



Statistical interpretation of Chemical Evidence Pertaining to Fire Debris  
*M. Lopatka*

The chemical analysis of forensic traces can be crucial to the overall forensic investigative process at complex crime scenes. Few crime scenes may be deemed more complex than the site of an arson investigation. As technology to perform sophisticated chemical analyses progresses, the forensic science community encounters opportunities to incorporate new sources of information into a rigorous scientific process aimed at providing objective and accurate assertions pertaining to forensic hypotheses.

Research in statistical methods for processing chemical evidence contributes to the evolution of chemical trace forensics in general. With each new generation of analytical technologies, limits of detection are driven ever lower and the specificity of measurements ever greater. One characteristic in which the development of new technologies places the most challenge is in the handling of an ever increasing abundance of data associated with each analysis performed. These challenges have already created the need for chemometrics and statistical involvement in many industries that use chemical data. Forensic science as a discipline imposes even more stringent constraints on the kinds of techniques that may or may not be employed, especially on the rigor with which those methods are critically evaluated. The framework of forensic trace evidence evaluation demands an understanding of the cause and impact of different sources of variability on a chemical measurement and the manner in which these variances may potentially bias any conclusions made on the basis of evidence observed. The task of forensic scientists is the interpretation and evaluation of evidence as objectively as possible, so that others can make informed decisions on the basis of the observed traces. During a fire scene examination the observation of evidence indicating the presence of ignitable liquid residues may completely alter the course of the investigation; even more so if evidence can be presented to indicate characteristics of the ignitable liquid suspected.

This thesis presents a collection of tools for the processing of complex chromatographic data. The development of these tools was carried out as part of the COMFOR research project, developing better solutions for forensic problems pertaining to the assessment of the evidential value of analytical data derived from the chemical analysis of fire debris. This task has been approached starting with the most raw, unprocessed, unfiltered data from a series of different analytical techniques employed when performing chemical analyses. Efforts have been made to assess the full information available from a chemical analysis in determining the informative portions of the signal for sample characterization, sample classification, and inter-sample comparison.

The process begins with a reevaluation of the classic peak detection paradigm for chromatographic chemical analyses. Rather than beginning with a binary decision pertaining to the presence or absence of particular usable features, a probabilistic view is adopted wherein each point in the original signal is evaluated for its likelihood of being affected by a chromatographic compound, and therefore a potentially interesting characteristic of the data for forensic examination. This fundamental deviation from a peak-table based examination of chemical samples exemplifies the unifying theme of the remaining research contained in this dissertation. As much as possible, the full data available from chemical analyses is used, and reduction in dimensionality is performed always keeping in mind the distinction between information and data. In this way feature selection is performed based on the context of the forensic problem being addressed. The conventional method for automated comparison of chemical analyses relies on constraints imposed by the a priori selection of informative regions in chemical data: the detection of pre-selected peaks, the integration of their areas, and the registration of these values into a table. Many complete chemical analysis workflows (in a variety of scientific applications) do not return to the original data and treat the feature selection as a separate and distinct step from the mul-

tivariate statistical analysis. Following this process one functionally discards the majority of the measured signal before any problem-specific analysis begins.

The probabilistic peak detection method instead presents the probability that each region in the original signal may contain a chromatographic peak, this information may be potentially propagated through subsequent processing steps. This method avoids prematurely eliminating candidate features in chromatographic data prior to multivariate analysis. The first successful application of this is presented in the task of baseline correction for complex chromatographic samples. By propagating the probabilistic peak detection through subsequent data processing steps, very good performance is achieved in reducing the effect of unwanted and forensically uninformative artifacts from chromatographic signals while avoiding the loss of trace compounds (frequently overlooked in automated data preprocessing methods).

Another approach to examining chromatographic data is also explored, based on existing methodologies, developed for the automation of fire debris detection. Again, a variation on full data utilization manifests itself by using the Total Ion Spectra (TIS) representation of chemical data from fire debris traces. This work relies on developing a more sophisticated statistical method for extrapolating relatively few samples to a general model for the classification of ignitable liquids and their residues. A novel approach is implemented to perform digital recombination of chemical characteristics of fire debris samples with known composition in order to formulate a statistical model that may approximate the complexity of the variation of fire debris samples in a reduced feature space. This method used chemical analyses of ignitable liquids from the publicly accessible Ignitable Liquids Reference Collection (ILRC)<sup>1</sup> in combination with substrate materials from the publicly available Substrate Database<sup>2</sup> and tested against real fire debris samples. The results achieved demonstrate a promising extrapolation between laboratory data collection and real-world application in a forensic investigation. The nature of the data used in this research has great advantages in terms of the robustness to measurement variations, but at the cost of information loss.

Given the success achieved by the complete removal of chromatographic features, subsequent investigation is aimed at achieving a hybrid approach between the TIS method and the use of comprehensive two-dimensional gas chromatography (GC×GC). The drastic increase in the amount of data produced by one analytical measurement both allows and necessitates a reduced representation of the sample before statistical modeling of the features can be performed. In theory, the orthogonal retention mechanisms employed by GC×GC mean that compounds eluting at different locations in a two-dimensional separation space will have distinct chemical characteristics. The accumulation of ion abundances for spatially defined regions aims to discard the maximum amount of data while retaining the most informative characteristics for a forensic assessment. In an effort to maintain the full multidimensional characteristic of the data, a processing method was conceived that combines the robustness and reliability observed in early TIS work but also makes use of the significantly improved separation space afforded by comprehensive chromatography. Total ion intensity values are summed over distinct spatially defined regions in the GC×GC chromatogram yielding multiple consolidated characteristic summed ion traces. These new reduced features are introduced as Local Ion Signature (LIS). Subsequent feature reduction is performed in an objective evaluation of the potential forensic relevance of reduced

---

<sup>1</sup>Ignitable Liquids Reference Collection Database, National Center for Forensic Science, University of Central Florida. [cited 24-03-2016], Available from: <http://ilrc.ucf.edu/>.

<sup>2</sup>Substrate Database. National Center for Forensic Science, University of Central Florida. [cited 24-03-2016], Available from: <http://ilrc.ucf.edu/substrate/>

representation of this measurement to eventually yield a highly informative representation of the data. The findings demonstrate that GC×GC can be harnessed without having to rely on peak detection and integration methods. Preliminary results show that it is possible to evaluate the evidence (i.e. calculate likelihood ratios) in a fully automated manner. Furthermore, a visualization tool has been developed in order to assist forensic scientists in identifying which chemical compounds are the basis of the likelihood ratios obtained. This promotes a practical interaction between forensic expertise and statistical methodology in assessing the evidential value of fire debris samples.

The forensic problem of ignitable liquid residue detection and subsequent classification are investigated in the space of comprehensive chromatographic analytical data. In order to evaluate the new technique, a series of laboratory burn experiments mimicking the digital mixing procedure were performed by the COMFOR team. Performance was assessed using the resulting representative and realistic fire debris collection, in which a forensic classification paradigm could be evaluated for a set realistic and complex fire debris samples wherein the ground truth pertaining to ignitable liquids used as accelerants was known.

The amalgamation of analytical chemistry, chemometrics, and forensic statistics provided herein presents an exciting prospect for further development of methods targeted towards the statistical interpretation of chemical evidence pertaining to fire debris.