

©Copyright 2014
Benjamin Kehimkar

Fundamental studies of Rocket propellant fuel using GC \times GC – TOFMS instrumentation with
chemometric data analysis.

Benjamin Kehimkar

A dissertation submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2014

Reading Committee:

Robert E. Synovec, Chair

František Tureček

Xiaosong Li

Program Authorized to Offer Degree:

Chemistry

University of Washington

Abstract

Fundamental studies of Rocket propellant fuel using GC \times GC – TOFMS instrumentation with chemometric data analysis.

Benjamin Kehimkar

Chair of the Supervisory Committee:

Professor Robert E. Synovec

Department of Chemistry

For most of the research presented herein, samples of complex composition, specifically rocket propellant (RP) fuels, were analyzed. The performance of a fuel is directly linked to composition; understanding the connection between composition and physical properties will lead to better control of said properties and overall performance of the fuels. Samples of RP

fuels were analyzed using comprehensive two dimensional gas chromatograph coupled to a time of flight mass spectrometer (GC \times GC – TOFMS) using a reverse column configuration; a long primary column (polar) followed by a short (non polar) secondary column. This instrumentation setup was used to obtain multidimensional chromatograms and achieve great temporal separation of many compounds (including various isomers) listed herein as alkanes, cycloalkanes and aromatics which can be found in RP fuels. GC \times GC – TOFMS chromatograms are extremely information rich, making the manual process of interpretation unwieldy; therefore various algorithms are used in the analysis, many of which are known as chemometrics. Chemometrics are a specific class of analysis tools that use linear algebra to analyze data to achieve various goals. There are numerous chemometric techniques; some even have variations, therefore choosing the appropriate technique for a given data set is important. Principal component analysis (PCA) was used to analyze the sources of compositional variance between samples. PCA is extremely sensitive to variance between samples and is used to glean the relevant information and summarize the variance as succinct data matrices (scores and loadings). Partial least squares (PLS) is used to draw connections between two separate sets of measurements that are somehow related to each other; such as the chemical information found in the chromatograms and the measured physical properties such as density or kinematic viscosity. Often the intent of using PLS is to approximate values of interest using other related measurements for various reasons (cost, practicality, convenience). Of course to have significant confidence in the PLS models, they must undergo validation, to ensure these predicted values are representative/reasonable. Finally, parallel factor analysis (PARAFAC) is used to deconvolute (mathematically resolve) signals associated with different analytes that were poorly separated in the chromatogram. Successful deconvolution of analytes allows for better identification of

analytes when searching for a match in a mass spectral library. Analysis of data, i.e. the interpretation of data, obtained from an instrument is an inseparable part of research. An algorithm of analysis and detection of a specific list of compounds that may potentially be found in a sample, presented herein, is temporal mass spectral ratio analysis method (TMSRA) which was developed for analyzing peaks in GC-MS chromatograms and identifies pure mass channels (m/z); this was achieved via analysis of the ratios between m/z . The application of these analysis tools and interpretation of their respective results are discussed.

TABLE OF CONTENTS

	Page
List of Figures	v
List of Tables	vii
Acknowledgements	viii
Dedication	ix
Chapter 1: Introduction	1
1.1 The Instruments	1
1.1.1 The Gas Chromatograph	1
1.1.2 The Quadruple Mass Spectrometer	3
1.1.3 The Time of Flight Mass Spectrometer	4
1.2 The Chromatograms, the separation mechanism, and the columns	5
1.2.1 The GC-MS Chromatogram	5
1.2.2 The GC \times GC – TOFMS Chromatogram	6
1.2.3 The Separation Mechanism	7
1.2.4 The Reverse Column Configuration	8
1.3 Introduction to Chemometrics	10
1.3.1 Preprocessing	10
1.3.2 Principal Component Analysis	13
1.3.3 Partial Least Squares Regression	15
1.3.4 Parallel Factor Analysis	18

1.3.5	Targeted Techniques Overview	20
1.3.6	Non-targeted Techniques Overview	25
1.4	Conclusions.....	26
Chapter 2: Temporal Mass Spectral Ratio Analysis Method (TMSRA)		34
2.1	Introduction.....	34
2.2	Theory	38
2.3	Experimental	43
2.4	Results and Discussion	46
2.4.1	PFTBA Study.....	46
2.4.2	Case Study 1: 1-Heptene and Heptane.....	48
2.4.3	Case Study 2: Octane and 1,4-Dioxane	50
2.4.4	Study Case 3: 1,4-Dioxane and Octane	51
2.4.5	TMSRA Computation Time	52
2.5	Conclusions.....	53
Chapter 3: Analysis of RP-1 Fuel Content and Physical Properties using PLS Regression.....		69
3.1	Introduction.....	69
3.2	Experimental	73
3.3	Results and Discussion	77
3.3.1	GC × GC Separation of RP-1 Using Reversed Column Format.....	77
3.3.2	Chemical Composition Studies Using PLS	80
3.3.3	Physical Properties Studies Using PLS.....	82
3.4	Conclusions.....	86
3.5	Supplemental Material	100

Chapter 4: Analysis of RP-2 Fuel Using PCA.....	119
4.1 Introduction.....	119
4.2 Experimental.....	120
4.3 Results and Discussion	120
4.3.1 RP-2 Chromatograms.....	120
4.3.2 PCA Results.....	121
4.4 Conclusions.....	123
Chapter 5: Analysis of RP-1 Fuel and Advanced Distillation Curves using PLS Regression	133
5.1 Introduction.....	133
5.2 Experimental.....	137
5.3 Results and Discussion	141
5.4 Conclusions.....	146
Chapter 6: Analysis of RP-1 Fuel and Physical Properties Over a Temperature Range	154
6.1 Introduction.....	154
6.2 Experimental.....	156
6.3 Results and Discussion	157
6.3.1 Modeling Density as a Function of Temperature	157
6.3.2 Modeling Speed of Sound as a Function of Temperature.....	158
6.3.3 Kinematic Viscosity as a Function of Temperature.....	159
6.3.4 Dynamic Viscosity as a Function of Temperature.....	160
6.3.5 Adiabatic Compressibility as a Function of Temperature	161
6.4 Conclusions.....	162
Chapter 7: Conclusions and Future of Data Analysis.....	172

7.1	Conclusions.....	172
7.2	Future directions	173
Appendix	Further Investigation Into the Sulfur Content of RP-1 Fuels	175
	Bibliography	184
	Curriculum Vitae	200

LIST OF FIGURES

Figure Number	Page
1.1	The GC instrument, the GC × GC instrument, and respective chromatograms27
1.2	The quadrupole MS, the TOFMS, and a mass spectrum of adamantane28
1.3	GC-MS chromatogram, and the multidimensional data cube.....29
1.4	2D TIC chromatograms of fuels, normal vs. reverse column30
1.5	Principal component analysis (PCA).....31
1.6	Partial least squares (PLS)32
1.7	Parallel factor analysis (PARAFAC).....33
2.1	Processes used in construction of the Connectivity matrices54
2.2	PFTBA spectra.....55
2.3	Temporal mass spectral ratio analysis (TMSRA) results for PFTBA study.....56
2.4	Statistical results (of TMSRA) for PFTBA study.....57
2.5	Case Study 1: 1-heptene and heptanes simulated peak overlap.....58
2.6	TMSRA results for Case Study 1.....59
2.7	Quantification results for Case Study 160
2.8	Case Study 2: octane and 1,4-dioxane simulated peak overlap61
2.9	TMSRA results for Case Study 2.....62
2.10	Quantification results for Case Study 263
2.11	Case Study 3: 1,4-dioxane and octane peak overlap and TMSRA results.....64
2.12	Quantification results for Case Study 365
3.1	RP-1 GC × GC – TOFMS TIC chromatograms89
3.2	LOOCV results for PLS models of compositional properties90
3.3	Results for PLS model for density.....91
3.4	Results for PLS model for kinematic viscosity.....92
3.5	Results for PLS model for net heat of combustion (MJ/l).....93
3.6	Scree plots of highlighted PLS models.....105
3.7	LRVs of highlighted PLS models.....106
3.8	Results for PLS model for iso-alkanes107
3.9	Results for PLS model for tri-cycloalkanes.....108
3.10	Results for PLS model for aromatics.....109
3.11	Results for PLS model for hydrogen content.....110
3.12	Results for PLS model for net heat of combustion (MJ/kg)111
3.13	Results for PLS model for sulfur content112
3.14	Results for PLS model for sustained boil113
3.15	Results for PLS model for vapor rise.....114
4.1	RP-2 GC × GC–TOFMS chromatograms125
4.2	PCA scores plots using the averaged RP-2 chromatograms.....126
4.3	PCA (unfolded) loadings plots (TIC) using the averaged chromatograms (PCs 1 & 2).....127

4.4	PCA (unfolded) loadings plots (TIC) using the averaged chromatograms (PCs 3 & 4)	128
4.5	PCA scores plots using all chromatograms replicates	129
5.1	TIC and select ion chromatograms of RP-1 fuel LB073009-08	148
5.2	ADC data and PLS results	149
5.3	Select LRVs or PLS models for ADC data	150
5.4	LOOCV results and scree of select ADC % distillation	151
6.1	PLS 5LV model results for density as a function of temperature	163
6.2	Select LRVs of PLS (5LV) models for density	163
6.3	PLS 5LV model results for speed of sound as a function of temperature	164
6.4	Select LRVs of PLS (5LV) models for speed of sound	164
6.5	PLS 5LV model results for kinematic viscosity as a function of temperature	165
6.6	Select LRVs of PLS (5LV) models for kinematic viscosity	165
6.7	PLS 5LV model results for dynamic viscosity as a function of temperature	166
6.8	Select LRVs of PLS (5LV) models for dynamic viscosity	166
6.9	PLS 5LV model results for adiabatic compressibility as a function of temperature	167
6.10	Select LRVs of PLS (5LV) models for adiabatic compressibility	167
A.1	Positive portion of the LRV for the PLS model for sulfur content	179
A.2	Noted subregions used in creating PLS models (large, medium, small)	180
A.3	LRV from 5LV PLS model using the 'Large' subregion	181
A.4	LRVs from PLS models using the 'Medium' subregion	181
A.5	LRVs from PLS models using the 'Small' subregion	182

LIST OF TABLES

Table Number	Page
2.1	TMSRA ratio matrices66
2.2	The index key for PFTBA containing <i>m/z</i> of intensity above the minimum threshold66
2.3	The index key for sorted mass channels of 1-heptene67
2.4	The index key for sorted mass channels of octane67
2.5	The index key for sorted mass channels of 1,4-dioxane68
3.1	RP-1 Fuel Set94
3.2	Identified compounds (template data)95
3.3	Summary of PLS model metrics for measured properties analyzed97
3.4	Contributing compounds identified in LRVs for density (positive)98
3.5	Contributing compounds identified in LRVs for density (negative)99
3.6	Contributing compounds identified in LRVs for kinematic viscosity (positive)115
3.7	Contributing compounds identified in LRVs for kinematic viscosity (negative)116
3.8	Contributing compounds identified in LRVs for net heat of combustion (MJ/l) (positive)117
3.9	Contributing compounds identified in LRVs for net heat of combustion (MJ/l) (negative)117
3.10	Contributing compounds identified in LRVs for hydrogen content (positive)118
3.11	Contributing compounds identified in LRVs for hydrogen content (negative)118
4.1	RP-2 Fuel Set130
4.2	Compounds of interest, based on the PCA loadings of RP-2Fuels, PC1 positive values130
4.3	Compounds of interest, based on the PCA loadings of RP-2Fuels, PC1 negative values131
4.4	Compounds of interest, based on the PCA loadings of RP-2Fuels, PC2 positive values131
4.5	Compounds of interest, based on the PCA loadings of RP-2Fuels, PC2 negative values132
5.1	Contributing compounds identified in the LRVs that contribute positively152
5.2	Contributing compounds identified in the LRVs that contribute negatively152
5.3	Compounds of interest identified in the LRVs that exhibit a sign change across the ADC153
6.1	Contributing compounds identified in the speed of sound LRV that contribute positively169
6.2	Contributing compounds identified in the speed of sound LRV that contribute negatively169
6.3	Contributing compounds identified in the dynamic viscosity LRV that contribute positively170
6.4	Contributing compounds identified in the dynamic viscosity LRV that contribute negatively170
6.5	Compounds of interest identified in adiabatic compressibility LRV (contribute positively)171
6.6	Compounds of interest identified in adiabatic compressibility LRV (contribute negatively)171
A.1	Identified sulfur containing compounds183
A.2	Results for the PLS models for sulfur content using smaller subregions183

ACKNOWLEDGEMENTS

Thanks to my advisor, Rob Synovec, for both the mentorship and the opportunity of being in his research group. Thanks to all members of Synovec group who helped make the experience in graduate school a pleasant one.

My thanks to my former advisor, Norm J. Dovichi, though my time in Dovichi labs was limited, it was still an invaluable experience in my first two years in graduate school.

My thanks to the Air Force Research Laboratory (AFRL) for both providing the funding and the opportunity to analyze rocket propellant blends, for my research.

Special thanks to Jamin C. Hoggard, for all the invaluable feedback, advice. Also for teaching me all those useful computer and programming skills that came in handy more than one occasion.

My thanks to my parents, Elisha and Jessica Kehimkar, for your love and support throughout the years, and giving me the best education you could afford.

A special thanks to my beloved wife, Galit Kehimkar. For marrying me when I just started graduate school, and adamantly standing beside me through both the brightest and darkest of times.

DEDICATION

For my parents, my grandparents,
and my beloved wife, Galit.

Thank you for your support and for believing in me.

Chapter 1: Introduction

1.1 The Instruments

1.1.1 The Gas Chromatograph

The Gas Chromatograph (GC) is one of the most widely used instruments in separation science. A diagram of a GC can be seen in Figure 1.1A. In short, separation of analytes using chromatography is based on the interactions of analytes with the stationary phase relative to their respective interaction with the mobile phase. In the discussions presented the stationary phases used were fused silica capillary columns. The autoinjector introduces the sample (liquid in these cases) into the heated inlet, where it is vaporized into the gas phase (nitrogen gas). The sample is loaded onto the column located inside an oven; the separation of analytes is based on their respective times spent in the stationary phase relative to the mobile phase. At the end of the column, the eluted analytes are introduced to the detector. The detector could be single channel such as a flame ionization detector (FID) or multi-channel such as a mass spectrometer (MS) as used in the research presented herein^{1,2}.

For comprehensive two dimensional separation ($GC \times GC$), two columns are coupled in tandem (see Figure 1.1B). The second column is significantly shorter of the two in order to minimize band broadening which would result over time. Also to help minimize band broadening, small concentrated ‘plugs’ or modulations are transferred from the first column to

(1) J.C. Giddings, *Unified Separation Science*, John Wiley & Sons, New York, 1991.

(2) G.D. Christian, *Analytical Chemistry*, 6th ed., John Wiley & Sons, New York, 2004.

the second column using a valve system or a thermal modulator³. Ideally all analytes in a modulation would have reached the detector, before the next modulation is introduced.

Occasionally some analytes are retained longer than the set modulation period and appear in a subsequent modulation: this undesirable occurrence is called wrap around. Wrap around may pose a problem if wrapped around peaks overlap with other analytes. Having two columns with different separation mechanisms must be used in order to maximize the separation power between unresolved compounds from the first column. A secondary oven set at a higher temperature for the second column may also be used to help achieve this goal^{4,5}.

The recorded data, a chromatogram, has the recorded signal over the duration of the separation. The x-axis corresponds to time (usually in units of minutes or seconds) while the y-axis corresponds to response or signal intensity. In the case of GC × GC chromatograms the x-axis corresponds to time on the first column separation, also known as the primary column (displayed in units of minutes or seconds) and the y-axis is corresponds to time on the second column separation, also known as the secondary column (displayed in units of seconds). The intensity in GC × GC chromatograms is expressed either via the z-axis, the peak color intensity, or both. Examples of these chromatograms can be seen in Figures 1.1C and 1.1D, respectively.

(3) Z. Liu, M. Zhang, J.B. Philips, *J. Chromatogr. Sci.* 28 (1990) 567-571.

(4) R.E. Mohler, B.P. Tu, M. Dombek, J.C. Hoggard, E.T. Young, R.E. Synovec, *J Chromatogr. A.* 1186 (2008) 401-411.

(5) B.D. Fitz, R.B. Wilson, B.A. Parsons, J.C. Hoggard, R.E. Synovec, *J. Chromatogr A.* 1266 (2012) 116-123.

1.1.2 The Quadrupole Mass Spectrometer

The basic concept behind MS is analyte(s) are ionized and (reproducibly) fragment into smaller ions, and the ions undergo a sorting process as they travel to the detector. Often the detector is an electron multiplier where the relative abundance is expressed by signal intensity. The sorting of the ions themselves is based on their mass to charge ratio (m/z). Though ionization could be either positive or negative, in all research projects herein, the ion source was set to produce positively charged ions. The fragmentation process can be achieved either by soft or hard ionization; for all research projects herein, hard ionization was used. Hard ionization entails vaporizing the analyte off the filament and its charge causes the analyte to break into many ionized fragments. The ions are then focused into the mass analyzer.

In the case of a quadrupole MS, the mass analyzer is a hyperbolic quadrupole chamber containing two pairs of metal rods. A diagram of a quadrupole MS can be seen in Figure 1.2A. The ions are separated by applying superposed direct current (DC) fields and an oscillating radio frequency field on the metal rods. The signs change in the electric potential causes ions to resonate between the poles as they travel to the detector. The trajectory of the ions depends on both the ions m/z value and the RF field. Therefore by altering the RF frequency, the voltage and the 'zero-to-peak' amplitude, one can cause undesired ions to collide with the quadrupoles while allowing desired ions to traverse through the quadrupole chamber and reach the detector. The two main modes a quadrupole can run are select ion monitoring (SIM) and scan mode. In the research projects herein, when used, the quadrupole MS was set to scan mode. In scan mode the quadrupole RF field parameters cycle allowing ions of a wide range of m/z to reach the detector over a short period of time; however as the range of m/z widens so does the time to scan through

them. To maximize the number of ions reaching the detector, the interior is generally held under a vacuum^{2,6,7}.

1.1.3 The Time of Flight Mass Spectrometer

The time of flight mass spectrometer (TOFMS), differs from the quadrupole MS by the means of which the ions are sorted as they reach the detector. For the TOFMS, the separation of ions is also based on their m/z but in a manner analogous to chromatography. A diagram of a TOFMS instrument can be seen in Figure 1.2B. The ions are initially energized with approximately an equal amount of energy received from an electric field and are introduced in pulses to a field free region known as the flight tube where they are allowed to ‘drift’ and resolve from one another: Since the velocities of ions are determined by their masses, the heavier ions travel slower while the lighter ions travel faster. To help minimize the variance in velocities between ions with the same m/z , a reflectron (or an ion mirror) is often used. With respect to ions of the same m/z , an ion with a higher velocity will penetrate the reflectron deeper than an ion with a lower velocity. A reflectron consists of a series of disc shaped electrodes that produce an electric field which slows incoming ions and reverses their direction. The reflectron is often tilted at a slight angle to direct the ions at the detector located adjacent to the ion source^{2,7}.

The recorded data is called a mass spectrum, often viewed as a bar graph with the x-axis corresponding to the m/z values and the y-axis corresponding to the abundance. An example of a

(6) F.W. McLafferty, F. Tureček, Interpretation of Mass Spectra, 4th ed., University Science Books, Sausalito Ca USA, 1993.

(7) E. de Hoffmann, V. Stroobant, Mass Spectrometry: Principles and Applications, 3rd ed., John Wiley & Sons, New York, 2007.

mass spectrum can be seen in Figure 1.2C. Note the measurement is relative; to associate ion abundance with respect to absolute values such as concentration in parts per million (ppm), a set of standard concentrations must be measured for a calibration curve.

1.2 The Chromatograms, the separation mechanism, and the columns

1.2.1 The GC-MS Chromatogram

The chromatograms collected from a GC instrument coupled to an MS is known as a GC-MS. The recorded chromatogram is a two dimensional matrix, with both a chromatographic and a mass spectral dimension respectively. Note that in all research presented herein, a quadrupole MS was used in a GC-MS instrument. The data can be seen as either a collection of mass spectra collected at each and every data point on the chromatogram or as a collection of chromatograms recorded for each m/z in the mass spectrum; an illustration is shown in Figure 1.3A. Regardless of the perspective on the GC-MS data, the signal is simultaneously a function of the retention time of analytes in the chromatographic dimension and the abundance of ions of various m/z in the mass spectral dimension; hence the GC-MS chromatogram is two-way data. In ideal or close to ideal conditions, the chromatographic data is said to be bilinear (i.e. the separation of analytes is linearly correlated to their respective interaction with the stationary phase and the signal in the mass spectra is linearly proportional to the abundance of ions reaching the detector). Alternatively, one can say a GC-MS chromatogram is bilinear if the analyte peaks can be accurately modeled as the outer product of a concentration profile and a

mass spectrum. When multiple chromatograms are unified into a three dimensional matrix, the format presented herein is $\text{Samples} \times \text{GC} \times \text{MS}$ ⁸.

1.2.2 The $\text{GC} \times \text{GC} - \text{TOFMS}$ Chromatogram

A $\text{GC} \times \text{GC}$ instrument coupled to a TOFMS is known as a $\text{GC} \times \text{GC} - \text{TOFMS}$, and a chromatogram collected from a $\text{GC} \times \text{GC} - \text{TOFMS}$ is a three dimensional matrix with two chromatographic dimensions and one mass spectral dimension. In the research shown, $\text{GC} \times \text{GC} - \text{TOFMS}$ chromatogram dimensions are formatted as $\text{Column2} \times \text{Column1} \times \text{MS}$ and is known as a ‘data cube’⁴ (shown in Figure 1.3B). Being a function of both the separations in the first and second column dimensions and the relative abundance of ions for each m/z channel, a single $\text{GC} \times \text{GC} - \text{TOFMS}$ chromatogram is an example of three-way data. Analogous to GC-MS, the $\text{GC} \times \text{GC} - \text{TOFMS}$ chromatogram under ideal or near ideal conditions is said to be trilinear. The trilinearity of the data can be observed when the retention times of an analyte captured in multiple modulations are the same. Alternatively, one can say a $\text{GC} \times \text{GC} - \text{TOFMS}$ chromatogram is trilinear if the analytes can be accurately modeled as the outer product of two concentration profiles (one for each column dimension) and a mass spectrum. Deviations from trilinearity could be caused by instrumentation issues such as drift. A set of $\text{GC} \times \text{GC} - \text{TOFMS}$ chromatograms is formatted as $\text{Samples} \times \text{Column2} \times \text{Column1} \times \text{MS}$ and can be considered as four-way data (i.e. as a function of the samples as well).

(8) S. Yang, J.S. Nadeau, E.M. Humston-Fulmer, J.C. Hoggard, M.E. Lidstrom, R.E. Synovec, J. Chromatogr. A. 1240 (2012) 156–164.

1.2.3 The separation mechanism

If one were to observe the separation on the macroscale, the migrations of analytes through a column would appear as smooth and continuous bands that travel at different velocities, on the nanoscale however, the individual molecules travel in a more erratic motion. The velocity of analytes is dictated by the temperature, the flow velocity of the mobile phase, and the analyte interaction with the stationary phase (sorption/desorption). Different stationary phases will produce different results with respect to the separation of analytes. Due to variability in retention times (t_R) between individual molecules of the same analyte, the resulting peak is often Gaussian or a Gaussian-like shape. A few notable issues one may encounter that may affect the separation negatively include column overloading and longitudinal diffusion. Column overloading occurs when the concentration of the analyte is too high and molecules interfere or compete with one another for active sites to interact with the stationary phase, forcing some of the molecules to move further down the column where they may encounter vacant sites in the stationary phase. The result of this is undesired band broadening, in addition the signal at the detector could be overloaded due to too high of a concentration, in which case the solution is simply to inject the sample in smaller concentrations (either by dilution or split injection). Longitudinal diffusion, i.e. bands widening over time, can be minimized by performing the separation over shorter time intervals^{1,2}. The separation performance is often expressed in terms of the Van Deemter equation:

$$H = B/v + C_m v + C_s v \quad (1.1)^1$$

where H is the plate height (the zone where the molecules of an analyte establish an equilibrium between the mobile and stationary phase), B is the longitudinal diffusion, v is the mobile phase

flow velocity, C_m is the concentration of analytes in the mobile phase and C_s is the concentration of analytes in the stationary phase. Calculating plate height can be used to evaluate the overall efficiency of the separation in a system by calculating the number of theoretical plates (N):

$$N = X/H \quad (1.2)^1$$

where X is column length. If one considers H a function of v , then N could be maximized for a system when H is at a minimum. In practice however separations are run at a higher flow velocity for the purpose of increasing the sample throughput¹.

1.2.4 The Reverse Column Configuration

Traditional GC \times GC instruments often employ what is considered a ‘normal column configuration’. Typically a long non-polar column is followed by a short polar column; this is also known as ‘orthogonal separation’. It is called orthogonal since the separation on the first column is based on volatility while separation on the second column is based on both polarity and volatility, i.e. a new separation mechanism is introduced (in addition) in the second column. Normal column configuration has been used in the separation of hydrocarbon containing samples. An example of a GC \times GC – TOFMS chromatogram using a normal column configuration (20 m RTX-5 column followed by a 2 m RTX 200 column) can be seen in Figure 1.4A where a jet fuel was analyzed. The normal column configuration offers superior separation of aromatics, as they are spread across a relatively huge GC \times GC – TOFMS chromatographic region. In the second column dimension, alkanes and cycloalkanes co-elute first while aromatics elute later. The GC \times GC – TOFMS using a normal column configuration is able to separate more peaks when compared to GC-MS separation, an analogue is shown in Figure 1.4B. In a

GC-MS chromatogram many peaks would overlap and appear as a huge amount of signal, especially when viewing the total ion current (TIC) chromatogram, i.e the summation of the signal of all m/z channels. If the second column dimension chosen offered no real benefit in the separation, the analyte peaks would elute on the diagonal from the lower left corner (the start of the separation) towards the upper right side (the end of the separation).

In contrast to normal column configuration, the reverse column configuration uses a long polar column followed by a short non-polar column. It is also known as ‘non-orthogonal separation’ because the separation in the first column dimension is based on both volatility and polarity while the separation in the second column dimension is based only on volatility. As shown in Figure 1.4C (using a 30 m RTX-wax column followed by a 1.2 m RTX-1 column), the order of separation of a alkane rich rocket propellant (RP) fuel sample is reversed on the second column dimension; aromatics elute first, cycloalkanes elute second, and alkanes elute last. The separation between classes of compounds is considered better as alkanes are more visually separated from the cycloalkanes, while the cycloalkanes themselves show somewhat of a separation from di- & tri-cycloalkanes. The separation of alkanes and cycloalkanes is vastly improved at the expense of the separation of the aromatic compounds. This compromise is considered acceptable in the case of the RP fuels as the aromatic concentrations are relatively low and most mono-aromatics and di-aromatics are visually well separated from each other. Moreover, the reverse phase column configuration offers an additional advantage over normal column configuration with respect to the separation of isomers. Since the separation in the first column would be affected by polarity, isomers would elute at slightly different times. The later eluting isomers would separate earlier in the second column dimension due to an increase in temperature over time caused by the applied temperature program hence the apparent diagonal

separation of isomers from the upper left to the lower right⁹⁻¹¹. Therefore, the separation of analytes using a reverse column configuration appears to be well suited for the study of RP fuels.

