technische universität
dortmund

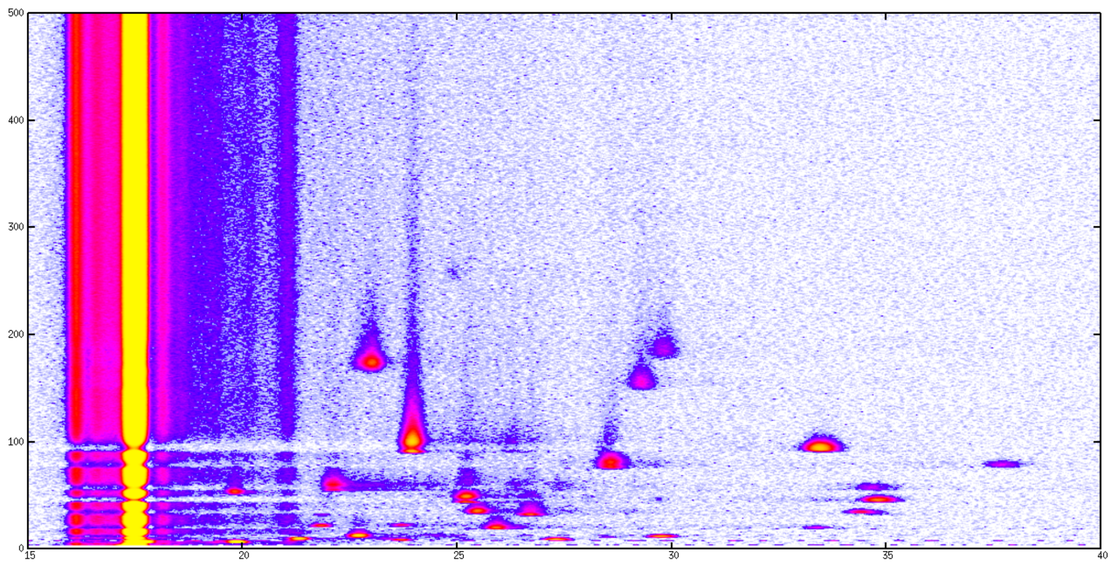
Resource-Constrained Analysis of Spectrometry Data

Dominik Kopczynski* Sven Rahmann†

*Bioinformatics for High-Throughput Technologies, Algorithm Engineering, Computer Science XI, TU Dortmund, Germany
dominik.kopczynski@tu-dortmund.de

† Genome Informatics, Institute of Human Genetics, Faculty of Medicine, University of Duisburg-Essen, Germany
sven.rahmann@uk-essen.de

Measurement

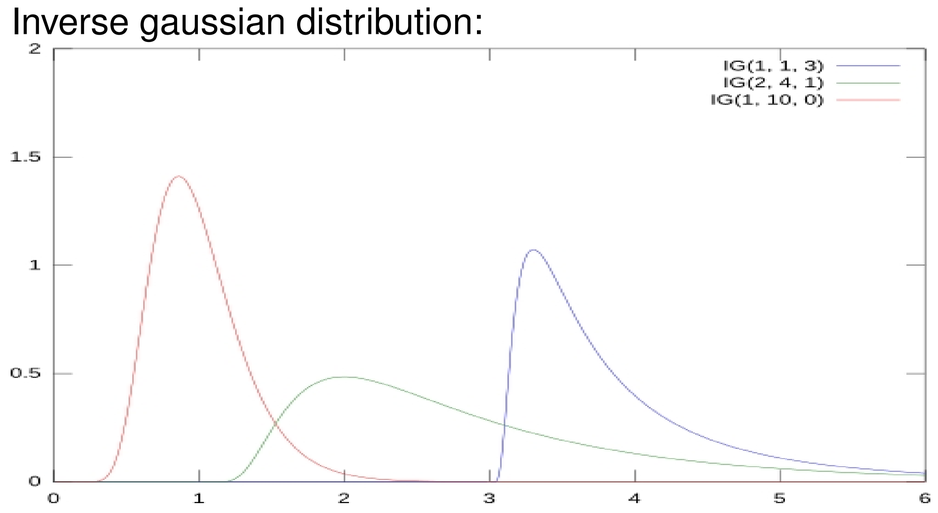


- Ion mobility spectrometry (IMS) device coupled with a multi-capillary column (MCC)
- Depending on resolution amount of data points from 3 to 75 million
- Visualisation as heatmap (X-axis: drift-time d in ms; Y-axis: retention-time r in s)
- Regions with high signal intensity called peaks

