Supplementary Material

Detailed Analytical Methodology

Sample preparation

Fire debris samples were first analysed according to the standard procedure to get an impression on what kind of components are present in the fire debris samples and to determine the headspace volume that should be sampled to make extracts for the GC×GC-MS analyses. First headspace analysis was carried out on gas chromatography coupled to a flame-ionization detector (GC-FID) by heating the jars for at least 4 hours at70°C and then injecting 0.5 mL of headspace. After this screening measurement the sampling volume was determined. Extracts for GC×GC-MS analysis were prepared by drawing a headspace volume from the jar through a sorbent tube containing activated charcoal using a 100 mL syringe. Samples were collected by removing both ends of the sorbent tube using a glass cutter and connecting a needle on one end and the syringe on the other, using two pieces of silicon tubing. After pulling sample headspace through the tube, it was cut through and the charcoal contents were collected in plastic microcentrifuge vials. Into each of the vials 1 mL of dichloromethane (DCM) (Biosolve, Valkenswaard, The Netherlands) spiked with 0.01 mg/mL chlorobenzene (Grace-Alltech, Breda, The Netherlands) was added as an internal standard (IS). After centrifuging for 15 min (at 13000 rpm) the supernatant was transferred using a glass Pasteur pipette to a GC vial, capped and stored for further analysis. The complete composition of each experimental burn is provided as supplementary material.

GC-FID screening

The fire debris samples were first analysed by GC-FID after heating the sample storage jars for 4 hours at 70°C. 0.5 mL was injected; based on the signal intensity observed, the appropriate sampling volume was determined for subsequent GC×GC-MS analyses. Thus headspace was sampled from the same jars on a sorbent tube containing activated charcoal. GC-FID screening analyses were performed on an Agilent Technologies 6890A Network GC System equipped with an Agilent Technologies 7683B Series Injector and an Agilent Technologies 7683B Series Autosampler (all from Agilent, Santa Clara, CA, USA). An Agilent medium-polarity DB-624 (6% cyanopropyl-phenyl, 94% dimethyl polysiloxane) column (30 m × 320 μm i.d.; 1.8 μm film thickness) was used. The injection volume was 0.5 mL of headspace with a split ratio of 20 : 1. Helium was used as the carrier gas at a constant flow rate of 2 mL/min. The following temperature program was applied for the separation: 80°C with a 2 min hold, followed by a linear ramp of 40°C/min to 255°C with a 5 min hold. Both the injector temperature and the detector temperature were set to 250°C. GC-FID screening analysis allowed the determination of how much headspace volume should be sampled to get extracts for the subsequent GC×GC-MS analyses. The sampling

volume was determined by looking at the signals from the screening methods.

GC×GC-MS analysis

GC×GC-MS analyses were carried out on an Agilent Technologies 6890N GC System with a LECO dual-stage, quad-jet thermal modulator coupled to a Pegasus III ToF-MS (LECO, St. Joseph, MI, USA). The sample injection order was randomized and samples were run in batches over the course of 4 weeks. A Gerstel (Mülheim an der Ruhr, Germany) autosampler was used for all injections. An Agilent DB-1 (100% dimethyl polysiloxane) first-dimension column (30 m × 250 μm i.d.; 0.5 μm film thickness) was used in combination with an Agilent DB-17 ((50%-Phenyl)-methyl polysiloxane) second-dimension column (1 m × 100 μm i.d.; 0.2 μm film thickness).The columns were connected via a Siltek, universal press-tight connector (Restek, USA). For the analysis, extracts were injected as such. Splitless injections were carried out for the extracts and performance test mixtures with an injection volume of 1 μL. The performance test mixture contained the following compounds: chlorobenzene, p-xylene (Fluka, Zwijndrecht, The Netherlands), o-xylene (Fluka, Zwijndrecht, The Netherlands), 1-ehtyl-2-methylbenzene (Chem Service, Alltech, Dieren, The Netherlands), n-nonane and n-decane (Acros Organics, Geel, Belgium), n-undecane (Merck, Darmstadt, Germany), n-dodecane and n-tridecane (Acros Organics, Geel, Belgium). Each compound had a final concentration of 0.005 mg/mL in DCM.

