# Comparing the Performance of LC-MS Processing Software

**Ralf Tautenhahn, Steffen Neumann**

Stress & Developmental Biology,
IPB Halle, Weinberg 3, 06120 Halle, Germany {rtautenh|sneumann}@IPB-Halle.DE

## Introduction

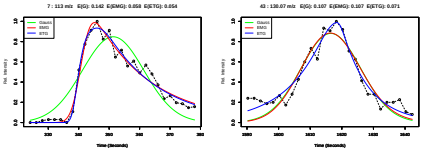
- Metabolomics is the quantification of a system's metabolites
- Mass spectrometry is a powerful analysis method for (secondary) plant Metabolites
- Most metabolites are still uncharacterised
- Peak picking and alignment are critical steps in the analysis of the LC-MS raw data
- Verification of the results is expensive
- Quantity and Quality of peaks in the measured sample is unknown

## Aim

- Compare performance of LC-MS processing software
- Assess the quality of peak picking and alignment
  - Tolerance against chemical & physical noise
  - Influence of peak shape on peak picking
  - Correctness of quantification
  - Information about rectifiable retention time differences
- Feedback & Improvement

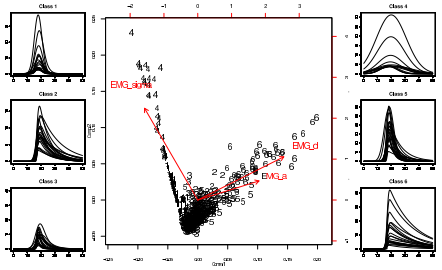
## Fitting LC-MS Peaks

- Peaks selected manually and by XCMS
- Raw data fitted to model using
  - Normal Gaussian (2 Parameters)
  - Exponentially Modified Gaussian (EMG, Marco et Bombi 2001) with 4 Parameters
  - Empirically Transformed Gaussian (ETG, Li 1997) with 11 Parameters
- Goodness-of-fit for experimental peaks is generally good for EMG and ETG



## Clustering of Model Parameters

- Machine Operators have a "feeling" for Peaks
- Clustering of EMG Parameters shows clusters with distinct shapes

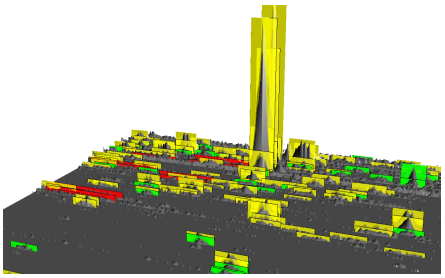



## Synthetic Peaks

- Synthetic Peaks are generated corresponding to the peak shapes found in the real data
- Different types of noise can be added, reflecting chemical and physical noise in the experimental data
- Other Distortions like time-shifts observed between different LC-MS runs can be introduced
- Quantity and Quality of peaks in the synthetic sample is *known*

## Performance

- Visualisation of peaks found by different programs and/or settings



## XCMS

- Part of `www.bioconductor.org`
- Peak Picking starts in the extracted ion chromatograms
- Matched filter approach
- Alignment through analysis of peak distribution for 'well behaved' peak groups
- `metlin.scribbs.edu`

## MetAlign

- Commercial, closed source
- Windows only
- Restricted to nominal masses
- `www.metabolomics.nl`

## MZmine

- Multiple filtering and normalisation methods
- Peak Picking starting from the spectra
- Simple successive alignment
- `mzmine.sourceforge.net`

## OpenMS

- Advanced binning algorithms
- 2D Peak Picking using peak models
- Alignment using 2D - 'Map Matching'
- `www.openms.de`

## Conclusion

- Quality and Quantity of peaks picked differ by program
- Broad range of processing speeds, depending on accuracy
- Parameter selection is sometimes crucial
- Some programs have special features e.g.
  - OpenMS - extensive preprocessing
  - XCMS - advanced alignment techniques
  - MZmine - support for parallel processing
  - MetAlign - long standing development

⇒ Robust Peak Picking still needed
⇒ Robust Alignment still needed
⇒ Hierarchical combination of e.g. XCMS and OpenMS

## Acknowledgements

- Dierk Scheel, Jürgen Schmidt and Christoph Böttcher for their discussions on mass spectrometry and biology

bmb+f The work is supported under Grant 0312706G