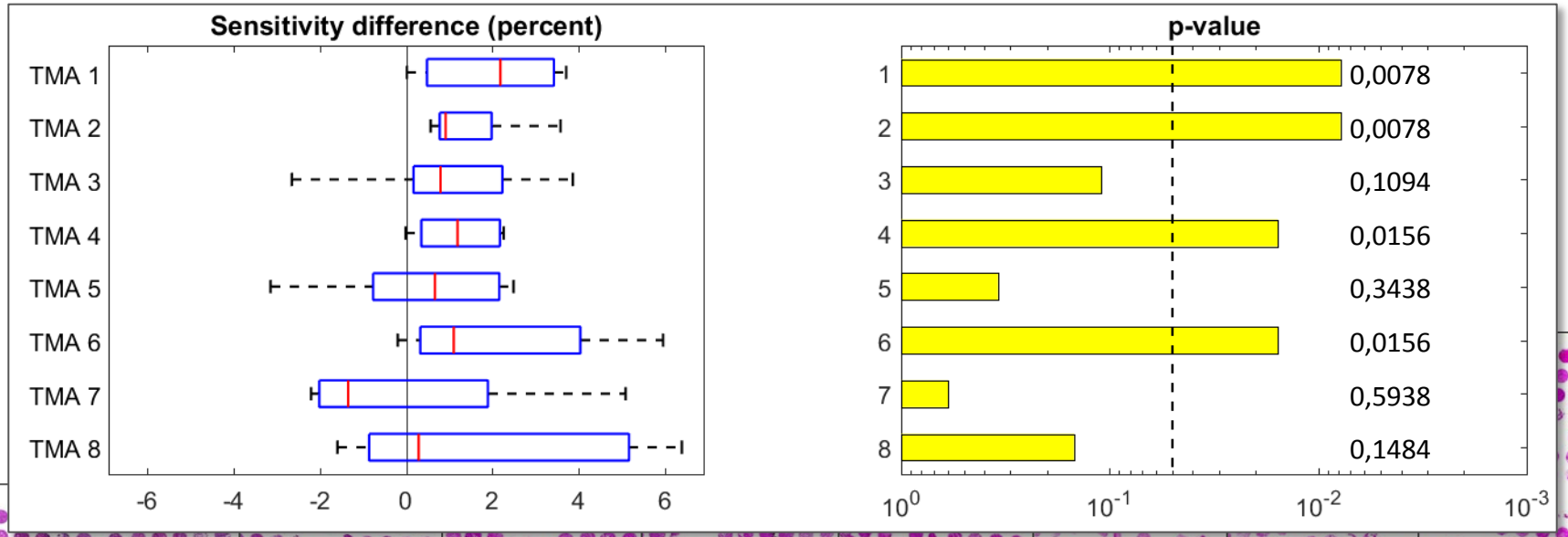
Peptide	m/z
CK5	1410.700
CK7	1406.600
CK15_1	1821.840
CK15_2	1877.850
HSP27	1905.940

