



## Simulation and Data Treatment in Chemical Analysis of Fossil and Regenerative Resources

**M. Otto<sup>1</sup>, P. Rathsack<sup>1</sup>, M. Kreuzer<sup>2</sup>, M. Kroll<sup>1</sup>**

<sup>1</sup>TU Bergakademie Freiberg, Institute of Analytical Chemistry, D-09596 Freiberg, Germany

<sup>2</sup>University of Applied Sciences, Bioinformatics Group, D-09648 Mittweida, Germany

### Abstract

This work describes new methods for data analysis in comprehensive gas chromatography and ultra-high resolution mass spectrometry in cases where samples contain hundreds or thousands of organic compounds. In two-dimensional gas chromatography classes of organic compounds can be modelled on the basis of mass spectral features and unknown compounds can be assigned to those classes by Bayesian discriminant analysis. High molecular weight compounds were analysed by ultra-high resolution mass spectrometry and their feasible fragmentation patterns are simulated on the basis of graph theory and validated by the corresponding mass spectra.

### Introduction

Chemical analysis of organic constituents in fossil and regenerative resources (coal, crude oil, biomass, scrap tires) and their conversion products remains a challenging task, because of the sheer numbers of compounds which have to be identified and quantified. Although novel developments of high performing analytical methods, such as, multidimensional analytical separations and ultra-high resolving mass spectrometry (MS) provide new platforms to analyse samples of this complex nature. However, the amount of data generated by this instrumentation can only be handled by extensive use of chemometric methods.

Over the years the number of organic compounds in the environment increased with the in-depth understanding about possible isomers. Kerber et. al. [1] computed all mathematically possible constitutional isomers up to a molecular mass of 150 Da and compared it to the number of compounds given in the Beilstein data base and the number of spectra in the *NIST* database of Mass Spectra. As an example for the molecular formula  $C_6H_6N_4O$  they calculated 151,838,122 isomers whereas there are only 273 compounds given in the *Beilstein* and just 11 compound spectra are found in the *NIST* database. In two-dimensional separations and in ultra-high resolution MS you will encounter many more of those isomers that cannot be processed by hand in reasonable time. In this contribution a classification method is presented to address organic compound classes in comprehensive gas chromatography mass spectrometry (GG×CG-MS). As a second approach graph theory is used for mathematical representation of complexly composed chemical mixtures and applied to fragmentation schemes in ultra-high resolution mass spectrometry.

### Materials & Methods

Pyrolysis liquids of brown coals from central Germany (Schöningen) were obtained from a fixed bed laboratory reactor at temperatures between 400 and 700 °C. A Pegasus 4D System (Leco) was used for analysis of pyrolysis oils. High resolution mass spectra were recorded by using a 15 T Solarix FT-ICR-MS from Bruker Daltonics, equipped with an electrospray ionisation source in negative ion mode. Bruker *Data Analysis* software was used for peak detection and molecular formula assignment. All in-house-scripts were run in Matlab (Mathworks).

### Results

Organic compounds up to molecular masses of 400 Da can be separated in multi-constituents samples, in principal, by two-dimensional gas chromatography. However, in the two-dimensional chromatogram the different classes of chemical compounds overlap in many cases and, therefore, the

compound clusters in the separation plane cannot be described by simple polygons. If electron impact mass spectra are available for each separated compound a more sophisticated classification model can be developed. In this work the mass spectra are transformed into different features, such as, *modulo-14 summation* [2]. The resulting matrix is decomposed into principal components and the scores from PCA are modeled on the compound classes by *linear discriminant analysis*. Prediction of unknown compounds is feasible after pretreatment of the peak and mass spectral information in the same way as for the reference compounds (Fig. 1). The *likelihood* of class assignments is calculated from Bayesian statistics as the posterior probability [3].

Compounds of higher molecular masses can be investigated by means of Fourier transform ion cyclotron resonance mass spectrometry (FT-ICR-MS). In case of the analysis of fossil and regenerative resources, mass spectra contain thousands of peaks which correspond to a large number of molecular formulas that represent an unknown amount of isomers.

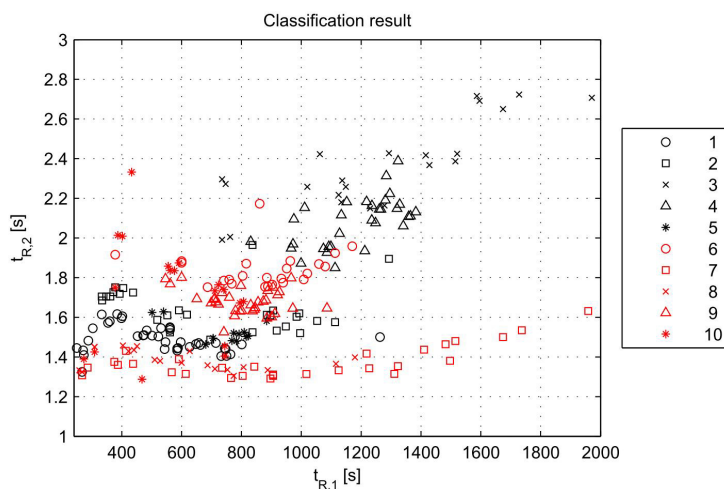


Fig. 1; Classification of compounds in pyrolysis oil from a scrape tire

In order to derive the exact compound structure, fragmentation of the compounds by collision induced decay (CID) is needed. The mathematically possible fragments were simulated by graph theory and validated by the experimental mass spectra. The complete CID-graph of a compound with the molecular formula  $C_{40}H_{62}O_{19}$  is depicted in Fig. 2 in dependence on the collision energy applied.

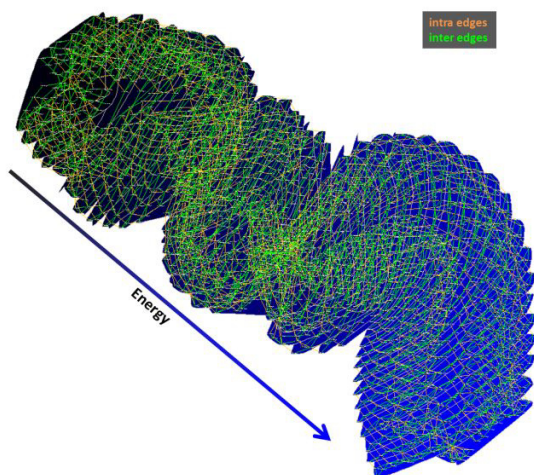


Fig. 2; Complete CID-graph of a compound with molecular formula  $C_{40}H_{62}O_{19}$

develop chemometric methods in order to cope with the big data streams and to extract optimum information from those data.

## References

- 1) A. Kerber, R. Laue, M. Meringer, C. Rücker, E. Schymanski, *Mathematical Chemistry and Chemoinformatics – Structure Generation, Elucidation and Quantitative Structure – Property Relationships*, Berlin: De Gruyter, 2014, ISBN 978-3-11-030007-9.
- 2) K. Varmuza, P. Filzmoser P., *Introduction to multivariate statistical analysis in chemometrics*, Boca Raton: Taylor & Francis, 2009, ISBN 978-1-4200-5947-2.
- 3) M. Otto, *Chemometrics*, 2<sup>nd</sup> Ed., Weinheim: Wiley-VCH, 2007, ISBN 978-3-527-31418-8