1.3 Introduction to Chemometrics

1.3.1 Preprocessing

Prior to the use of chemometrics, the raw data is often subjected to one or several preprocessing steps to remove undesirable sources of variance in order to minimize the error introduced in the analysis. The origins of undesired variance can include but not limited to sample preparation and instrumentation. This crucial step can more than often mean the difference between a successful and an unsuccessful application of chemometrics: poor preprocessing leads to poor processing (models)¹². Preprocessing includes but is not limited to the following practices:

Deskewing

Deskewing, is a preprocessing step often used in chromatography where a quadrupole MS is used. Since the sample is being forwarded to the detector at a constant stream and due to the nature of the scan mode of quadrupole MS, the ion concentrations can change significantly

(9) B. Omais, M. Courtiade, N. Charon, D. Thiébaud, A. Quignard, M.-C. Hennion, J. Chromatogr. A. 1218 (2011) 3233–3240.

(10) C. Vendevre, R. Ruiz-Guerrero, F. Bertoncini, L. Duval, D. Thiébaud, M.-C. Hennion, J. Chromatogr. A. 1086 (2005) 21–28.

(11) M. Adahchour, J. Beens, R.J.J. Vreuls, A.M. Batenburg, U.A.T. Brinkman, J. Chromatogr. A. 1054 (2004) 47–55.

(12) K.R. Beebe, R.J. Pell, M.B. Seasholtz, Chemometrics A practical guide, John Wiley & Sons, Inc., New York, 1998.

from the beginning to the end of a single mass spectral scan, skewing the signal of the mass spectra. The solution often used to solve the skewing problem is an interpolation of the intensities to approximate the true ion concentrations¹³.

Noise Reduction

With every measurement regardless of instrument, it is assumed the recorded signal will include an amount of high frequency fluctuations in the signal, known as noise. To enhance the signal to noise ratio (S/N) of data with low S/N, smoothing can be applied, such as the Savitzky-Golay smoothing algorithm^{12,14}.

Baseline Correction

While smoothing removes the high frequency noise, baseline correction can rectify less than ideal baselines, including drift, bias, and low frequency fluctuations¹².

Normalization

To minimize the run-to-run variation in signal attributed to the differences in the amounts injected by the autoinjector, normalizing to an internal standard can be applied. This is done by dividing all the variables by a value (maximum signal or area of a peak of known concentration) designated as the standard. Alternatively, one may choose to normalize to the total summed signal. Normalizing can give better quantification results and lower standard deviation, though the effects of normalization should be inspected¹².

Peak Alignment

Though GC is a robust instrument with high reproducibility, run-to-run variation in retention times of peaks can still occur. Peak misalignment can be detrimental to the analysis,

(13) W.G. Pool, J.W. de Leeuw, B. van de Graaf, J. Mass Spectrom. 31 (1996) 213–215.

(14) A. Savitzky, M.J.E. Golay, Anal. Chem. 36 (1964) 1627–1639.

especially for techniques that are highly sensitive to variance. There are several algorithms made for peak alignment, an example of which is the piecewise correlation optimized warping (COW) algorithms^{12,15-17} which shift the peaks found in samples closer to their respective peaks found in a designated target sample.

Data reduction (Binning)

When chromatograms are huge and analyzing them using chemometrics would be too time consuming or impossible due to memory constraints, binning is often used to mitigate the problem. Consider a 1,000×1,000,000 data matrix, by binning the first dimension by every two data points would result in a 500×1,000,000 data matrix. If the 500×1,000,000 matrix were then binned in the second dimension by every two data points a 500×500,000 data matrix would be obtained, which is 25% the size of the original data matrix. Binning data also provides the benefits of retaining the values of area or volume underneath the peaks (for GC-MS and GC × GC – TOFMS, respectively) and correcting for slight misalignment of peaks at the cost of chromatographic resolution. It is a good practice to inspect the binned chromatogram to ensure that the loss of resolution is acceptable¹⁸.

Mean Center and Autoscale

For chemometric methods that are scale dependent, meaning methods that place more importance to variables with greater signal, may be susceptible to bias caused by the magnitude of these variables. Mean centering the data helps remove potential bias caused by the magnitude of

(15) J.S. Nadeau, R.B. Wilson, J.C. Hoggard, B.W. Wright, R.E. Synovec, *J. Chromatogr. A.* 1218 (2011) 9091–9101.

(16) J.S. Nadeau, B.W. Wright, R.E. Synovec, *Talanta.* 81 (2010) 120–128.

(17) K.M. Pierce, B.W. Wright, R.E. Synovec, *J. Chromatogr. A.* 1141 (2007) 106–116.

(18) R.A. Davis, A.J. Charlton, J. Godward, S.A. Jones, M. Harrison, J.C. Wilson, *Chemom. Intell. Lab. Syst.* 85 (2007) 144–154.

signal by subtracting the respective means of each variable in each sample, which results in a matrix with a mean of zero. Autoscale is the division of the variables in the mean centered data with their respective standard deviations, thereby creating a matrix with a unit scale¹⁹. It is important to note that autoscale performs better using on distinct uncorrelated measurements as autoscale gives every datapoint equal weight. With respect to correlated data such as a chromatogram or a spectrum, mean centering is more appropriate as autoscale would otherwise apply equal weight to variables that are not chemically meaningful such as baseline.

1.3.2 Principal Component Analysis

Principal Component analysis (PCA) is an unsupervised chemometric technique designed to model the variance in a given X-block (X). PCA analyzes two-way data: the first dimension (rows, denoted as i) corresponds to the samples and the second dimension (columns, denoted as j) corresponds to the variables, i.e the spectrum or the signal in a chromatogram. PCA can be applied to both GC-MS and GC \times GC – TOFMS chromatograms by unfolding (or vectoring) the chromatographic and mass spectral dimensions, as performed in the study of GC \times GC – TOFMS chromatograms of RP-2 fuels using PCA – which is featured in chapter 4.

PCA analysis can be explained in short using the following equation:

$$X = 1x_{\text{mean}} + TP' + E \quad (1.3)^{20}$$

Where X is a ($i \times j$) data matrix, $1x_{\text{mean}}$ is an optional variable (included when X has been mean

(19) P. Geladi, B.R. Kowalski, Anal. Chim. Acta. 185 (1986) 1–17.

(20) S. Wold, K. Esbensen, P. Geladi, Chemom. Intell. Lab. Syst. 2 (1987) 37–52.

centered) and is composed of the outer product between a column vector of 1s and an averaged sample row vector x_{mean} . TP' is PCA's approximation of X using n factors known as principal components (PCs), where T ($i \times n$) is the sample scores for n PCs, and P' ($n \times j$) is the transposed orthogonal loadings or patterns exhibited in the variables for n PCs. E is the ($i \times j$) error matrix or residual of variance not explained by the PCA model. Alternatively, TP' can be expressed as a series of orthogonal components where each principal component is the outer product between a corresponding set of scores and (orthogonal) loadings vectors ($\sum t_n p_n$), as seen in Figure 1.5A. Note that PC1 will always describe the pattern that captures the most variance in X , and PC2 will capture the pattern with the second most variance in X orthogonal to PC1. The maximum number of components is determined by the rank of X , which is defined as the number of linearly independent vectors in X . A vector that is said to be linearly independent cannot be expressed as a combination of one or more of the other vectors in X . These vectors could technically be either row or column vectors. The rank is also dependent on the smallest dimension which is almost always i (samples) when analyzing chromatograms, of which examples can be seen in Figure 1.5B. P' could be considered as a redefinition of the axes with respect to the variance between variables and T corresponds to the magnitude. A visual representation of this can be seen in Figure 1.5C. Note that it is the responsibility of the user to inspect the PCA model(s) and determine the appropriate number of components, as the last possible components to use may only capture the variance attributed to noise. Fortunately, PCA modes are nested models; due to the orthogonality of the PCs, a 3 PC PCA model for example, contains the entire 2 PC model nested within plus the third PC, meaning the user has the leisure of picking a relatively higher number of PCs and choose to remove or ignore the later PCs at no consequence -assuming they contain no relevant information. Regarding the interpretation of T

and P' , T contain quantitative information on samples with respect to which samples are similar, while P' contain the information of the source of variance with respect to the variables themselves^{12,19–22}. For a detailed interpretation of PCA models, please see results and discussion in chapter 4.

1.3.3 Partial Least Squares Regression

Partial least squares regression (PLS) is a chemometric method related to PCA, with the purpose of correlating information between two data matrices, known as the X-block (X) and Y-block (Y), respectively. The intent of PLS is using X to predict Y , which is especially useful when the Y values are difficult, expensive, or time consuming to obtain. As with PCA, PLS can be used to analyze GC-MS and GC \times GC – TOFMS chromatograms after the chromatographic and mass spectral dimensions are unfolded. PLS can be explained using scores and loadings similar to PCA. Y is modeled using the following equation:

$$Y = UQ' + F \quad (1.4)^{19}$$

where Y is a ($i \times k$) data matrix, U ($i \times n$) is the Y scores, Q' ($n \times k$) is the transposed Y loadings, and F ($i \times k$) is the Y residuals. A visual representation of Eq. 1.4 can be seen in Figure 1.6A.

Note that mean centering or autoscaling can be applied to X , or Y individually but is omitted from the equation for the purpose of clarity. Similar to PCA, X is modeled as:

$$X = TP' + E \quad (1.5)^{19}$$

where X is a ($i \times j$) data matrix, T ($i \times n$) is the X scores, P' ($j \times n$) is the X loadings, and E ($i \times j$)

(21) A. Daffertshofer, C.J.C. Lamoth, O.G. Meijer, P.J. Beek, Clin. Biomech. 19 (2004) 415–428.

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

is the X residuals. A visual representation of Eq. 1.5 can be seen in Figure 1.6B. It is crucial to note that TP' in PLS differs from TP' in PCA, as the PLS model captures variance with maximum correlation between X and Y. For each corresponding factor, the n vectors of P' are iteratively rotated until the covariance (linear relationship) between T and U is at a maximum. Once the PLS model is constructed, T can be used to estimate U, as shown in the following equation:

$$U_{\text{pred}} = BT \quad (1.6)^{19}$$

where B is the regression coefficients matrix (of n values). A visualization of Eq. 1.6 can be seen in Figure 1.6C and a visual representation of PLS can be seen in Figure 1.6D. Eq. 1.6 implies that one must simply be able to estimate U to be able to estimate Y itself. Combining Eqs. 1.4 and 1.6 results in the following equation:

$$Y = BTQ' + F \quad (1.7)^{19}$$

Y can be estimated by using B, T, and Q'. The factors in PLS are named latent variables (LVs) to help differentiate from principal components.

Testing the overall accuracy of the PLS model and obtaining statistical information on the residuals is important for making reliable Y predictions. There are three common means of validating the model. The first is validation using a new set of samples (a different X) not used in the calibration and using the PLS model to predict their respective Y. The second type is called cross validation, using a training set vs. validation sets, where X is broken into subsets, that are used in creating a model which is tested (validated) using the other sub sets and statistical information is gathered from the results. However, the number of subsets and the distribution of samples between the subsets can be of significance and it is up to the user to decide with *a priori* knowledge the appropriate number of sets and the distribution of samples.

For example, a model calibrated from a subset composed entirely of replicate samples would perform poorly in predicting values for other subsets compared to a model constructed of a subset composed of a wide variety of samples. This method of cross validation is best suited for X matrices with a large number of samples. The third validation type is called leave one out cross validation (LOOCV). In LOOCV for n -samples, each combination of $(n-1)$ samples is used to construct models which are used to predict against the sample left out. LOOCV has the advantages of being comprehensive and working best on X with a small number of samples, though the drawbacks are that LOOCV is both memory intensive and time consuming so using it on large sets of samples may not be practical.

As with PCA, it is ultimately up to the user to determine the number of LVs for the PLS model. Note however, in the case of PLS, the LVs are not necessarily orthogonal and therefore the models are not nested as with PCA. The user must exercise caution in selecting the most appropriate number of LVs, and inspect the performance of each of the PLS models varying in the number of LVs with respect to the residuals and error (specifically the root mean squared of error of cross validation), and the linear regression vectors (LRVs). The LRVs can be derived as expressed in the following equation:

$$\text{LRV} = \text{PBQ}' \quad (1.8)$$

LRVs can be used to inspect which variables appear most influential with respect to the correlation and when multiplied using an inner product operator with a new X, the predicted Y is obtained. Moreover, it is advisable to follow the rule of parsimony, i.e. when encountering two sequential PLS models (with respect to the number of LVs) that offer somewhat similar results, if the difference is not statistically significant it is better to choose the PLS model with fewer LVs, as to minimize variance captured that may be attributed to noise^{12,19}. For detailed

interpretations of various PLS models, please see the results and discussion sections in chapters 3, 5 and 6.

1.3.4 Parallel Factor Analysis

Parallel factor analysis (PARAFAC) is classified as a decomposition method, also known as canonical decomposition (CANDECOMP). PARAFAC is used to mathematically resolve (or deconvolute) overlapping signals in X . Unlike PCA and PLS which are designed for bilinear data, PARAFAC is designed to handle trilinear data without the need for unfolding the dimensions. In the studies mentioned in subsequent chapters, PARAFAC is used as a tool for purifying signals of badly resolved analytes of interest (based on results from PLS and PCA) and identifying those analytes (PARAFAC is not a main focus of discussions presented herein).

PARAFAC model can be explained using the following equation:

$$X = \sum A_f B_f C_f + E \quad (1.9)^{22}$$

where X is a $(i \times j \times m)$ data matrix, A $(i \times f)$, B $(j \times f)$ and C $(m \times f)$ are the loadings composed of f factors, and E $(i \times j \times m)$ is the residual matrix. Alternatively, one may view each individual factor as an outer product between three loadings vectors (in PARAFAC there is no distinction for the scores, rather it is another loadings), of which a visual representation can be seen in Figure 1.7. PARAFAC resolves signals using the assumption that multi-dimensional signals of pure component (analytes in chromatograms) can be sufficiently defined by the outer product of three vectors (trilinear). Plainly speaking, the PARAFAC algorithm is a multi-way version of alternating least squares (ALS); a deconvolution algorithm designed for two-way data. In ALS,

X is a two-way data matrix that can be modeled using two matrices A and B. The process involves:

- 1) Initializing B,
- 2) Using B and X to estimate A,
- 3) Using A and X to recalculate B
- 4) Continue iterating steps 2-4 until convergence, i.e. the residuals $\|X-AB\|$ reaches a minimum.

For PARAFAC an algorithm derived from ALS is applied to multi-way data. The main difference in the PARAFAC algorithm is B and C are initialized are used along with X to estimate A, then the roles of A, B and C are alternated until convergence, meaning the sum of the squared residuals (calculated from E) reaches a minimum.

In practice, PARAFAC is analyzed on chromatogram sub-regions as the analysis of an entire GC \times GC – TOFMS chromatogram would require a huge number of factors and would be time consuming²³. As long as an analyzed multi-way (including but not limited to GC \times GC – TOFMS) chromatographic sub-region has sufficient physical separation (selectivity) of each of the overlapped peaks on two of the three dimensions, then PARAFAC can resolve overlapping peaks and remove noise as independent factors, thus providing both the pure peak concentrations and mass spectral profiles without the need of standards²⁴. A potential pitfall for iteratively minimizing the sum of the squared residuals is when the algorithm hangs in a local minimum. Due to the (default) random initialization settings of PARAFAC, running the analysis several times and perhaps adjusting the stop criteria parameters with respect to the sum of the squared residuals can help ensure the sum of squared residuals converge on the absolute minimum. In

(23) J.C. Hoggard, W.C. Siegler, R.E. Synovec, J. Chemom. 23 (2009) 421–431.

(24) K.M. Pierce, B. Kehimkar, L.C. Marney, J.C. Hoggard, R.E. Synovec, J. Chromatogr. A. 1255 (2012) 3–11.

situations where the model struggles, better preprocessing, introduction of standards and stricter model constraints can also help. There are two constraints worth mentioning: nonnegativity and unimodality. Nonnegativity restricts the PARAFAC model to have loadings with variables being ≥ 0 , while unimodality can be thought of as the signal of an analyte peak that can only appear once. Moreover, it is crucial to inspect the model loadings to ensure they are logical. All of the above (preprocessing, constraints, sum of the squared residuals, loadings inspection) helps in selecting the appropriate number of factors for the model. As with PLS, PARAFAC models are not nested and selecting the correct number of factors is extremely important; too few will result in failure to mathematically resolve some or all of the overlapping peaks, and too many will result in degeneracy. For PARAFAC models using no constraints, the degeneracy manifests itself as two factors that appear very similar but differ in sign. When nonnegativity constraint is enforced on a PARAFAC model, degeneracy appears as two or more factors each partially capturing the signal of a single peak. As with PLS, one must parsimoniously choose the proper number of factors for the PARAFAC model^{22–25}.

1.3.5 Targeted Techniques Overview*

Specifically, targeted analysis refers to analytical projects where the objective is the identification or quantification of a specific list of analytes that may appear in a complex sample, while ignoring the remaining compounds²⁴. These methods often require *a priori* knowledge

* Large portions of this section published in: K.M. Pierce, B. Kehimkar, L.C. Marney, J.C. Hoggard, R.E. Synovec, J. Chromatogr. A. 1255 (2012) 3–11.

(25) K.M. Pierce, J.C. Hoggard, R.E. Mohler, R.E. Synovec, J. Chromatogr. A. 1184 (2008) 341–352.

such as the retention times reference mass spectra of targeted analytes. The following paragraphs are a brief overview of various noteworthy techniques used in targeted analysis, recent advancements, and studies.

PARAFAC, as stated earlier, is used to deconvolute select peaks in sub chromatographic regions containing several overlapped peaks and/or baseline into a series of factors²². For more detail, please see section 1.3.4 above. Examples of recent studies using PARAFAC include, but are not limited to, resolving and quantifying targeted analytes (metabolites) in GC × GC – TOFMS chromatograms of bacterial samples²⁶, isotopically labeled (12-C & 13-C) analytes^{27,28}, dimethyl methylphosphonate (DMMP) forensic signatures²⁹, and yeast metabolites^{4,30}. For PARAFAC to produce accurate results, the data analyzed must be trilinear. PARAFAC2 is a notable robust variant on the PARAFAC algorithm designed to handle deviations from trilinearity (i.e. deviation from linearity in one of the dimensions). In the case of within-run retention time shifts in the second column dimension, PARAFAC2 has been used to restore trilinearity to the data and provide more accurate quantitative results³¹.

ChromaTOF (by LECO) is a software package for various LECO 'separation science' instruments, specifically the GC × GC – TOFMS. ChromaTOF can not only operate LECO instruments and visualize chromatograms but also contains both preprocessing tools such as

(26) M.J. Gómez, S. Herrera, D. Solé, E. García-Calvo, A.R. Fernández-Alba, *Anal. Chem.* 83 (2011) 2638–2647.

(27) S. Yang, M. Sadilek, R.E. Synovec, M.E. Lidstrom, *J. Chromatogr. A.* 1216 (2009) 3280–3289.

(28) E.M. Humston, J.C. Hoggard, R.E. Synovec, *Anal. Chem.* 82 (2010) 41–43.

(29) J.C. Hoggard, J.H. Wahl, R.E. Synovec, G.M. Mong, C.G. Fraga, *Anal. Chem.* 82 (2010) 689–698.

(30) E.M. Humston, K.M. Dombek, J.C. Hoggard, E.T. Young, R.E. Synovec, *Anal. Chem.* 80 (2008) 8002–8011.

(31) T. Skov, J.C. Hoggard, R. Bro, R.E. Synovec, *J. Chromatogr. A.* 1216 (2009) 4020–4029.

baseline correction and processing tools such as the mathematical peak resolution tool. ChromaTOF can also integrate with other software such as the NIST spectral matching software, allowing access to the NIST mass spectral library for a list of the best matches for any mass spectrum of interest. Though originally designed for non-targeted applications, PARAFAC can be used for targeted analysis. For instance, wastewater and river water samples were submitted to stir bar sorptive extraction (SBSE) followed by GC × GC – TOFMS separation, and ChromaTOF was used to deconvolute and identify the spectra of the targeted pollutants such as personal care products and polycyclic aromatic hydrocarbons²⁶. Other studies using ChromaTOF software for targeted analysis include toxic waste samples from factories stationed in developing countries³², identifying compounds found in pyrolyzed oils³³, and analysis of drug metabolites found in hair³⁴. ChromaTOF currently allows the export of processed peak table data, however to perform many chemometric techniques, it is crucial to have the ability to export the raw data (lossless transfer) from the instrument software and import it into data analysis software tools such as MATLAB or Excel. A function has recently been developed to help import data from ChromaTOF to MATLAB called ‘peg2mat’^{35–38}.

(32) J. de Vos, R. Dixon, G. Vermeulen, P. Gorst-Allman, J. Cochran, E. Rohwer, J.-F. Focant, *Chemosphere*. 82 (2011) 1230–1239.

(33) T. Sfetsas, C. Michailof, A. Lappas, Q. Li, B. Kneale, *J. Chromatogr. A*. 1218 (2011) 3317–3325.

(34) B. Guthery, T. Bassindale, A. Bassindale, C.T. Pillinger, G.H. Morgan, *J. Chromatogr. A*. 1217 (2010) 4402–4410.

(35) R.E. Mohler, K.M. Dombek, J.C. Hoggard, K.M. Pierce, E.T. Young, R.E. Synovec, *Analyst*. 132 (2007) 756–767.

(36) K.M. Pierce, J.C. Hoggard, J.L. Hope, P.M. Rainey, A.N. Hoofnagle, R.M. Jack, B.W. Wright, R.E. Synovec, *Anal. Chem.* 78 (2006) 5068–5075.

(37) J.C. Hoggard, The Synovec Research Group at UW, (<http://depts.washington.edu/synlab/people/dr-jamin-hoggard-2/>).

Match algorithms can also be used for both target analysis and comprehensive non-target analysis. NIST is a commercial mass spectral library for a wide variety of compounds.

SEQUEST and MASCOT are smaller protein and peptide web based mass spectral libraries.

Researchers wanting to use a small mass spectral library specifically for targeted analytes used both SEQUEST and MASCOT to construct the ZFIN database, designed for identifying and quantifying targeted proteins found in Zebra fish³⁹.

The ability to target groups of chemicals (compound classes) can be useful, especially in chromatograms that are extremely rich with chemical information and targeting individual compounds is either impractical or less relevant. Targeting groupings of compound (or compound classes) is essentially targeted analysis, only the scope of the analysis is larger. One such approach to target compound classes is using partial least squares (PLS) analysis. In a sense, identifying compounds of interest that exhibit correlation with respect to measured properties, followed by deconvoluting the peaks using PARAFAC as needed and finally forward their respective mass spectra to a search library for identification. n-PLS is a PLS variant designed for n-way data exhibiting ‘n-linearity’; as a result, the n-PLS has stricter criteria in model construction in comparison to PLS. Both PLS and/or n-PLS have been used to analyze GC × GC – TOFMS chromatograms of metabolic samples⁴⁰, adulterated gasoline⁴¹, gasoline and

(38) K.M. Pierce, J.C. Hoggard, Chromatographic data analysis. Part 3.3.4: handling hyphenated data in chromatography, *Anal Methods* (2014), DOI 10.1039/C3AY40965A

(39) M.B. Lucitt, T.S. Price, A. Pizarro, W. Wu, A.K. Yocum, C. Seiler, M.A. Pack, I.A. Blair, G.A. FitzGerald, T. Grosser, *Proteomics*. 7 (2008) 981–994.

(40) T. Gröger, R. Zimmermann, *Talanta*. 83 (2011) 1289–1294.

(41) M.P. Pedroso, L.A.F. de Godoy, E.C. Ferreira, R.J. Poppi, F. Augusto, *J. Chromatogr. A*. 1201 (2008) 176–182.

kerosene blends⁴², and blends of conventional diesel with biodiesel⁴³.

Additionally for targeting compound classes, one can use supervised statistical classification based algorithms such as Fisher ratios (F-ratios). F-ratios highlight the differences between sample classes via a ratio between the class-to-class variation and within class variation. F-ratios have been used to differentiate types of cacao bean samples and the biomarkers were identified by forwarding their respective mass spectrum to the NIST library search tool in ChromaTOF. The peak analytes were also quantified using an in-house GUI of targeted PARAFAC^{44,45}. Another report used F-ratios, PARAFAC and PCA for quantifying potential biomarkers for perinatal asphyxia⁴⁶.

Also in-house processing tools should be mentioned. These various collections of code labeled as 'in-house software' can include but are not limited to custom scripts for managing commercial software, applying known methods to higher order data, or altogether new algorithms for analyzing data for either targeted or non-targeted analysis. One example is the study of valley to peak ratio on GC × GC chromatograms to measure the resolution of the targeted analytes and estimate the error of quantification⁴⁷. The Dotmap algorithm is an in-house software that can be used for both custom mass spectral libraries or NIST^{48,49}. An algorithm

(42) L.A.F. de Godoy, E.C. Ferreira, M.P. Pedroso, C.H. de V. Fidélis, F. Augusto, R.J. Poppi, *Anal. Lett.* 41 (2008) 1603–1614.

(43) K.M. Pierce, S.P. Schale, *Talanta*. 83 (2011) 1254–1259.

(44) E.M. Humston, J.D. Knowles, A. McShea, R.E. Synovec, *J. Chromatogr. A*. 1217 (2010) 1963–1970.

(45) E.M. Humston, Y. Zhang, G.F. Brabeck, A. McShea, R.E. Synovec, *J. Sep. Sci.* 32 (2009) 2289–2295.

(46) A.C. Beckstrom, E.M. Humston, L.R. Snyder, R.E. Synovec, S.E. Juul, *J. Chromatogr. A*. 1218 (2011) 1899–1906.

(47) S. Peters, G. Vivó-Truyols, P.J. Marriott, *J. Chromatogr. A*. 1146 (2007) 232–241.

(48) J.L. Hope, A.E. Sinha, B.J. Prazen, R.E. Synovec, *J. Chromatogr. A*. 1086 (2005) 185–192.

called alternating weighted residue constraint quadrilinear decomposition (AWRCQLD) is a third order calibration algorithm designed for analyzing quadrilinear data⁵⁰, i.e such as GC × GC × GC – TOFMS chromatograms.

1.3.6 Non-targeted Techniques Overview

Non-targeted analysis is a comprehensive analysis of the entire set of data such as chromatograms for compounds of significance with minimal to no *a priori* knowledge and user input, often analytes that are either unique to specific samples or that vary in concentration across samples. As implied earlier, manychemometric techniques can be used both in a targeted and non-targeted manner²⁴.

