Environmental Science & Technology

Subscriber access provided by Helmholtz Zentrum Muenchen - Zentralbibliothek

Article

Untargeted identification of wood type-specific markers in particulate matter from wood combustion

Benedikt Alexander Weggler, Saray Ly-Verdú, Maximilian Karl Jennerwein, Olli Sippula, Ahmed Abd Alhadi Reda, Jürgen Orasche, Thomas M. Gröger, Jorma Jokiniemi, and Ralf Zimmermann

Environ. Sci. Technol., **Just Accepted Manuscript** • DOI: 10.1021/acs.est.6b01571 • Publication Date (Web): 23 Aug 2016

Downloaded from http://pubs.acs.org on September 6, 2016

Just Accepted

"Just Accepted" manuscripts have been peer-reviewed and accepted for publication. They are posted online prior to technical editing, formatting for publication and author proofing. The American Chemical Society provides "Just Accepted" as a free service to the research community to expedite the dissemination of scientific material as soon as possible after acceptance. "Just Accepted" manuscripts appear in full in PDF format accompanied by an HTML abstract. "Just Accepted" manuscripts have been fully peer reviewed, but should not be considered the official version of record. They are accessible to all readers and citable by the Digital Object Identifier (DOI®). "Just Accepted" is an optional service offered to authors. Therefore, the "Just Accepted" Web site may not include all articles that will be published in the journal. After a manuscript is technically edited and formatted, it will be removed from the "Just Accepted" Web site and published as an ASAP article. Note that technical editing may introduce minor changes to the manuscript text and/or graphics which could affect content, and all legal disclaimers and ethical guidelines that apply to the journal pertain. ACS cannot be held responsible for errors or consequences arising from the use of information contained in these "Just Accepted" manuscripts.

ACS Publications

Environmental Science & Technology is published by the American Chemical Society.
1155 Sixteenth Street N.W., Washington, DC 20036

Published by American Chemical Society. Copyright © American Chemical Society. However, no copyright claim is made to original U.S. Government works, or works produced by employees of any Commonwealth realm Crown government in the course of their duties.

Untargeted identification of wood type-specific markers in particulate matter from wood combustion

Authors:

Benedikt A. Weggler[1,2] Saray Ly-Verdu[1,2], Maximilian Jennerwein[1,2,3], Olli Sippula[4], Ahmed A. Reda[1,2], Jürgen Orasche[1,2], Thomas Gröger[1,2,], Jorma Jokiniemi[4], Ralf Zimmermann[1,2]*

[1] Cooperation Group Comprehensive Molecular Analytics, Helmholtz Zentrum München, German Research Center for Environmental Health, 85764 Oberschleißheim, Germany

[2] Analytical Chemistry, Institute of Chemistry, University of Rostock, Rostock 12, Germany

[3] ASG Analytik Service Gesellschaft mbH, Neusäß, Germany

[4] University of Eastern Finland, Department of Environmental and Biological Sciences, Fine Particle and Aerosol Technology Laboratory, P.O. Box 1627, FIN-70211 Kuopio, Finland

* Corresponding author, Helmholtz Zentrum München, Deutsches Forschungszentrum für Gesundheit und Umwelt (GmbH), Cooperation Group „Comprehensive Molecular Analytics (CMA)“, Gmunder Strasse 37,81379 Munich, Germany
Mail: thomas.groeger@helmholtz-muenchen.de
Phone: +49 (0)89 3187 – 4554

KEYWORDS

Comprehensive two-dimensional gas chromatography, wood combustion, aerosol data mining, direct thermal desorption, in situ derivatization

ABSTRACT

Residential wood combustion emissions are one of the major global sources of particulate and gaseous organic pollutants. However, the detailed chemical compositions of these emissions are poorly characterized due to their highly complex molecular compositions, non-ideal combustion conditions and sample preparation steps. In this study, the particulate organic emissions from a masonry heater using three types of wood logs, namely, beech, birch and spruce, were chemically characterized using thermal desorption in situ derivatization coupled to a GCxGC-ToF/MS system. Untargeted data analyses were performed using the comprehensive measurements. Univariate and multivariate chemometric tools, such as analysis of variance (ANOVA), principal component analysis (PCA) and ANOVA simultaneous component analysis (ASCA), were used to reduce the data to highly significant and wood type-specific features. This study reveals substances not previously considered in the literature as meaningful markers for differentiation among wood types.

INTRODUCTION

According to the German Federal Environment Agency, 80-90% of the total particulate emissions of small combustion plants originate from wood combustion [1, 2]. The total emission totals approximately 20,000 tons of particulate matter (PM) per year and thus represents an important factor of the total aerosol burden. From a global perspective, biomass burning contributes the highest anthropogenic PM burden worldwide[3].

Small-scale wood combustions are generally, highly fluctuating processes. The resulting emission aerosols are affected by variables such as wood type, combustion technology (wood stove type) and stove operating practices. The batch-wise combustion means repeated, complex processes of drying, gasification, ignition and burning, accompanied by pyrolysis[4-6]. All these processes contribute to the formation of gaseous compounds and fine PM, containing soot, salts and a complex mixture of, more or less, hazardous organic compounds.

Several studies suggest a direct link between the presence of fine PM in aerosol and severe health effects, such as allergies, the prevalence of cardiovascular and respiratory diseases and increased mortality. These studies also support the hypothesis that the toxic and genotoxic carcinogenic mechanisms of these particulates are affected by both the physical properties of the particle (size, shape and surface) and their chemical loading (dissolved and adsorbed chemicals, surface or catalytic reactions)[7-10].

The Helmholtz Virtual Institute for Complex Molecular Systems in Environmental Health (HICE) has taken up the task of linking biological responses to physical and chemical properties[11]. The general procedure of HICE is exposing lung cell tissue to aerosol. In order to disclose the source of the biological response, comprehensive chemical insight into the aerosol composition is needed. A broad field of analytical techniques has to be considered for the highly diverse mixture of molecular components.

Comprehensive techniques, such as GC×GC-ToF/MS, allow for a more holistic approach with detailed, untargeted insight into the particulate phase of anthropogenic organic aerosols. By coupling to an in situ derivatization, direct thermal desorption system (IDTD), the sample preparation steps are shortened and a decrease in measurability constraints can be achieved. IDTD uses only small amounts of derivatization agent and allows for fast sample treatment under relatively mild conditions. In addition, IDTD provides transfer of a broad range of polar and non-polar organic compounds, which are not accessible with commonly used sample preparation techniques, to the gas chromatographic separation[12, 13]. Because of this complex molecular mixture GCxGC-ToF/MS is needed to obtain a better separation, especially for untargeted assignment of significant marker substances.

Several attempts to utilize GC×GC-ToF/MS techniques for PM analysis have been published in recent years[14-16], but these methods are not widely applied to aerosol science due to the vast amount of data produced by GC×GC-ToF/MS, although various chemometric data mining and reduction techniques have been reported[17-19] Special attention should be given to methods that reduce data, based on statistic values, regarding anomalies or significant differences between each measurement. Such methods are commonly used for the evaluation of metabolomic datasets[20, 21].