For the analysis of the diluted intact ignitable liquids split injections were carried out with an injection volume of 1 μL and a split ratio of 30 : 1. Neat ignitable liquids were diluted to a final concentration of 0.01 mg/mL. Here the use of a split ratio injection was deemed likely to cause less severe discrimination of compounds compared to further dilution of the samples. The temperature program of the PTV injector started with a temperature of 40°C with a hold time of 0.1 min, followed by a linear ramp of 12°C/s to a final temperature of 250°C. Helium was used as the carrier gas at a constant pressure of 110 kPa. The temperature for the first-dimension separation was initially set at 45°C for 0.5 min, followed by a linear ramp of 1°C/min to a temperature of 80°C, followed by a second linear ramp of 3°C/min to a temperature of 130°C, and finally a linear ramp of 5.5°C/min to a final temperature of 255°C with a hold time of 10 min. An offset of +5°C was used for a parallel temperature program in the second column oven. The modulator temperature offset was 15°C. The inlet temperature was held at 250°C. A modulation time of 4 s was used during the entire run with a hot-pulse duration of 400 ms. The MS transfer-line temperature was maintained at 225°C. The ion-source temperature was 250°C with an electron ionization (EI) energy of 70 eV. A solvent delay of 350 seconds was used. The acquisition rate was 200 scans per second, covering a 35-500 m/z range. Example chromatograms are provided for each of the three ILs of interest in the main manuscript.

Algorithmic representation of the data processing workflow

Figure 1: Workflow for classifying test samples based on features selected using a tuning data partition and a model generated from training data

Data: $\mathbf{x}$ = Reduced form representation (RFR) dataset $[n \times [m \times t]]$
$\mathbf{y}$ = Class labels $[1 \times n]$
$\mathbf{n}$ = Number of experimental samples [155]
$\mathbf{m}$ = Number of nominal m/z values measured [466]
$\mathbf{t}$ = Number of tiles dividing the chromatogram [12]
$\mathbf{k}$ = Number of stratified data partitions [10]
Result: $\hat{\mathbf{y}}$ = Predicted class labels $[1 \times n]$
for *j=1:k* **do**
  Divide data into $\mathbf{k}$ blocks;
  k_1 = *testing* block;
  k_2 = *parameter tuning* block;
  $k_{\notin(1,2)}$ = *training* block $[k_3 \dots k_{10}]$;
  for $i = 1 : (m \times t)$ **do**
    ssDistances$_i$ = calculate pairwise distances for $x \in k_{\notin(1,2)}$ with same y;
    dsDistances$_i$ = calculate pairwise distances for $x \in k_{\notin(1,2)}$ with different y;
    modelSame$_i$ = fit normal distribution to ssDistances$_i$;
    modelDiff$_i$ = fit normal distribution to dsDistances$_i$;
    CLLR$_i$ = evaluate LRs of $x \in k_2$ using: modelSame$_i$, modelDiff$_i$, $y \in k_2$;
    topFeats = sort in ascending order:CLLR$_i$, select features $1 : 25$;
  end
  projectedData = transform($x_{k_{\notin(1,2)}}$, $y_{k_{\notin(1,2)}}$, topFeats) to new space by LDA;
  for h = *unique element* $\in y$ **do**
    classModel$_h$ = fit normal distribution to projectedData$_{(y=h)}$
  end
  for *each element in* k_1 **do**
    $\hat{y}_{k_1}$ = max(*projectedData* $\in k_1$, classModel$_{1:h}$)
  end
  circular shift data blocks by 1 index (i.e. k becomes k+1, 10 becomes 1)
end

All calculations were carried out on an Apple Inc. MacBook Pro^{TM}laptop computer with a 3.0GHz Dual-core Intel Core i7 processor and 16GB 1600MHz DDR3 SDRAM. The methods described herein were programmed in the MATLAB^{TM} programming language using MATLAB version 8.4.0.150421, 2014b release (MathWorks Inc. Natick, Massachusetts, U.S.A.).