Journal of Chromatography A, 1364 (2014) 241–248

Contents lists available at ScienceDirect

Journal of Chromatography A

journal homepage: www.elsevier.com/locate/chroma

Advanced scripting for the automated profiling of two-dimensional gas chromatography-time-of-flight mass spectrometry data from combustion aerosol☆

Benedikt A. Weggler[a,b,c,1], Thomas Gröger[a,b,*], Ralf Zimmermann[a,b,c,1]

[a] Joint Mass Spectrometry Centre, Cooperation Group "Comprehensive Molecular Analytics" Helmholtz Zentrum Muenchen, D85764 Neuherberg, Germany
[b] Joint Mass Spectrometry Centre, Institute of Chemistry, Chair of Analytical Chemistry UNiversity of Rostock, D18057 Rostock, Germany
[c] Helmholtz Virtual Institute of Complex Molecular Systems in Environmental Health – Aerosol and Health (HICE)

ARTICLE INFO

Article history:
Received 21 March 2014
Received in revised form 28 July 2014
Accepted 27 August 2014
Available online 1 September 2014

Keywords:
Comprehensive two dimensional gas chromatography
Data mining
Ship fuel combustion
Aerosol

ABSTRACT

Multidimensional gas chromatography is an appropriate tool for the non-targeted and comprehensive characterisation of complex samples generated from combustion processes. Particulate matter (PM) emission is composed of a large number of compounds, including condensed semi-volatile organic compounds (SVOCs). However, the complex amount of information gained from such comprehensive techniques is associated with difficult and time-consuming data analysis. Because of this obstacle, two-dimensional gas chromatography still receives relatively little use in aerosol science [1–4]. To remedy this problem, advanced scripting algorithms based on knowledge-based rules (KBRs) were developed in-house and applied to GCxGC-TOFMS data. Previously reported KBRs and newer findings were considered for the development of these algorithms. The novelty of the presented advanced scripting tools is a notably selective search criterion for data screening, which is primarily based on fragmentation patterns and the presence of specific fragments. Combined with "classical" approaches based on retention times, a fast, accurate and automated data evaluation method was developed, which was evaluated qualitatively and quantitatively for type 1 and type 2 errors. The method's applicability was further tested for PM filter samples obtained from ship fuel combustion. Major substance classes, including polycyclic aromatic hydrocarbons (PAH), alkanes, benzenes, esters and ethers, can be targeted. This approach allows the classification of approximately 75% of the peaks of interest within real PM samples. Various conditions of combustion, such as fuel composition and engine load, could be clearly characterised and differentiated.

1. Introduction

1.1. Aerosol and health/HICE

Epidemiological studies indicate that fine air particulate matter and traffic-related air pollution are correlated with severe health effects, including increased mortality and prevalence of cardiovascular and respiratory diseases and allergies [5]. These studies also support the hypothesis that both the physical (particle size, shape, surface) and chemical (dissolved and adsorbed chemicals, surface catalytic reactions) properties of the particles are involved in the toxic, genotoxic and carcinogenic mechanisms of inhaled particulates.

In the Helmholtz Virtual Institute of Complex Molecular Systems in Environmental Health (HICE), local and global competencies in the fields of analytical chemistry, aerosol science and chemical toxicology are joined to investigate the long-term effects of environmentally related diseases. One of the main research points is the investigation of reactive organic compounds in environmental aerosols and the synergistic effects of the gaseous and particulate phases. Many of these effects are supposed to be induced by the organic PM compounds or corresponding ageing products. Unfortunately, the elucidation of the organic fraction is rather complex.

Depending on the sampling sites, the contribution of organic material is between 20 and 90% of the total aerosol fine PM mass [6] and often consists of several thousand compounds. For a

---

☆ Presented at the 13th International Symposium on Hyphenated Techniques in Chromatography and Separation Technology, Bruges, Belgium, 29–31 January 2014.

* Corresponding author at: Kooperationsgruppe "Comprehensive Molecular Analytics (CMA)", Helmholtz Zentrum München, Deutsches Forschungszentrum für Gesundheit und Umwelt (GmbH), Ingolstädter Landstraße 1, 85764 Oberschleißheim/Deutschland, Germany. Tel.: +49 89 3187 4554; fax: +49 89 3187 3371.
E-mail address: thomas.groeger@helmholtz-muenchen.de (T. Gröger).
URL: http://www.helmholtz-muenchen.de/en/neu-cma/ (T. Gröger).
[1] www.hice-vi.eu.

http://dx.doi.org/10.1016/j.chroma.2014.08.091
0021-9673/© 2014 Elsevier B.V. All rights reserved.

better understanding of aerosol-related health issues and to link the biological response to specific compounds, a non-targeted, comprehensive screening of the aerosol components, especially the organic fraction, is needed.

Hyphenated techniques with great separation power, such as two-dimensional gas chromatography-time-of-flight mass spectrometry (GCxGC-TOFMS), are well suited for aerosol research. However, these techniques suffer from time-consuming analysis due to the vast amount of data obtained using non-targeted comprehensive screening. To address this issue, automated compound classifications (ACCs) were further developed and applied.

1.2. State of the art

With the advent of one-dimensional GC–MS techniques, various chemometric data-mining techniques for mass spectral information have been developed.

The commonly used compound classification in GCxGC-TOFMS is not applied directly. Usually, a NIST or in-house library search is performed, and substances are identified according to the spectral matching factors and retention times. Depending on the identification, the compound classes can be assigned. However, this indirect method is not suitable for untargeted screening in complex matrices with thousands of compounds and a large variety of compound classes.