For non-targeted techniques there are mainly two flavors: supervised and unsupervised. When a technique is classified as supervised it means either a standard is used in the analysis or classification of the samples is given and unsupervised means no such information is forwarded to the technique²⁴.

PCA is an example of unsupervised technique (see section 1.3.4). PCA has been used in the study of the effects of peak alignment¹⁶. Other supervised techniques worth mentioning are: Fisher ratios¹⁶ and analysis of variance (ANOVA, which uses the F-test to analyze and highlight locations of variance between classes of samples on a pixel or a data point level)^{24,51}. PLS is

(49) A.E. Sinha, J.L. Hope, B.J. Prazen, E.J. Nilsson, R.M. Jack, R.E. Synovec, J. Chromatogr. A. 1058 (2004) 209–215.

(50) H.-Y. Fu, H.-L. Wu, Y.-J. Yu, L.-L. Yu, S.-R. Zhang, J.-F. Nie, S.-F. Li, R.-Q. Yu, J. Chemom. 25 (2011) 408–429.

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

another supervised technique (discussed in section 1.3.3) and its use in studies are presented herein chapters 3, 5 and 6.

PARAFAC, which has been previously mentioned in section 1.3.5 (targeted techniques overview), can also be used as an unsupervised non-targeted analysis technique; an algorithm exists that calculates the best number of components to use by comparing the extracted mass spectrum from different components from the same and different models to each other⁵². It was further developed to allow the analysis of an entire chromatogram or a set of chromatograms by breaking the task into a series of sub region chromatograms to be analyzed^{23,24}.

ChromaTOF as stated earlier can be used for non-targeted analysis by applying some of the included tools such as PCA, and F-ratios^{44,45}.

1.4 Conclusions

Data analysis will always be a part of research, especially with information-rich data such as chromatograms (GC-MS and GC \times GC – TOFMS). The overview included the basics of the separation in chromatography, the instrumentations used, and the data formats used. The data analysis process was explained including preprocessing and processing of PCA, PLS, and PARAFAC. Furthermore, recent advancements in both targeted and non-targeted analysis of multi-dimensional chromatograms were also briefly discussed. This chapter sets the ground work for the following chapters and serves as a brief introduction.

(52) J.C. Hoggard, R.E. Synovec, *Anal. Chem.* 80 (2008) 6677–6688.

Figure Captions

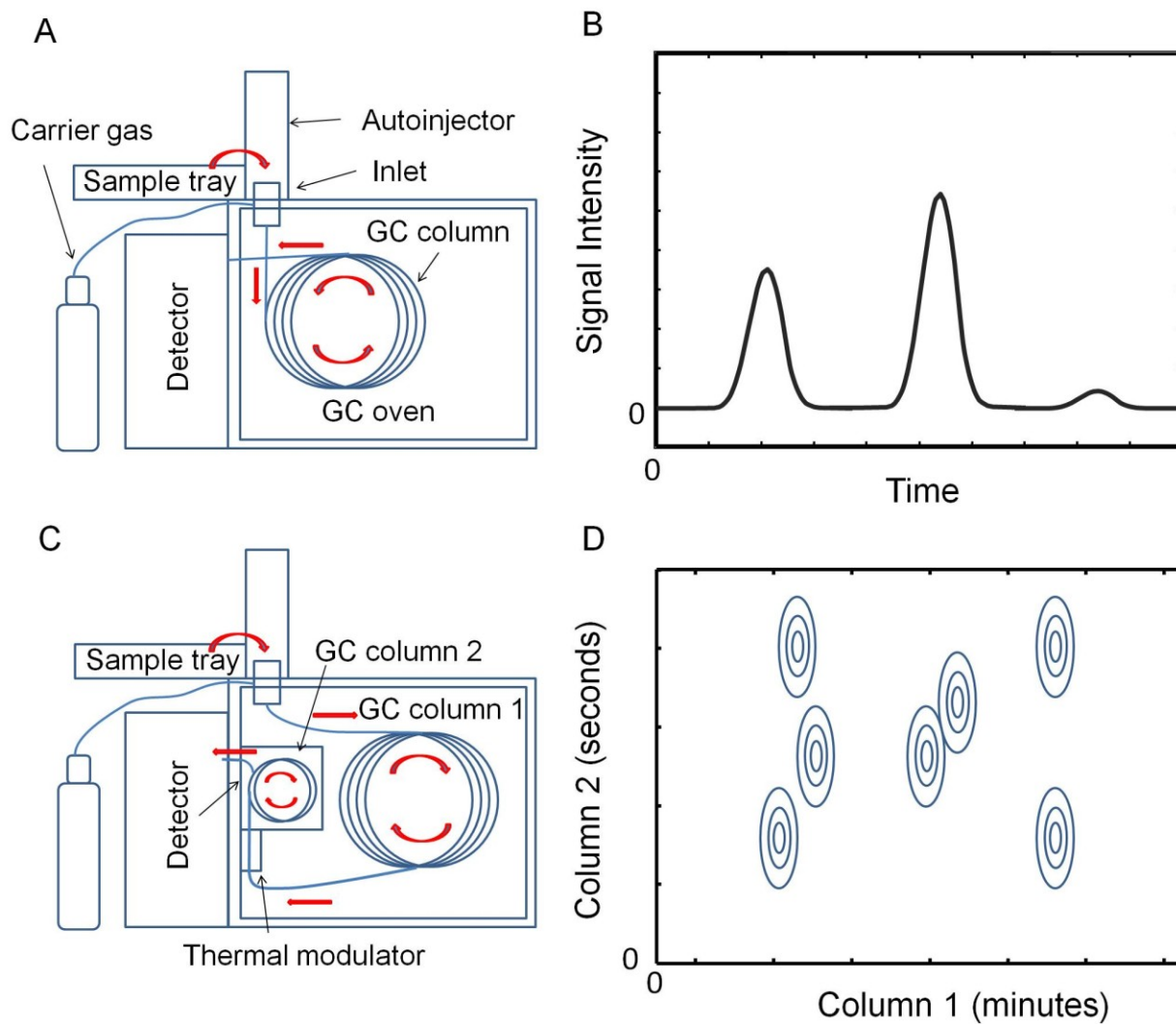


Figure 1.1. (A) An illustration of a gas chromatograph (GC) instrument. (B) An illustration of a GC chromatogram showing three separated peaks. (C) An illustration of a GC instrument setup for GC \times GC separation. (D) An illustration of a GC \times GC chromatogram containing seven demonstrating the separation power of two columns as using only the primary column would fail to separate all seven peaks.

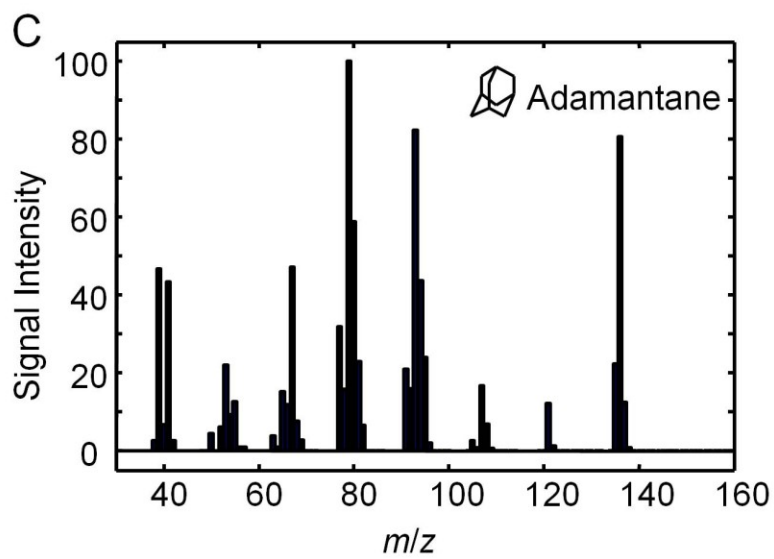
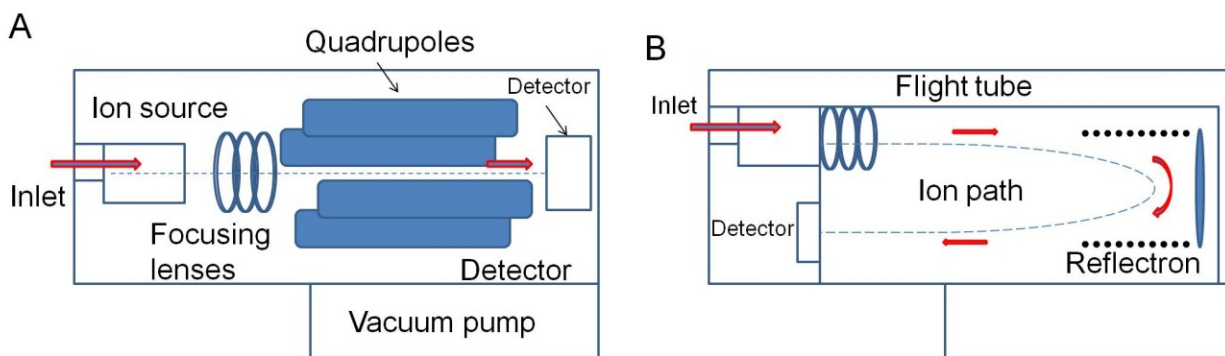


Figure 1.2. (A) An illustration of a quadrupole mass spectrometer. (B) An illustration of a time-of-flight mass spectrometer (TOFMS or TOF). (C) Mass spectrum of adamantane.

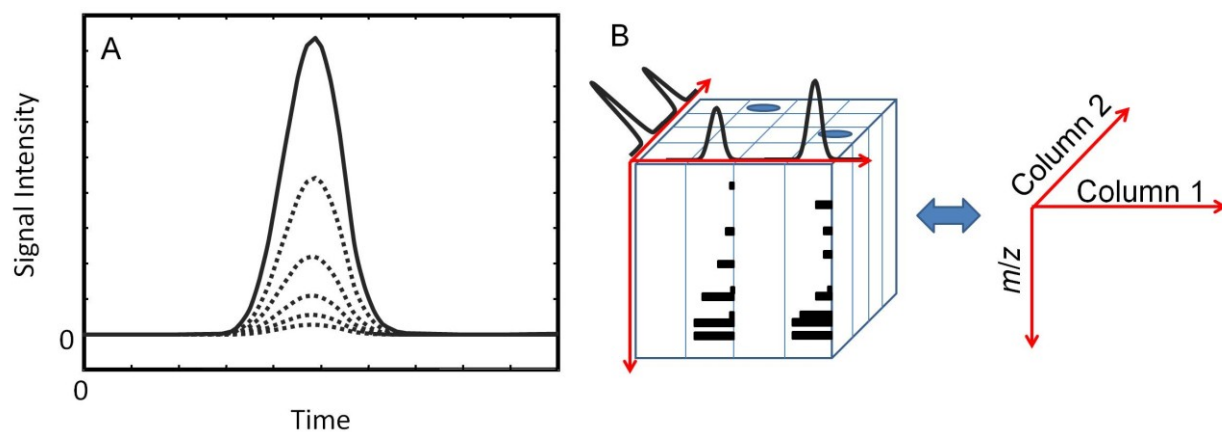


Figure 1.3. (A) An illustration of a GC-MS chromatogram, the dashed lines represent individual selective m/z chromatogram, the solid line represents the total ion current (TIC) chromatogram which is the summation of all selective m/z chromatograms. (B) As a way of representing the GC \times GC – TOFMS chromatogram, an illustration of the data cube is used. Each specific data point is a function of the two chromatographic dimensions and the mass spectral dimension.

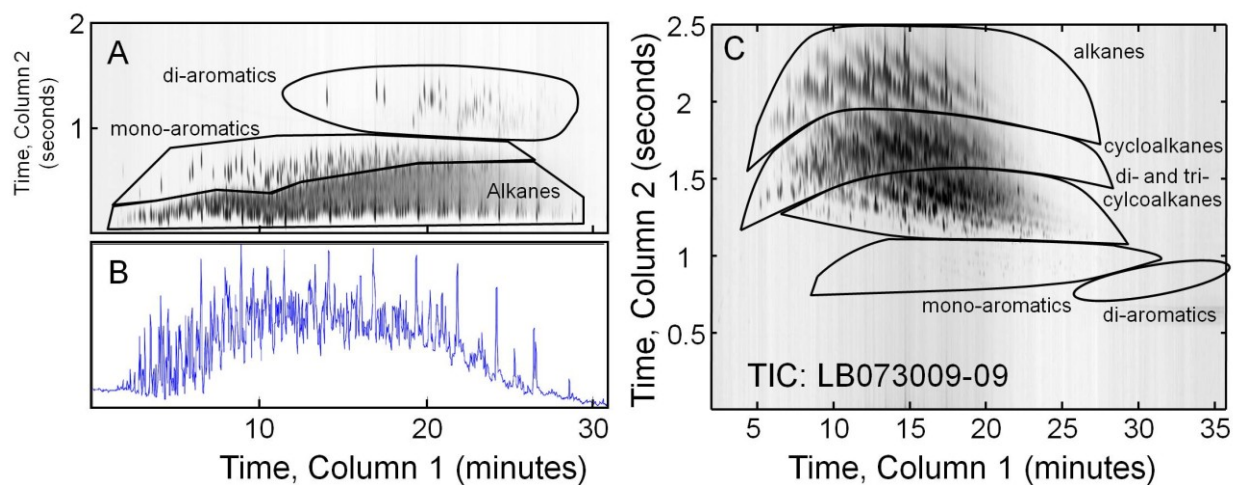


Figure 1.4. (A) GC \times GC – TOFMS TIC chromatogram using a normal column configuration. The sample used was a jet fuel (POSF54658), regions are encircled and labeled; note the major separation is between aromatics and alkanes (with the alkanes group containing both alkanes and cycloalkanes), courtesy of Dr. Jamin C. Hoggard. (B) The simulated 1D TIC chromatogram analogue of Figure 1.4A by summing the second dimension, the apparent rise in baseline is due to the huge number of overlapping peaks, courtesy of Dr. Jamin C. Hoggard. (C) Raw GC \times GC – TOFMS TIC chromatogram of RP fuel LB073009-09, using a reverse column configuration. The alkanes and cycloalkane groups are well separated, along with the di- & tri-cycloalkanes group and the aromatic groups (where a few faint peaks can be seen), isomers are separated in diagonal patterns.

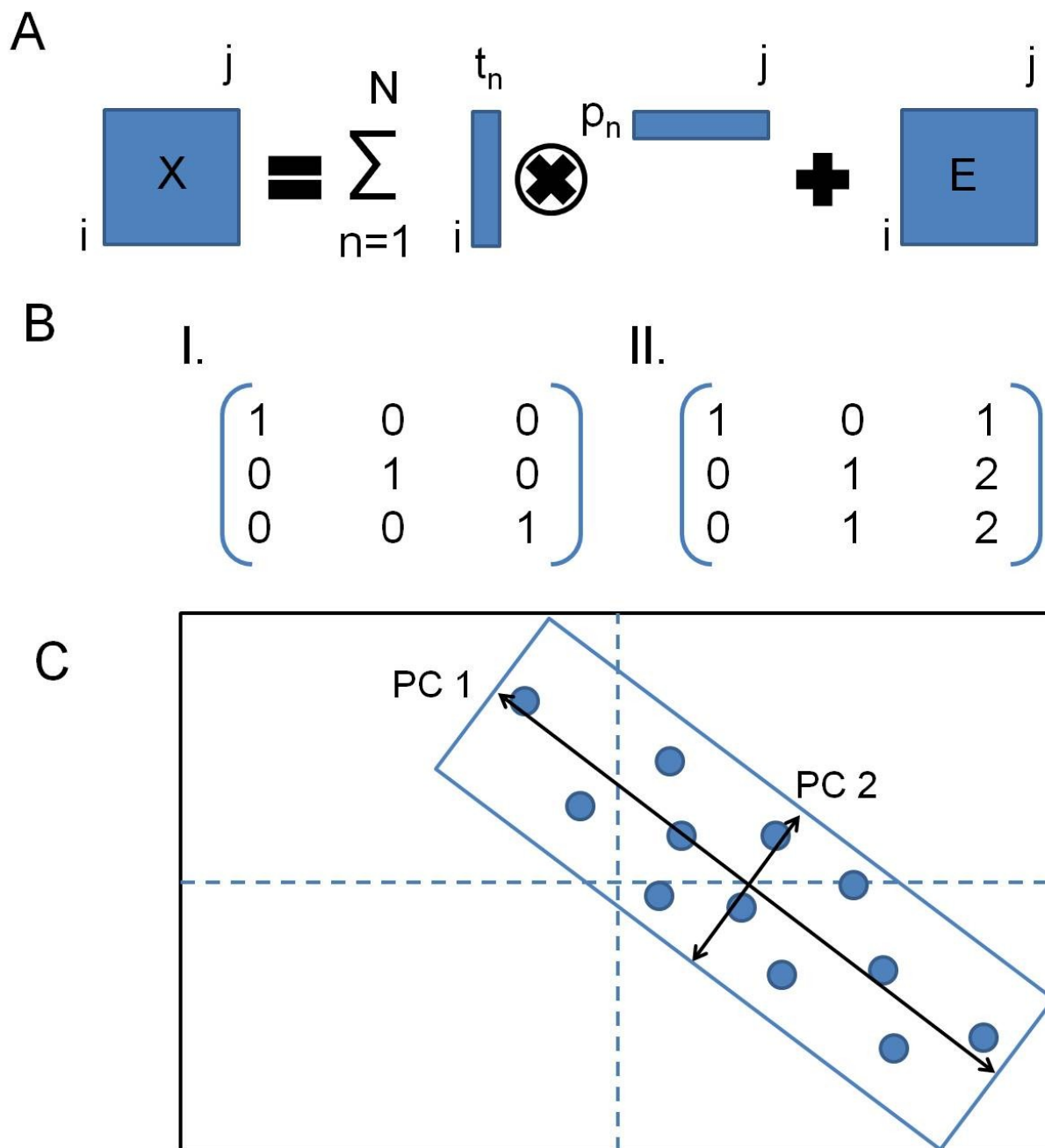


Figure 1.5. (A) A visual representation of the PCA model, where X is approximated by a series of components (outer product between scores and loadings) and the remainder of unexplained (uncaptured) variance remains in the residual matrix E . (B) A visual example for explaining the rank of a matrix. (I) A 3×3 identity matrix I , is an example of a matrix with full rank. (II) A rank deficient matrix as the third column vector can be expressed as a combination of the first two column vectors. (C) An illustration of captured variance of data in using PCA, the PCs could be seen as a redefinition of the axes for the data, in this illustration the origin of the PC is at the center of the data, implying that mean centering was performed.

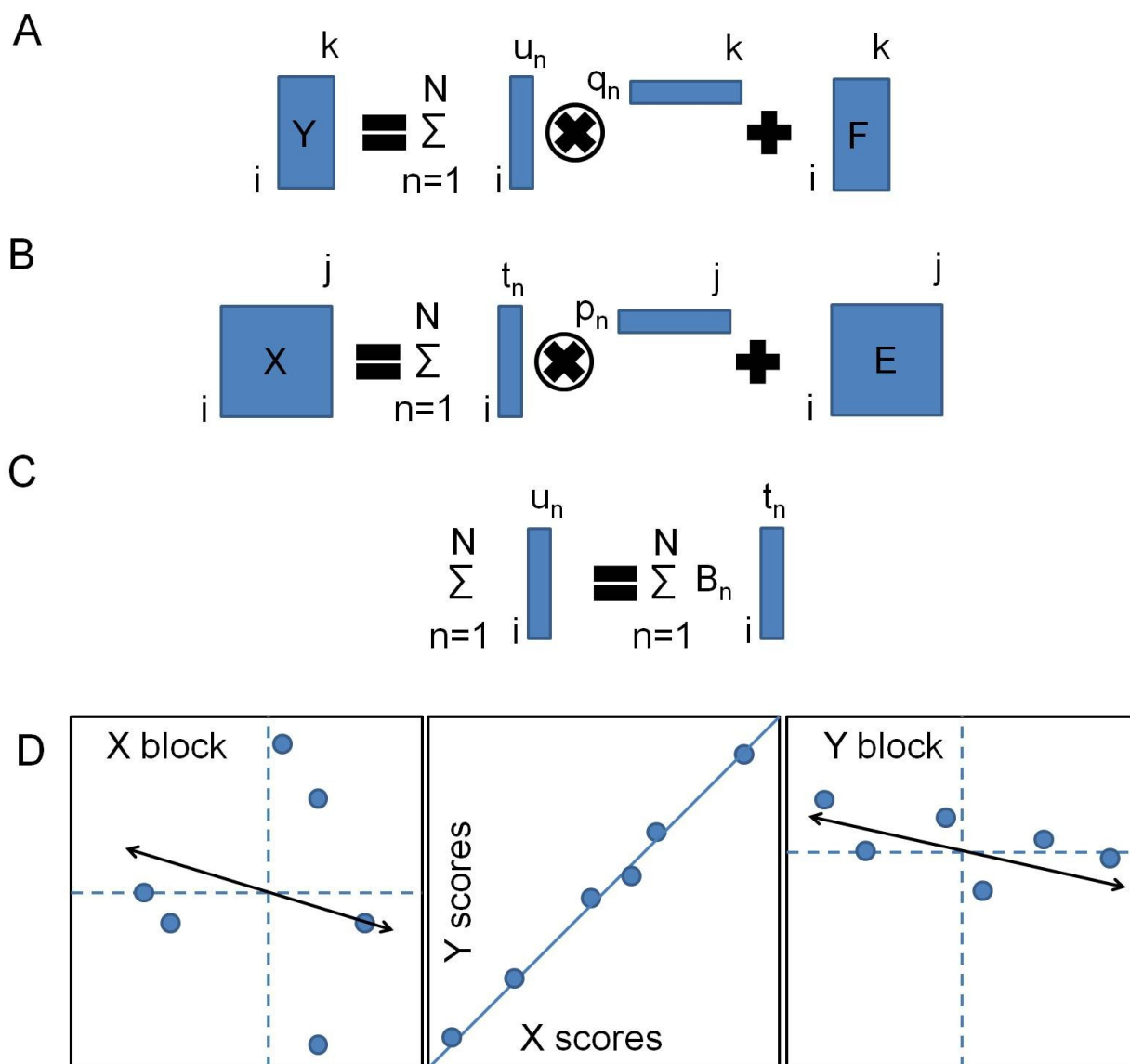


Figure 1.6. A visual representation of the PLS model. (A) The variance modeled for Y, similar to PCA. (B) The variance of X captured each latent variable (outer product of scores and loadings) has a maximum correlation with the corresponding component in Y. (C) Using the scores from X and its respective regression coefficient, the scores of Y can be estimated. (D) A visual representation of PLS; the Loadings of X is rotated so that the X scores achieve the maximum correlation with the Y scores.

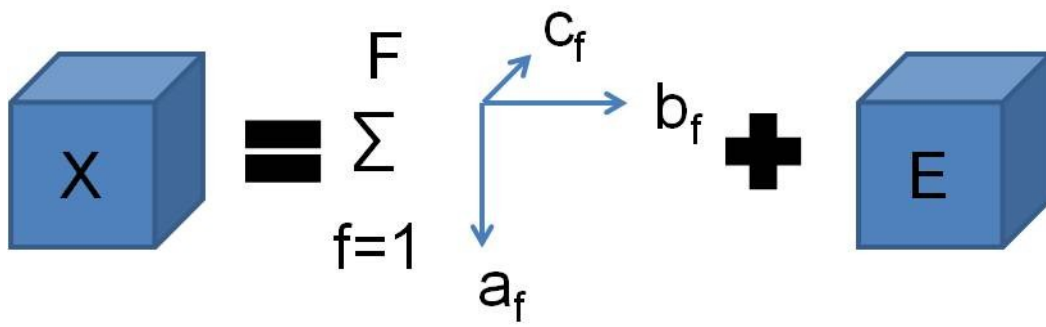


Figure 1.7. A visual representation of modeling a trilinear X using a series of factors (PARAFAC).

Chapter 2: Temporal Mass Spectral Ratio Analysis Method (TMSRA)^{53*}

2.1 Introduction

Gas chromatography (GC) is a widely used analytical technique^{16,44,54–58}, for many studies including, but not limited to, the following areas: pesticides analysis^{16,54,55}, the petroleum industry¹⁶, for toxicology⁵⁶, the food and beverage industry^{44,54,57}, and for metabolomic studies^{44,54,58}. As with any chemical analysis technique, data analysis plays a necessary role. It is critical to rapidly and confidently analyze the data. In many studies, the analyst is interested in specific compounds, referred to as targeted analysis. In contrast, if the analyst is searching for initially unknown analytes; this is often referred to as untargeted analysis. This report is aimed at improving GC data analysis for targeted analysis.

The algorithm presented herein is designed for processing data acquired from a GC coupled to mass spectrometry (MS). The algorithm is also designed for situations where high throughput is needed. The samples may be complex (the total number of compounds may far exceed that of targeted analytes) and chromatographically and mass spectrally interfering compounds are likely. Taking these factors into account, it is expected that the chromatographic

* Large portions of this chapter published in: B. Kehimkar, J.C. Hoggard, J.S. Nadeau, R.E. Synovec, *Talanta*. 103 (2013) 267–275.

(53) B. Kehimkar, J.C. Hoggard, J.S. Nadeau, R.E. Synovec, *Talanta*. 103 (2013) 267–275.

(54) F.L. Dorman, J.J. Whiting, J.W. Cochran, J. Gardea-Torresdey, *Anal. Chem.* 82 (2010) 4775–4785.

(55) W. Zhang, P. Wu, C. Li, *Rapid Commun. Mass Spectrom.* 20 (2006) 1563–1568.

(56) M.R. Meyer, F.T. Peters, H.H. Maurer, *Clin. Chem.* 56 (2010) 575–584.

(57) A. Genovese, R. Dimaggio, M.T. Lisanti, P. Piombino, L. Moio, *Ann. Chim.* 95 (2005) 383–394.

(58) E. Jellum, O. Stokke, L. Eldjarn, *Anal. Chem.* 45 (1973) 1099–1106.

resolution, R_s , between compounds could be compromised for targeted analytes.

There are many methods one may use to quantify targeted analytes⁵⁹⁻⁶¹. Some MS quantification methods use the selective ion monitoring (SIM) mode. By preselecting the mass channels (m/z) to monitor, and using the intensities to quantify the targeted analyte(s), there are benefits of greater sensitivity, greater signal-to-noise ratio (S/N), and lower limit-of-detection (LOD) when compared to scan mode. However, the SIM mode is less comprehensive than the scan mode, since the m/z range collected is generally small, thus it may be difficult to confidently identify all compounds of interest⁵⁵. Also, the m/z used in the SIM mode may be interfered by signals from non-targeted compounds, which may change from run-to-run. While it may be possible to detect interfering compounds in many situations, this cannot be corrected without either a loss of confidence in results or additional experiments required to provide confidence.