The scope of this study is the untargeted assignment of significant marker substances to individual woodtypes. Therefore, variant and unsteady PM samples from wood combustion were investigated by GC×GC-ToF/MS. The measurements were enhanced by sophisticated sample treatment techniques. Univariate and multivariate chemometric techniques were applied to identify significant features of the semi-volatile organic fraction of the PM from beech, birch and spruce wood combustion in a modern soapstone masonry heater. These features are not considered in literature so far and provide valuable information when tracing back potential biological risks to individual substances.

MATERIAL & METHODS

AEROSOL SAMPLING

The combustion experiments and sampling were performed during an HICE-field campaign in the winter of 2013 at the University of Eastern Finland. A modern soapstone masonry heater (Hiisi 4, Tulikivi Ltd., Finland), which was described in detail by Reda et al., [22] was used for aerosol generation. Beech, birch and spruce log wood were used as feedstock.

A total of 6 batches of log wood were burnt over 4 h. One batch weighed approximately 2.5 kg, with a maximum log size of 30 cm. After the sixth batch, stoking was performed without adding more fuel. The total weight of the consumed wood was approximately 15.0 kg per experiment. Properties, e.g. water content, of the used wood can be found in Table S1 in the supporting information. A graphical overview of the combustion procedure and the wood stacking can be found in supporting material, figure S1.

Prior to the filter sampling a PM2.5 impactor was placed on the sampling system. Filter samples were obtained for off-line analyses with a modified speciation sampler (Rupprecht & Patashnik 2300, Thermo Scientific, Waltham, USA), which was able to sample four filters in parallel. Samples were collected after an initial dilution step via a porous tube diluter directly followed by the first ejector diluter. This process results in an average dilution ratio of ~40. PM samples were collected on quartz fiber filters (QFF, T293, Munktell, Sweden). A more detailed overview of the experimental setup can also be found in the supporting information, Figure S2. The QFF were preconditioned by baking at 500°C for at least 12 h and were then stored in sealed glass containers prior to sampling. After collection of the PM, the filter samples were immediately frozen at -25°C and stored at this temperature until analysis. Per filter, an average total volume of 1.8 m^3/mm^2 diluted combustion aerosol was sampled. Four repetitions per combustion were performed.

SAMPLE PREPARATION

The IDTD described by Orasche et al.[13] and Schnelle-Kreis et al. [12, 23] was applied prior to the GC×GC-ToF/MS measurements. The following describes the enhancements made to this method. A graphical representation of the general automated procedure is shown in Figure S3. Custom-made goose-neck glass liners were used as inserts for the GC injector system. To eliminate impurities, an extensive purification process was applied. The liners were placed in a saturated KOH:isopropanol solution followed by cleaning in aqua regia for five hours for each cleaning step. After repeated rinsing with water, the liners were deactivated overnight using trimethylchlorosilane (Supelco). A conditioning step under a protective gas atmosphere at 360°C was performed before use.

Round filter punches of the collected PM samples with a diameter of 5 mm were prepared for an average volume of sampled combustion aerosol of 0.035 m^3. The filter aliquots were placed in the liner. Each filter was infused with isotopically labeled standards. The composition and of the internal standard is described by Orasche et. al[13]and added to the supporting information Table S2.

Samples were set aside for 30 min at room temperature to evaporate the solvent of the internal standard mixture. After the evaporation step, the pieces of filter inside the liner were soaked in 9 µL N-methyl-N-(trimethylsilyl)-trifluoracetamide (MSTFA, Macherey Nagel Düren, Germany) and placed in the GC-injection system.

The IDTD method used the following parameters. The GC injector was kept at 70°C during the liner insertion step. After insertion and closing of the liner head, the injector was flushed with Helium, the carrier gas of the GC, with a flow of 1 mL/min and a split flow of 100 mL/min for 60 sec. The gas flow was increased to 2.8 mL/min after flushing the injector. A bypass system allowed for the saturation of the carrier gas with MSTFA for a total reaction time of 15 minutes while the injector was heated to 330°C at a heating rate of 16°C/sec. The desorbed components were then collected in the GC oven which was held at 70°C during this desorption step. After this reaction time, the bypass was closed and the flow was set to 1 mL/min with a split flow of 50 mL/min.

GC×GC-ToF-MEASUREMENTS

GC×GC measurements were conducted using an Agilent 6890 gas chromatograph equipped with a LECO Pegasus 4D ToF MS (LECO, USA) using helium as the carrier gas. The ToF/MS was operated at 200 Hz with an acquisition range from *m/z* 35 to 600. Electron ionization occurred at 70 eV at a source temperature of 250°C.

An Optic 3 injection system and flow control (ATAS GL, Netherlands) was also used. Data collection was performed using LECO ChromaTOF® software v.4.50.8.0. The method is described elsewhere[17].Chromatographic separation was performed on a 58 m × 0.25 mm i.d. × 0.25 µm df BPX5 (SGE, Australia) column coupled to a 2.75 m × 0.1 mm i.d. × 0.10 µm df BPX50 (SGE, Australia) column. The modulation period was adjusted to 5 s with a modulator temperature offset of 30°C. During the thermal desorption step, the oven was kept at 70°C for 15 minutes before increasing to 140°C at 80 K/min followed by a slow temperature ramp of 3 K/min to 360°C, which was held for 20 min. This oven configuration, combined with IDTD allows a reliable chromatographic separation of compounds ranging from C14 to C40.

The measurement of a sample series was a reference sample followed by five combustion aerosol samples. Every six samples the reference measurement was repeated. After a total of 14 samples an empty liner was measured as blank. No considerable memory effects were observed in the reference samples or the blank runs.

DATA PROCESSING

The data processing consisted of three; namely, basic processing (peak detection and peak alignment), statistical and multivariate data analysis, e.g., analysis of variance (ANOVA), principle component analysis (PCA), ANOVA simultaneous component analysis (ASCA) and mass spectral interpretation of the spectra. Basic processing and library searches of the MS spectra were performed using the LECO ChromaToF software ver. 4.50.8.0 and NIST v.2.0, and the data analysis was performed with MATLAB R2011a, including the PLS toolbox v8.1 (Eigenvector Research, Manson, WA). Figure 1 shows a graphical representation of the data workflow. The baseline was adjusted directly above the noise as an initial processing step. For peak assignment, peak broadening was considered with 25 seconds in the first dimension and 1 second in the second dimension. Only signals exceeding a signal-to-noise (S/N) ratio of 100 were considered. Subpeaks for the second dimension required an S/N greater than 50 and were combined when the spectral matching factor was higher than 75%.

ALIGNMENT

For a direct statistical comparison and further multivariate data analysis, the peak lists of individual GC×GC analyses were aligned using the peak alignment function of ChromaToF’s statistical compare (SC). Whether a peak was retained after the pairwise sample comparison depended on how often it appeared and in which samples it was detected. A peak had to be present in at least 12 (out of 36) samples (regardless of the wood type) or in more than 50% (6 samples) of the wood type-specific samples. The mass spectra of a peak in question had to have a spectral similarity of 75%. The limiting retention windows were set to a deviation of one modulation period in the first dimension and to 1 sec in the second dimension. With these settings, 2779 peaks were found and used as further processing variables.