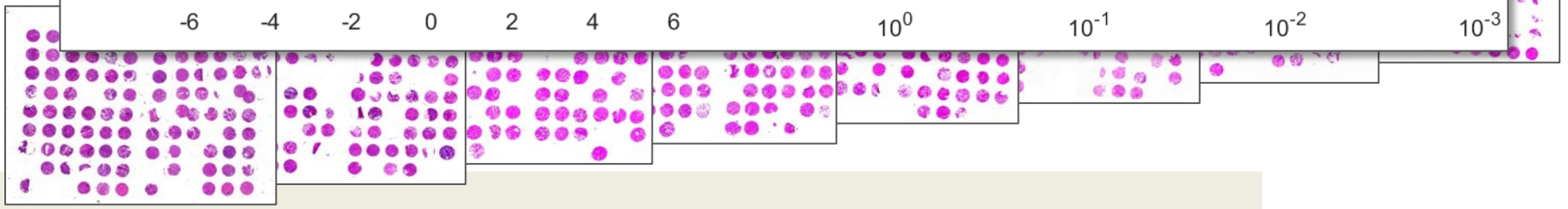


Universität Bremen*

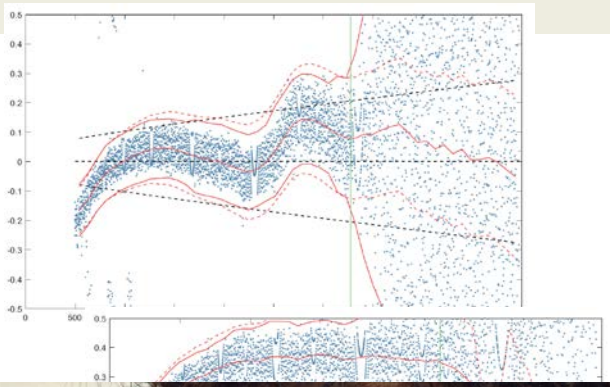
Real World Relevance: Classification

- Overall median performance Variant B: 77.6%
  - Five m/z features, one TMA for training
- Variant B (mass shift adjusted) better in 7 of 8 training TMAs
  - Statistically significant ($p < 0.05$) in 4 of 8 TMAs





Conclusion



Mass shift profiles useful for

- detecting data anomalies
- comparing measurements
- leveling differences between datasets

All data is dirty!

Data is not information.

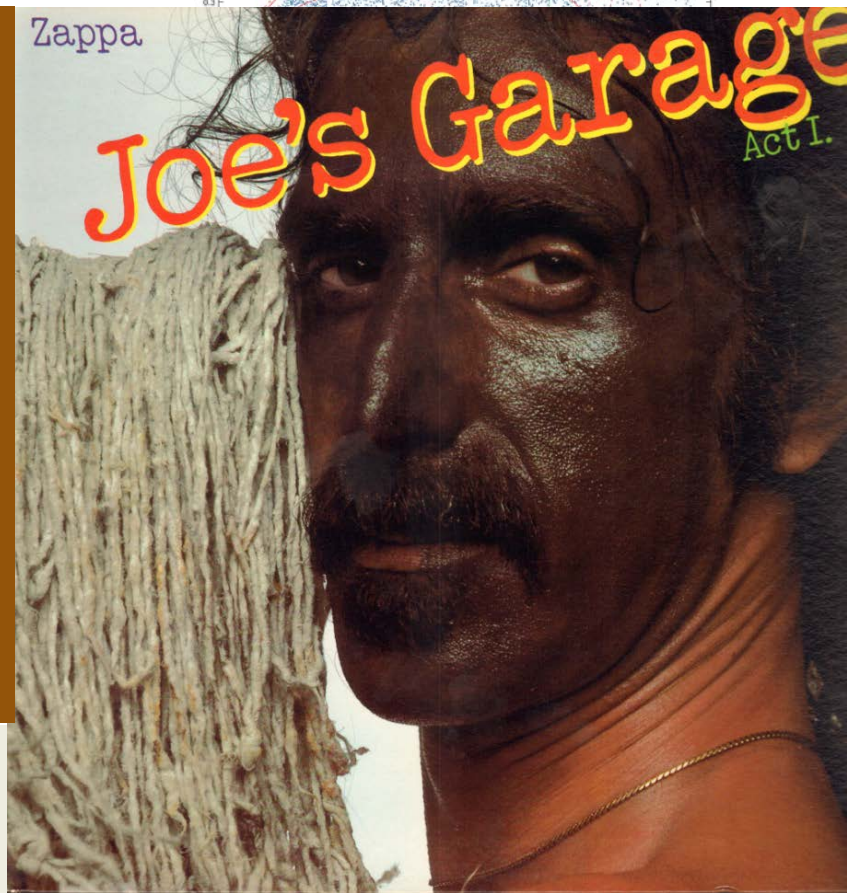
Information is not knowledge.

Knowledge is not wisdom.

Wisdom is not truth.

…

(Frank Zappa, Packard Goose, Joe‘s Garage)



Thank You

Center for Industrial Mathematics

- Delf Lachmund
- Jens Behrmann
- Yovany Cordero
- Christian Etmann
- Jost Vehmeyer
- Peter Maaß

MALDI Imaging Lab

- Janina Oetjen

SCiLS GmbH

- Jan-Hendrik Kobarg
- Orlando Galashan
- Dennis Trede

Proteopath, Trier

- Rita Casadonte
- Rémi Longuespée
- Jörg Kriegsmann

More tissue and data contributed by

- Mark Kriegsmann, UK Heidelberg
- Birte Beine, ISAS Dortmund
- Sebastian Huss, UK Münster
- Oliver Klein, BCRT Berlin
- Axel Wellmann, Celle

GEFÖRDERT VOM
Bundesministerium für Bildung und Forschung
ZIM Zentrales Innovationsprogramm Mittelstand
DFG Deutsche Forschungsgemeinschaft