There is also software used to deconvolute (mathematically resolve) and quantify analytes such as AMDIS^{55,56,59}, or ChromaTOF⁶¹. Though both AMDIS and ChromaTOF can be used for targeted analysis, they are primarily non-targeted analysis methods, and can be time consuming for data computation and interpretation steps. Alternatively, one could choose chemometric tools to deconvolute and quantify targeted analytes. Parallel factor analysis (PARAFAC)²² and trilinear decomposition (TLD)⁶² are two examples; both however, are multi-way methods, so to work on GC-MS data they require simultaneously analyzing several temporally aligned GC-MS runs, and this process can be time consuming. Another chemometric

(59) J.M. Halket, A. Przyborowska, S.E. Stein, W.G. Mallard, S. Down, R.A. Chalmers, *Rapid Commun. Mass Spectrom.* 13 (1999) 279–284.

(60) W.G. Pool, J.W. de Leeuw, B. van de Graaf, *J. Mass Spectrom.* 32 (1997) 438–443.

(61) Y. Koh, K.K. Pasikanti, C.W. Yap, E.C.Y. Chan, *J. Chromatogr. A.* 1217 (2010) 8308–8316.

(62) E. Sanchez, B.R. Kowalski, *J. Chemom.* 4 (1990) 29–45.

tool to consider is the generalized rank annihilation method (GRAM)⁶³. GRAM is used for targeted analysis and can be applied to a single GC-MS run when data from a standard exists (i.e., requiring another GC-MS run with the standard, typically spiked into the original sample). A major drawback is that GRAM may fail if the targeted analyte is severely overlapped with another compound, or if the analyte peak in the standard (spiked sample) is misaligned with the analyte peak in the original sample, or if there is a difference between the shape of the peaks of the standard and the sample. Alternatively, multivariate curve resolution-alternating least squares (MCR-ALS)^{64,65} could be used, though ambiguities can pose problems in “un-mixing” the targeted analyte. A daunting drawback for applying chemometrics is that in most instances the algorithms require their respective input parameters to be selected on a case by case basis, so having them run automatically is generally not possible⁵². Of course, one may achieve a better separation for a particular analyte by adjusting instrumental parameters; this approach requires running the sample more times, may compromise the separation of other analytes, and may not always be possible, especially when time and/or resources are limited. Hence, often modifying the GC separation conditions may not be an appealing solution.

In contrast, the data analysis method presented herein is designed to rapidly analyze GC-MS data, one run at a time, even with significant chromatographic peak overlap. The method and algorithm presented is referred to as targeted mass spectral ratio analysis (TMSRA), and it runs automatically with the following inputs and specifications: (1) a standard or representative sample chromatogram for alignment, (2) the processed spectra of the targeted analytes (see

(63) C.G. Fraga, *J. Chromatogr. A.* 1019 (2003) 31–42.

(54) M. Daszykowski, B. Walczak, *TrAC Trends Anal. Chem.* 25 (2006) 1081–1096.

(65) M. Garrido, F.X. Rius, M.S. Larrechi, *Anal. Bioanal. Chem.* 390 (2008) 2059–2066.

Theory section), (3) the respective retention times and the m/z range scanned for each targeted analyte, and (4) code to automate the initialization process of the algorithm and storing the results. The basis of TMSRA is the following: as long as there are a few m/z of a given targeted analyte that are objectively determined to be sufficiently pure (not interfered significantly so bias is minimal), quantification is possible even at low R_s . The determination of sufficient mass spectral purity is done by evaluating the ratios between intensities of m/z in a spectrum and comparing the results relative to those from a library spectrum to identify all m/z sufficiently pure to ensure accurate quantification. While quantifying targeted analytes using selective m/z is not new, the approach we utilize to evaluate m/z is novel by requiring “internal consistency” with the list of sufficiently pure m/z that are found.

For the first study, the targeted analyte was perfluorotributylamine (PFTPA), commonly used to calibrate mass spectrometers. The objective of the first study was to evaluate the mass spectrometer reproducibility and aid in selecting a mass spectral threshold for the subsequent case study. For the subsequent case study the targeted analyte was 1-heptene, and the interferent was n-heptane. They were selected due to their mass spectra being very similar with signal at many m/z in common. Chromatographic data regions containing a pure GC-MS peak of these analytes were extracted (for 5 replicates), and mathematically added at various temporal offsets to generate various degrees of peak overlap to facilitate evaluation of the TMSRA algorithm as the R_s between the two compounds was controlled, spanning a R_s range of 0 to 1.5. The quantitative deviation (and bias) and precision for 1-heptene was studied as a function of R_s .

2.2 Theory

When chromatographic resolution, R_s , is compromised between a targeted analyte, T , and an interfering unknown compound, U , the challenge is to determine which m/z of the target analyte are sufficiently pure for quantification. For this discussion, resolution is defined as

$$R_s = ({}^T t_R - {}^U t_R) / (({}^T W_b + {}^U W_b) / 2) \quad (2.1)^1$$

where t_R and W_b are the retention time and peak width at the base, respectively. W_b is defined as four times the standard deviation of the peak using a Gaussian peak model for simplicity.

For targeted analysis, the ratios between intensities of paired m/z channels serve as a guide to identify sufficiently pure i.e. uninterfered, m/z (or mass channels), meaning the m/z in a sample are compared to those in a library spectrum, L , which is a designated standard of a given target analyte. For subsequent clarity during visualization purposes, the m/z of L are sorted (indexed) according to intensity in descending order. Taking the respective indexed m/z for L , each intensity is divided by other intensities as shown in Eq. (2.2) to produce the ratio matrix, ${}^L R_{i,j}$. Where ${}^L I_i$ is the signal intensity of indexed m/z i in L , and ${}^L I_j$ is the signal intensity of indexed m/z j in L ; the elements of the ratio matrix are given by,

$${}^L R_{i,j} = {}^L I_i / {}^L I_j \quad (2.2)$$

Alternatively, one may see this procedure as treating the intensities of indexed m/z values as a vector and taking the outer product between the vector and its respective (element by element) inverse. Calculating ${}^L R_{i,j}$ (for all i and j) generates a matrix of ratios between all m/z as shown in Table 2.1A. ${}^L R_{i,i}$ values on the diagonal are equal to 1, (because there, $i = j$, thus ${}^L I_i = {}^L I_j$). In the lower left triangle of the matrix, ${}^L R_{i,j}$ is less than a value of 1 (since there, ${}^L I_i < {}^L I_j$). The upper

right portion of the matrix is simply the inverse of the lower left triangle (${}^L I_i > {}^L I_j$). To remove redundancy in ${}^L R_{i,j}$, it is useful to exclusively use the lower left triangle.

Likewise, the m/z index of L is used to sort the m/z of the given targeted analyte in the sample spectrum, S , resulting in the sorted m/z of S corresponding with the sorted m/z of L .

Similar to Eq. (2.2), the ratio matrix ${}^S R_{i,j}$ is calculated as follows,

$${}^S R_{i,j} = {}^S I_i / {}^S I_j \quad (2.3)$$

where ${}^S I_i$ is the signal intensity of indexed m/z i in S , and ${}^S I_j$ is the signal intensity of indexed m/z j in S . Next, the deviation between ${}^L R_{i,j}$ and ${}^S R_{i,j}$ relative to ${}^L R_{i,j}$ is calculated by Eq. (2.4), and the results are the elements of matrix $X_{i,j}$ (see Table 2.1B),

$$X_{i,j} = ({}^S R_{i,j} - {}^L R_{i,j}) / {}^L R_{i,j} \quad (2.4)$$

where $X_{i,j}$ is the indexed mass purity matrix and the division is done element by element. Pure m/z in S , relative to each other, have corresponding $X_{i,j}$ values close to 0, affected primarily by noise. If an m/z in S is being interfered, the corresponding $|X_{i,j}|$ absolute values should be greater since the intensity relative to other m/z would not match those in L . Looking at Table 2.1B, the elements on the diagonal have a value equal to 0 (because ${}^S R_{i,j} = {}^L R_{i,j} = 1$). Another noticeable pattern is that when the values of i and j in $X_{i,j}$ are transposed, then the sign changes, as confirmed by Eq. (2.4). Again, to remove redundancy, only the lower left triangle is used.

As the number of indexed m/z values increases, so does the complexity of ${}^S R_{i,j}$, ${}^L R_{i,j}$ and $X_{i,j}$. To help interpret the results embedded in plots of $X_{i,j}$, *connectivity (or adjacency) matrices* $C_{i,j}$ are derived from the lower left triangle of $X_{i,j}$. Detailed explanation on connectivity matrices

can be found elsewhere^{66,67}. Briefly, as used herein, connectivity matrices describe the relationship (i.e., the connection) between the elements of two vectors (here, indexed m/z) as a binary map. The matrix $C_{i,j}$ is constructed by setting a tolerance value for the lower left triangle of $|X_{i,j}|$, the most deviation the analyst is willing to accept from ${}^S R_{i,j}$ (analyte in a sample) with respect to ${}^L R_{i,j}$ (the target analyte standard), with the result being a logical matrix containing a value of 1 where $X_{i,j}$ elements are within the desired tolerance, and a value of 0 where they are not. A lax tolerance will have many elements with a value of 1 in $C_{i,j}$, while a strict tolerance will have few. The aforementioned tolerance is a value between 0 and 1, and can alternatively be seen as a percent deviation from the standard; for example a value of 0.1 can be seen as a tolerance of up to 10% deviation from ${}^L R_{i,j}$.

The $C_{i,j}$ matrix is transposed $C_{i,j}^T$ and added to itself $C_{i,j} + C_{i,j}^T$, producing an 'uncondensed' connectivity matrix $A_{i,j}$. In interpreting the matrix $A_{i,j}$, each element can be thought of as indicating the purity of pairs of indexed m/z , or alternatively, each row i , having a length equal to the number of indexed m/z being analyzed, indicating the purity of the m/z indexed by i with respect to those m/z indexed by j . Each element whose value is 1, meets the tolerance criteria with respect to the ratio of indexed m/z i and j , meaning the relative intensities of the m/z are consistent (to within tolerance) with those found in L .

One often finds patterns or rows in $A_{i,j}$ that appear identical to each other; this can be interpreted as a confirmation (from a different indexed m/z as a function of i) that the ratios of the particular set of indexed m/z indicated in $A_{i,j}$ are consistent. The diagonal is made equal to

(66) Y. Marrero-Ponce, E.R. Martínez-Albelo, G.M. Casañola-Martín, J.A. Castillo-Garit, Y. Echevería-Díaz, V.R. Zaldivar, J. Tygat, J.E.R. Borges, R. García-Domenech, F. Torrens, F. Pérez-Giménez, *Mol. Divers.* 14 (2010) 731–753.

(67) M. Randić, *J. Chem. Inf. Comput. Sci.* 37 (1997) 672–687.

elements with a value of 0, because these elements will always give a value of 1 for tolerances greater or equal to a value of 0, (since $X_{i,j} = 0$ when $i = j$). This also clears rows that only have a value of 1 on the diagonal as they can be interpreted as “the only indexed m/z as a function of j that are consistent with indexed m/z as a function of i is when $i = j$.” The resulting matrix, after setting the diagonal of $A_{i,j}$ to a value of 0, was denoted as $D_{i,j}$. However, slight differences between sets of m/z in $D_{i,j}$ indicated as pure complicate the logic to determine which set of m/z are consistent. To proceed, the $D_{i,j}$ matrix is taken to the second matrix power, $D_{i,j}^2$. The resulting effect is to soften the strict logical tolerance criteria of the connectivity matrix (as explained below) in order to achieve the goal of including at least three indexed m/z in each valid set. A minimum of three indexed m/z are needed; this is because it is not possible to recover an original set of m/z having fewer than three members found in $A_{i,j}$ after setting the diagonal to a value of 0 (as seen in Figure 2.1A) and then squaring the resulting matrix, $D_{i,j}$ to $D_{i,j}^2$ (seen in Figure 2.1B). The matrix multiplication of $D_{i,j}$ to itself annexes or connects the indexed m/z that are consistent with at least one indexed m/z within the set (known as including or connecting through the nearest neighbor) hence moderately increasing the size of the indexed m/z set and softening the connectivity matrix criteria (as seen in Figure 2.1B). All nonzero elements in $D_{i,j}^2$ are set to a value of 1 and plotted as a matrix of dots, each dot represents $X_{i,j}$ values that meet the set tolerance (including through the nearest neighbor). All sets (rows) with three or more elements equal to a value of 1 are extracted (excluding duplicate sets); the result is a collection of various sets of m/z that could (in principle) be used to quantify the target analyte T .

Before explaining the decision process to determine which set of m/z to use for quantification, the process of how the peak height is calculated needs to be explained. Referring

back to L , the sum of intensities of a set of m/z in L can be thought of as $\sum^L I_{\text{set}}$, or a fraction, q , of the summed intensities of all m/z in L ,

$$q = \sum^L I_{\text{set}} / \sum^L I_{\text{total}} \quad (2.5)$$

The peak height of analyte T , in S (the sample) is calculated by taking the sum of intensities of a set of m/z multiplied by its respective reciprocal of q (calculated from L), and is given by,

$$\sum^S I_{\text{total (calc)}} = \sum^S I_{\text{set}} / q \quad (2.6)$$

The calculated TIC equivalent peak height $\sum^S I_{\text{total (calc)}}$ for the sample is a close approximation of a pure TIC peak height of S ; this procedure facilitates comparison across different sets of m/z .

Implementing this procedure for every set of m/z extracted from $D_{i,j}^2$, gives a list of possible quantitative answers for the TIC equivalent peak height of T . At this point one may ask “which peak height is the best answer?” Logically, if T is present, and if there are several m/z of T sufficiently unaffected by interfering compounds U , then any interference from U would only add to the signal of the T 's peak height. Therefore, the most correct set of m/z to use for quantifying T is the set of m/z producing the lowest TIC equivalent peak height (i.e., which minimizes bias). Indeed, using a SIM mode approach, while fast, does not provide quantitative information to minimize bias, such as done in the method we employ.

Finally, for the purposes of this report to study quantitative accuracy and precision as a function of R_s , per the experimental design for the case study with 1-heptene and n-heptane, the calculated peak heights for each run are divided by their corresponding pure peak heights (same sample replicate prior to undergoing simulated peak overlap). Thus, the concentration ratio, F , of the analyte in the sample relative to the analyte in the library standard, is determined using the ratio of TIC equivalent peak heights, following from Eqs. (2.5) and (2.6),

$$F = \sum^S I_{\text{total (calc)}} / \sum^{S(\text{pure})} I_{\text{set}} \quad (2.7)$$

thereby 'normalizing' the quantitative result for determining concentration. For the case study, ideally, $F = 1$ should be experimentally obtained.

2.3 Experimental

The instrument used was an Agilent 6890 GC equipped with a 5973A Mass Spectrometer with unit mass resolution and a 7683B auto injector (Agilent Technologies, Palo Alto, CA, USA). Samples were run through a 30 m J&W 122-503E DB-5 capillary column, with 250 μm i.d., and 0.5 μm film thickness (Agilent Technologies, Palo Alto, CA, USA). The inlet temperature was 250 $^{\circ}\text{C}$. The temperature program utilized an initial temperature of 35 $^{\circ}\text{C}$ for 1.15 min, then ramped at a rate of 5 $^{\circ}\text{C}/\text{min}$ to 60 $^{\circ}\text{C}$, then ramped at 35 $^{\circ}\text{C}/\text{min}$ to 300 $^{\circ}\text{C}$ with this final temperature maintained for 1.5 min. The flow rate was held at 1.4 ml/min with an initial head pressure of 2.48 psig (17.1 kPa above ambient pressure) using H_2 as carrier gas. The mass spectrometer parameters were as follows: the electron energy (filament voltage) parameter was set to 70 eV, and the m/z scan range was 50 to 220 for the PFTBA study and 40 to 140 for the case study with 1-heptene and n-heptane. The 'sample' parameter (the number of times the signal abundance of each m/z is recorded then averaged) was set to 4, and the data acquisition rate was set to 20 Hz. For the case study, 0.5 μl of each sample was injected via the auto injector using split-less injection mode.

Since the proposed computational method relies heavily on the ratios between the intensities of m/z , the reproducibility of the spectrum and the mass spectral threshold needed investigation. Perfluorotributylamine (PFTBA) (Agilent Technologies, Palo Alto, CA, USA)

was used to evaluate these issues since it is the tuning compound for the quadrupole MS. Since perfluoroalkanes are commonly used in tuning MS instruments⁶⁸, it is assumed that for the specific vial of PFTBA the spectrum is fairly reproducible, and therefore could be used to gauge the run-to-run variation of m/z over a wide range of intensities. The intensities of the aforementioned m/z in the range 50 to 220 were measured for each spectrum. Twelve PFTBA spectra were collected over three weeks, in order to assess longer term reproducibility issues. The spectra were then manually imported into Matlab2009b (Mathworks, Natick MA, USA) where the pool of spectra were analyzed using principal component analysis (PCA) software from PLS_Toolbox Version 6.2, (Eigenvector Research, Inc., Wenatchee WA, USA), and the scores of each PFTBA spectrum were used to calculate the Mahalanobis distance to determine the spectrum best suited, i.e., that with the smallest average Mahalanobis distance, for use as the library⁶⁹⁻⁷¹. The remaining spectra were then analyzed with respect to the library using a subsection of the TMSRA algorithm; the subsection entails from the start of the original algorithm to the calculation of the index mass purity plot (Eq. (2.4)). The reason for omitting the quantification portion of the algorithm is because the focus of this particular study was to qualitatively analyze the spectra and to validate that the run-to-run variance was relatively small.

The test compounds reported herein for the case study are 1-heptene and n-heptane, (Sigma-Aldrich, St. Louis MO, USA) were used to make mixtures (by mass between 450 to 750 mg). A 100 to 150 mg aliquot from the stock of standard mixtures was diluted in 5000 mg of acetone, resulting in seven standards of various mixtures of compounds of known

(68) E. Estrada, J. Chem. Inf. Comput. Sci. 35 (1995) 31–33.

(69) T.I. Dearing, J.S. Nadeau, B.G. Rohrback, L.S. Ramos, R.E. Synovec, Talanta. 83 (2011) 738–743.

(70) M. Fransson, S. Folestad, Chemom. Intell. Lab. Syst. 84 (2006) 56–61.

(71) M. Daszykowski, B. Walczak, J. Chromatogr. A. 1176 (2007) 1–11.

concentration between 2500 to 3700 ppm. In the case study, 1-heptene was designated as the target while n-heptane was used as the interferent. Other case studies with other compounds were performed, with results not presented for brevity, since the 1-heptane : n-heptane case was the most challenging and interesting case. Samples were run in sets of 6 replicates and the regions containing the target analytes were extracted. One 1-heptene replicate was arbitrarily chosen as the reference and its mass spectrum was designated as the library spectrum, the remaining five peaks were used for simulated peak overlap with the interfering n-heptane peak.

Data analysis was performed on an Athlon II X2 (AMD, 3.1 GHz) with 4 GB of RAM. Prior to analysis, GC-MS data files were imported into Matlab2009b (Mathworks, Natick MA, USA) using an in-house developed Matlab function. GC-MS data were subjected to m/z channel 'de-skewing' (details found elsewhere¹³) and the chromatograms were aligned using in-house software^{16,17,72,73}. For simulated chromatographic overlap, regions of the chromatograms containing well resolved analyte peaks (1-heptene and n-heptane) were extracted. For each replicate, the two chromatographic regions, one containing the targeted 1-heptene peak, with the other containing the n-heptane peak serving as the interferent, underwent baseline correction, and then were summed along the time axis to achieve peak overlap at a desired R_s as calculated in Eq. (2.1). The R_s range studied was 0 to 1.0 in increments of 0.1, and included the R_s values 1.25 and 1.5. The 'simulated peak overlap data' underwent data analysis processing following the algorithmic steps described in the Theory section. All the necessary information (the chromatographic regions used to extract the library spectrum (L), m/z range, expected t_R) was

(72) K.M. Pierce, J.L. Hope, K.J. Johnson, B.W. Wright, R.E. Synovec, J. Chromatogr. A. 1096 (2005) 101–110.

(73) N.E. Watson, M.M. VanWingerden, K.M. Pierce, B.W. Wright, R.E. Synovec, J. Chromatogr. A. 1129 (2006) 111–118.

input into the algorithm, which was allowed to process without further user intervention. The spectrum of S used for analysis was collected (via summation) across $\pm\sigma$ of the expected ${}^T t_R$ of the 1-heptene peak from which L was derived. The mass spectral threshold used was 1%. This threshold was chosen because of the results of the PFTBA experiment (as will be described later herein), which indicated the relative intensities of m/z with an intensity greater than 1% (of the most intense m/z) are sufficiently reproducible. Since there was no prior knowledge of what tolerance would be best to use in generating C_{ij} (and the eventual D_{ij}^2) with respect to the calculation of the peak height for a given T , a broad range of tolerances were investigated, ranging from 0.025 to 0.25, in increments of 0.025 (corresponding to deviation of 2.5% to 25%) for each R_s investigated. For brevity and clarity, only tolerances ranging from 0.05 to 0.2 are presented. Results (calculated peak heights and corresponding sets of m/z) obtained using TMSRA, for each R_s and at each tolerance value were plotted in Excel 2007 (Microsoft, Redmond WA, USA). After analysis, for each run, calculated peak heights were divided by their corresponding pure peak heights (same run prior to undergoing simulated peak overlap with an interfering peak), thereby 'normalizing' the quantitative result that should ideally have a value of 1 based upon Eq. (2.7). This process was performed to minimize the error due to instrumentation and help study the method presented in terms of both precision and accuracy.

2.4 Results and Discussion

2.4.1 PFTBA Study

The purpose of the PFTBA study was to assess the precision of the TMSRA algorithm and to determine suitable algorithmic tolerance levels. The library spectrum and a representative

sample spectrum are shown in Figure 2.2. The index key for PFTBA, indexed m/z (ordered by intensity) is shown in Table 2.2. PFTBA exhibited appreciable signal at 14 m/z values, and for example, $m/z = 219$ is the most intense so it has an index position of 1 ($m/z = 69$ is normally the most intense, but the tune parameters were set to emphasize higher, more selective m/z).

Figure 2.2 shows the spectra of the library and the sample and how they directly compare. In this case there are a few m/z in the targeted analyte PFTBA in the sample (Figure 2.2B) that are lower in intensity when compared to the library (Figure 2.2A); the variation is subtle but noticeable. Though it helps to look at the spectra directly it is important to go back to the original premise of this report and focus on the intensity ratios of the m/z as they offer a more sensitive means of comparing spectra.

Results for comparing the spectra of the sample to the library are shown in Figure 2.3. In Figure 2.3A, the indexed mass ratio matrix ${}^L R_{i,j}$ of the library spectrum is presented. In Figure 2.3B, the first column of ${}^L R_{i,j}$ ($j = 1$) is shown and using Eq. (2.2), the mass ratio matrix becomes easier to interpret: for the first column, the indexed m/z are divided by the first indexed m/z . Figure 2.3C is analogous to Figure 2.3A, only the sample spectrum is analyzed and ${}^S R_{i,j}$ is obtained, using Eq. (2.3). Likewise, Figure 2.3D presents the first column of ${}^S R_{i,j}$ in the same manner as described in Figure 2.3B (only Eq. (2.3) is used instead). Figure 2.3E is the indexed mass ratio matrices or 'Mass Purity Plots' obtained by applying Eq. (2.4); the differences in the intensity ratios between the library and the sample are subtle, which is expected for the PFTBA replicates. Another reason to use the intensity ratios instead of the intensities directly, is that the ratios also serve as a means of normalization across the various m/z channels. In Figure 2.3E, the 'Indexed Mass Purity Plot', one also observes that most m/z have $X_{i,j}$ values that fall below 0.05, (corresponding to 5% deviation) from the library. Figure 2.3F presents the first column of $X_{i,j}$;

only three m/z deviate beyond the arbitrary 0.05 tolerance value. More importantly there are no indexed mass channels with a large m/z ratio variance, a result that would indicate the existence of a bias in the m/z intensities.

Finally, statistical results for all 11 replicates of PFTBA showing the variance with respect to the library are shown in Figure 2.4. Figure 4A is the result of averaging $X_{i,j}$, i.e. the results after using Eqs. (2.2, 2.3 and 2.4) for the aforementioned 11 replicates with respect to the library standard spectrum. Figure 2.4B complements Figure 2.4A by showing the RSD for $X_{i,j}$. One may notice that the ratios between m/z are generally reproducible. The more intense m/z (top 9) typically have a RSD below 5%, and the remaining m/z have RSD values typically below 10%. These results suggest that based upon using PFBTA for assessing tolerances, 1% of the maximum intensity appears to be a suitable threshold for analyzing mass spectra. This threshold was applied in the subsequent case study demonstration in the next section of this report. It is important to note that the performance of the TMSRA algorithm depends heavily on the accuracy of the library spectrum as a representation of a given target analyte.

2.4.2 Case Study 1: 1-heptene and heptanes

For the case study we computationally generated a full range of chromatographic overlap (from a R_s of 0 to 1.5, using Eq. (2.1)) of the target analyte, 1-heptene, with the interferent, n-heptane. Their respective spectra are shown in Figure 2.5A. One can see that this is a relatively challenging case study since the two compounds exhibit significant signal intensity at many of the same m/z . Since there are few m/z that are selective for 1-heptene, its quantification would be a difficult feat at lower R_s , especially if the analyst does not know *a priori* which m/z

may or may not be sufficiently selective. The chromatographic peak heights of the two compounds are similar, as seen in Figure 2.5B. The sorted mass index information is provided in Table 2.3. The R_s values between 1-heptene and n-heptane that were analyzed were 1.5, 1.25 and from 1.0 to 0 at 0.1 increments. As stated in the Theory section, there was no *a priori* knowledge of which tolerance values were best suited to apply on the lower left triangle of $|X_{i,j}|$ for constructing the connectivity matrices ($C_{i,j}$ and eventually $D_{i,j}^2$, shown in Figure 2.1B) to use in the m/z selection process for the eventual quantification of the target analyte peak. Moreover, it was conceivable that the mass spectral run-to-run variation would result in different runs having different optimal tolerance values. Therefore, a range of different tolerance values were studied for each R_s ; the range of tolerance values were from 0.025 to 0.25 at 0.025 increments (corresponding to the maximum deviation and/or bias) the sample spectrum is allowed to deviate from the library in order to be considered as ‘pure’). An example of data encountered in intermediate steps in the algorithm is shown in Figure 2.6 at $R_s = 0.3$. The core of the process is using Eq. (2.2) and Eq. (2.3) to generate the intensity ratio matrices ${}^L R_{i,j}$, ${}^S R_{i,j}$, (such as Figures. 2.6A, 2.6B, respectively), and analyzing the differences between them using Eq. (2.4), resulting in $X_{i,j}$ (shown in Figure 2.6C). Setting the right tolerance values for $C_{i,j}$ is a balancing act; too strict and some good m/z will be thrown out, too lax and contributions from an interfering compound will add to the sum of intensity, $\sum I_{\text{set}}$. Fortunately, the algorithmic computation time is short, so many tolerance values can be rapidly evaluated. An example of $D_{i,j}^2$ resulting from a good tolerance selection is shown in Figure 2.6D. The final results (TIC equivalent peak height as a function of tolerance values) can be compared either by the user, or ultimately by an automated tolerance value selection process.