The first of these approaches are reviewed in [7]. Chemometric classifiers are of special interest for ACC. Based on the properties of the mass spectra, it is possible to determine substructures based on fragment prediction. Attempts to use such classifiers for two-dimensional data are published in [8]. However, the development of chemometrics classifiers is rather complex. It is also mentioned in [8] that the substances classified in this way exhibit poor matching factors with the confirmation system used.

Further possibilities, such as retention time prediction according to quantitative structure-retention relationship (QSRR) in GCxGC-TOFMS, have been reported by D'Archivio et al. [9] and Kiralj et al. [10].

The approaches described above are mainly based on information that is directly accessible from the mass spectra. There are two main principles in ACC. To classify a peak, Welthagen et al. and Vogt et al. [1,8,11] used so-called prominent features, e.g., m/z 74 as the base peak and m/z 87 with an abundance of at least 40% of the base peak's abundance for the substance class of n-alkane acid methyl esters. This approach leads to easily implementable algorithms for some substance classes. For two-dimensional GCxGC-TOFMS data, the most common way to apply this approach is using the LECO ChromaTOF® scripting and classification tool, which are part of a special package not included in the main basic software distribution.

However, even in the recent literature, this approach has only yielded rough algorithms for the characterisation of, e.g., tobacco smoke and fuel products [12,13]. Despite the simplicity of this approach, the accuracy of the delivered results is insufficient. To refine these algorithms, the retention regions have to be drawn directly and manually in the chromatogram. Depending on the extent of the differentiation, this demands sophisticated, artistic drawing skills and must be performed for every matrix. Nevertheless, warping and overlapping compounds with similar prominent features are still classified in this manner. It also must be mentioned that the retention shift between chromatographic runs also influences the accuracy and usage of these scripts.

The second main approach was published in 2010 by Hilton et al. [14]. Hilton uses a complex algorithm, based on textbook criteria, to determine the molecular ion in the mass spectrum. Based on the molecular ion, they were able to identify chlorinated and brominated compounds out of 10,000 substances based on the isotopic ratios. The use of this algorithm leads to good results for compounds with strong and abundant molecular ions and distinct isotopic patterns. Nevertheless, this approach always produces "molecular ions", even for substances with weak or absent molecular ions, e.g., alcohols. Due to the complexity and the massive drawback of always producing molecular ions, this approach is limited to a small number of compounds and is not yet well known or commonly used.

This paper presents a fast automated approach to classifying substances in large two-dimensional GC-TOFMS datasets in a non-targeted, comprehensive way.

2. Material and methods

2.1. Sampling

Anthropogenic aerosols were generated during the HICE field campaign in 2012. For this study, a single-cylinder diesel research engine was provided by the Institute of Piston Machines and Internal Combustion Engines in Rostock. It was operated with diesel fuel (LFO) according to DIN EN 5900 and heavy fuel oil 188 (HFO). Publications containing further information on the HICE Project and the related field campaigns are in submission.

$PM_{2.5}$ filter samples for off-line analyses were obtained with a modified speciation sampler (Rupprecht & Patashnik 2300, Thermo Scientific, Waltham, USA), which was able to sample four filters in parallel. The samples were collected after a first dilution step via a porous tube diluter directly followed by the first ejector diluter. This process results in an average dilution ratio of 40 for LFO feedstock and 100 for HFO feedstock. PM samples were collected on quartz fibre filters (QFF, T293, Munktell, Sweden) and PTFE membrane filters (PFF, Zefluor 1 μm, Pall, USA) for each of the experiments. The QFF were conditioned by baking at 500 °C for at least 12 h and then stored in sealed glass containers until sampling. After collection of PM, the filter samples were immediately frozen at −25 °C and stored at this temperature until analysis.

2.2. Extraction

The filter samples were extracted in an ultrasonic bath using 25 mL of methanol/dichloromethane (50/50, v/v) for 15 min. The extracts from two extraction steps were combined, filtered (PTFE filter, 0.2 μm, 25 mm, VWR, Germany) and divided into three identical aliquots for targeted and non-targeted analyses.

2.3. Chromatography

For the non-targeted analysis by GCxGC-TOFMS, 10 mL of the obtained extract was reduced to 500 μL and infused with 10 μL of isotope-labelled internal standard solution (acenaphthene d10, benz(a)anthracene d12, benzo(b)fluoranthene d12, chrysene d12, fluoranthene d12, fluorene d10, perylene d12, phenanthrene d10, pyrene d10; each 0.3 mg L^{-1}; Cambridge Isotope Laboratories, Inc., USA). GCxGC measurements were carried out using an Agilent 6890 gas chromatograph equipped with a LECO Pegasus 4D detector, including a Pegasus III TOF-MS (LECO, USA) using helium as the carrier gas according to Welthagen et al. [1]. An Optic 3 inlet system (ATAS GL, Netherlands) was used. Data collection was performed using LECO ChromaTOF® software v. 2.00. Chromatographic separation was performed on a 60 m × 0.25-mm i.d. × 0.25-μm df BPX5 (SGE, Australia) column and then a 1.9 m × 0.1-mm i.d. × 0.10-μm df BPX50 (SGE, Australia) column. The modulation period was adjusted to 3 s with a modulator temperature offset of 75 °C. The oven was heated from 40 °C to 160 °C at 15 K/min followed by a temperature ramp of 2 °C min^{-1} to 320 °C. In addition, 18-μL splitless injection with temperature-controlled evaporation was used.