DATA SCREENING

Spectra similarity searches were performed based on NIST v.2.0., GOLM and different in-house libraries. Names were assigned to substances when the spectral similarity match exceeded 70%. To obtain information about structure-related chromatographic allocation areas, the peak list from SC was investigated using an in-house searching method. This initial screening was based on the systematic naming of organic substances, and it screened the peak list for name fragments, for instance, “furanoses”. The presence of compounds known from the literature[4] was examined. Typical wood combustion markers for broad leaf wood or coniferous wood combustion, e.g., abietic acid, β-sitosterol or α-amyrin, were present, as well as decomposition products from celluloses and hemicelluloses, e.g., levoglucosan, mannosan and galactosan.

OUTLIER CORRECTION

The values of the peak areas obtained from the SC were normalized to an internal standard that was spiked on the filter during the sample preparation step. For each wood type, all peaks detected by the SC procedure were used in the principal component analysis (PCA) based on the peak areas. Experiments falling outside of the 95% confidence interval were investigated more closely using the Nalimow and Shapiro-Wilk outlier testing on each individual peak. When at least two-thirds of the peaks showed outlier indications, the corresponding measurement was considered an outlier and was excluded from consideration.

ANOVA

To elucidate the molecular differences among the various types of wood, the SC feature list was subjected to analysis of variance (ANOVA) with the peak areas normalized to the internal standard.

Several significance thresholds (p-values) were set to obtain only the highly significant features among the different wood types. The smaller the p-value, the higher the assurance that the selected feature was significantly different.

PCA

The features provided by the ANOVA were log-transformed and mean-centered before they were analyzed by PCA. The obtained features were assigned to each individual wood type based on their loadings for the different principal components (PC).

ANOVA-SIMULTANEOUS COMPONENT ANALYSIS (ASCA)

The number of simultaneously measured inputs in untargeted studies is of the order of thousands[24] because each variable corresponds to a specific substance. ANOVA aims to identify whether a significant difference exists among groups, making it possible to estimate the effect of the experimental factors on the data. However, for multivariate data, ANOVA does not explain the inter-relation between variables, and it is inadequate when the number of variables exceeds the number of measurements[25].

ANOVA-simultaneous component analysis (ASCA) overcomes these handicaps by combining the best parts of ANOVA and PCA[26]. The goal of ASCA is to combine the capacity of ANOVA, to separately analyze different sources of experimental variation, and PCA, to explain the correlation amongst variables. An advantage of ASCA over PCA is that the model is less affected by outliers because the matrices are divided according to the different contributions to the variation in the data.

In the linear model used in ANOVA that serves as the basis of ASCA, each measurement is decomposed into contributions to the variation caused by one treatment and variation caused by other sources, as described by Smilde et al.[25]. Once the different factors have been isolated, the behavior of the dependent variables under the different treatment levels is examined by PCA, from which the scores and loadings can be used to explain the decomposed data matrix, allowing the factors to be estimated separately.

To enable ASCA modeling, mean-centering was applied to the data[27], fulfilling the necessary and sufficient conditions for subsequent ASCA decomposition of the data matrix, which requires normalized but not log-transformed data[28].

FEATURE INTERPRETATION

Following the assignment of the individual features to a specific wood type, a library search was performed. The names were assigned only to those spectra that exceeded a similarity match of 70% compared to the reference spectra in NIST v.2.0, the Golm Database (GMD) or an in-house developed database.

RESULTS AND DISCUSSION

INITIAL DATA SCREENING

Compound classes typically appear as curved bands in the GCxGC-separation space[29]. Consequently, pattern recognition approaches[30] combined with automatic compound classifications[18, 19, 31]can be applied for rapid data screening. However, in IDTD, the elution pattern and mass spectra are mainly determined by derivatization. In brief, the mentioned approaches are complicated. In addition, distinguishable retention patterns were revealed by the initial screening approach.

It was possible to determine the chromatographic regions for pyranoses, furanoses, tetracyclic and pentacyclic triterpenoids, and PAH. In addition, with a total runtime of less than one minute, the initial screening is efficient. This procedure provides quantitative and untargeted insight into the molecular composition of the combustion aerosols in less than one minute. In Figure S4, a GC×GC chromatogram of birch combustion aerosol is overlaid with some of the screened results.

IDENTIFICATION OF WOOD TYPE-SPECIFIC FEATURES

The combustion markers discussed in the literature are generally wood type-independent substances. Sometimes, markers are assigned to wood families (coniferous wood and broad leaf wood). However, PM samples differ in molecular composition due to the presence of wood type-specific substances. For the identification of these molecular differences, ANOVA was applied to the peak list obtained from SC.

ANOVA provides a list of features that exhibit significant differences among the samples. The number of features decreases with a decreasing significance threshold. Table S2 shows the number of peaks at the corresponding p-values. A threshold of p = 5% is commonly applied. In this study, a 5% threshold yields 817 features. For this significance threshold, substances such as levoglucosan or 1,4,3.6 dianhydro-beta-D-glucopyranose are not present in the feature list. The cellulose contents for broad leaf wood and coniferous wood range from 42-51% and 42-49%[32], respectively. Therefore, significant differences were not expected.

Mannosan and galactosan, compounds that are related to the combustion of hemicellulose, disappear from the feature list at a significance threshold of 0.1%. Because the hemicellulose content ranges between 24 and 30% in broad leaf wood and between 27 and 40% in coniferous wood,[32] it is reasonable to expect these substances to show higher significance.

Following this approach, significant differences are expected between broad leaf and coniferous wood for substances derived from lignin. The lignin content ranges from 25-30% in coniferous wood and from 18-24% in broad leaf wood[32]. Nevertheless, the lignin components e.g., coniferyl and sinapyl alcohol, are present in only the feature list up to a threshold of p = 1%.

Consequently, discrimination by wood type is possible only to a certain degree when using the substances listed in literature. Strictly speaking, discrimination between only broad leaf wood and coniferous wood can be achieved. Table S3 shows the occurrence of typical marker substances at corresponding p-values. In the following section, features with values of p <= 0.001% are examined. This significance threshold yields 84 features. Additionally, the feature list contains marker substances or corresponding homologues that differentiate broad leaf and coniferous wood[4, 6].

However, ANOVA does not provide an assignment of the significant features to different wood types. When using the provided features as the input for PCA, the first and secondPC capture the greatest variance between samples, accounting for 68.64% of the variance.

The results of the PCA can be displayed using a biplot of the scores and loadings, as shown in Figure 2. Clear clustering according to the measured wood types and separation between softwood and hardwood is observed on the first PC. Further discrimination of the samples from softwood combustion occurs on the second PC.

Features were assigned to the wood type according to their location in the biplot. This assignment is shown in Figure 3. The size of the features represents the corresponding loading. Therefore, the loading is a measurement of how the PCA is influenced by the corresponding feature.

Features assigned to particulate matter from spruce or birch combustion show higher loadings than those of beech combustion, which indicates that the features produced during beech combustion are less specific than those produced during birch and spruce combustion. Additionally, the features from beech PM are present in the birch and spruce PM.

The library matching and manual inspection of the mass spectra of these features are shown in Table 1. Name suggestions and retention times corresponding to the library and matching factors are shown.