Statistical results of quantifying 1-heptene using Eq. (2.5) can be found in Figure 2.7A

and B. The sets of m/z used for quantification were recorded and the average size of sets of mass channels is shown in Figure 2.7C. One can observe in Figure 2.7A that for all tolerances at R_s between 0.6 to 1.5, quantitative results are in good agreement, with the concentration ratios, F , having RSD values typically below 2% (most being below 1%) and deviating generally within $\pm 1\%$ from the expected value of 1. At low R_s (0 to 0.5, see Figure 2.7B) as the R_s decreases, both the deviations in F from a value of 1 and the RSD increase (for the best tolerance values, which are typically 0.1 and 0.125, the deviation in F is $\pm 1\%$ to 9%, and the RSDs is between 2% to 9%). Looking at Figure 2.7C one can see that as the R_s decreases, the average size of the set of m/z used for quantification decreases. Also as the tolerance values decrease, the average size of the set of m/z shrinks. Note that if the size of the best set of m/z becomes too small (below three m/z) due to a decrease in tolerance or R_s , the next set of m/z (regardless of its size) that produces the smallest $\sum^S I_{\text{total (calc)}}$ (and therefore the smallest F) is chosen for quantification.

Regarding the tolerance value for the connectivity matrix for this case study (and others not shown for brevity), upon comparing results for individual replicates, no single tolerance value repeatedly performed best in every replicate sample, though in the case study presented the tolerance value of 0.1 generally fared best, it did not produce the best results in every replicate run and at every R_s value (sometimes the tolerance value 0.125 performed better). For practical use, an algorithm that converges on the most correct tolerance value on a run by run basis should be considered.

2.4.3 Case Study 2: Octane and 1,4-Dioxane

For the second case study, the effects of chromatographic overlap as described in the

previous case study were investigated with respect to octane as the target and 1,4-dioxane as the interferent. A comparison between the spectra of octane and 1,4-dioxane is shown in Figure 2.8A, and an example of simulated overlap as well as comparing the relative peak heights of the target analyte relative to the interferent is shown in Figure 2.8B. The sorted mass index information is provided in Table 2.4. Since octane exhibits more selective m/z compared to 1-heptene in the previous case study and the fact that the interfering peak is relatively small (about a third of the height), this case study was considered easier. ${}^L R_{i,j}$, ${}^S R_{i,j}$, $X_{i,j}$, and $D_{i,j}^2$ for this case are shown in Figures 2.9. The calculated peak maximum at various R_s values and the average size of the set of m/z used for quantification are shown in Figure 2.10.

Figures 2.10A, and B show that at resolutions above 0.4 the laxest tolerance values fared best. Note however at resolutions ≤ 0.4 the most stringent of tolerances (typically 0.05) fared better, being within ± 0.015 of 1 and with an RSD of around 1%. Overall the results were all within ± 0.02 of 1 with a RSD below 2.5%. Figure 2.10C shows that at lower tolerances the size of the set of m/z used for quantification decreases with R_s , though not as pronounced as in Case 1. These results further confirm that a tolerance value must be selected on a case by case basis for optimum results.

2.4.4 Study Case 3: 1,4-Dioxane and Octane

For the last case study, the effects of chromatographic overlap were investigated with respect to the 1,4-dioxane as the target and octane as the interferent. The sorted mass index information is provided in Table 2.5. Once again, the target analyte was quantified at various resolutions from its respective interferent and at a variety of tolerance values. The peak height of

the target analyte is smaller relative to the interferent, likewise there are fewer selective m/z for 1,4-dioxane (due to the size of the molecule and its fragmentation), making the final case study the most difficult of all cases studied. The ${}^L R_{i,j}$, ${}^S R_{i,j}$, $X_{i,j}$, $D_{i,j}^2$ for case 3 are shown in Figure 2.11. The calculated peak maximum at various R_s values and the average size of the set of m/z used for quantification are shown in Figure 2.12.

For tolerance values of 0.025 and 0.05, especially at resolutions below 1, the algorithm would not always find a set of three m/z to use for quantification. For this reason, a 0.05 tolerance value showed a strong negative bias and huge RSD in the statistical results. Figures 2.12A, and B show once again as with the previous case studies: when quantifying the peak maximum, there is no one tolerance value that works best for all R_s , though generally all tolerance values achieve results within 0.03 of 1. At higher R_s values, the % RSD (1.5-0.6) was typically around 2% and at lower R_s values (0.5-0.0) RSD was always below 4%, typically below 3%. One exception is for tolerance 0.25 at R_s of 0.8, where the best set of m/z used in a single run to quantify was severely interfered therefore the average peak maximum and RSD is skewed with positive bias. Figure 2.12C shows that at lower tolerances, the size of the set of m/z used for quantification decreases primarily due to R_s .

2.4.5 TMSRA Computation Time

Finally, the computation time of the TMSRA algorithm was investigated. We measured the time required for the algorithm to analyze a portion of a given sample GC-MS chromatogram that contains a target analyte. For one hundred runs the average computation time was 0.027 s with a standard deviation of about 7.5 ms. Note that the computation time could potentially be

shortened by further optimization of the code.

2.5 Conclusions

The TMSRA algorithm has the ability to quantify, using sufficiently selective m/z , a targeted analyte over a wide range of R_s relative to an interfering peak, with a fast computation time. At this stage the basic features of TMSRA have been presented with method development (PFTBA study) and method demonstration (case study with 1-heptene and n-heptane). For the targeted analyte 1-heptene in the case study, at the higher R_s studied ($0.6 \leq R_s \leq 1.5$) a deviation in the concentration ratio, F , within $\pm 1\%$ and a RSD generally below 1% were achieved. As the R_s decreased, the deviation and RSD both increased. At a $R_s = 0$, a deviation of $\sim 9\%$ and a RSD of $\sim 9\%$ were achieved. Regarding the tolerance value for the connectivity matrix, based on these study cases it appears that there is no "one value fits all". In the first study case, tolerance 0.1 fared best while in the second study case tolerance 0.05 yielded the best results, and in the third study case there was no one tolerance that worked best and the strictest tolerances often did not find any set of m/z . For practical use, an algorithm that converges on the most correct tolerance value on a run by run basis should be considered. While other approaches to use selective m/z with GC-MS have been reported, we utilized a novel approach requiring "internal consistency" with the list of sufficiently pure m/z that are found. This internal consistency should result in less deviation (more accurate) quantification for high speed analysis of targeted analytes with GC-MS.

Figure Captions

A

$$\begin{array}{ccc}
 \begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix} & \rightarrow & \begin{bmatrix} 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix} \\
 A_{i,j} & & D_{i,j}
 \end{array}$$

B

$$\begin{array}{ccc}
 \begin{bmatrix} 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix} & = & \begin{bmatrix} 2 & 1 & 1 & 0 & 0 \\ 1 & 2 & 1 & 0 & 0 \\ 1 & 1 & 2 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \\
 & & D_{i,j}^2
 \end{array}$$

Figure 2.1. (A) An example of the diagonal being set to a value of 0 (removed) in $A_{i,j}$ resulting in $D_{i,j}$. (B) The second matrix power of $D_{i,j}$ results in $D_{i,j}^2$. In the off diagonal, nonzero elements of $D_{i,j}^2$ indicates sets of at least three m/z that are candidates for quantification.

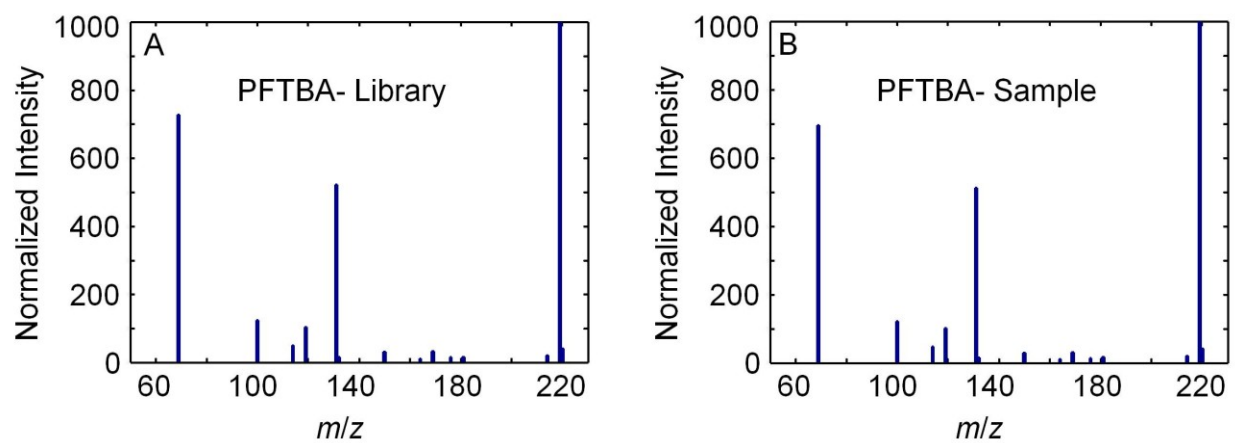


Figure 2.2. (A) The spectrum of PFTBA (library), m/z range of 50 to 220. (B) The spectrum of a PFTBA run (sample), m/z range of 50 to 220.

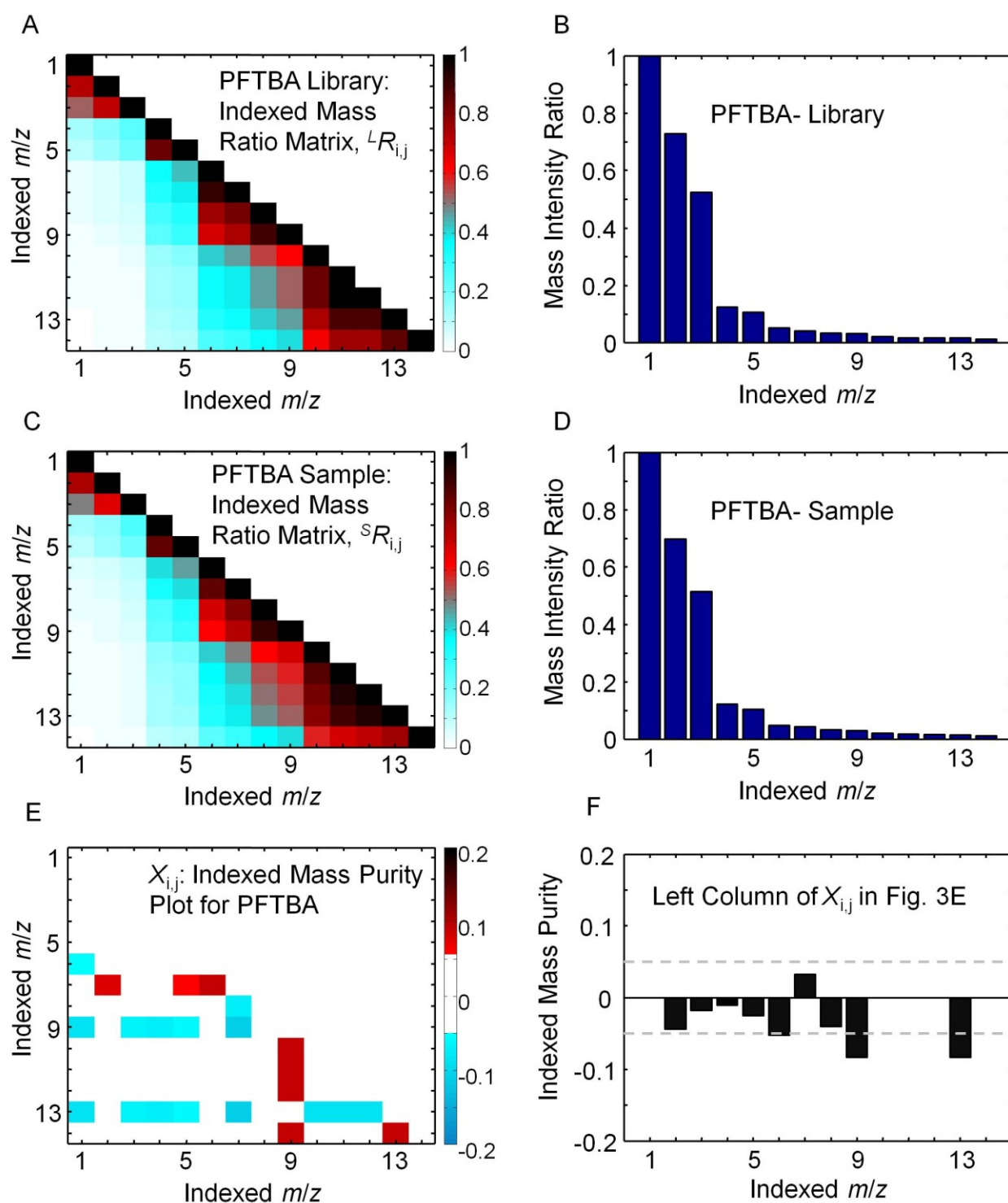


Figure 2.3. (A) Indexed Mass Ratio Matrix, ${}^L R_{i,j}$, for the PFTBA indexed library spectrum, obtained by taking the outer product of the spectrum with respect to its inverse as calculated in Eq. (2.2). (B) Indexed Mass Ratios of the first column of ${}^L R_{i,j}$ seen in Figure 3A. (C) Indexed Mass Ratio Matrix, ${}^S R_{i,j}$, for the PFTBA indexed sample spectrum, obtained by taking the outer product of the spectrum with respect to its inverse as calculated in Eq. (2.3). (D) Indexed Mass Ratios of the first column in ${}^S R_{i,j}$ seen in Figure 3C. (E) Indexed Mass Purity $X_{i,j}$ for PFTBA. (F) Indexed Mass Purity of the left column of $X_{i,j}$ in Figure 3E.

(E) Indexed Mass Purity Plot results from calculating the difference between ${}^L R_{i,j}$ and ${}^S R_{i,j}$, as shown in Eq. (2.4). (F) Indexed Mass Purity Ratios of the first column in $X_{i,j}$ seen in Figure 3E. The dashed lines mark the $\pm 5\%$ deviation.

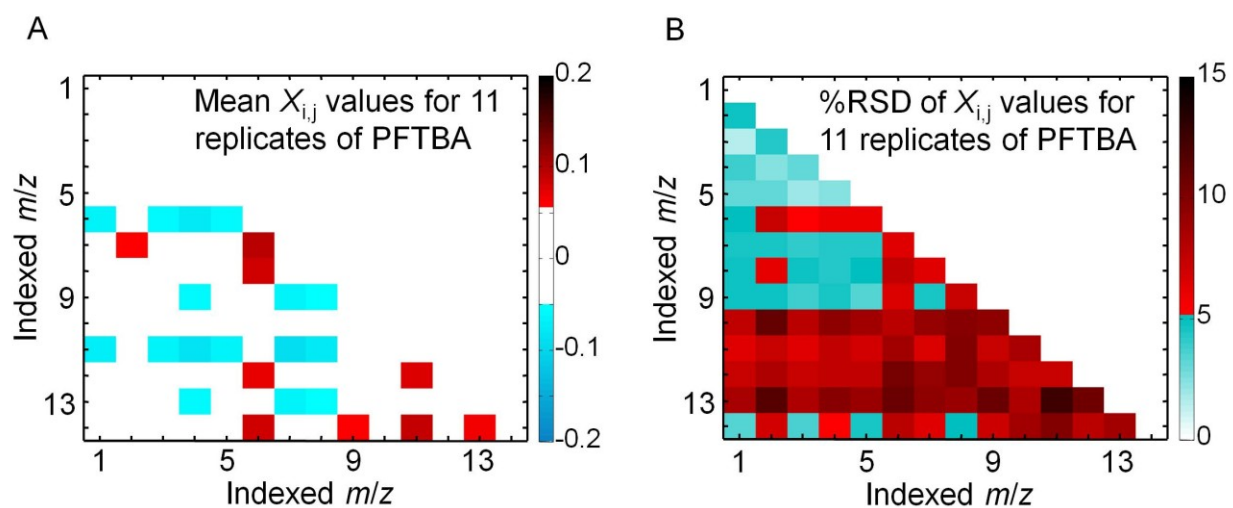


Figure 2.4. (A) Statistical results (average of PFTBA pairing of 11 samples with the single library standard) for the Indexed Mass Purity plots of PFTBA. (B) Statistical results (%RSD) for Indexed Mass Purity Plots of PFTBA.

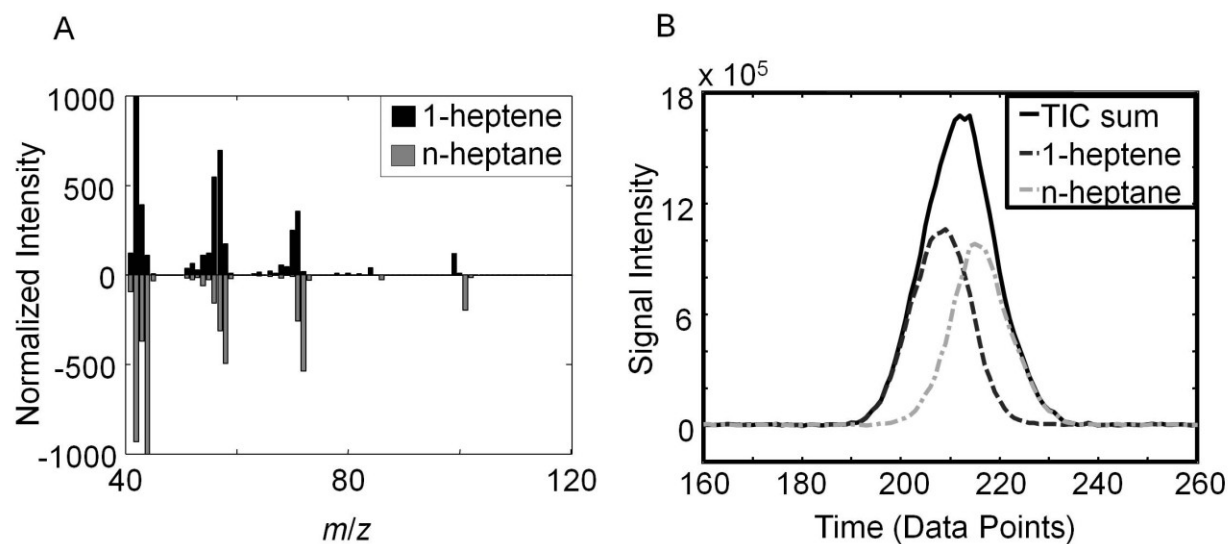


Figure 2.5. (A) Case study with the mass spectrum of 1-heptene compared to n-heptane. (B) Total ion current (TIC) signal illustrates the overlap resulting from overlaying and summing two GC-MS chromatographic portions, 1-heptene (on the left) at $R_s = 0.3$ from n-heptane.

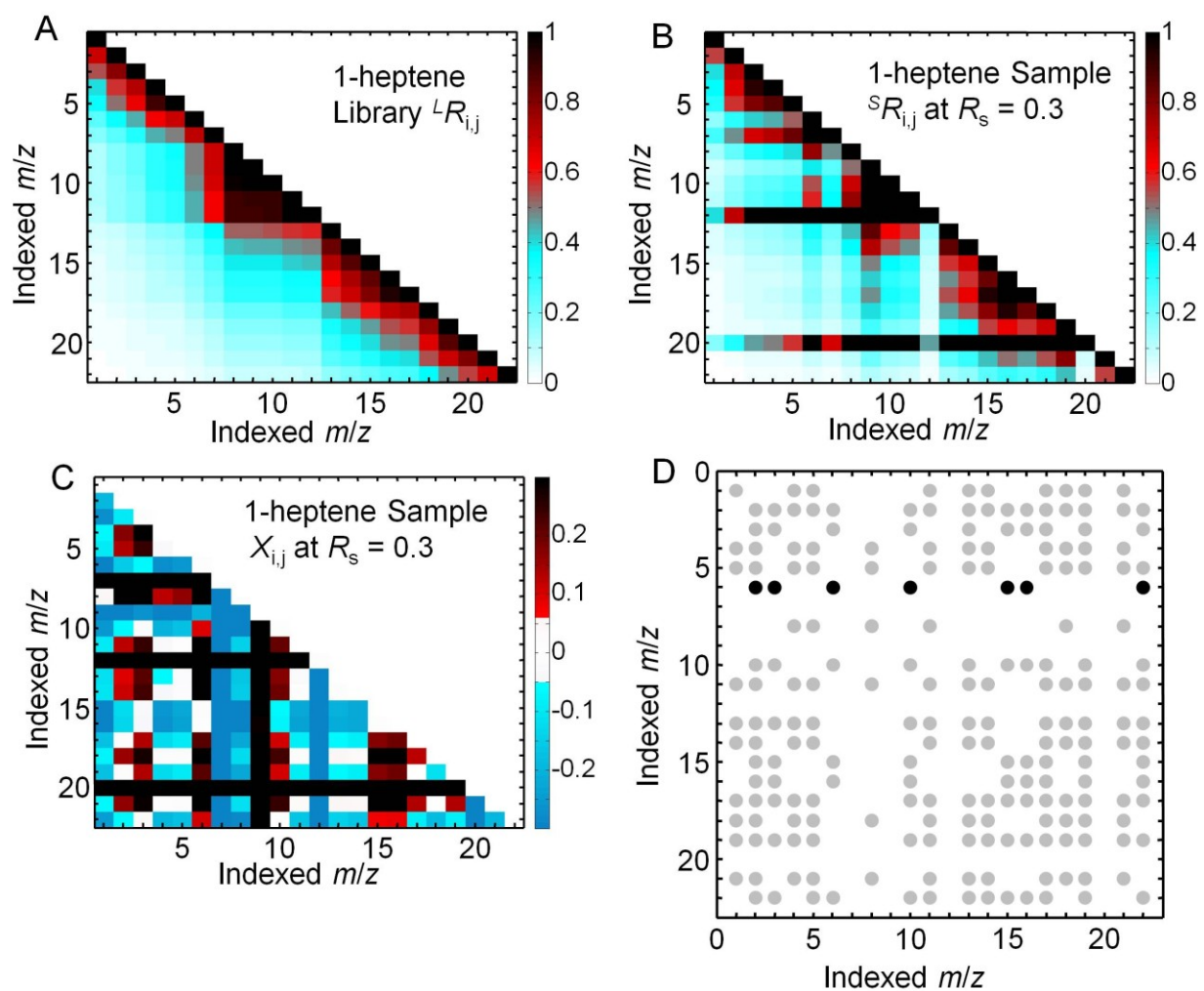


Figure 2.6. (A) Indexed Mass Ratio Matrix ${}^L R_{i,j}$ for the case study (data in Figure 5). Obtained using the 1-heptene indexed library spectrum and using Eq. (2.2). (B) Indexed Mass Ratio Matrix ${}^S R_{i,j}$ of the case study (at $R_s = 0.3$). Obtained by using the sample spectrum indexed according to ${}^L R_{i,j}$, and using Eq. (2.3). (C) Indexed Mass Purity Plot ($X_{i,j}$) result from calculating the difference between Figures. 6A and 6B, using Eq. (2.4). (D) $D_{i,j}^2$ connectivity matrix with tolerance at 0.10 (10%), derived from Indexed Mass Purity Plot (Figure 6C) for the condition that $R_s = 0.3$.

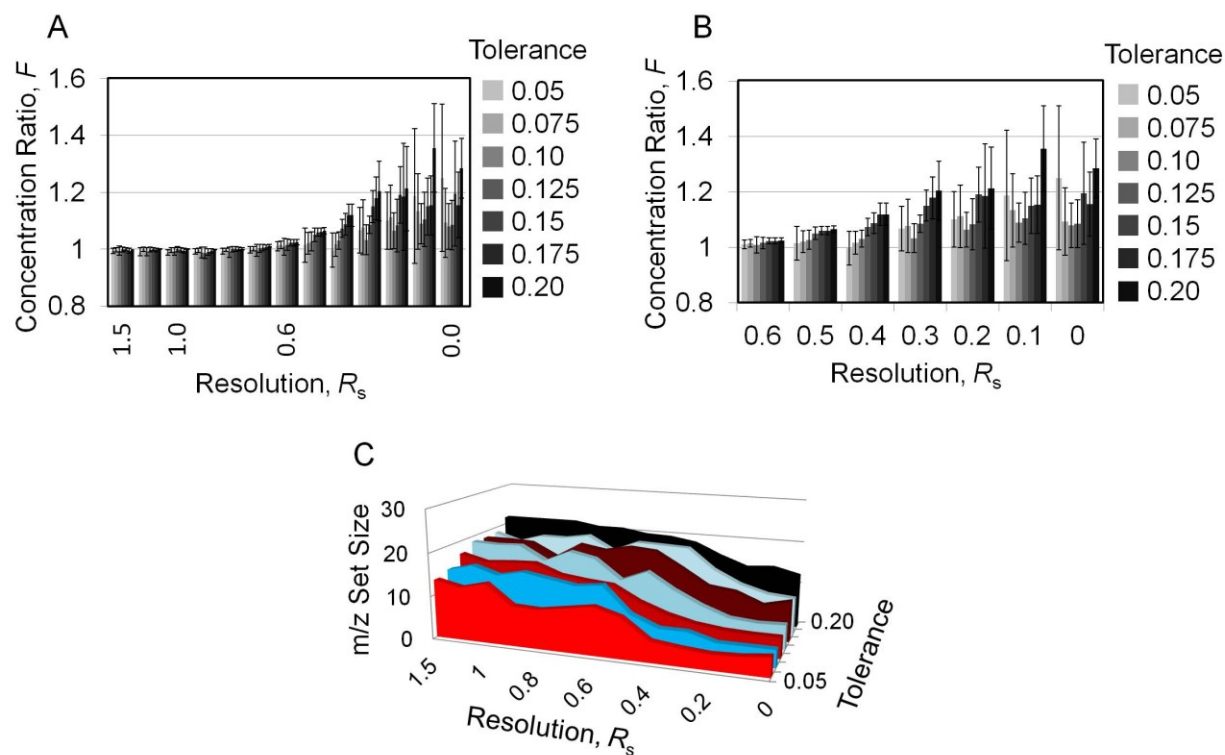


Figure 2.7. (A) Quantitative results for the concentration ratio, F , for 1-heptene at various R_s , by selecting the best m/z set at various tolerance values, using Eqs. (2.5, 2.6, and 2.7). The best m/z set provides the minimum quantification value (lowest potential for bias). Tolerances 0.025, 0.225 and 0.25 were omitted, and only R_s values 1.5, 1.0, 0.6 and 0.0 were indicated on the x-axis for clarity. The full range of R_s values are 1.5, 1.25, and from 0.0 to 1.0 by increments of 0.1. (B) Quantitative results, F , for 1-heptene at various tolerances, at R_s ranging from 0 to 0.6. Tolerances 0.025, 0.225 and 0.25 were omitted for clarity. (C) Average size of set of m/z used to calculate the peak maximum shown in Figures. 7A, 7B as a function of R_s and tolerance for C_{ij} . The tolerance values shown, in ascending order are 0.05 0.075, 0.10, 0.125, 0.15, 1.175, and 0.20.

