

Monitoring Diesel Fuel Degradation by Gas Chromatography–Mass Spectroscopy and Chemometric Analysis

Kevin J. Johnson, Susan L. Rose-Pehrsson, and Robert E. Morris*

Naval Research Laboratory, Code 6112, Washington, DC 20375-5320

Received October 6, 2003

Recent advances in the field of chemometrics have provided us with an opportunity to determine if it is possible to extend numerical multivariate pattern recognition techniques beyond simple fuel characterization, to the identification of important compositional features that are related to fuel quality. This can include unique combinations of normally benign constituents that exert an impact on fuel stability. Fuels are ideal candidates for chemometric analysis, because we are often concerned with minute features within a complex compositional matrix. Two potential benefits of this approach are the development of diagnostic and predictive models that can relate fuel composition to quality. We have begun our investigation with studies of gas chromatography–mass spectrometry (GC–MS) data from fuels that have undergone various levels of thermally induced autoxidation. An analysis of variance (ANOVA) feature selection technique has been applied to locate features in the GC–MS data that change from sample to sample, thus allowing for a quick evaluation of how fuel composition is altered during stress. In this manner, evaporative losses, rather than fuel degradation, have been observed to dominate the chemical variations that are produced in naval distillate fuels (NATO F-76) during oven stress at 60 °C. Thermal stress in a closed low-pressure reactor (LPR) eliminates evaporative losses, and the chemical changes have been readily observed and modeled. A progressive change in composition during both oven and LPR stress is revealed from multi-way principal component analysis. Decomposition of windowed regions of the GC–MS data via parallel factor analysis provides a means of extracting the mass spectra of individual fuel constituents that change during stress. This illustrates the potential diagnostic capability of multi-way chemometric analysis of GC–MS data.

Background

Hydrocarbon fuels are complex mixtures of organic compounds that are manufactured to comply with performance specifications on the basis of properties and not composition. Thus, although it is true that fuels obtained in conformance with a particular specification may be similar in their physical properties and performance, their composition may differ in both obvious and subtle ways. These compositional differences are a consequence of many factors, which include refining and finishing methods, crude sources, handling methods, contamination, and blending with other fuels. Not only does the chemical composition of each fuel differ, but the composition of any particular fuel can also change with time. It is usually the formation of insoluble reaction products or changes in a critical property that bring fuel stability into question. Very often, the chemical processes that lead to stability problems are due to the presence (or absence) of constituents at trace concentration levels.

Considering the complexity and variability of fuels, it is not surprising that success has often eluded researchers in their efforts to understand and provide practical solutions to operational problems attributed

to undesirable fuel chemistry. As a consequence, methodologies to characterize and study fuel stability have focused on producing and quantifying insoluble reaction products. However, it is not always realistic to attach any mechanistic significance to quantities of insoluble products formed to define relationships between composition and liquid-phase chemistry.

The development of predictive models that are based on critical relationships between fuel composition and thermal stability would enable the assessment of fuel performance on the basis of composition. Moreover, this approach would not necessarily entail stressing or other treatments that could change the chemistry. Recent advances in analytical techniques, computer technology, and the science of chemometrics have made it possible to bring these technologies together to address the task of developing a useful predictive model that correlates fuel stability with fuel composition.

The science of chemometrics has grown from a need to identify hidden relationships in complex data, and it is used extensively in spectroscopy. Because the compositional features that can influence critical fuel properties are typically obscured by an abundance of irrelevant information, some of the techniques of chemometrics would seem well-suited to reveal hidden information in complex data from fuel compositional analysis. Chemometric analysis of near-infrared spectroscopy (NIR) data

* Author to whom correspondence should be addressed. E-mail address: morris4@ccs.nrl.navy.mil.

is successfully used to obtain estimates of octane number and aromatic content of gasoline.¹ NIR has also been investigated as a method to obtain rapid field evaluation of critical middle distillate fuel properties, using partial-least-squares (PLS) analysis. Spectral features of midband (650–4000 cm⁻¹) IR spectra from middle distillate fuels have been correlated with various physical properties,^{2,3} demonstrating the potential of chemometrics of NIR spectra to predict those fuel properties that can be linearly related to chemical structural features that produced the IR resonance bands. These properties included density, viscosity, cetane index, heat of combustion, cloud point, distillation properties, and aromatic content.