When comparing the features for the individual wood types, two substance classes stand out, namely, tetra/pentacyclic triterpenoids and diterpenoid carboxylic acids. The former are exhibited mainly in the PM from birch combustion, and the latter are present in the PM from spruce combustion. These substances have a strong resemblance to phytosteroids in the case of birch combustion, which exhibit high biological activity, such as alteration in cell membranes[33], induction of apoptosis in cancer cells[34, 35]and anti-inflammatory effects[36, 37].

The substances found in the PM in spruce combustion, resinoic acids retene etc., are known to be part of colophony, which, when burned, increases asthma risk, chronic bronchitis, and chemical hypersensitivity [38].

By gradually decreasing the p-value in the ANOVA from 5% to 0.001%, the PCA clustering becomes influenced by less significant features. An increase in distinct chromatographic regions is observed by plotting these features as a pseudo chromatogram, as shown in Figure 4. In the supporting information the density maps for all p-values are shown, Figures S5-7.

A more thorough investigation of these regions supports the above findings. PM generated from beech wood exhibits significant differences, with high loadings in the chromatographic regions, which contain small polar components. Features assigned to the PM from birch combustion are located in two regions, which contain mainly the previously mentioned methylated naphthalenes and tetra/pentacyclic triterpenoids. For spruce PM, the region containing retene and resinoic acid derivatives, such as abietic acid, is clearly enhanced. The substances responsible for the differences in spruce wood are typical products from the combustion process of coniferous wood[10].

Class	Name	RT1	RT2	Similarity	Reverse	Library	Class
Beech	6H-Dibenzo[b,d]-pyran	2747.9	1.9	811	841	NIST mainlib	dibenzopyran
	#9H-Xanthene	2714.8	1.9	794	886	NIST replib	dibenzopyran
	#Benzene, 1,2,3-trimethoxy-5-methyl-	2219.7	1.4	895	917	NIST replib	methoxy benzene
	Phenol, 2,6-dimethoxy-, acetate	2346.2	1.9	821	865	NIST replib	methoxybenzene
	#Hydrocaffeic acid (3TMS)	3144.7	1.1	762	776	GOLM DB	Cinnamic acid derivatives
	D-Erythro-pentofuranose (TMS)	2336.9	0.7	801	823	NIST mainlib	furanoses
	Lupa-2,22(29)-dien-28-ol	5513.6	3.3	936	940	IH Reference	tetracyclic triterpene
	#Octadecen-1-ol, 9-(Z)- (1TMS)	3517.2	1.1	793	848	GOLM DB	unsaturated alcohol
	Benzene, (1-methylundecyl)-	2979.5	1.2	854	883	NIST mainlib	benzene derivatives
Birch	Cycloartenol (1TMS)	4650	1.8	753	746	GOLM DB	tetracyclic triterpene
	Cycloeucalenol (1TMS)	5133.8	2.5	808	829	GOLM DB	tetracyclic triterpene
	Stigmasta-3,5-dien-7-one	5067.2	2.7	928	929	IH Reference	tetracyclic triterpene
	Oleanolic acid (2TMS)	5538.1	3.3	918	943	GOLM DB	pentacyclic triterpene
	Oleanolic acid, 3-acetoxy- (3-beta-)- (1TMS)	5504.7	3.8	761	776	GOLM DB	pentacyclic triterpene
	Dibenzofuran	2523.8	1.8	922	926	NIST mainlib	hetero PAH
	Cembrene	4520.3	2	763	812	GOLM DB	diterpene
	#1-Undecene, 4-methyl-	2630	0.8	759	789	NIST mainlib	unsaturated, branched
	#9,12-Octadecadiynoic acid, trimethylsilyl ester	2765.7	1	718	738	NIST mainlib	alkyne
	2,6-Di-tert-butylquinone	2349.4	1.1	885	895	NIST replib	quinone

	Benzene, [(tetramethylcyclopropylidene)methyl]	2984.3	1.4	713	773	NIST mainlib	benzene
	(-)-Isolongifolol, trimethylsilyl ether	2981.7	1.1	704	714	NIST mainlib	tricyclic sesquiterpene
Spruce	Abietic acid (1TMS)	4051.2	1.8	889	900	GOLM DB	diterpene carboxylic acid
	6-Dehydrodehydroabietic acid(TMS)	4337.1	2.4	760	829	IH Reference	diterpene carboxylic acid
	7-Oxodehydroabietic acid, methyl ester	4083.3	2.3	751	767	IH Reference	diterpene carboxylic acid
	Dehydroabietic acid(TMS)	3939.5	1.9	722	743	IH Reference	diterpene carboxylic acid
	Dehydroabietylamine	3862.9	1.7	700	710	NIST mainlib	diterpene carboxylic acid
	Methyl dehydroabietate	3959.4	2.3	807	908	NIST replib	diterpene carboxylic acid
	Dehydroabietal	3874	2.3	779	797	NIST mainlib	diterpene carboxylic acid
	Estradiol, 17alpha- (2TMS)	4106.2	1.7	702	708	GOLM DB	steroid
	#Benzo[k]fluoranthene	4690.7	4	913	930	IH Reference	PAH
	Perylene	4814.3	4.3	933	940	IH Reference	PAH
	8-Isopropyl-1,3-dimethylphenanthrene	3949.1	2.4	863	880	NIST mainlib	retenoid
	Retene	3803.3	2.4	927	937	NIST replib	retenoid
	Galactonic acid-1,4-lactone (4TMS)	2598	0.7	707	722	GOLM DB	lactone
	Pentadecanoic acid (1TMS)	3085	1	715	748	GOLM DB	carboxylic acid
	#Heptadecanoic acid (1TMS)	3436	1.1	846	879	GOLM DB	carboxylic acid
	Octadecenoic acid, 9-(E)- (1TMS)	3665.9	1.2	906	930	GOLM DB	unsaturated carboxylic acid
	#Tetracosan-1-ol (1TMS)	4434.3	1.2	758	728	NIST mainlib	alcohol

Table 1: The matched peaks for a p-value of 0.001%, retention times, matching factors and corresponding library. Gray highlighted features are related to the isolated factor “wood type” using the ASCA model. Substances marked with # are related to combustion processes where at least two burning events were observed, refer to the ASCA section for further information.

ISOLATING WOOD TYPE-SPECIFIC MARKERS

The previously described ANOVA and PCA consider all factors that influence the genesis of the PM. Significant differences are therefore not attributed solely to the wood types. In contrast, despite the PCA patterns highlighting their similarities, PCA does not distinguish between the factors that induce differences among samples. Therefore, ASCA is applied to the data set. An improvement of ASCA over the PCA for the entire data set is the certainty that the variation captured by different submodels does not originate from any other source than the studied factor.

To determine these factors, in addition to the easily identified factor “wood type”, a criterion for the quality of the combustion was introduced. For this criterion, carbon monoxide (CO) and the organic gaseous carbon (OGC) values were plotted against each other. These values were logged in a time-resolved fashion during each combustion experiment. Therefore, CO vs OGC plots were produced for each batch introduced into the oven. These plots were then compared to the corresponding plot of the other experiments within a wood type. Assuming ideal combustion conditions, small OGC vs CO values are expected. If the values were high, the corresponding batch was marked as conspicuous. The monitored events were considered in the ASCA model for batches 1-3 of each wood type. In the supporting information, the CO vs OGC plots are shown with additional information, Figure S8.