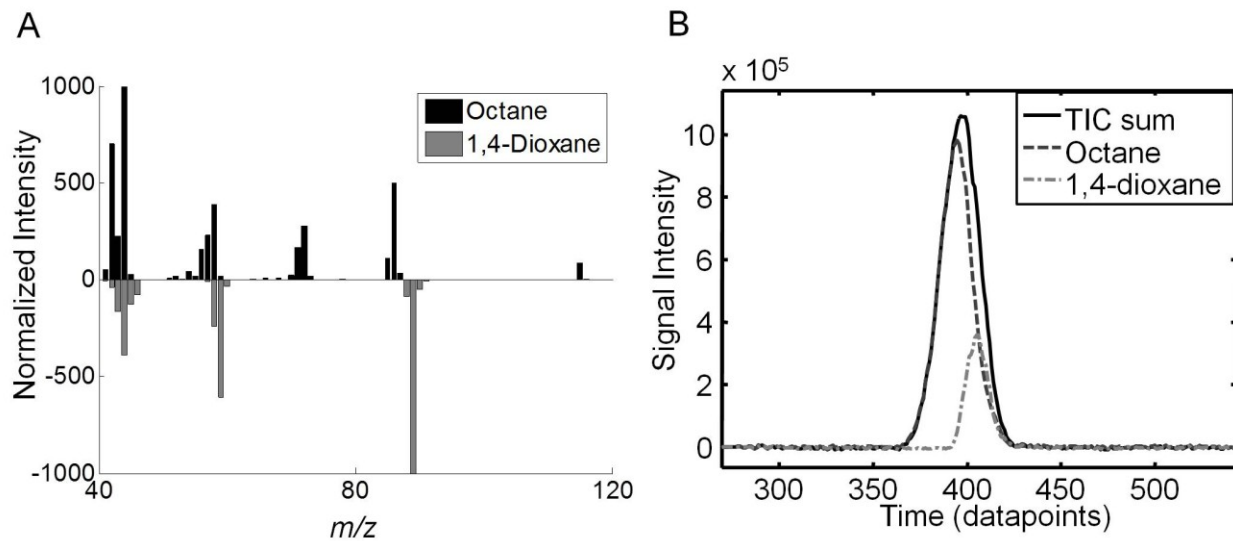


Figure 2.8. (A) Case study 2. Mass spectrum of octane compared to 1,4-dioxane. (B) ‘TIC overlap’, resulted from overlaying and summing two sub-chromatograms, octane (on the left) at R_s 0.3 from 1,4-dioxane.

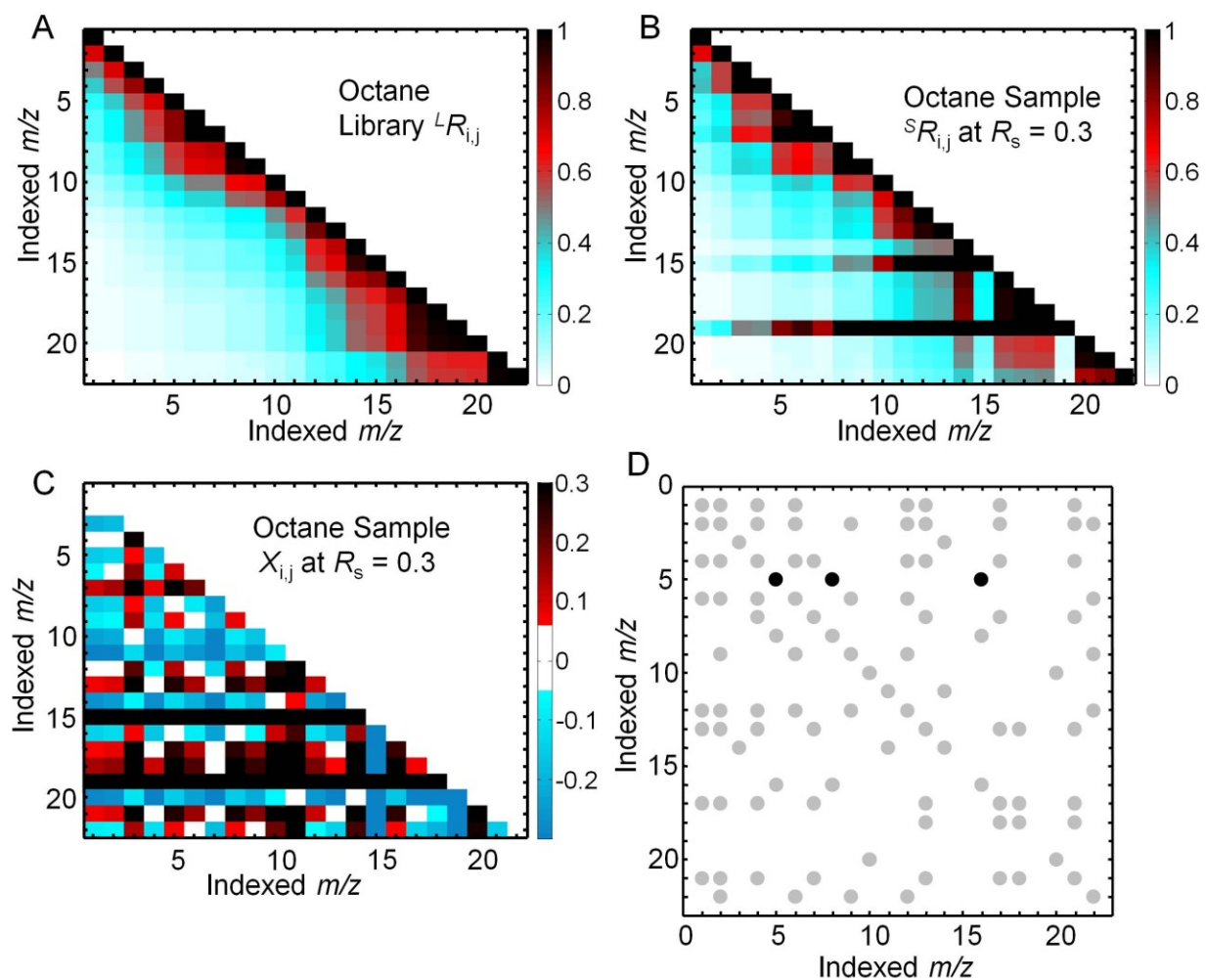


Figure 2.9. (A) Indexed Mass Ratio Matrix ${}^L R_{ij}$ for case study 2. Obtained using octane's indexed library spectrum, and using Eq. (2.2). (B) Indexed Mass Ratio Matrix ${}^S R_{ij}$ of case study 2 (at R_s 0.3). Obtained by using the sample spectrum indexed according to ${}^L R_{ij}$, and using Eq. (2.3). (C) Indexed Mass Purity Plot (X_{ij}) results for calculating the difference between Figures. 9A and 9B, using Eq. (2.4). (D) D_{ij}^2 connectivity matrix with tolerance at 0.05 (5%), from Indexed Mass Purity Plot (Figure 9C) in a scenario of R_s 0.3.

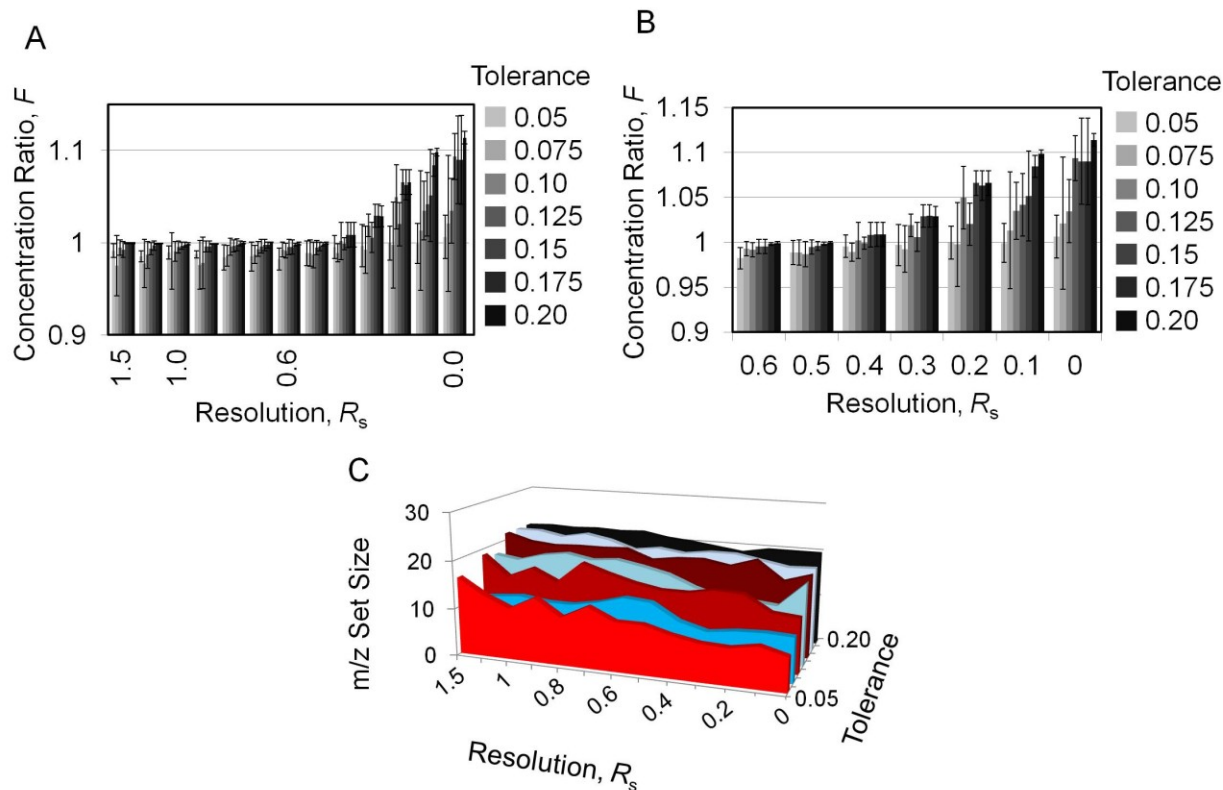


Figure 2.10. (A) Average results for peak maximum for octane at various R_s by selecting the best m/z set at various tolerance values. Tolerances 0.025, 0.225 and 0.25 were omitted for brevity. (B) Average results for peak maximum for octane, at various tolerances, at R_s ranging from 0.6 to 0. Tolerances 0.025, 0.225 and 0.25 were omitted for brevity. (C) Average size of set of masses used to calculate the peak maximum shown in Figures 10A, 10B as a function of R_s and tolerance for C_{ij} . Tolerances 0.025, 0.225 and 0.25 were omitted for brevity.

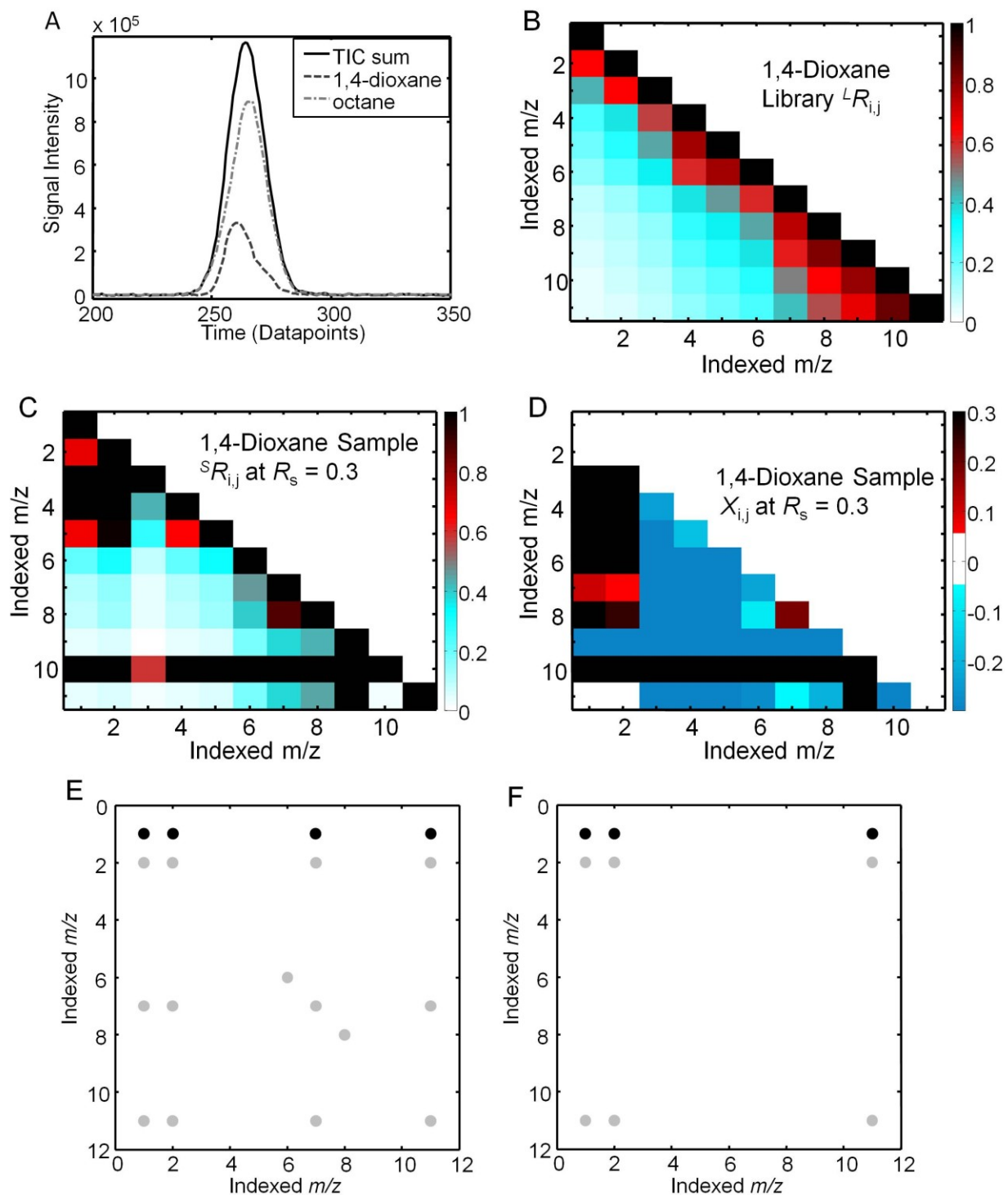


Figure 2.11. Case study 3. (A) spectra shown in Figure 8A. ‘TIC overlap’, resulted from overlaying and summing two sub-chromatograms, 1,4-dioxane (on the left) at R_s 0.3 from octane. (B) Indexed Mass Ratio Matrix ${}^L R_{i,j}$ for case study 3. Obtained using 1,4-dioxane’s indexed library spectrum, and using Eq. (2.2). (C) Indexed Mass Ratio Matrix ${}^S R_{i,j}$ of case study 3 (at R_s 0.3). Obtained by using the sample spectrum indexed according to ${}^L R_{i,j}$, and using Eq. (2.3). (D) Indexed Mass Purity Plot ($X_{i,j}$) results for

calculating the difference between Figures. 9A and 9B, using Eq. (2.4). (E) D_{ij}^2 connectivity matrix with tolerance at 0.05 (5%), from Indexed Mass Purity Plot (Figure 12C) in a scenario of R_s 0.3. (F) D_{ij}^2 connectivity matrix with tolerance at 0.10 (10%), from Indexed Mass Purity Plot (Figure 12C) in a scenario of R_s 0.3. Showing in this case the set of masses is limited.

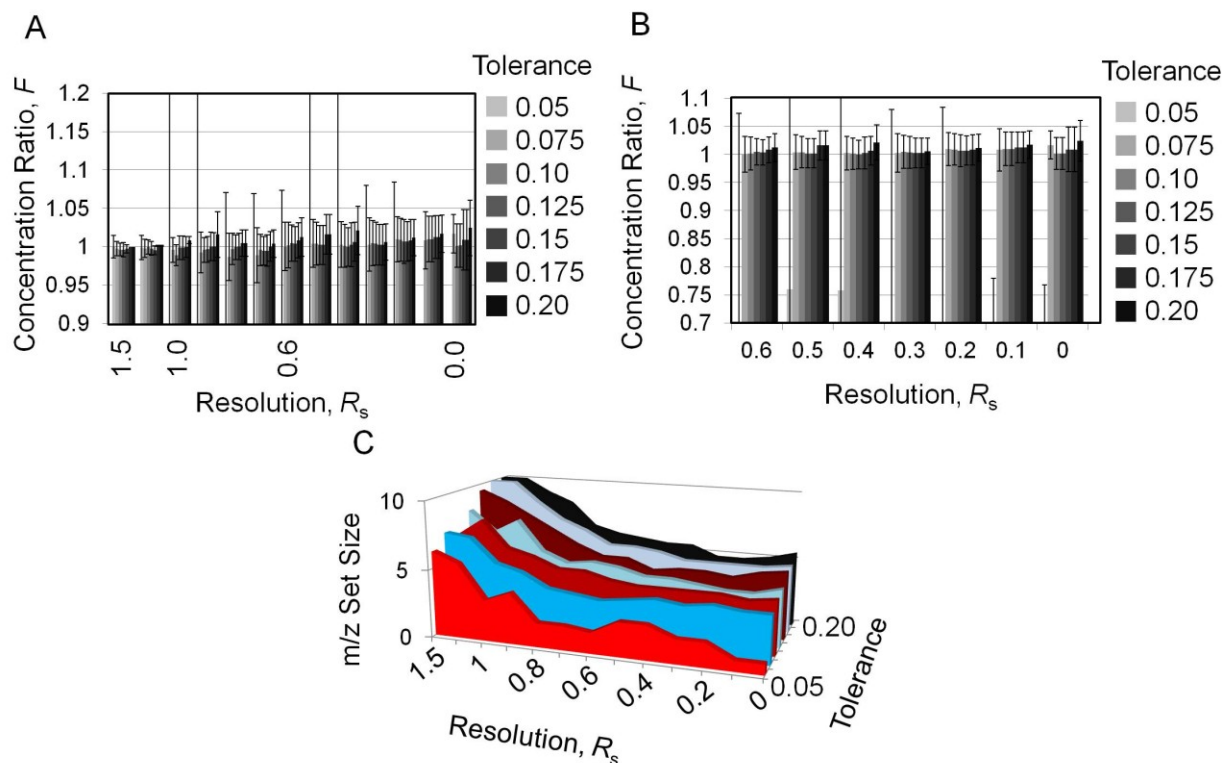


Figure 2.12. (A) Average results for peak maximum for 1,4-dioxane at various R_s by selecting the best m/z set at various tolerance values. Tolerances 0.025, 0.225 and 0.25 were omitted for brevity. (B) Average results for peak maximum for 1,4-dioxane, at various tolerances, at R_s ranging from 0.6 to 0. Tolerances 0.025, 0.225 and 0.25 were omitted for brevity. (C) Average size of set of masses used to calculate the peak maximum shown in Figures. 13A, 13B as a function of R_s and tolerance for C_{ij} . Tolerances 0.025, 0.225 and 0.25 were omitted for brevity.

Table Captions

Table 2.1. TMSRA ratio matrices. (A) A resulted matrix when calculating $R_{i,j}$ for all i and j in the spectrum. (B) The resulted matrix for calculating $X_{i,j}$ for all i and j values.

(A)

$R_{i,j}$ Intensity Ratios for two m/z values			
i=1	$R_{(1,1)}=1$	$R_{(1,2)}>1$	$R_{(1,3)}\gg 1$
i=2	$R_{(2,1)}<1$	$R_{(2,2)}=1$	$R_{(2,3)}>1$
i=3	$R_{(3,1)}\ll 1$	$R_{(3,2)}<1$	$R_{(3,3)}=1$
	j=1	j=2	j=3

(B)

$X_{i,j}$ Relative Mass Purity			
i=1	$X_{(1,1)}=0$	$X_{(1,2)}\approx\bar{1}0$	$X_{(1,3)}\approx\bar{1}0$
i=2	$X_{(2,1)}\approx\pm 0$	$X_{(2,2)}=0$	$X_{(2,3)}\approx\pm 0$
i=3	$X_{(3,1)}\approx\pm 0$	$X_{(3,2)}\approx\bar{1}0$	$X_{(3,3)}=0$
	j=1	j=2	j=3

Table 2.2. The index key for PFTBA containing m/z of intensity above the minimum threshold.

Index	m/z
1	219
2	69
3	131
4	100
5	119
6	114
7	220
8	169
9	150
10	214
11	132
12	176
13	181
14	164

Table 2.3. The index key for sorted mass channels of 1-heptene.

Index	<i>m/z</i>	Index	<i>m/z</i>
1	41	12	43
2	56	13	51
3	55	14	67
4	42	15	68
5	70	16	83
6	69	17	50
7	57	18	53
8	40	19	65
9	54	20	71
10	98	21	63
11	53	22	66

Table 2.4. The index key for sorted mass channels of octane.

index	<i>m/z</i>	index	<i>m/z</i>
1	43	12	40
2	41	13	53
3	85	14	86
4	57	15	44
5	71	16	69
6	42	17	54
7	56	18	51
8	70	19	58
9	55	20	72
10	84	21	67
11	114	22	50

Table 2.5. The index key for sorted mass channels of 1,4-dioxane.

index	m/z
1	88
2	58
3	43
4	57
5	42
6	44
7	45
8	87
9	89
10	41
11	59

Chapter 3: Correlation of rocket propulsion fuel properties with chemical composition using comprehensive two-dimensional gas chromatography with time-of-flight mass spectrometry followed by partial least squares regression analysis^{74*}

3.1 Introduction

The chemical composition of a kerosene fuel is complex, and changes in composition have been widely demonstrated to impact fuel properties and performance^{75–78}. However, achieving precise control over the chemical composition of distillate fuels such as RP-1 (MIL-DTL-25576E) is challenging due to variations in crude oil composition and place of origin, refinery and post-refinery operating conditions, or even the date and time the material was refined, treated, and formulated to meet the detail specification requirements. A better understanding of fuel composition and how it relates to fuel performance and properties is expedient for a number of reasons. Indeed, it has become increasingly important to gain a better

* Large portions of this chapter published in: B. Kehimkar, J.C. Hoggard, L.C. Marney, M.C. Billingsley, C.G. Fraga, T.J. Bruno, Synovec RE, J Chromatogr. 1327 (2014) 132140.

(74) B. Kehimkar, J.C. Hoggard, L.C. Marney, M.C. Billingsley, C.G. Fraga, T.J. Bruno, Synovec RE, J Chromatogr. 1327 (2014) 132140.

(75) D.J. Cookson, B.E. Smith, Energy Fuels. 4 (1990) 152–156.

(76) G. Liu, L. Wang, H. Qu, H. Shen, X. Zhang, S. Zhang, Z. Mi, Fuel. 86 (2007) 2551–2559.

(77) M.L. Huber, E.W. Lemmon, T.J. Bruno, Energy Fuels. 23 (2009) 5550–5555.

(78) M.J. DeWitt, T. Edwards, L. Shafer, D. Brooks, R. Striebich, S.P. Bagley, M.J. Wornat, Ind. Eng. Chem. Res. 50 (2011) 10434–10451.

understanding of fuel composition, and an assessment of the potential sources of fuel composition variation is paramount to maintain control of fuel performance^{79–85}. It is also beneficial to relate new chemical analysis technologies to the benchmarking ASTM methods for characterizing properties and compositions of fuels such as RP-1. For such assessments, it is often beneficial to evaluate special laboratory blends (where the analyst has some control over the chemical composition, see Table 1) as well as to assess the performance of “field” fuels^{79–83}.

Gas chromatography coupled with mass spectrometry (GC-MS) is an established instrumental platform of the chemical analysis laboratory. GC-MS has proven itself as a powerful tool for the study of the chemical composition of complex samples including but not limited to fuels^{16,79–84,86,87}. Nevertheless, GC-MS can be made substantially more powerful by adding another dimension of separation, that is, by performing comprehensive two-dimensional (2D) gas chromatography prior to time-of-flight mass spectrometry detection (GC × GC –

(79) M.C. Billingsley, T. Edwards, L.M. Shafer, T.J. Bruno, *Proc. 46th AIAA/ASME/SAE/ASEE Joint Propulsion Conference*, 2010.

(80) T.M. Lovestead, B.C. Windom, J.R. Riggs, C. Nickell, T.J. Bruno, *Energy Fuels*. 24 (2010) 5611–5623.

(81) R.V. Gough, T.J. Bruno, *Energy Fuels*. 27 (2013) 294–302.

(82) P.Y. Hsieh, K.R. Abel, T.J. Bruno, *Energy Fuels*. 27 (2013) 804–810.

(83) J.L. Burger, T.J. Bruno, *Energy Fuels*. 26 (2012) 3661–3671.

(84) N.J. Begue, J.A. Cramer, C. Von Bargaen, K.M. Myers, K.J. Johnson, R.E. Morris, *Energy Fuels*. 25 (2011) 1617–1623.

(85) T.J. Bruno, L.S. Ott, T.M. Lovestead, M.L. Huber, *J. Chromatogr. A*. 1217 (2010) 2703–2715.

(86) R.B. Wilson, W.C. Siegler, J.C. Hoggard, B.D. Fitz, J.S. Nadeau, R.E. Synovec, *J. Chromatogr. A*. 1218 (2011) 3130–3139.

(87) J.H. Christensen, G. Tomasi, A.B. Hansen, *Environ. Sci. Technol.* 39 (2005) 255–260.

TOFMS)^{24,25,48,51,52,88-91}. GC × GC – TOFMS is well suited for the analysis of complex mixtures of volatile (and/or semi-volatile) compounds, such as those that are present in fuels^{9,24,25,48,51,88-91}. With GC × GC – TOFMS, the 2D separation commonly involves the first separation dimension performed using a non-polar stationary phase column and a separation run time of 30 to 60 minutes^{9,52,88} followed by a polar stationary phase column, providing a complementary separation relative to the first column, so chemical compounds not separated on the first dimension at a given retention time have the opportunity to be separated on the second dimension. However, for the current study of RP-1 fuels, which contain primarily *n*-alkanes, *iso*-alkanes, cyclic alkanes, and relatively fewer aromatics, it was deemed necessary to apply a “reversed column” GC × GC format in order to provide better selectivity⁹, with the first dimension separation using a polar phase (RTX-wax) and the second dimension separation using a non-polar phase (RTX-1). GC × GC – TOFMS provides a considerable amount of data for a given complex sample (e.g., typically ~ 300 to 400 MB compressed per sample run). It has become clear there are also significant challenges to readily glean useful information from this significant amount of data, which is why powerful chemometric software methods are used for analysis^{9,23-25,35,36,48,51,52,88-91}.