Several research groups have also successfully used chemometric methods to discriminate between different fuel types from gas chromatography (GC) analyses. In 1979, Clark and Jurs⁴ described automated classification algorithms that successfully characterized GC data from crude oil samples according to their origin. Neural networks have also been used⁵ to develop multivariate pattern recognition models for classifying jet fuel samples by type (e.g., JP-4 versus JP-5). Lavine et al.⁶ noted that mathematical pattern recognition methods offer a better approach to GC fuel analysis than visual inspection, because of the complex nature of the processed fuel samples. They demonstrated that pattern recognition methods could successfully identify fingerprint patterns in the GC data that were characteristic of fuel type, even in the presence of severe weathering. In each of these applications, the information needed to discriminate the fuel samples consisted of subtle variations in peak intensities distributed across multiple peaks in the chromatograms. To improve the selectivity and predictive power of principal component analysis of GC data, Johnson and Synovec⁷ used analysis of variance (ANOVA) calculations to restrict the analysis to only those features in the data that were relevant to the classification. In this manner, they were able to discriminate between JP-5, JP-8, and JP-TS in mixtures with as little as 1% variation in volume, using comprehensive two-dimensional GC (GC × GC), and eliminate the impact of geographical variances in the jet fuel samples.

We have previously investigated the use of chemometrics to obtain a predictive model during an incident where a jet engine combustor had experienced failures when using fuels from a particular source. The fuels that caused these failures had passed the standard laboratory tests required by the applicable specifications. In addition, a suite of nonstandard tests did not reveal the presence of abnormal or highly reactive constituents that could have been identified as responsible for these failures. Discriminant partial-least-

Table 1. Summary of Fuel Stress Conditions Imposed on the 23 Samples of Naval Distillate Fuel (NATO F-76)

sample numbers	stress conditions
1–6	unstressed
7–11	LPR, 16 h
12–17	LPR, 42 h
18–23	LPR, 42 h, copper added

squares regression (DPLS) was used to classify the sample chromatograms as belonging to either engine-passing or engine-failing fuel samples. Interval partial-least-squares regression (i-PLS) was used to define an early portion of the gas chromatograms that was observed to be the most significant in discriminating between the failed and good fuels in this application. DPLS cross-validation and contiguous blocks cross-validation confirmed that the subtle differences in composition, as indicated by this GC retention time window, were more highly correlated with the performance of these fuels in the engine test than the entire chromatogram.

This illustrated the potential of chemometrics to correlate fuel composition with performance. However, a robust predicative model that is derived from GC data is limited, because of the fact that differences between fuels can be much greater than the subtle compositional differences of interest. In the present work, we expand beyond the scope of these previous studies to include multi-way chemometric techniques and gas chromatography–mass spectrometry (GC–MS) data as a possible means of overcoming these limitations.

Experimental Section

A specification naval distillate fuel (NATO F-76) was used for these studies. In storage stability testing by the low-pressure reactor (LPR) method in accordance with ASTM D 5304,⁸ this fuel produced 0.7 mg sediment per 100 mL. Two methods of thermally stressing the fuel were used to provide fuel samples with varying levels of thermal degradation for analysis by GC–MS. In the first study, 100 mL of fuel was placed in a vented 250-mL borosilicate glass bottle and stored in an oven, in darkness, at 60 °C. Aliquots (1 mL) were periodically withdrawn without cooling for GC–MS analysis at 0, 7, 14, 24, 31, and 37 days of oven stress. In a second study, fuels were subjected to stress in a closed LPR. LPR stress was applied on aliquots of fuel in accordance with ASTM D 5304 (i.e., at a temperature of 100 °C and under an atmosphere of 100 psig oxygen) for both the standard duration of 16 h and also an extended duration of 42 h. The extended duration LPR conditions were chosen to guarantee significant oxidative changes in the fuel. After the fuel samples were removed from the LPR, they were cooled in darkness and 2 mL aliquots were filtered through a 0.2- μ m nylon Millipore filter for analysis. A summary of the LPR-stressed samples analyzed via GC–MS is presented in Table 1.

GC–MS data were obtained using a Hewlett–Packard model HP 5890 Series II gas chromatograph coupled to a Hewlett–Packard model HP 5971 mass-selective detector. Replicate dilutions were prepared and analyzed for each fuel sample: three replicates per oven-stressed sample and six replicates per LPR-stressed sample. Each replicate was prepared by dissolving 7.5 μ L of fuel in 1500 μ L of dichloromethane. A Hewlett–Packard model HP 6890 injector and

(1) Swarin, S. J.; Drumm, C. A. Prediction of Gasoline Properties with Near-Infrared Spectroscopy and Chemometrics. *SAE Tech. Paper Ser.* **1991**, 912390.

(2) Fodor, G. E.; Kohl, K. B. *Energy Fuels* **1993**, 7, 598–601.

(3) Westbrook, S. R. Army Use of Near-Infrared Spectroscopy to Estimate Selected Properties of Compression Ignition Fuels. *SAE Tech. Paper Ser.* **1993**, 930734.

(4) Clark, H. A.; Jurs, P. C. *Anal. Chem.* **1979**, 51, 616–623.

(5) Long, J. R.; Mayfield, H. T.; Henley, M. V.; Kromann, P. R. *Anal. Chem.* **1991**, 63, 1256–1261.

(6) Lavine, B. K.; Mayfield, H. T.; Kromann, P. R.; Faruque, A. *Anal. Chem.* **1995**, 67, 2846–3852.