In Figure 5, the ASCA score plot for the factor wood type is shown. Instead of clustering similar to the ANOVA procedure, all data from one individual wood type are coincident, which is expected because only the variance caused by the isolated factor “wood type” is considered. In addition, highly correlated features can be observed. By minimizing the distance to an original straight line that passes through the wood score point, the features show high mass spectrometric response. In addition, these features were detected almost exclusively in the measurements of the respective wood type. Similar to the previously described PCA, the features were assigned to the individual wood types. In Table 1, the features that were influenced only by the wood type are shaded in light gray.

Not all the previously identified features are wood type-related. Therefore, the factor of the CO and OGC events was investigated. Figure S9 show the score plot for the observed CO and OGC events during the first batch is shown. Clear separation between samples is observed between samples with none, one or two CO/OGC events for the first batch of feedstock. A correlation of features similar to the isolated factor wood type is not observed. Features were assigned to the events similarly to the PCA. Features were classified by the occurrence of no event, one event or both events. This assignment is indicated in Table 1 by superscript indices.

Of the 38 features found by the ANOVA/PCA approach, only 5 were not assigned to a distinct wood type or to a combustion event. Although the aerosols generated from combustion present a challenge for metrology and data evaluation, the combination of highly sophisticated sample preparation, comprehensive separation power and chemometric analysis is versatile in elucidating the molecular composition of aerosols.

In summary, sequential analyses of data by data screening, ANOVA, PCA and ASCA allows for statistical discrimination among PM samples from beech, birch and spruce wood combustion. In brief, small polar compounds were enhanced in the PM from beech combustion. Large differences in methylated naphthalenes and tetra/pentacyclic triterpenoids were found in the PM from birch combustion, whereas the differences for PM from spruce combustion were attributed to derivatives from retene and resinoic acid.

ACKNOWLEDGEMENTS

This work was supported by the Virtual Institute of Complex Molecular Systems in Environmental Health (HICE), the Helmholtz Association and the German Federal Ministry of Education and Research (BMBF). This support is gratefully acknowledged.

SUPPORTING INFORMATION AVAILABLE

Supporting Information.

- Parameter of the used wood feedstock // Table S1
- Wood stacking and stove operating procedure // Figure S1
- Sampling setup // Figure S2
- IDTD Schematic // Figure S3
- Example for Peak List Scrubbing // Figure S4
- Peak Number vs. p Value // Table S2
- Marker Occurrence vs p-value // Table S3
- Density Map different p-values beech/birch/spruce // Figure S5-7
- CO/OGC Event // Figure S8
- ASCA Plot // Figure S9
- Internal Standards // Table S 4

This information is available free of charge via the Internet at http://pubs.acs.org.

1. Behnke, A., Novellierung der Verordnung über kleine und mittlere Feuerungsanlagen. In 03/2010 ed.; Agency, G. F. E., Ed. German Federal Environmental Agency, Bressestelle, Wörlitzer Platz 1 06844 Dessau-Roßlau: Germany, 2010; p 9.

2. Agency, G. F. E., Emissionsminderung bei Kleinfeuerungsanlagen. In German Federal Environmental Agency: http://www.umweltbundesamt.de/, 2014.

3. Meyer, N. K., Particulate, black carbon and organic emissions from small-scale residential wood combustion appliances in Switzerland. *Biomass and Bioenergy* **2012,** *36*, 31-42.

4. Simoneit, B. R. T., Biomass burning - A review of organic tracers for smoke from incomplete combustion. *Applied Geochemistry* **2002,** *17*, (3), 129-162.

5. Klass, D., Biomass for Renewable Energy, Fuels, and Chemicals. In 1st Edition ed.; ACADEMIC PRESS, 1998; p 651.

6. Medeiros, P. M.; Simoneit, B. R. T., Analysis of sugars in environmental samples by gas chromatography-mass spectrometry. *Journal of Chromatography A* **2007,** *1141*, (2), 271-278.

7. Heal, M. R.; Kumar, P.; Harrison, R. M., Particles, air quality, policy and health. *Chemical Society Reviews* **2012,** *41*, (19), 6606-6630.

8. Mauderly, J. L.; Chow, J. C., Health effects of organic aerosols. *Inhalation Toxicology* **2008,** *20*, (3), 257-288.

9. Poschl, U., Atmospheric aerosols: Composition, transformation, climate and health effects. *Angewandte Chemie-International Edition* **2005,** *44*, (46), 7520-7540.

10. Cassee, F. R.; Heroux, M. E.; Gerlofs-Nijland, M. E.; Kelly, F. J., Particulate matter beyond mass: recent health evidence on the role of fractions, chemical constituents and sources of emission. *Inhalation Toxicology* **2013,** *25*, (14), 802-812.

11. Oeder, S.; Kanashova, T.; Sippula, O.; Sapcariu, S. C.; Streibel, T.; Arteaga-Salas, J. M.; Passig, J.; Dilger, M.; Paur, H. R.; Schlager, C.; Mulhopt, S.; Diabate, S.; Weiss, C.; Stengel, B.; Rabe, R.; Harndorf, H.; Torvela, T.; Jokiniemi, J. K.; Hirvonen, M. R.; Schmidt-Weber, C.; Traidl-Hoffmann, C.; Berube, K. A.; Wlodarczyk, A. J.; Prytherch, Z.; Michalke, B.; Krebs, T.; Prevot, A. S. H.; Kelbg, M.; Tiggesbaumker, J.; Karg, E.; Jakobi, G.; Scholtes, S.; Schnelle-Kreis, J.; Lintelmann, J.; Matuschek, G.; Sklorz, M.; Klingbeil, S.; Orasche, J.; Richthammer, P.; Muller, L.; Elsasser, M.; Reda, A.; Groger, T.; Weggler, B.; Schwemer, T.; Czech, H.; Ruger, C. P.; Abbaszade, G.; Radischat, C.; Hiller, K.; Buters, J. T. M.; Dittmar, G.; Zimmermann, R., Particulate Matter from Both Heavy Fuel Oil and Diesel Fuel Shipping Emissions Show Strong Biological Effects on Human Lung Cells at Realistic and Comparable In Vitro Exposure Conditions. *Plos One* **2015,** *10*, (6).

12. Schnelle-Kreis, J.; Welthagen, W.; Sklorz, M.; Zimmermann, R., Application of direct thermal desorption gas chromatography and comprehensive two-dimensional gas chromatography coupled to time of flight mass spectrometry for analysis of organic compounds in ambient aerosol particles. *Journal of Separation Science* **2005,** *28*, (14), 1648-1657.

13. Orasche, J.; Schnelle-Kreis, J.; Abbaszade, G.; Zimmermann, R., Technical Note: In-situ derivatization thermal desorption GC-TOFMS for direct analysis of particle-bound non-polar and polar organic species. *Atmospheric Chemistry and Physics* **2011,** *11*, (17), 8977-8993.