Even though GC × GC – TOFMS is a powerful instrumental platform for fuels analysis, it is critical to develop and apply data analysis software to convert the immense data into readily interpretable and useful information. For this purpose, multivariate data analysis methods have

(88) R.E. Mohler, K.M. Dombek, J.C. Hoggard, E.T. Young, R.E. Synovec, *Anal. Chem.* 78 (2006) 2700–2709.

(89) C.G. Fraga, B.J. Prazen, R.E. Synovec, *Anal. Chem.* 72 (2000) 4154–4162.

(90) B.J. Prazen, K.J. Johnson, A. Weber, R.E. Synovec, *Anal. Chem.* 73 (2001) 5677–5682.

(91) K. J. Johnson, R. E. Synovec, *J. Chemom. Intell. Lab. Syst.* 60 (2002) 225.

been developed, broadly referred to as “chemometrics”. Chemometrics is ideally suited to reveal similarities and/or differences between sets of GC \times GC – TOFMS data^{23,25,48,51,88,90,91}. Specific to this study, the partial least-squares (PLS) regression analysis was used to associate differences in measurable information for each fuel sample, in this case chemical and physical property data, to chemical composition differences as provided by the signal intensity differences, from one compound peak to the next, in the GC \times GC – TOFMS data. Details on PLS theory can be found elsewhere^{19,92–95}. Briefly, PLS analyzes two data matrices (X-block and Y-block, respectively) and calculates loadings referred to by the number of latent variables (LVs). Using PLS, models are constructed to account for variance (ideally, the relevant chemical differences) in both the GC \times GC – TOFMS data, i.e., the signal intensities (which constitute the X-block) and the respective measured property values (which constitute the Y-block). Thus, PLS modeling ideally provides a one-to-one correspondence between the measured values (plotted on the x-axis as the benchmark method) relative to the predicted values (plotted on the y-axis using the GC \times GC – TOFMS data in the PLS model).

In this study, we sought to demonstrate the potential of the GC \times GC – TOFMS instrumental platform, combined with PLS analysis, to provide useful information in the chemical analysis of RP-1 samples. By doing so, our goals were to demonstrate the feasibility of being able to (1) build PLS models to relate chemical composition data obtained from the GC \times GC – TOFMS to measured fuel performance quantities (e.g., density, kinematic viscosity, net heat of combustion, and so on), and then to (2) assess the quality of those models for

(92) K.J. Johnson, B.J. Prazen, D.C. Young, R.E. Synovec, *J. Sep. Sci.* 27 (2004) 410–416.

(93) F. Westad, N.K. Afseth, R. Bro, *Anal. Chim. Acta.* 595 (2007) 323–327.

(94) T. Rajalahti, O.M. Kvalheim, *Int. J. Pharm.* 417 (2011) 280–290.

(95) A.A. Gowen, G. Downey, C. Esquerre, C.P. O’Donnell, *J. Chemom.* 25 (2011) 375–381.

subsequent prediction of fuel chemical and physical characteristics without making direct measurements. Eventually, this chemical analysis approach will provide insight into addressing (3) the overall goal of optimizing fuel composition to meet desired fuel property and performance characteristics. To begin to address these goals by use of GC \times GC – TOFMS with PLS analysis, a key focus is to be able to elucidate chemical compounds or compound classes responsible for observed differences between fuels (e.g. type, feedstock) or differences in their measured physical properties. Specifically we report the use of GC \times GC – TOFMS with PLS to model and predict measured fuel properties (density, kinematic viscosity, net heat of combustion)⁸⁰.

3.2 Experimental

Several RP-1 fuel samples were obtained from the Air Force Research Laboratory (AFRL), and have been studied in prior reports^{79,80}, as listed in Table 1. All chromatographic data were obtained using the GC \times GC – TOFMS consisting of an Agilent 6890N GC (Agilent Technologies, Palo Alto, CA, USA), a thermal modulator (4D upgrade, LECO, St. Joseph, MI, USA), and a Pegasus III TOFMS (LECO, St. Joseph, MI, USA). Aliquots of the RP-1 fuel samples were introduced to the GC \times GC – TOFMS instrument via a 7683B auto-injector (Agilent Technologies, Palo Alto, CA, USA). The following experimental conditions were applied. The auto-injector was set to a 1 μ l injection, using a 200:1 split injection with helium carrier gas. Acetone was used for the solvent rinse. Isobaric mode was used with an inlet pressure of 35 psig (241 kPa). The GC oven initial temperature was set to 40 $^{\circ}$ C for 2 min and ramped to 225 $^{\circ}$ C at a rate of 6 $^{\circ}$ C/min where the final temperature was maintained for 3 min.

The GC inlet temperature was set to 225 °C and the transfer line temperature was set to 235 °C. The thermal modulator offset was 20 °C, with a hot pulse time of 0.59 s and a cool time of 0.35 s. The primary column (first separation dimension) for the GC × GC used a RTX-wax (polar) stationary phase: 30 m length, 250 µm i.d., and 0.5 µm film thickness. The modulation period was set to 2.5 s (i.e. the secondary column separation time). The secondary column (second separation dimension) used a RTX-1 (non-polar) stationary phase: 1.2 m length, 100 µm i.d., and 0.18 µm film thickness. The secondary column oven temperature control was not applied, so the secondary column oven was open and at the same nominal temperature as the primary column oven. The mass spectrometer electron energy was set to -70 eV and the detector voltage was set to 1600 V. The ion source temperature was set at 225 °C. The data acquisition parameters were set with a 120 s acquisition delay, mass-to-charge ratio (m/z) scan range of 35-334 amu at unit resolution, and acquisition rate set of 100 Hz.

There were replicate GC × GC – TOFMS chromatographic data collected for each RP-1 sample, and each set of replicates were analyzed separately using the PLS procedure (described below). Results for both replicates are provided herein, overlaid in figures, similar to a previous report⁹¹. Chromatographic runs were imported to MATLAB2009b (MathWorks, Natick MA) using the in-house ‘peg2mat’ function^{35–37} and underwent preprocessing and analysis using a combination of both in-house and commercial software. Once the data was in the MATLAB workspace, it underwent baseline correction. The baseline correction was performed individually on the signal at every m/z . First, a rolling minimum of the data is calculated to estimate the low frequency noise, which is then subtracted. Second, the mode of the noise distribution is subtracted to numerically center the noise around 0. To help save memory and computation time, the data underwent a condensing procedure. First, data-point summation or

binning was performed; every two points (i.e. data pixels) were added together along both separation dimensions, leading to a 4-fold reduction in the size of the data set. The binning procedure also mitigated any slight run-to-run misalignment⁹⁶. Next, m/z channels that were either unselective (such as m/z 44) or exhibited insufficient signal at any time during the chromatographic run were omitted (i.e., signal below a user-specified signal-to-noise S/N threshold, which was signal > 5 standard deviations of baseline corrected noise); the excluded m/z were the following: 35-37, 43-47, 51, 58-62, 73-76, 87-90, 101-103, 115-118, 133, 207, 214-334. Hence, only signal above the S/N threshold could contribute to the chromatographic peaks in the PLS analysis. Finally, chromatographic regions that contained no peaks, i.e. only contained either noise or column bleed at all m/z were inspected and later zeroed (while taking into consideration the variability between chromatograms to avoid the possibility of removing analytes exclusive to certain samples); the overloaded acetone peaks were also removed in this fashion. All of these preprocessing steps contributed to reducing noise from being introduced into the PLS analysis, which otherwise would adversely impact the constructed models.

The 2D chromatographic and mass spectral dimensions of the GC \times GC – TOFMS data were unfolded (from 10 samples \times 125 secondary column data points \times 405 primary column data points \times 148 mass channels, to 10 samples \times 7492500 vectorized data points) prior to being forwarded to PLS along with the measured RP-1 properties, and underwent mean centering²⁴. PLS analysis was initially performed using PLS Toolbox 6.02 (Eigenvector Research Inc., Wenatchee WA, USA), and subsequent data analysis work was performed using PLS Toolbox 6.7. Many of the measured compositional and physical properties for the RP-1 fuels have been

(96) L. C. Marney, W. C. Siegler, B. A. Parsons, J. C. Hoggard, B. W. Wright, R. E. Synovec, *Talanta* 115 (2013) 887–895.

reported^{79,80,97}. The complete list of physical and compositional properties measured and methods used to obtain values are the following: Density (g/ml) – ASTM D4052 (at 15°C)⁹⁷, Kinematic Viscosity mm²/s – ASTM D445 (at -10°C)⁹⁷, Net Heat of Combustion (MJ/kg) – ASTM D4809, Net Heat of Combustion (MJ/l) – ASTM D4809 (density at 30°C), Hydrogen Content (weight%) – Perkin Elmer Elemental Analyzer, Model EA2400, Total Sulfur (ppm) – SCD, Sustained Boiling Temperature (°C), and Vapor Rise Temperature (°C). The following compound class analyses (mass%) were performed using ASTM D2425: paraffins, cycloparaffins, dicycloparaffins, and tricycloparaffins. In the results reported herein, the designation “alkanes” is used in place of “paraffins”, hence alkanes, di-cycloalkanes, and so on. The ASTM-based *n*-alkane analysis involved further categorization, with *iso*-alkane content assumed by subtraction. Total aromatics content (mass%) was determined according to ASTM D6379. The aforementioned measured chemical and physical property data underwent auto-scaling (subtracting the mean and dividing by the standard deviation to achieve unit scale) for preprocessing prior to analysis.

For the constructed PLS models, all underwent leave-one-out-cross-validation (LOOCV) for testing the overall predictive ability of the models, and to determine the most appropriate number of LVs. The PLS linear regression vectors (LRVs), as defined in PLS toolbox, were investigated for the purpose of introspectively analyzing the samples used in constructing the PLS models and to discover correlation between the GC × GC – TOFMS chemical information and compositional and/or physical measurements. The second set of GC × GC – TOFMS replicates were analyzed in separate PLS models. The validation of the PLS models is an important part of the analysis, and it was performed via LOOCV and root mean squared error of

(97) T.J. Fortin, Energy Fuels. 26 (2012) 4383–4394.

cross validation (RMSECV) for each replicate model were obtained and reported. RMSECV is essentially calculating RMSE values with respect to the cross validation predicted values obtained from the validation procedure². The LOOCV procedure for N samples is performed by constructing an (N-1) PLS model and using the *i*th sample (omitted from the model for prediction, and after all N combinations of (N-1) models are analyzed the residuals are analyzed via RMSE. In short, the RMSECV was calculated according to²:

$$\text{RMSECV} = [(1/N) * \sum (y_{i,cv} - y_{i,meas})^2]^{0.5} \quad (3.1)$$

where $y_{i,cv}$ is the cross validation predicted value of sample (*i*) from LOOCV of the PLS model, $y_{i,meas}$ is the measured value for the same chromatogram, N is the number of chromatograms used in construction of the PLS model, and the summation is from *i* equals 1 to N. Identification of analytes of interest as indicated in the LRV in the GC × GC – TOFMS data was performed using ChromaTOF V.3.32 (LECO Corporation, St. Joseph, MI, USA) and in-house software for non-target PARAFAC⁵² (for well resolved and badly resolved peaks, respectively), and the mass spectra for analyte compounds of interest were identified using the NIST11 V2.0g mass spectral library (National Institute of Standards and Technology, Boulder CO, USA).

3.3 Results and Discussion

3.3.1 GC × GC Separation of RP-1 Using Reversed Column Format

Using the “reversed” GC × GC column format, we achieved an excellent 2D separation of the compound classes in the RP-1 fuels. Figures 3.1A-B show total ion current (TIC) chromatograms of two representative fuels: LB073009-05 and LB073009-09. The signal contrast and scaling is identical for these two chromatograms so the signal intensities can be

directly compared. The TIC chromatograms of the other RP-1 fuels listed in Table 3.1 are omitted for brevity. For the purpose of discussion, LB073009-05 will be used as the primary benchmark with respect to describing, comparing and contrasting all of the fuel samples studied in Table 3.1. Notably, separation patterns of closely eluting peaks were achieved, as seen in Figures 3.1A. Many well resolved, closely related compounds are separated diagonally, with each subsequent row (from left to right) pertaining to an increasing compound size. With the TIC, the diagonal spread of related compounds is most easily seen in the region that has been identified and labeled as "alkanes" in Figures 3.1A, in a 2D region enclosed between 1.84 to 2.50 s in the second column dimension. Directly below the alkanes are two tightly clustered compound classes: the cycloalkanes and the di- & tri-cycloalkanes (at around 1.50 to 1.84 s and 1.08 to 1.50 s on the second column dimension, respectively), are combined together as "cycloalkanes" (from 1.08 to 1.84 s). Finally, two smaller compound classes, the mono-aromatics and di-aromatics, have been combined as "aromatics" found between 0.68 to 1.08 s on the second column dimension. For all of the enclosed regions for the compound classes, the left boundary is the same at 4.85 min, while the right boundary is 25.9 min, 26.7 min and 36.3 min for the alkanes, cycloalkanes, and the aromatics, respectively. These qualitative boundaries for the compound classes are based on the identification of specific compounds in the 2D chromatogram, listed in Table 3.2 (identified using the LB073009-05 chromatogram which visually contains the most diverse number of compounds), as well as experience in visually interpreting chromatograms; the locations of the compounds in Table 2 and the subsequent boundaries form the compound class template shown in Figure 3.1C. The numbers noted in Figure 3.1C correspond to selected compounds listed in Table 3.2 (for example, 1 is methylnonane, 11 is trimethylcyclohexane, and so on). Not all compounds identified and used in

the making of the compound class template are labeled in Figure 3.1C for clarity. Furthermore, since the emphasis of this study is not the exhaustive identification of all chromatographic peaks, but rather the elucidation and identification of those compounds and classes primarily responsible for differences in measured fuel chemical and physical characteristics, we present the compound class boundaries and compounds identified primarily as additional information external to the PLS models to enable qualitative comparisons between fuels and to guide interpretation of specific compositional impacts on fuel properties resulting from PLS analysis.

We now make general comparisons of the two representative RP-1 fuels in Figures 3.1A-B, to provide insight into the types of variation readily observed with the fuels in Table 3.1. Upon inspection of the GC \times GC chromatogram of LB073009-09 shown in Figure 3.1B there are some significant differences with respect to the benchmark chromatogram of LB073009-05 in Figure 3.1A. The most noticeable features are the mere traces of mono-aromatics present around 20 min and a lack of di-aromatics. Also for LB073009-09, with respect to the alkane class, the peaks at the 20 min mark are generally lower in concentration, and there is a lack of the later eluting alkanes beyond the 21 min mark. Moreover, large peaks that appear in the alkane class at around 2 s in the second column dimension also are generally not present with LB073009-09. With respect to the cycloalkane(s) class, many of the peaks eluting around 10 to 20 min are more concentrated due to the higher signal of these peaks. Moreover, there are very few peaks eluting after about 23 min, suggesting LB073009-09 lacks the heavier cycloalkanes present in LB073009-05. While the general descriptions of these fuels provide a broad overview of the RP-1 fuel variation, chemometric analysis of the GC \times GC – TOFMS data provides a much deeper insight into the relationship between chemical composition and fuel properties, as we report herein.

The excellent resolving power of this reversed column GC \times GC configuration for RP-1 fuels is further discussed in chapter 5.

3.3.2 Chemical Composition Studies Using PLS

The GC \times GC – TOFMS data were analyzed using PLS with the previously measured mass% content of the following compound classes (using ASTM methods mentioned⁷⁹ in the Experimental section): *n*-alkanes, *iso*-alkanes, cycloalkanes, di-cycloalkanes, tri-cycloalkanes, and aromatics. PLS models were constructed using only relevant 2D chromatographic regions (alkanes region for the *n*-alkanes and *iso*-alkanes models, cycloalkanes region for the cycloalkanes, di-cycloalkanes and tri-cycloalkanes models, and the aromatics region for the aromatics model). The results of three PLS models (i.e., for *n*-alkanes, cycloalkanes, di-cycloalkanes) are shown in Figures 3.2A-C and summarized in Table 3.3, the others are only summarized in Table 3.3 for brevity. Indeed, a figure key, summary of the RMSECV values, for all the PLS models (compositional and physical properties) are presented in Table 3.3. While the LRVs could be used to draw connections between compound class and actual measured mass%, they are omitted for brevity. For measured chemical content, the PLS models used were between 2-4 latent variable (LV) models. The selection process for the number of LVs includes inspecting the RMSECV results for models of increasing number of LVs (see the scree plots provided in the Supplemental Material) which comprehensively assesses the predictive power of the various PLS models (using increasing number of LVs) with respect to the measured properties. When the difference between two favorable PLS models is small or ambiguous, the rule of parsimony is adhered to and the model with fewer LVs was chosen.

For *n*-alkanes, the PLS model results using 3 LV are shown in Figure 3.2A, with the predicted mass% plotted relative to the measured mass%. The models match the measured mass% well with a RMSECV of 1.26 and 1.41 mass%, respectively, for the two GC × GC – TOFMS data sets. The measured mass% range is a little over one order of magnitude. The majority of mass% values are clustered between 0.7% and 2.7% with three samples between 6.8% and 9.2%, and one sample at 14%. Only one RP-1 sample was significantly under predicted (LB073009-08); since it is a sample with an extreme mass%, using LOOCV, PLS may struggle in predicting samples at extreme ends of the range.

For cycloalkanes, Figure 3.2B provides the PLS model predicted mass%, using 3 LV, plotted relative to the measured mass%. The mass% values fall in a narrow range, from about 27% to 38%, making the PLS modeling challenging, resulting in a relatively weak linear pattern. The RMSECV values are 2.65 and 2.26 mass% for the two GC × GC – TOFMS data sets. The samples PLS modeling appears to struggle with the most are those with extreme measured mass% such as LB073009-08, LB073009-09, as well as samples with scores on the outskirts of the scores cluster (such as LB073009-02, and LB073009-03), however the sample with the shortest distance from the mean centered origin (XC2521HW10) is under predicted, suggesting uncertainty and/or inaccuracy in the measured values (x-axis) may also be significantly affecting the PLS models.

For di-cycloalkanes, PLS modeling with 2 LV yielded a predicted mass% comparison to the measured mass% shown in Figure 3.2C. The mass% range is 13.6% to 18.7%, meaning the chemical variation between any two samples is at most 5.1% with respect to the content of di-cycloalkanes, which is also very challenging for PLS modeling. LB073009-01 was the most under predicted sample (highest mass%) while LB073009-02 was the most over predicted

sample (scores on the outer rim of the scores cluster). Even with this challenging PLS modeling situation, the RMSECV values are 0.99 and 0.69 mass% for the two GC \times GC – TOFMS data sets.

To summarize this section on chemical composition studies, we have demonstrated the use of GC \times GC – TOFMS chromatographic regions to readily correlate, via PLS modeling, to previously collected ASTM measured quantitative compound class data for RP-1 fuels. The results show promise in that PLS modeling of the GC \times GC – TOFMS data should be able to address more challenging and interesting studies, specifically, correlations to physical properties with direct inference back to chemical compound identification related to physical property relationships, as are explored in the next section.

3.3.3 Physical Properties Studies Using PLS

PLS modeling was performed on several measured chemical and physical properties (density, kinematic viscosity, net heat of combustion, hydrogen content, sulfur content, sustained boiling temperature, and vapor rise temperature). The results of three PLS models are provided in Figures 3.3-3.5 and summarized in Table 3.3, while the others are only summarized in Table 3.3 for brevity. The PLS model prediction using LOOCV for the designated measured properties density, kinematic viscosity and net heat of combustion (MJ/l) and their respective LRVs are provided. In these PLS models, the entire chromatograms were forwarded to PLS rather than the compound class 2D regions (as was performed for the analyses in the previous section). Referring to the PLS model plots of Figures 3.3A, 3.4A, and 3.5A, the x-axis displays the measured property values and the y-axis displays the PLS model predicted values. The scree

plots in the Supplemental Material provide information regarding the selection process for the number of LVs for the PLS models. The LRVs of PLS models offer complementary information to the values predicted by the PLS models relative to the measured results. Each LRV was imported into the base MATLAB workspace (from PLS_Toolbox) where it was refolded to recover the dimensions of the 2D separation. The LRV, also known as the regression coefficients, is derived by multiplying the X-block loadings, the regression coefficients between the X-block and Y-block loadings and the transposed Y-block loadings. By dot product multiplication of an unfolded LRV with a given unfolded $GC \times GC$ chromatogram, the result is the respective predicted 'single' value for the modeled property. Preprocessing steps such as mean centering performed on the calibrated data must also be taken into consideration when doing predictions using new data. Thus, the intensity per m/z at a specific chromatographic peak in the LRV is a score which indicates the contribution (sign and magnitude) of the compound(s) with respect to predicting a given property. To view the LRVs more easily, they are refolded into a 2D separation format, with the mass spectral dimension summed, forming a LRV TIC similar to the shown $GC \times GC$ TIC chromatograms. With the LRVs, we can begin to explore, in detail, the contribution of the chemical composition with respect to the physical properties. Enclosed compound classes (in Figures 3.3B, 3.3C, 3.4B, 3.4C, 3.5B, and 3.5C) are compared to specific 2D regions in various LRV plots. Using these LRV plots, Tables 3.4 and 3.5 list identified, prominent compounds based on the LRVs for density (Figures 3.3B-C), while Tables 3.6 and 3.7 (kinematic viscosity) and Tables 3.8 and 3.9 (net heat of combustion) provide identified, prominent compounds based on their LRVs (Figures 3.4B-C, and Figures 3.5B-C, respectively).

For the density study, a 4 LV PLS model was used. Comparison between the PLS

predicted densities to the measured densities as well as the LRV are provided in Figures 3.3A-C. Measured densities fall between 806 and 816 g/l. The RMSECV values are 1.83 and 1.88 g/l for the two GC \times GC – TOFMS data sets; the cross validation (expectedly) struggled to predict the samples with the lowest and highest density (LB073009-08 and LB073009-01, respectively). With regard to the noticeably under predicted LB073009-05 sample (813.4 g/l measured and 809.7 g/l predicted density), the scores are relatively extreme (almost isolated from the general sample cluster). In general, inspection of the LRVs (Figures 3.3B-C) for density indicate many earlier eluting peaks (before 15 min for the alkanes class, and before 12 min for cycloalkanes class) have negative scores, i.e., their contribution lowers the overall predicted fuel density. The later eluting peaks have a positive score, suggesting they act to increase the predicted fuel density. In general, the results for modeling density indicate smaller more volatile compounds decrease the overall predicted fuel density, while the larger, less volatile compounds tend to increase the overall predicted fuel density. Identified compounds of interest (regions in the LRVs with very large positive or negative values) are reported in Tables 3.4 and 3.5.

For the kinematic viscosity study, a 4 LV model was constructed, and the predicted kinematic viscosities are plotted relative to the measured kinematic viscosities as well as the LRV are provided in Figures 3.4A-C. Kinematic viscosities tended to fall into two distinct groups: low (4.8-5.1 mm²/s), and high (above 5.3 mm²/s). This behavior is most likely due to specific low- or high-viscosity materials used in fuel preparation. The PLS model appears to handle the variation adequately, with RMSECV values of 0.120 and 0.147 mm²/s for the two GC \times GC – TOFMS data sets. The LRV for kinematic viscosity is similar to the LRV for the measured density (see the positive and negative portions in Figures 3.4B and 3.4C), however with some noticeable differences. In Figure 3.4B there are significantly less positive

contributing compounds, with those appearing primarily between 15 and 25 min. Meanwhile in the negative portion of the LRV shown in Figure 3.4C, there are significantly more contributing peaks between 5 to 15 min on the first column). Therefore, the PLS modeling indicates earlier eluting compounds contribute to lowering the kinematic viscosity, while heavier compounds eluting after about 15-17 min have positive scores and tend to raise the overall predicted kinematic viscosity. Interestingly, the aromatics class shows little contribution. The aromatics compounds generally did not contribute significantly to the PLS models of the various physical properties, likely because they were at relatively low mass % (range of 0.2 to 3.7). Identified compounds of interest (based on regions in the LRVs with dominating positive/negative values) are identified in Tables 3.6 and 3.7 of the Supplemental Material.

The results for the PLS model (2 LV) constructed for volume-based net heat of combustion (MJ/l) are provided in Table 3, where the predicted net heat of combustion is plotted relative to the measured net heat of combustion as well as the LRV are provided in Figures 3.5A-C. Using the LRV plots, identified compounds of interest are reported in Tables 3.8 and 3.9 of the Supplemental Material. Note the scale for net heat of combustion ranges from 34.47 to 34.89 MJ/l (less than 0.5 MJ/l), making it technically one of the more challenging properties to model with PLS. The RMSECV values are 0.064 and 0.066 MJ/l for the two GC × GC – TOFMS data sets. The samples with the greatest cross validation residuals are the samples with the lowest and the highest net heat of combustion (LB073009-08 and LB073009-01, respectively). The corresponding LRV, in Figures 3.5B and 3.5C, bares some similarity to the LRV for the density model. In general, the positive contribution appear to have more emphasis (greater magnitude) on peaks between 15 and 25 min, while the negative contributions appear nearly identical with the exception of a few alkane peaks eluting after 15 min; though this could

be an artifact of covariance between those peaks and the aromatics as they too have a slight negative contribution.

Other measured chemical and physical properties were also analyzed using the method described herein, including hydrogen content, sulfur content (ppm), mass based net heat of combustion (MJ/kg), sustained boiling point temperature (°C), and vapor rise temperature (°C). Their respective modeling results are reported in Table 3.3, but further discussion has been omitted for brevity.

3.4 Conclusions

The study presented herein, has demonstrated separation not only between compound classes, but also between many distinct analyte compounds in RP-1 fuel using GC × GC – TOFMS with a reversed column configuration (RTX-wax primary column coupled to a RTX-1 secondary column), followed by chemometric analysis using PLS and the identification of analyte peaks using ChromaTOF, nontarget PARAFAC, and then the NIST-MS library. PLS modeling was performed on three relevant 2D regions of the GC × GC – TOFMS data to analyze chemical composition. With respect to a targeted focus of drawing connections between the compounds separated in the 2D chromatograms and measured physical properties of the fuel, the entire chromatograms were used. Even though some compound classes exhibited compositional correlation (likely an artifact of fuel preparation involving blending of available feed stocks) with other compound classes for certain measured values in the PLS models, the overall results of this study were promising. Ideally, a greater collection of different fuel samples with broader

diversity in chemical composition would serve to minimize correlation between compound classes, and produce more robust PLS models. This analytical platform could ultimately be used as an adequate estimation of ASTM methods; standards containing mixtures of specific groups of chemicals would be extremely helpful in constructing PLS models (by helping to break some of the covariance across compound classes). Any observed correlation may suggest the LRVs need to be inspected more carefully and caution should be exercised when drawing connections between measured property values and chromatographic/chemical information. Identifying compounds contributing positively and negatively to the property in question can aid in the interpretation of LRVs by substantiating the identification of influencing compound classes and confirming the compositional correlation between modeled fuel samples. We also note the RMSECV results presented in Table 3.3 are in good agreement for the replicate GC \times GC – TOFMS data sets for a given measured property, suggesting some imperfections in the modeling (i.e., results that deviate significantly from the line with a slope of unity) are possibly due to inaccuracies in the measured data (i.e., ASTM values), but further study is needed to draw any conclusions on this matter.