(7) Johnson, K. J.; Synovec, R. E. *Chemom. Intell. Lab. Syst.* **2002**, 60, 225–237.

(8) Standard Test Method for Assessing Middle Distillate Fuel Storage Stability by Oxygen Overpressure, ASTM D 5304-03. In *2003 Annual Book of ASTM Standards*; American Society for Testing and Materials (ASTM): Philadelphia, 2003; Vol. 05.03.

autosampler delivered 1.0 μL aliquots of each replicate sample to the gas chromatograph in a random order. A split/splitless injector at 250 °C with a split flow ratio of 60:1 was used along with a 50 m \times 0.2 mm Agilent HP-1 (dimethylpolysiloxane) capillary column. The oven temperature profile was 60 °C to 288 °C at a rate of 3 °C/min, giving a run time of 76 min. A solvent delay of 4.40 min was used, which reduced the data acquisition time to 71.6 min per run.

Following acquisition, GC–MS chromatograms were imported into MATLAB version 6.5 (Mathworks, Inc., Natick, MA) for all subsequent chemometric analyses. After being imported into MATLAB, GC–MS chromatograms were pre-processed to minimize any undesired variation between chromatograms. First, all chromatograms were normalized to a unit area basis, to minimize effects that were due to variation in the injected sample volume. Next, the chromatograms were retention-time-aligned, to minimize retention-time variation between chromatograms that is due to unavoidable fluctuations in instrument parameters (e.g., oven temperature, flow rate, etc.) during the course of the experiments.

Retention-time alignment was accomplished via a two-step procedure. First, a total ion current (TIC) chromatogram for each GC–MS chromatogram was constructed by taking the sum along the mass spectral axis in the data. The set of all TIC chromatograms was then subjected to a manual peak-alignment procedure, where corresponding peaks from chromatogram to chromatogram were visually identified and tabulated in Matlab. Taking the first chromatogram as the target for alignment, the required shifts for the remaining chromatograms were calculated. In the second step, these required shifts were applied to the retention-time axis of the GC–MS chromatograms via interpolation.

Because the data acquired for this work were accumulated over a relatively short time period on a single instrument, the instrumental parameters could be more closely controlled than is generally possible in daily routine analyses. Given the relative stability of the GC–MS data acquired, it was unnecessary to add internal standards for retention-time alignment and normalization purposes in this experiment. In general practice, under conditions where carrier-gas flow rate variations are encountered, it certainly may be necessary to use internal standards. In previous studies,⁹ we have successfully corrected nonlinear variations in GC carrier-gas flow rates by aligning chromatograms within portions of the data bracketed by the major hydrocarbon peaks. Such a windowed approach to baseline peak alignment is a viable approach to minimizing instrumental variations that could be encountered in actual use and when the best possible quantitative precision is required.

Data sets for each experiment consisted of a series of two-dimensional GC–MS chromatograms, one for each sample analyzed, stacked atop each other to form a three-dimensional array, or cube, of data. These cubes were then submitted to multivariate chemometric analysis algorithms.

The data sets were probed for chromatographic features that described the difference between stressed and unstressed fuels using an ANOVA-based feature selection algorithm, as described by Johnson et al.⁷ and implemented by an in-house-written Matlab program. The algorithm functions by grouping a set of data into defined classes and looking for data points that are adept at describing the differences between classes, while remaining the same within a given class. This is accomplished by performing an ANOVA *F*-ratio calculation for each data point in the two-dimensional space of the GC–MS data across the set of samples examined. In this manner, a large data set can be rapidly and automatically scanned for features important for a given classification.

Multivariate models of GC–MS data were constructed to describe compositional changes that occur in the fuel samples due to stress. Modeling was accomplished via Matlab routines that were provided in the PLS toolbox (version 3, Eigenvector Research, Inc. Manson, WA) and the Nway toolbox (version 2.1, Andersson and Bro, available at <http://www.models.kvl.dk/source/nwaytoolbox/>).¹⁰ Two separate techniques were used to model changes in fuel composition: multiway principal components analysis (MPCA) and parallel factor analysis (PARAFAC).^{11–14} MPCA amounts to unfolding each two-dimensional chromatogram into a vector of data and subjecting the resulting set of vectors to standard principal components analysis (PCA).¹³ PCA is a commonly used chemometric technique for data reduction.¹⁵ It functions by constructing a basis set of orthogonal vectors in the original data space that most efficiently describes the variation present in the data set. The vectors are known as loadings, and the projections of each sample onto a particular loadings vector are known as scores. Each pair of scores/loadings comprises a principal component, and the number of principal components required to model a data set adequately is dependent on the number of independent sources of variation present in the data set. The loadings vector of the first component is constructed so that it lies in the direction of the greatest variation in the data space, that of the second component is constructed in the direction of the second greatest variation, orthogonal to the first, and so on. Prior to application of MPCA, the GC–MS data in this work were subjected to an eight-point boxcar average along the chromatographic dimension of the data to reduce the size of the data set, because of memory limitations of the personal computer on which the analysis was performed.

