A176 Comparison of SIMCA With LDA and QDA for the Identification and Classification of Ignitable Liquid Residues

Erin E. Waddell, BS, Mary R. Williams, BS, and Michael E. Sigman, PhD, Univ of Central Florida, NCFS, PO Box 162367, Orlando, FL 32816*

The goal of this presentation is to establish a methodology with known error rates for identification and classification of ignitable liquids in fire debris samples.

This presentation will impact the forensic science community by discussing the benefits of each method when applied to ignitable liquid and pyrolysis product classification in fire debris analysis. The methods investigated are intended to provide statistical support for current laboratory practices.

This research compared chemometric methods for classifying ignitable liquids and ignitable liquid residues in fire debris samples. The methods emphasized identifying samples that contain ignitable liquid residues and assigning ignitable liquid residues to classes as defined by the American Society for Testing and Materials (ASTM) E1618 standard. ASTM classes include: aromatic, gasoline, petroleum distillates, isoparaffinics, naphthenic paraffinics, normal alkanes, miscellaneous, and oxygenates. This presentation will impact the forensic science community by discussing the benefits of each method when applied to ignitable liquid and pyrolysis product classification in fire debris analysis. The methods investigated are intended to provide statistical support for current laboratory practices. The goal of this research is to establish a methodology with known error rates for the identification and classification of ignitable liquids in fire debris samples.

Results using Soft Independent Modeling Of Class Analogies (SIMCA) will be compared to those obtained using Linear Discriminant Analysis (LDA) and Quadratic Discriminant Analysis (QDA). SIMCA is a soft classification technique which allows a sample to be classified into a single class, multiple classes, or to be unassigned. The option of classifying into multiple classes could be beneficial for fire debris samples since the ratio of ignitable liquid residue-to-substrate pyrolysis is unknown. Discriminant analysis is a hard classification technique which means that each sample must be assigned to a single class and failure to assign to a class is not an option. For both techniques, models were developed from libraries of Gas Chromatography-Mass Spectrometry (GC/MS) data for ignitable liquids and pyrolysis products. Models for assigning samples to the ASTM classes were based on the Total Ion Spectrum (TIS). Cross-validation techniques were also used for each chemometric method. A test set was created by randomly selecting 20% of the samples from each class, with the remaining 80% developing a model data set. The samples in the test set were then classified using the respective chemometric method. Cross-validation steps were repeated 100 times with a new test set being selected and classified each time. Total correct classification percentages were calculated as the sum of the cross-validation tests. All models developed for ignitable liquid and substrate library data were applied to fire debris samples.

Outlier samples within each class were first removed based on their orthogonal and score distances before the SIMCA model was developed. For each class, Principal Components Analysis (PCA) was performed and the cutoff distances for the orthogonal and score distances were calculated from the model data. The test data set was projected into the model PCA space for the respective classes and orthogonal and score distances were calculated. A sample was assigned to the given class if the calculated score value was less than or equal to one. If the value exceeded one, the sample was not assigned to that class. A sample was considered to be correct if it only assigned to the known class and partially correct if it assigned to multiple classes that included the correct class.

Preliminary results show that when using all ASTM classes, other than miscellaneous and oxygenate classes, most of the samples classified using SIMCA were correct, or partially correct, with very few samples being incorrectly assigned. A partially correct sample was assigned to multiple classes that included the correct class, but also included incorrect assignment(s). For discriminant analysis methods, a two-class multi-step classification scheme was used.[1] The optimal QDA model gave correct classification rates greater than 90% for all steps; the optimal LDA model's correct classification rates were greater than 75% for all steps.

This work was supported by the National Institute of Justice, Office of Justice Programs, award 2009-DN-BX-K227. The content of this publication does not necessarily reflect the position or the policy of the Government and no official endorsement should be inferred.

Reference:

1. Waddell EE, Song ET, Rinke, CN, Williams MR, Sigman ME. Progress toward the determination of correct classification rates in fire debris analysis. J Forensic Sci 2012, In-press.

Fire Debris, Chemometrics, Error Rates