14. Hamilton, J. F., Using Comprehensive Two-Dimensional Gas Chromatography to Study the Atmosphere. *Journal of Chromatographic Science* **2010,** *48*, (4), 274-+.

15. Kallio, M.; Jussila, M.; Rissanen, T.; Anttila, P.; Hartonen, K.; Reissell, A.; Vreuls, R.; Adahchour, M.; Hyotylainen, T., Comprehensive two-dimensional gas chromatography coupled to time-of-flight mass spectrometry in the identification of organic compounds in atmospheric aerosols from coniferous forest. *Journal of Chromatography A* **2006,** *1125*, (2), 234-243.

16. Ozel, M. Z.; Hamilton, J. F.; Lewis, A. C., New sensitive and quantitative analysis method for organic nitrogen compounds in urban aerosol samples. *Environmental science & technology* **2011,** *45*, (4), 1497-505.

17. Zeng, Z. D.; Hugel, H. M.; Marriott, P. J., Chemometrics in comprehensive multidimensional separations. *Analytical and Bioanalytical Chemistry* **2011,** *401*, (8), 2373-2386.

18. Weggler, B. A.; Groger, T.; Zimmermann, R., Advanced scripting for the automated profiling of two-dimensional gas chromatography-time-of-flight mass spectrometry data from combustion aerosol. *Journal of Chromatography A* **2014,** *1364*, 241-248.

19. Welthagen, W.; Schnelle-Kreis, J.; Zimmermann, R., Search criteria and rules for comprehensive two-dimensional gas chromatography–time-of-flight mass spectrometry analysis of airborne particulate matter. *Journal of Chromatography A* **2003,** *1019*, (1-2), 233-249.

20. Ly-Verdu, S.; Groger, T. M.; Arteaga-Salas, J. M.; Brandmaier, S.; Kahle, M.; Neschen, S.; de Angelis, M. H.; Zimmermann, R., Combining metabolomic non-targeted GCxGC-ToF-MS analysis and chemometric ASCA-based study of variances to assess dietary influence on type 2 diabetes development in a mouse model. *Analytical and Bioanalytical Chemistry* **2015,** *407*, (1), 343-354.

21. Almstetter, M. F.; Appel, I. J.; Dettmer, K.; Gruber, M. A.; Oefner, P. J., Comparison of two algorithmic data processing strategies for metabolic fingerprinting by comprehensive two-dimensional gas chromatography-time-of-flight mass spectrometry. *Journal of Chromatography A* **2011,** *1218*, (39), 7031-7038.

22. Reda, A. A.; Czech, H.; Schnelle-Kreis, J.; Sippula, O.; Orasche, J.; Weggler, B.; Abbaszade, G.; Arteaga-Salas, J. M.; Kortelainen, M.; Tissari, J.; Jokiniemi, J.; Streibel, T.; Zimmermann, R., Analysis of Gas-Phase Carbonyl Compounds in Emissions from Modern Wood Combustion Appliances: Influence of Wood Type and Combustion Appliance. *Energy & Fuels* **2015,** *29*, (6), 3897-3907.

23. Schnelle-Kreis, J.; Orasche, J.; Abbaszade, G.; Schafer, K.; Harlos, D. P.; Hansen, A. D. A.; Zimmermann, R., Application of direct thermal desorption gas chromatography time-of-flight mass spectrometry for determination of nonpolar organics in low-volume samples from ambient particulate matter and personal samplers. *Analytical and Bioanalytical Chemistry* **2011,** *401*, (10), 3083-3094.

24. Boccard, J.; Rudaz, S., Harnessing the complexity of metabolomic data with chemometrics. *Journal of Chemometrics* **2014,** *28*, (1), 1-9.

25. Smilde, A. K.; Jansen, J. J.; Hoefsloot, H. C. J.; Lamers, R. J. A. N.; van der Greef, J.; Timmerman, M. E., ANOVA-simultaneous component analysis (ASCA): a new tool for analyzing designed metabolomics data. *Bioinformatics* **2005,** *21*, (13), 3043-3048.

26. Zwanenburg, G.; Hoefsloot, H. C. J.; Westerhuis, J. A.; Jansen, J. J.; Smilde, A. K., ANOVA-principal component analysis and ANOVA-simultaneous component analysis: a comparison. *Journal of Chemometrics* **2011,** *25*, (10), 561-567.

27. Bro, R., Multivariate calibration - What is in chemometrics for the analytical chemist? *Analytica Chimica Acta* **2003,** *500*, (1-2), 185-194.

28. Jansen, J. J.; Hoefsloot, H. C. J.; van der Greef, J.; Timmerman, M. E.; Westerhuis, J. A.; Smilde, A. K., ASCA: analysis of multivariate data obtained from an experimental design. *Journal of Chemometrics* **2005,** *19*, (9), 469-481.

29. Harvey, P. M.; Shellie, R. A., Data Reduction in Comprehensive Two-Dimensional Gas Chromatography for Rapid and Repeatable Automated Data Analysis. *Analytical Chemistry* **2012,** *84*, (15), 6501-6507.

30. Johnson, K. J.; Synovec, R. E., Pattern recognition of jet fuels: comprehensive GC x GC with ANOVA-based feature selection and principal component analysis. *Chemometr Intell Lab* **2002,** *60*, (1-2), 225-237.

31. Jennerwein, M. K.; Eschner, M.; Groger, T.; Wilharm, T.; Zimmermann, R., Complete Group-Type Quantification of Petroleum Middle Distillates Based on Comprehensive Two-Dimensional Gas Chromatography Time-of-Flight Mass Spectrometry (GCxGC-TOFMS) and Visual Basic Scripting. *Energy & Fuels* **2014,** *28*, (9), 5670-5681.

32. Lohmann, U.; Blosen, M., *Holzlexikon das Standardwerk für die Holz- und Forstwirtschaft*. Genehmigte Lizenzausg. der 4. Aufl. ed.; Nikol: Hamburg, 2010; p 1423 S.

33. Halling, K. K.; Slotte, J. P., Membrane properties of plant sterols in phospholipid bilayers as determined by differential scanning calorimetry, resonance energy transfer and detergent-induced solubilization. *Bba-Biomembranes* **2004,** *1664*, (2), 161-171.

34. Choi, Y. H.; Kong, K. R.; Kim, Y. A.; Jung, K. O.; Kil, J. H.; Rhee, S. H.; Park, K. Y., Induction of Bax and activation of caspases during beta-sitosterol-mediated apoptosis in human colon cancer cells. *International journal of oncology* **2003,** *23*, (6), 1657-62.

35. Awad, A. B.; Roy, R.; Fink, C. S., Beta-sitosterol, a plant sterol, induces apoptosis and activates key caspases in MDA-MB-231 human breast cancer cells. *Oncology reports* **2003,** *10*, (2), 497-500.

36. Awad, A. B.; Toczek, J.; Fink, C. S., Phytosterols decrease prostaglandin release in cultured P388D(1)/MAB macrophages. *Prostag Leukotr Ess* **2004,** *70*, (6), 511-520.