PARAFAC is a factor analysis technique that is similar in concept to PCA, except that it functions on three-dimensional arrays.^{11,12,14} Thus, each independent factor in a PARAFAC model consists of three loadings vectors—one for each dimension of the original data set. The loadings vectors generated by PARAFAC are calculated in an iterative fashion to best describe the overall variation in the data set in a least-squares sense, given a specified number of underlying factors present, as well as any constraints that may be imposed on any of the modes, such as non-negativity or unimodality. The incorporation of a third dimension allows for unique solutions to the factor decomposition problem, and, provided that the underlying factors fit the PARAFAC model (i.e., can be modeled as the outer product of three loadings vectors, which is a condition known as trilinearity), concentration and spectral features can be extracted directly. PARAFAC has been previously applied to hyphenated systems that involve chromatographic separation, followed by spectral detection.^{16,17} According to the strength and weaknesses of each technique, bulk changes in fuel composition (i.e., changes reflected in the overall chromatogram) were modeled with MPCA, because of their deviation from trilinearity, whereas PARAFAC was used to model smaller local regions of GC–MS data to extract concentration profiles and mass spectra of pure fuel components.

Discussion

There are two basic premises for applying chemometric methods to instrumental data. The first is that the

(10) Andersson, C. A.; Bro, R. *Chemom. Intell. Lab. Syst.* **2000**, *52*, 1–4.

(11) Smilde, A. K. *Chemom. Intell. Lab. Syst.* **1992**, *15*, 143–157.

(12) Henrion, R. *Chemom. Intell. Lab. Syst.* **1994**, *25*, 1–23.

(13) Dahl, K. S.; Piovoso, M. J.; Kosanovich, K. A. *Chemom. Intell. Lab. Syst.* **1999**, *46*, 161–180.

(14) Bro, R. *Chemom. Intell. Lab. Syst.* **1997**, *38*, 149–171.

(15) Malinowski, E. R. *Factor Analysis in Chemistry*, 2nd ed.; Wiley: New York, 1991.

(16) Beltran, J. L.; Guiteras, J.; Ferrer, R. *J. Chromatogr., A* **1998**, *802*, 263–275.

(17) Andersson, G. A.; Dable, B. K.; Booksh, K. S. *Chemom. Intell. Lab. Syst.* **1999**, *49*, 193–213.

(9) Morris, R. E.; Hammond, M. H.; Shaffer, R. E.; Gardner, W. P.; Rose-Pehrsson, S. L. *Energy Fuels* **2004**, *18*, 485–489.

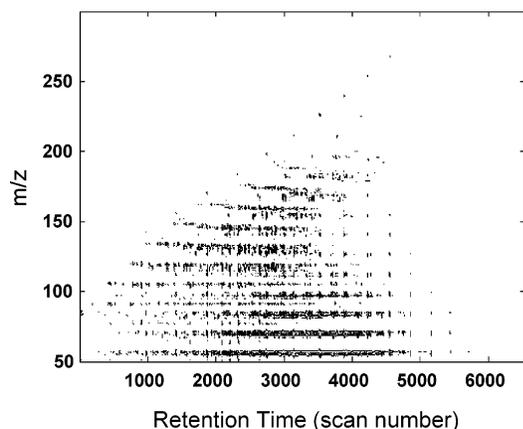


Figure 1. Typical gas chromatography–mass spectroscopy (GC–MS) chromatogram of an unstressed diesel fuel, presented as a topographical contour plot with only one contour line, drawn at a signal intensity just above the baseline. The *x*-axis is the retention time, in terms of MS scan number. The *y*-axis is the mass spectral axis from mass-to-charge (*m/z*) ratios of 50 to 300.

dataset constitutes an accurate numerical representation of the relevant chemical information, and the second is that the differences between the different classes of samples are statistically greater than the differences between samples within the same class. If these two criteria are met, then it should be possible to develop predictive and diagnostic models based on these numerical representations of fuel composition. Moreover, a mathematical discrimination of all the significant differences in fuel composition after stress could reveal subtle features that are highly correlated with fuel quality, which could have gone otherwise undetected.

The data resulting from a single GC–MS analysis of a typical unstressed diesel fuel can be depicted in a two-dimensional plot as shown in Figure 1, which is a topographical contour plot with a single contour line drawn at a value of signal intensity just above baseline noise. In this figure, the *x*-axis is retention time, in terms of mass spectral scan number, and the *y*-axis is the mass spectral axis and encompasses mass-to-charge (*m/z*) ratios of 50–300. Thus, each slice of this plot taken along the *x*-axis represents the mass spectrum recorded at a given retention time in the GC separation. A data set that results from multiple GC–MS analyses can thus be conceptualized as a series of two-dimensional GC–MS chromatograms, stacked one on top of another to form a cube. Therefore, in addition to dimensions that describe GC retention time and MS *m/z* ratios, such a data set also has a third dimension, reflecting the identity of the different samples that are subjected to GC–MS analysis.