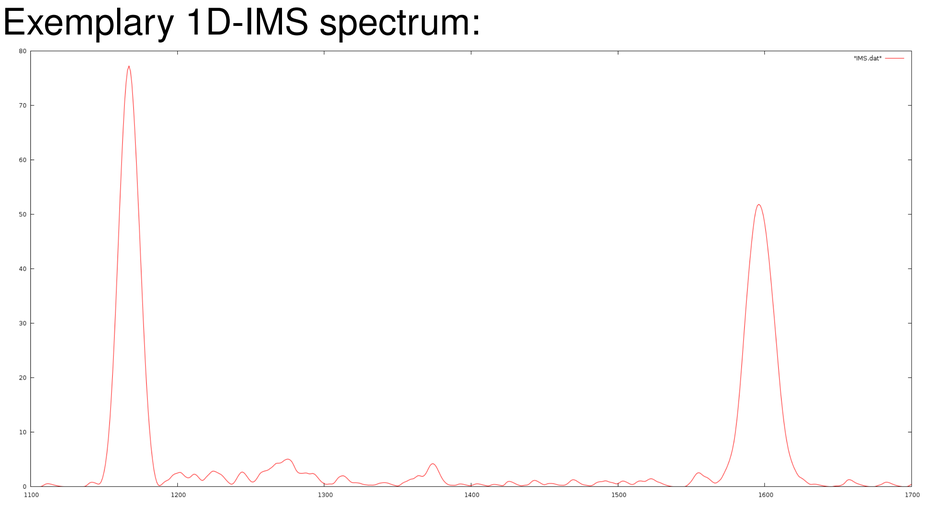
Feature Selection

- Modeling 1D-IMS spectra as weighted sum of inverse gaussian distributions
- $IG(x;\mu,\lambda,o) = \left(\frac{\lambda}{2\pi(x-o)^3}\right)^{\frac{1}{2}} \cdot \exp\left(-\frac{\lambda((x-o)-\mu)^2}{2\mu^2(x-o)}\right)$
- Use EM algorithm for estimation of all parameters for all models
- After processing one spectrum connect models to coherent models in previous spectrum
- Use EM algorithm again to estimate parameters for 2D peak model with already processed 1D model chains
- Seven parameters per 2D model
- $M(r,d;\mu_r,\lambda_r,o_r,\mu_d,\lambda_d,o_d,\omega) = \begin{cases} \omega \cdot IG(r;\mu_r,\lambda_r,o_r)\cdot IG(d;\mu_d,\lambda_d,o_d) & \text{if } r > o_r \wedge d > o_d, \\ 0 & \text{otherwise} \end{cases}$
- Data reduction factor ranges from 10,000 to 250,000.

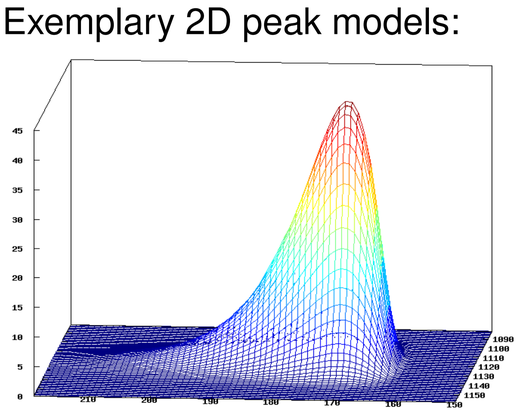
Inverse gaussian distribution:

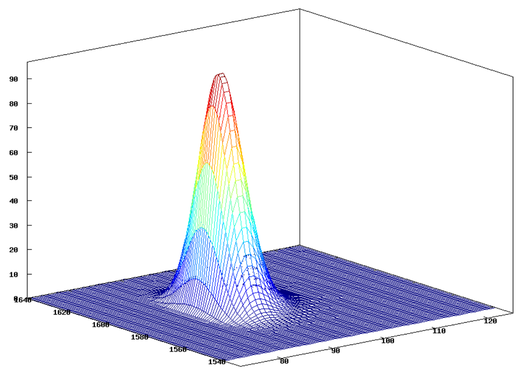


Exemplary 1D-IMS spectrum:



Exemplary 2D peak models:





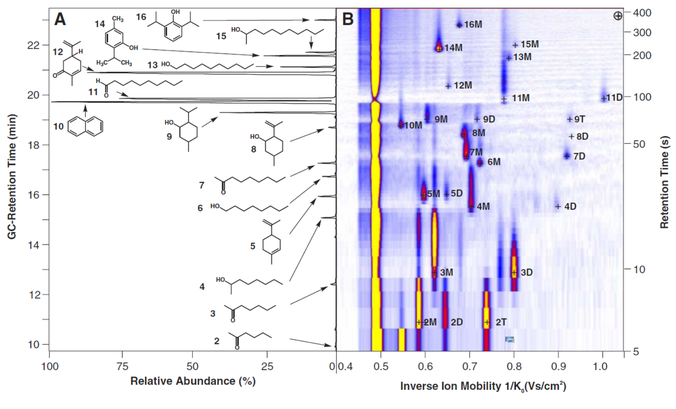
Further Work

Peak Alignment:

- Wastage parts, temperature or carrier/-drift gas flow lead to displaced peaks in measurement
- MCC/IMS devices not exactly calibrated
- necessary for correct peak comparison

Registrate Metabolites:

- Completing IMS metabolite database
- Use gas chromatography (GC) data to identify metabolites in IMS measurement with normalized data (alignment)



Prediction:

- Metabolite calibration very costly in terms of time
- simulation of MCC/IMS device with unregistrated compounds just by considering chemical structures