37. Navarro, A.; De las Heras, B.; Villar, A., Anti-inflammatory and immunomodulating properties of a sterol fraction from Sideritis foetens CLEM. *Biol Pharm Bull* **2001,** *24*, (5), 470-473.

38. Palmer, K.; Crane, G., Respiratory disease in workers exposed to colophony solder flux fumes: Continuing health concerns. *Occup Med-Oxford* **1997,** *47*, (8), 491-496.

CAPTIONS:

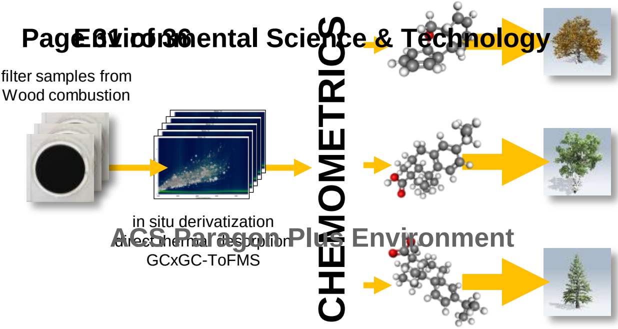
Figure 1: Graphical representation of the data workflow.

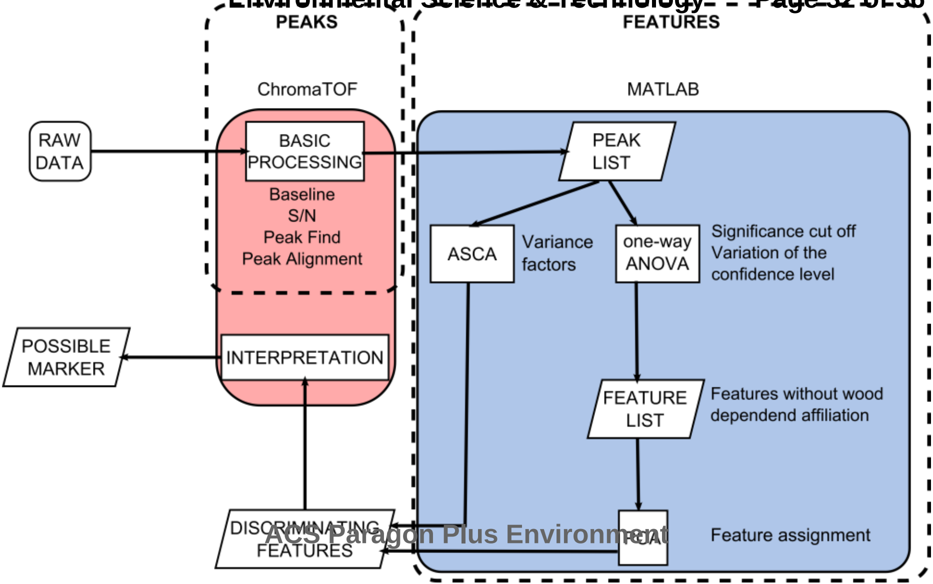
Figure 2: Scores and loading biplot for the peaks with p <= 0.001%

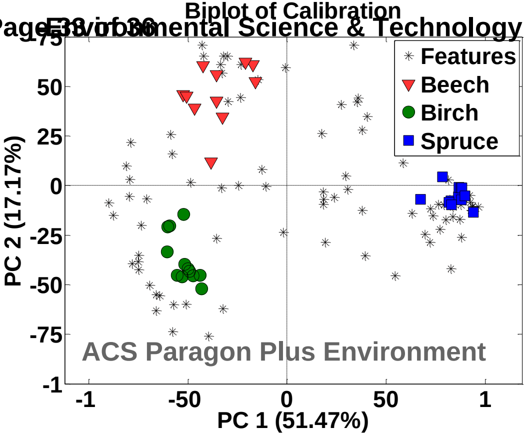
Figure 3: Assignment of the features to different wood types. The bubble size represents the loading of each peak. Features from birch and spruce have higher loading than features from beech combustion.

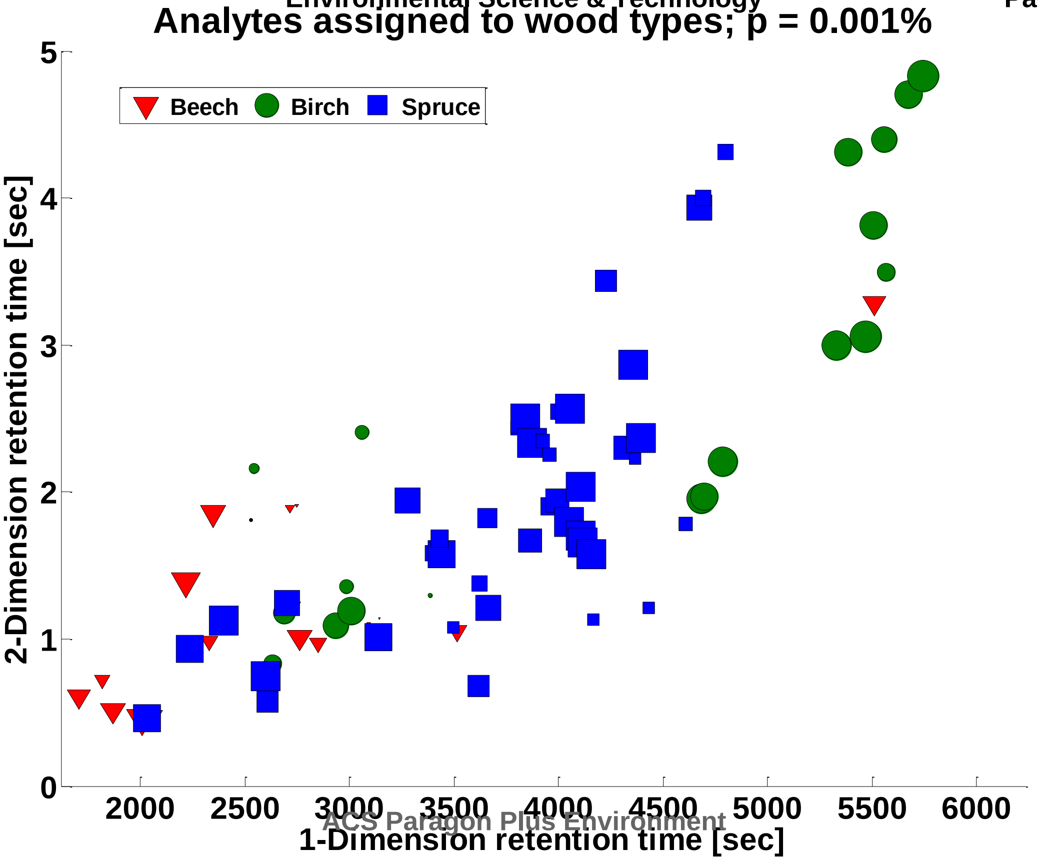
Figure 4: Density map of the occurrence of significant features. Decrease of the confidence interval for the beech (A & B), birch (C & D) and spruce (E & F) samples from 0.001% to 5%. The increase of significant features is mainly observed in distinct chromatographic regions, which are highlighted in the density maps for p = 5%. The * region in Figure 4B contains mainly small polar compounds, e.g., methoxy benzenes and furanoses. In Figure 4D, the * region contains mainly methylated naphthalenes, whereas the # region contains phytoseroids. In 4F, the # region has abundances of retene and resinoic acid derivatives, the diamond region contains mainly PAH, such as benzofluoranthenes. For the * region in 4F, a clear occurrence of one specific chemical class is not observed; phenylpropanoid derivatives and sugar components are found, in addition to naphthalene derivatives.