An initial examination of the GC–MS data indicated that the chromatographic retention-time precision and signal reproducibility were relatively high. For example, the data collected in the study of LPR-stressed fuels exhibited a mean standard deviation in peak position of 2.7 scan numbers across 286 peaks in the 18 chromatograms. This deviation is equal to only roughly one-fourth of the typical peak width and, thus, manual alignment of the chromatographic profiles was readily achieved. Turning to signal reproducibility, an examination of replicate total ion current chromatograms yielded

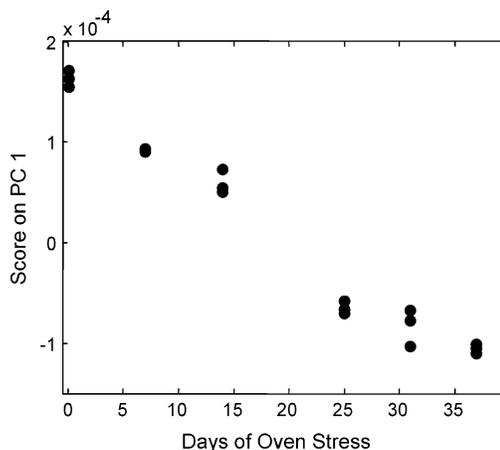


Figure 2. Multiway principal components analysis (MPCA) model of oven-stressed fuel, showing a clear progression in the scores on the first principal component as the duration of oven stress at 60 °C is increased.

a mean root-square deviation (RSD) of 9.6% across the entire chromatographic run for the raw data, and a mean RSD of 3.3% for data that had been area-normalized and aligned.

An MPCA model was constructed from data acquired in the analysis of the oven-stressed neat fuel samples. The scores on the first principal component of the MPCA model are shown in Figure 2. This plot indicates that there is a clearly defined progressive change in composition as the fuel was subjected to increasing durations of oven stress at 60 °C. The loadings vector (reshaped to matrix form) is associated with the first principal component of the MPCA model and represents the portions of the GC–MS data that are undergoing changes during oven stress. The loadings plot thus derived indicated that much of this change was associated with a decrease in the concentrations of early eluting (i.e., high-volatility) components, which is consistent with evaporative loss during stress, rather than chemical changes associated with fuel degradation. Models constructed from fuels spiked with copper demonstrated similar behavior.

The extent of the compositional changes associated with evaporative loss was further examined with the use of an ANOVA-based feature selection utility.⁷ ANOVA *F*-ratios were calculated by defining the replicates of each sample as a different class, thus probing for features that were different between different levels of oven stress but the same between replicate chromatograms of the same fuel sample. Thus, locations in the chromatogram that have a high *F*-ratio are good at describing the differences between fuel samples, whereas those with low *F*-ratios are not. The *F*-ratios from this analysis are shown in Figure 3 as a contour plot drawn with a single contour line at an *F*-ratio value just above the baseline noise. As the MPCA loadings plot indicated, the variation from sample to sample in this data set is dominated by decreasing concentrations of early eluting components, lending credence to the theory that it is evaporative loss rather than fuel degradation that is being modeled. To further illustrate this point, a feature-selected total ion current (TIC) chromatogram was constructed by including only data points with an ANOVA *F*-ratio of >80. This feature-selected TIC thus represents only those fuel components that had changed

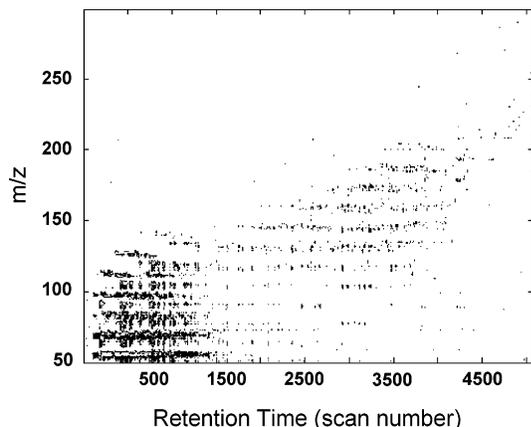


Figure 3. Analysis of variation (ANOVA) f -ratios for oven-stressed fuel, which provide a measure of how well each data point describes the differences as the oven stress increases. Clearly, as the oven-stress duration increased, the differences were associated with evaporative losses.

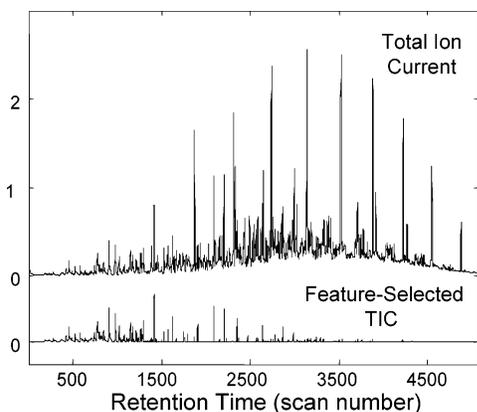


Figure 4. Feature-selected total ion content (TIC) of an unstressed fuel from the oven-stress study. The feature-selected TIC includes only those data points with an ANOVA f -ratio of >80 and shows that the major source of variation between fuels as they underwent oven stress was the evaporative loss of lighter components. Offset from the feature-selected TIC is the raw (i.e., non-feature-selected) TIC of the same fuel sample.

during the oven test. When compared with the TIC of the unstressed fuel (Figure 4), two things are evident: (i) the ANOVA f -ratio test is able to detect subtle changes in composition from GC-MS data, and (ii) the lighter components were being lost during oven testing.

To minimize evaporative losses, a second study was conducted in which neat and copper-doped fuel samples were thermally stressed in a LPR at 100 °C under 100 psi oxygen for 16 and 42 h. Figure 5 depicts ANOVA f -ratios for the LPR-stressed fuel data set. The f -ratio calculation was made by defining the classes of samples (e.g., unstressed; 16 h, LPR-stressed, neat; 42 h, stressed, neat; and 42 h, LPR-stressed with copper), thus providing a measure of how well each data point in the GC-MS chromatograms describes the difference between LPR-stressed and unstressed fuels. As shown in Figure 5, evaporative loss does not dominate the changes that are associated with LPR stress of this fuel, as they do in the first oven-stress study.

An MPCA analysis of the LPR-stressed fuel data set was conducted, and the scores on the first principal component are shown in Figure 6. As in Table 1, the first six samples are replicates of unstressed fuel, the

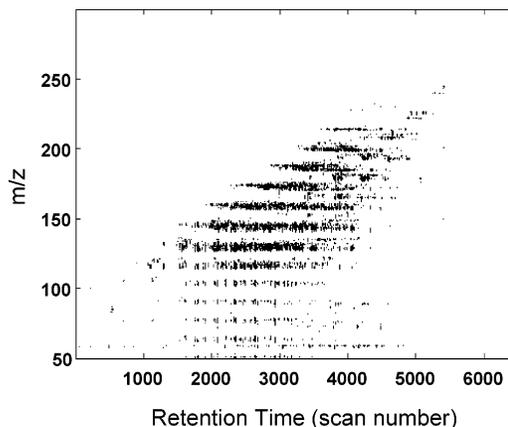


Figure 5. ANOVA f -ratios for changes in the composition of low-pressure reactor (LPR)-stressed fuels. In this case, evaporative losses do not dominate the changes associated with LPR stress, as they do with oven stress.

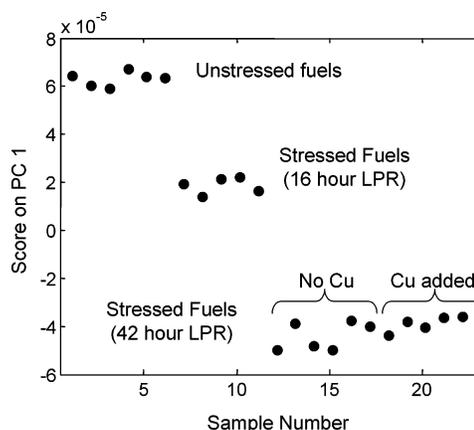


Figure 6. MPCA scores on the first principal component of LPR-stressed fuel, showing a clear delineation between stressed and unstressed fuels. It also indicates that no significant compositional differences were detected via GC-MS between fuel stressed with copper and fuel stressed without copper.

next five were 16 h LPR-stressed fuel, the next six were 42 h LPR-stressed fuel, and the final six were 42 h LPR-stressed fuel with added copper. From this plot, there is a clear differentiation between stressed and unstressed fuels (i.e., intrareplicate variation is much less than interreplicate variation), but essentially no difference is observed between the comparison of stressed fuels with and without copper. The loadings from this first principal component of the MPCA model agree very closely with the features defined by the ANOVA f -ratio calculation, and this indicates that the MPCA model is describing the difference between stressed and unstressed fuels without the need for feature selection prior to MPCA. A second ANOVA f -ratio calculation was performed, this time using only the 42 h LPR-stressed fuels and defining two classes: fuels stressed with copper and fuels stressed neat. This analysis failed to locate any chromatographic features that are capable of making this classification. This suggests that, even though copper can accelerate the rate of fuel autoxidation, in this instance, the presence of copper had little or no significant impact on the final composition of the LPR-stressed fuel.