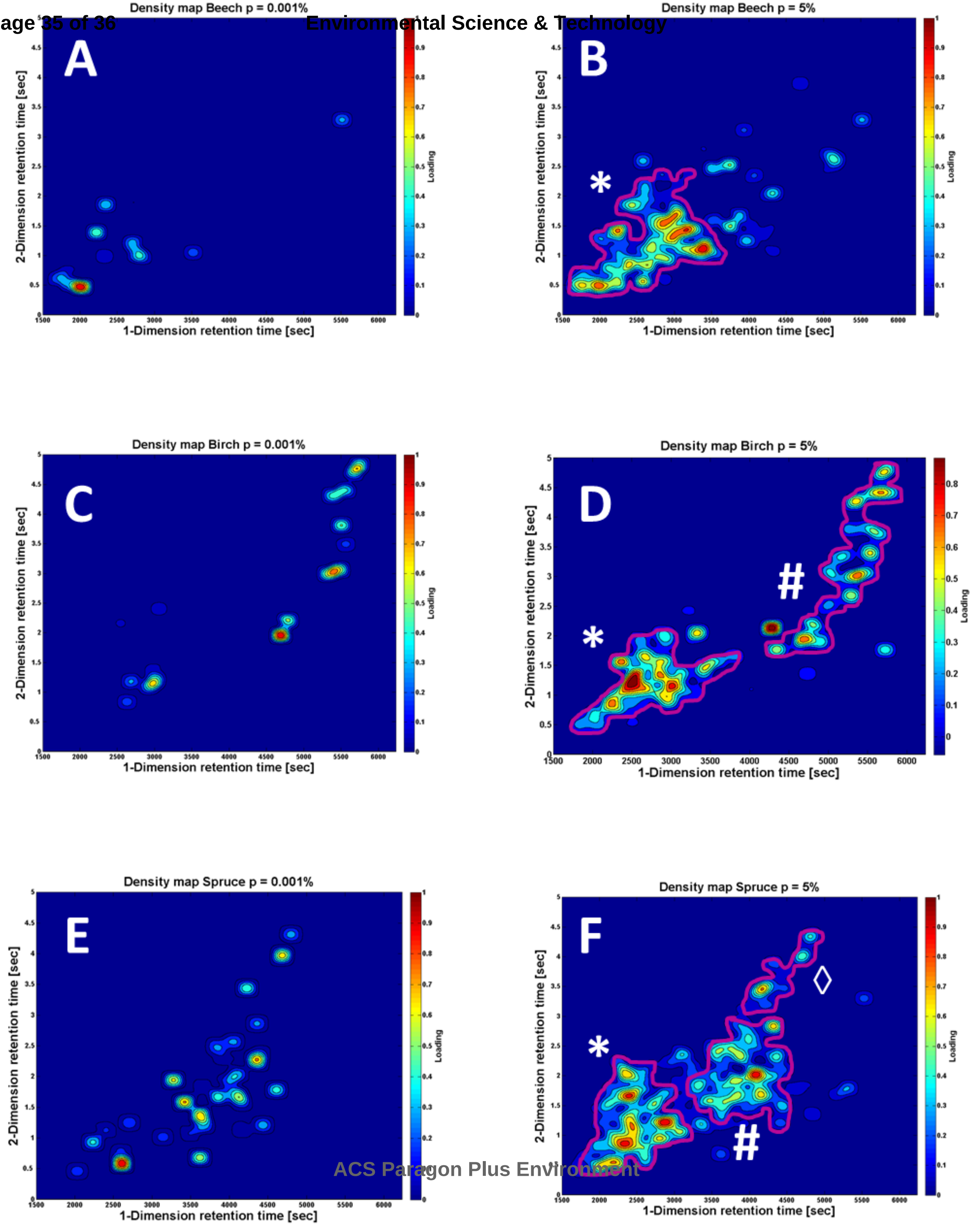
Figure 5: ASCA scores and loadings biplot for the wood types. All other variance factors were excluded; therefore, the different experiments are clustered at only one point. Highly correlated features indicate components having high mass spectrometric response.

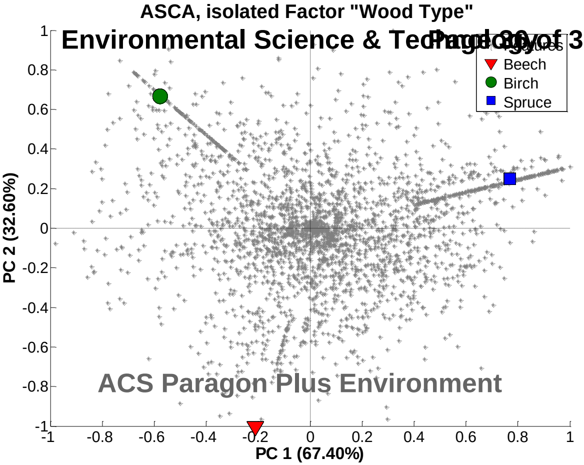

filter samples from
Wood combustion
in situ derivatization
direct thermal desorption
GCxGC-ToFMS
CHEMOMETRICS


PEAKS
FEATURES
ChromaTOF
MATLAB
RAW DATA
BASIC PROCESSING
Baseline
S/N
Peak Find
Peak Alignment
PEAK LIST
ASCA
Variance factors
one-way ANOVA
Significance cut off
Variation of the confidence level
FEATURE LIST
Features without wood dependend affiliation
PCA
Feature assignment
INTERPRETATION
POSSIBLE MARKER
DISCRIMINATING FEATURES


Biplot of Calibration
Features
Beech
Birch
Spruce
PC 2 (17.17%)
75
50
25
0
-25
-50
-75
-1
PC 1 (51.47%)
-1
-50
0
50
1


Analytes assigned to wood types; p = 0.001%
Beech
Birch
Spruce
2-Dimension retention time [sec]
1-Dimension retention time [sec]
0
1
2
3
4
5
2000
2500
3000
3500
4000
4500
5000
5500
6000


Density map Beech p = 0.001%
A
Density map Beech p = 5%
B
*
Density map Birch p = 0.001%
C
Density map Birch p = 5%
D
*
#
Density map Spruce p = 0.001%
E
Density map Spruce p = 5%
F
*
#
◊
2-Dimension retention time [sec]
1-Dimension retention time [sec]
Loading
1500
2000
2500
3000
3500
4000
4500
5000
5500
6000
0
0.5
1
1.5
2
2.5
3
3.5
4
4.5
5
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1


ASCA, isolated Factor "Wood Type"
Figures
Beech
Birch
Spruce
PC 2 (32.60%)
PC 1 (67.40%)
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
-1
-0.8
-0.6
-0.4
-0.2
0
0.2
0.4
0.6
0.8
1