The feature-selected TIC displayed in Figure 7 was computed using the ANOVA f -ratios shown in Figure 5

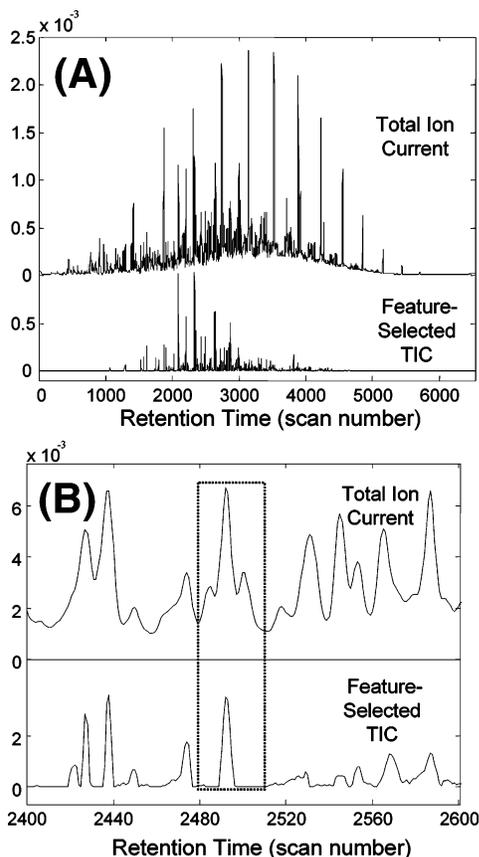


Figure 7. (A) Feature-selected TIC of unstressed fuel from the LPR stress study, including only data points with an ANOVA f -ratio of >100 . Offset from the feature-selected TIC is the raw (i.e., non-feature-selected) TIC of the same fuel sample. (B) Enlarged region of the feature-selected TIC.

to filter the chromatogram of an unstressed fuel, so that only the chromatographic features that differ significantly between the LPR-stressed and unstressed fuels are observed. The raw (i.e., non-feature-selected) TIC of the unstressed fuel is also shown, for comparison. In this instance, the feature selected TIC was calculated by including only data points with an ANOVA f -ratio of >100 . This threshold value was chosen empirically, based on a visual inspection of all of the calculated ANOVA f -ratios, which demonstrated that ANOVA f -ratios in the baseline areas of the chromatogram were all <100 . A more-rigorous optimization of this threshold, such as that performed by Johnson et al.,⁷ was not undertaken, because the threshold obtained by visual inspection performed adequately for the purposes of this study. Figure 7B shows an enlarged region of Figure 7A and demonstrates the degree to which fuel components that changed during LPR stress were intermingled with those that did not in the GC-MS chromatogram.

An attempt was made at spectral deconvolution of an ANOVA-selected feature from GC-MS data of LPR-stressed fuel. The boxed region in Figure 7B was subjected to PARAFAC decomposition with unimodality and non-negativity constraints on the GC mode, and non-negativity constraint on the mass spectral mode. A five-component PARAFAC model was then constructed and the chromatographic and sample-mode loadings are shown in Figure 8. As shown in Figure 8A, four discrete Gaussian peaks, as well as a fifth peak

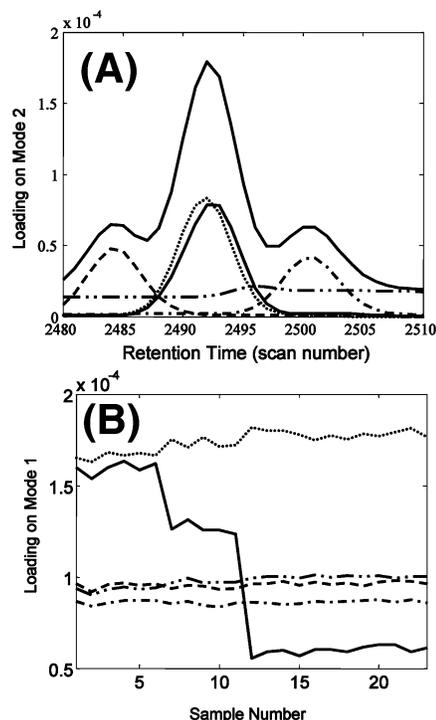


Figure 8. PARAFAC decomposition of the boxed region in Figure 7B. Unimodality and non-negativity constraints were imposed on the chromatographic mode, and a non-negativity constraint was imposed on the mass spectral mode ((A) resulting deconvoluted GC profiles, and (B) deconvoluted concentration profiles showing only one of the four discrete components decreasing in composition with stress).

that was somewhat less well-defined, were extracted from the region shown in Figure 7B. One of this set of four peaks occurs at the same retention time as the desired feature located by the ANOVA feature selection algorithm and is drawn with a solid line in Figure 8B.

Figure 8B shows the concentration profiles of these extracted components by sample number. As in Table 1, the first six samples are replicates of unstressed fuel, the next five were 16 h LPR-stressed fuel, the next six were 42 h LPR-stressed fuel, and the final six were 42 h LPR-stressed fuel with added copper. It can be observed that one extracted component in this region is decreasing in concentration with LPR stress whereas the others remain essentially the same. This component, which is depicted again with a solid line, corresponds with the chromatographic feature that has been located by ANOVA feature selection. The loadings in the mass spectral mode for this component of the PARAFAC model are shown in Figure 9, with major peaks at $m/z = 91, 115, 117, 131, 145,$ and 160 . Utilizing an in-house-built MATLAB program, these extracted MS data were submitted to a mass spectral library search for tentative identification. This was accomplished by writing the data to an appropriately formatted text file, which was then sent to the NIST Mass Spectral Search Program for the NIST/EPA/NIH Mass Spectral Library (version 2.0), which then returned automated library search results for the submitted data. A molecular structure assignment (5-ethyl-1,2,3,4-tetrahydronaphthalene) was obtained via the NIST mass spectral library for the PARAFAC model component in question. This illustrates the potential diagnostic capability of this ap-

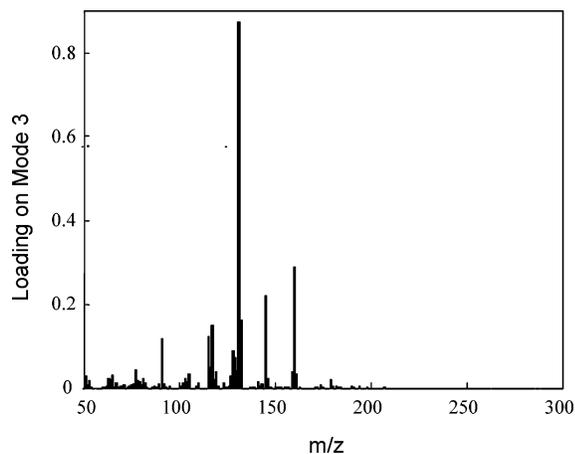


Figure 9. Mass spectral loadings of the parallel factor analysis (PARAFAC) model component that decreased in concentration during LPR stress. A tentative identification of this compound as 5-ethyl-1,2,3,4-tetrahydronaphthalene was made via a NIST Mass Spectral Library match.

proach to detect and characterize those trace fuel constituents that have changed during use or thermal stress.

We are currently focusing our efforts to optimize the data collection and analysis methods and to resolve other issues, including detection limits and model sensitivity to instrumental variations.

Conclusions

When using multidimensional chemometric techniques to perform discriminant analyses of complex fuel composition datasets, it is imperative that measures be taken to avoid unintentional weighting of results. Thus, we must be very circumspect in regard to how we define what constitutes a statistically significant change in composition, while avoiding interpretations that may be skewed by preconceived ideas of which chemical processes are relevant. Another aspect of this methodology that is critical is the establishment of the ability of the numerical data to accurately represent the magnitude of change, i.e., the linearity of instrumental response to compositional change.

The analysis of variation (ANOVA) has been shown to be useful as a means of extracting the significant information from complex compositional data. Analysis of the oven-stressed diesel fuel samples clearly showed that the chemical variations were dominated by evaporative loss of the more-volatile fuel components, rather than those that are lost or gained due to fuel degradation. Utilization of low-pressure reactor (LPR) stress conditions minimized this effect and allowed for a better

examination of changes in fuel composition due to degradation.

Fuel degradation in samples subjected to LPR stress under the conditions described was readily observed and modeled as changes in fuel composition, as monitored by gas chromatography–mass spectroscopy (GC–MS) analysis of fuel samples. The ANOVA-based feature selection was able to locate the features that changed from sample to sample, thus allowing for a quick evaluation of how fuel composition was altered during stress, as well as aiding in the construction of hypotheses about the mechanisms of this change. In this experiment, it seems that this diesel fuel stressed with and without high levels of copper approached the same chemical composition “end point”, although the mechanism and rate of change may have been different. This approach could be used with a variety of fuels, to determine if catalytic copper serves as an autoxidation accelerant without imposing significant mechanistic changes on the autoxidation mechanism. Such a finding would imply that copper could be used to accelerate fuel stability testing in the laboratory.

Identification of several individual fuel constituents that change significantly during LPR stress may be possible through parallel factor analysis (PARAFAC) decomposition of local regions of GC–MS data that have been identified as being significant by ANOVA-based feature selection. The potential of this approach is as a diagnostic tool, as well as a means of more completely understanding the complex processes that occur as a fuel degrades. This type of analysis could eventually provide a means to characterize compositional changes in fuels that are associated with degradation to an unprecedented level of detail. This level of detail and understanding of the fuel degradation process is necessary to develop reliable and robust models to predict fuel quality that are functional for more than one type of fuel. Moreover, by incorporating structure assignment functionality along with localized PARAFAC modeling and appropriate data preprocessing, it would theoretically be possible to provide a means of automating the process of analyzing the GC–MS datafiles to provide rapid compositional profiles for evaluating fuel samples.

Acknowledgment. The authors gratefully acknowledge the Office of Naval Research for funding support of this study and Thomas Evans of Geo-Centers, Inc., for conducting the GC-MS fuel analyses. Dr. Kevin Johnson is a National Research Council–Naval Research Laboratory Postdoctoral Research Associate.

EF030161Z