Acknowledgement

The work at the University of Washington (UW) was performed under subcontract to ERC, Incorporated, Air Force Research Laboratory, Edwards AFB, CA. The fuels were provided by the Air Force Research Laboratory/RQRC, Edwards AFB, CA. Certain commercial equipment, instruments or materials are identified in this paper in order to adequately specify the experimental procedure. Such identification does not imply recommendation or endorsement by

the University of Washington, the United States Air Force, or the National Institute of Standards and Technology, nor does it imply the materials or equipment identified are necessarily the best available for that purpose.

Figure Captions

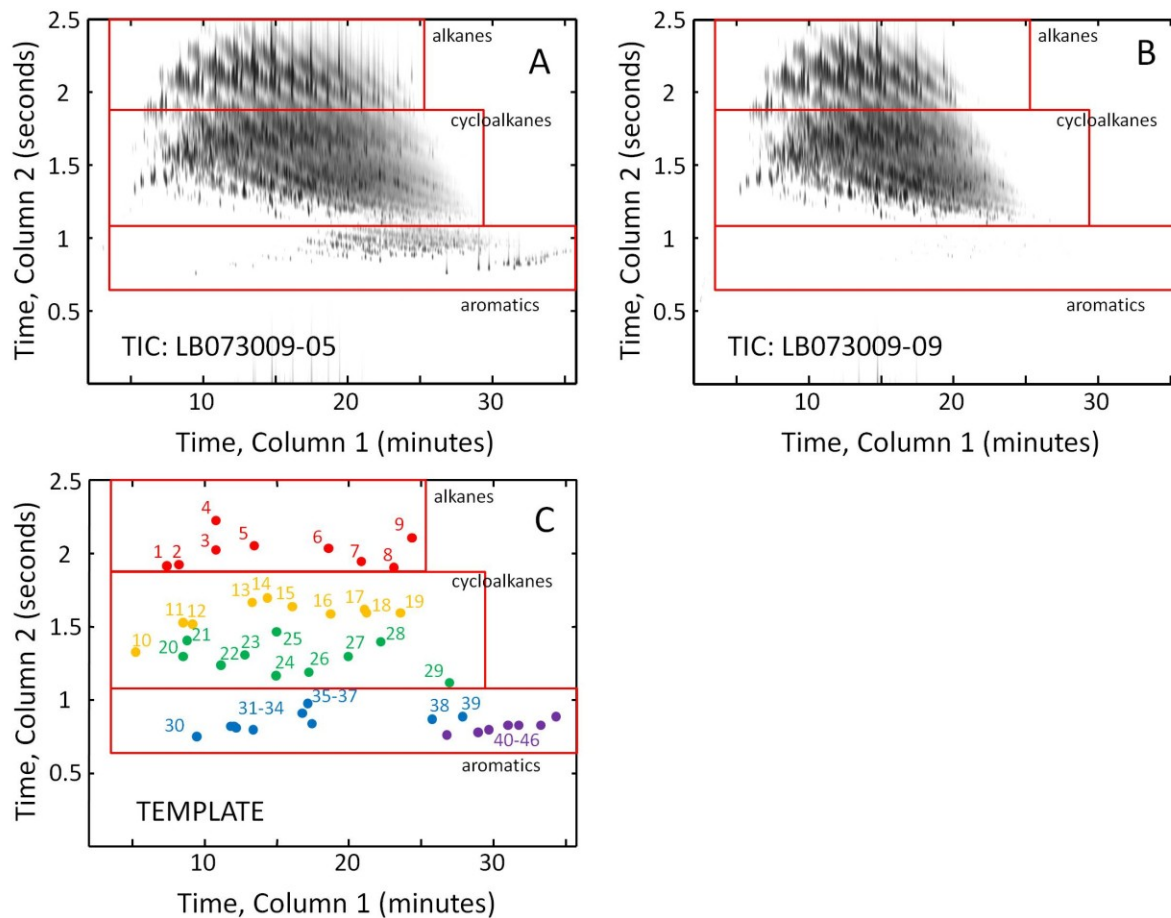


Figure 3.1. RP-1 GC \times GC – TOFMS total ion current (TIC) chromatograms, collected using a 30 m Rtx-wax column for the first separation dimension followed by a 1.2 m Rtx-1 column for the second separation dimension at a constant inlet pressure of 35 psig (241 kPa). Compound classes are indicated and annotated. (A) RP-1 LB073009-05. (B) RP-1 LB073009-09. (C) Template used for aforementioned compound class boundaries. Dots signify locations of identified compound peaks used in defining the class boundaries. The numbers correspond to analyte peaks specified in Table 3.2.

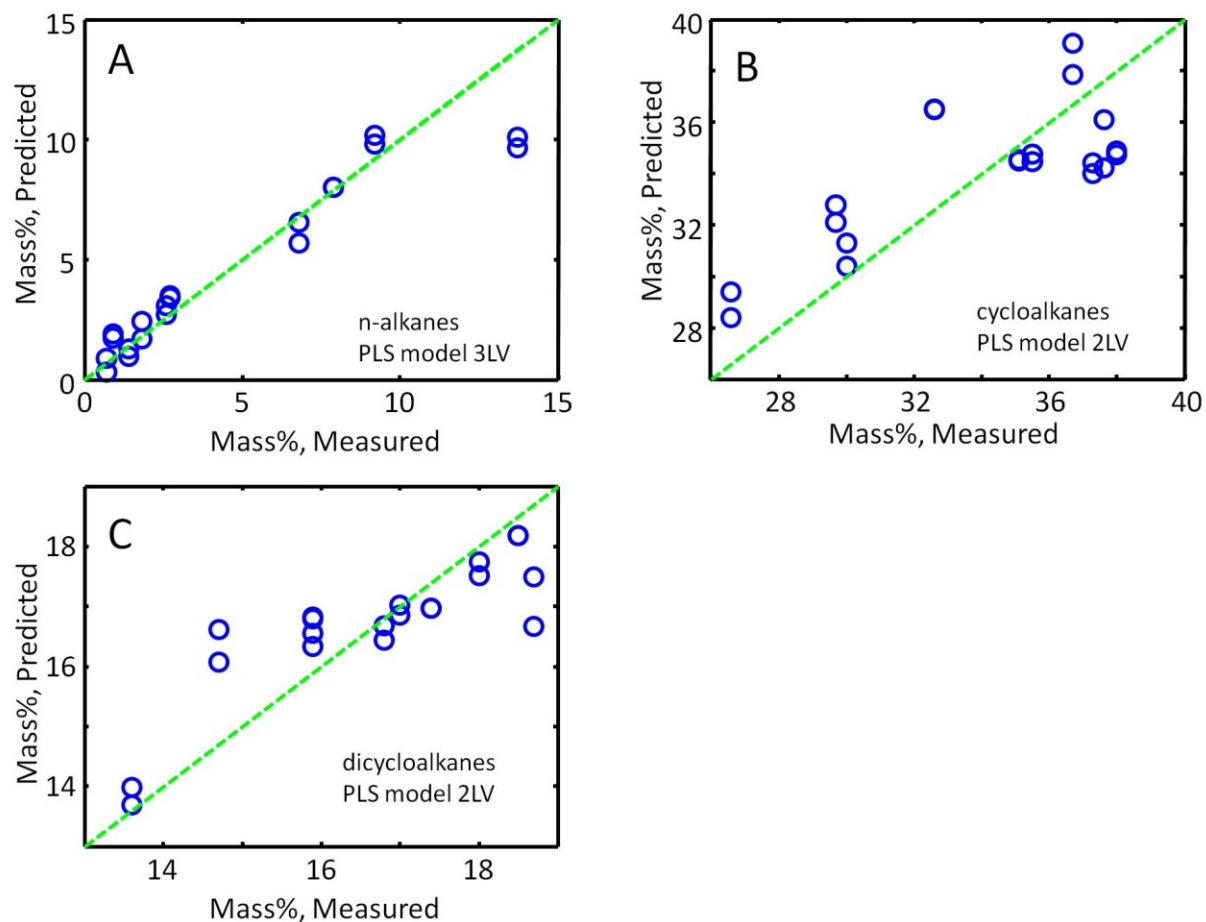


Figure 3.2. Leave-one-out-cross-validation (LOOCV) results for compound class mass% derived from the PLS modeling using the GC \times GC – TOFMS data replicate data sets and the respective measured mass% for the ten RP-1 fuels listed in Table 3.1. The dashed line represents an ideal agreement between the predicted and measured mass%, with results for both modeled replicate data sets overlaid. (A) *n*-alkanes (ASTM D2425 with ‘*n*-alkane analysis’) using a 3 LV PLS model. (B) cycloalkanes (ASTM D2425) using a 2 LV PLS model. (C) di-cycloalkanes (ASTM D2425) using a 2 LV PLS model.

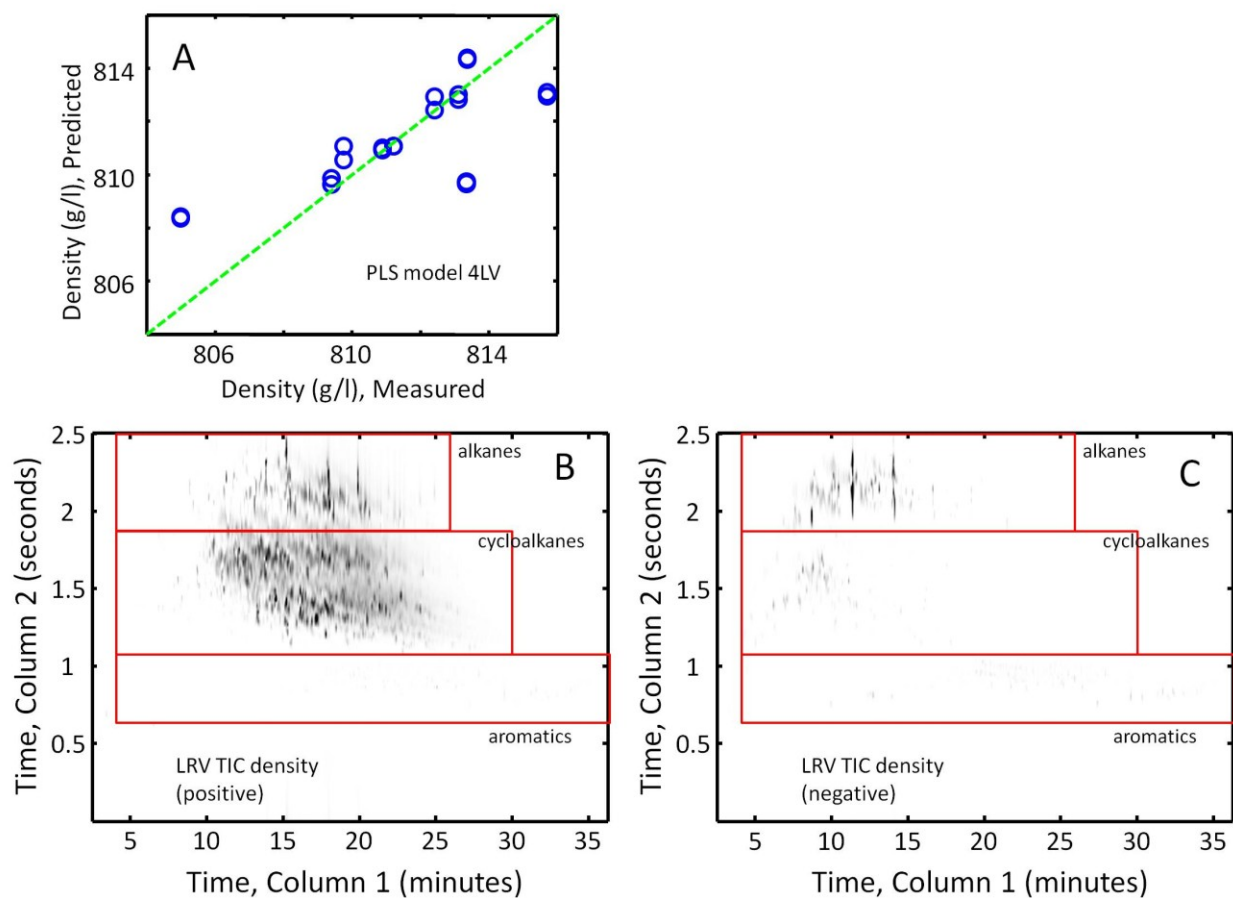


Figure 3.3. Results of the 4 LV PLS model for density (g/l) using LOOCV. (A) Predicted density plotted relative to the measured density (ASTM D4052). The dashed line represents an ideal agreement between the predicted and measured density, with both model replicates overlaid. (B) Density LRV, positive values only. (C) Density LRV, negative values only.

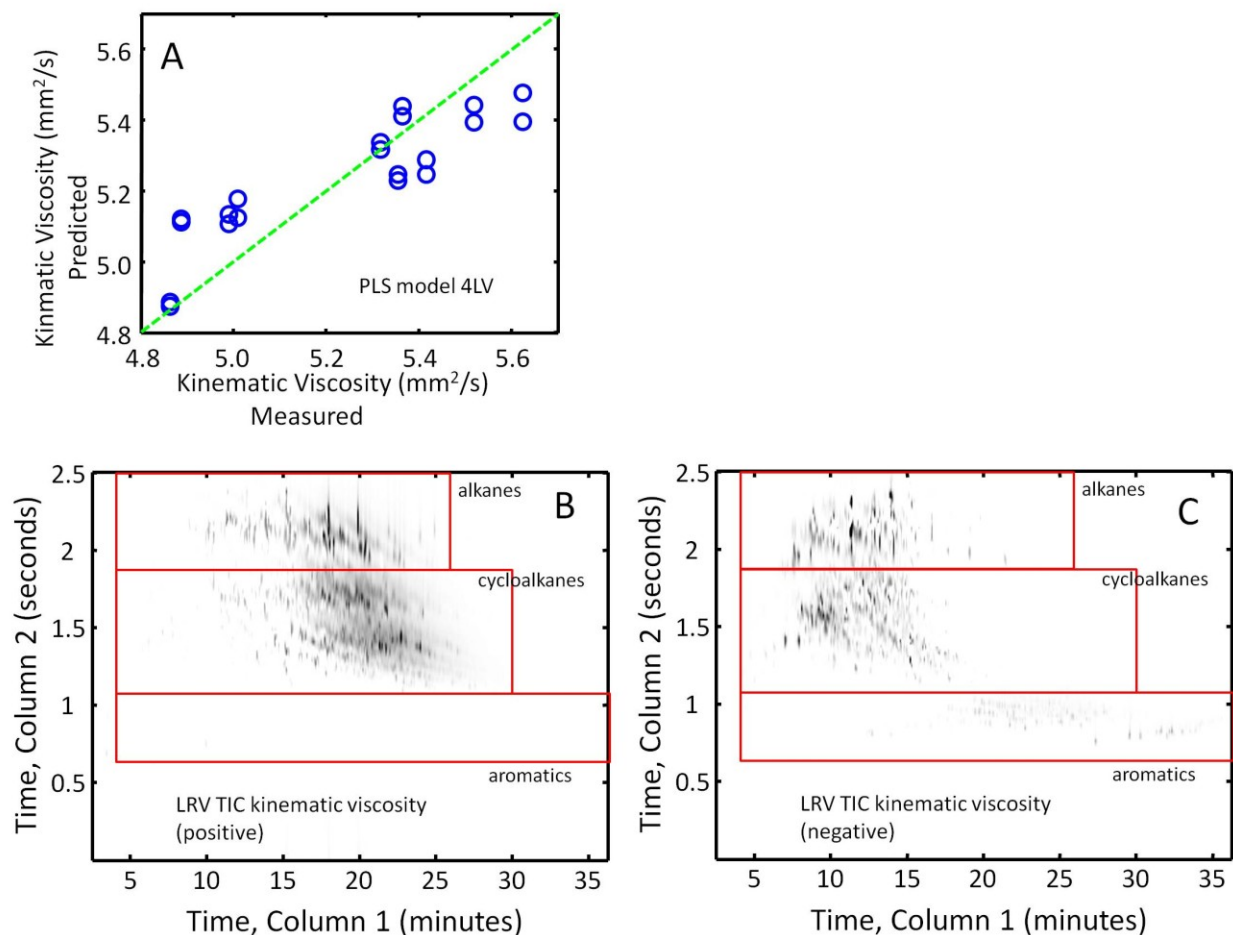


Figure 3.4. Results of the 4 LV PLS model for kinematic viscosity (mm²/s) using LOOCV. (A) Predicted kinematic viscosity plotted relative to the measured kinematic viscosity (ASTM D445). The dashed line represents an ideal agreement between the predicted and measured kinematic viscosity, with both model replicates overlaid. (B) Kinematic viscosity LRV, positive values only. (C) Kinematic viscosity LRV, negative values only.

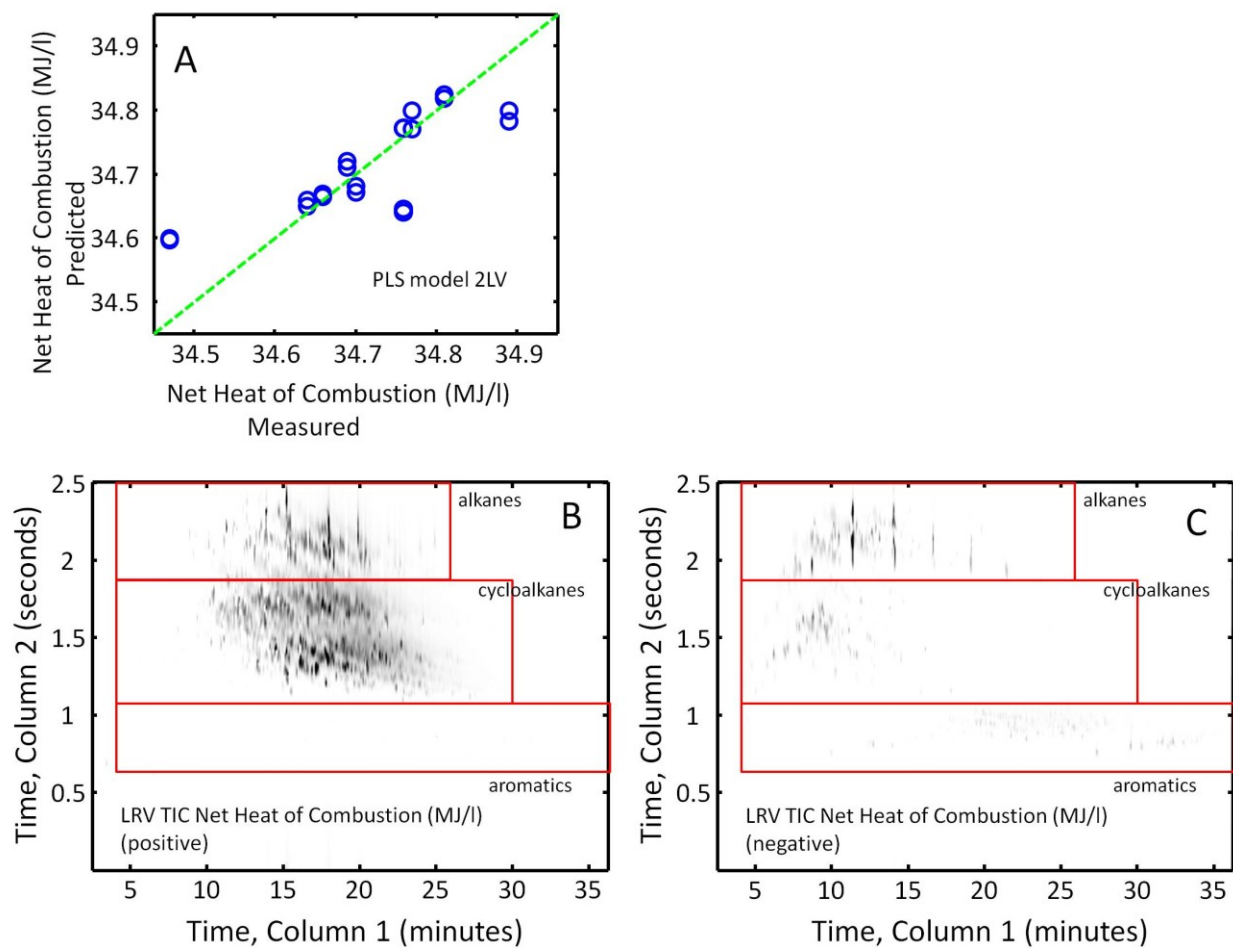


Figure 3.5. Results of the 2 LV PLS model for net heat of combustion (MJ/l) using LOOCV. (A) Predicted net heat of combustion plotted relative the measured net heat of combustion (ASTM D4809). The dashed line represents an ideal agreement between the predicted and measured net heat of combustion, both model replicates are overlaid. (B) Net heat of combustion LRV, positive values only. (C) Net heat of combustion LRV, negative values only.

Tables

Table 3.1. RP-1 Fuel Set, where the RP-1 Sample number is used herein, while the NIST and AFRL numbers are provided for reference to previous studies of interest.

RP-1 Sample	NIST Number ⁸⁰	AFRL Designation ⁷⁹
1	11	LB080409-01
2	10	LB073009-06
3	9	LB073009-08
4	8	LB080409-05
5	7	LB073009-05
6	5	LB073009-01
7	4	LB073009-09
8	1	LB073009-02
9	2	LB073009-03
10	3	XC2521HW10

Table 3.2. Template data with the assigned number (#) used in Figure. 3.1D, of each representative identified compound with their retention times on column one (1t_R) and two (2t_R) in s, and mass spectral match value (MV). This information was used to define the encircled various compound classes.

#	Compound Identification	1t_R (min)	2t_R (s)	MV	Compound Class
1	Methylnonane	7.42	1.92	925	alkanes
2	Decane	8.25	1.93	950	alkanes
3	Undecane	10.83	2.03	946	alkanes
4	Dimethyldecane	10.83	2.23	902	alkanes
5	Dodecane	13.5	2.06	941	alkanes
6	Tetradecane	18.63	2.04	959	alkanes
7	Pentadecane	20.92	1.95	938	alkanes
8	Hexadecane	23.167	1.91	920	alkanes
9	Pristane	24.42	2.11	921	alkanes
10	Trimethylcyclohexane	5.29	1.33	911	cycloalkanes
11	Cyclohexane, -tetramethyl isomer	8.58	1.53	881	cycloalkanes
12	Cyclohexane, -tetramethyl isomer	9.21	1.52	840	cycloalkanes
13	Cyclohexane, pentyl-	13.33	1.67	843	cycloalkanes
14	Cyclohexane, 1-methyl-4-(1-	14.38	1.70	881	cycloalkanes
15	Cyclohexane, hexyl-	16.13	1.64	901	cycloalkanes
16	Heptylcyclohexane	18.79	1.59	869	cycloalkanes
17	Cyclotetradecane	21.13	1.62	858	cycloalkanes
18	Cyclohexane, octyl-	21.29	1.60	882	cycloalkanes
19	n-Nonylcyclohexane	23.63	1.60	929	cycloalkanes
20	2-Methyloctahydropentalene	8.58	1.30	914	di- & tri-cycloalkanes
21	Bicyclo[2.2.1]heptane, trimethyl-(C ₁₀ H ₁₈)	8.83	1.41	806	di- & tri-cycloalkanes
22	1H-Indene, octahydro-, cis-	11.17	1.24	930	di- & tri-cycloalkanes
23	Naphthalene, decahydro-, trans-	12.83	1.31	944	di- & tri-cycloalkanes
24	Adamantane	15	1.17	962	di- & tri-cycloalkanes
25	Naphthalene, decahydro-2,6-dimethyl-	15.04	1.47	880	di- & tri-cycloalkanes
26	2-Methyladamantane	17.25	1.19	871	di- & tri-cycloalkanes
27	Bicyclohexyl	20	1.30	890	di- & tri-cycloalkanes
28	Trans-hexamethyl-octahydro-1H-Indene (C ₁₅ H ₂₈)	22.25	1.40	820	di- & tri-cycloalkanes
29	Tricyclo[4.2.2.0(2,5)]dec-7-ene, 7-butyl-	27.03	1.12	820	di- & tri-cycloalkanes
30	Toluene	9.5	0.75	952	mono-aromatics
31	Ethylbenzene	11.88	0.82	953	mono-aromatics
32	Xylene isomer	12.08	0.82	960	mono-aromatics
33	Xylene isomer	12.25	0.81	965	mono-aromatics
34	Xylene isomer	13.42	0.80	957	mono-aromatics
35	Ethyl trimethyl benzene	17.21	0.91	902	mono-aromatics
36	Isobutyltoluene	17.21	0.98	914	mono-aromatics
37	Trimethylbenzene	17.5	0.84	950	mono-aromatics
38	Methyltetralin	25.83	0.87	928	mono-aromatics
39	Dimethyltetralin	27.96	0.89	915	mono-aromatics
40	Naphthalene	26.83	0.76	905	di-aromatics

41	Methyl naphthalene isomer	29	0.78	900	di-aromatics
42	Methyl naphthalene isomer	29.75	0.80	902	di-aromatics
43	Dimethyl naphthalene isomer	31.08	0.83	931	di-aromatics
44	Dimethyl naphthalene isomer	31.83	0.83	921	di-aromatics
45	Methyldiphenyl (methyl phenyl benzene)	33.33	0.83	909	di-aromatics
46	Trimethyl naphthalene	34.42	0.89	939	di-aromatics

Table 3.3. Summary of PLS model metrics for measured properties analyzed. Metrics (slope, RMSECV for the first and second GC × GC – TOFMS replicates, referred to as rep1 and rep2) are provided in the context of the relevant Figures. (a) Figures originally omitted for brevity can be found in the supplementary section.

Measured Property	Figures	Range	RMSECV rep1	RMSECV rep2
<i>n</i> -alkanes (mass%)	2A	0.50-14.0	1.27	1.41
<i>iso</i> -alkanes (mass%)	— ^(a)	36.2-41.5	1.16	1.11
Cycloalkanes (mass%)	2B	26.5-38.0	2.65	2.26
Di-cycloalkanes (mass%)	2C	13.5-18.8	0.99	0.69
Tri-cycloalkanes (mass%)	— ^(a)	3.50-4.80	0.28	0.22
Aromatics (mass%)	— ^(a)	0.2-3.7	0.08	0.07
Hydrogen Content (wt%)	— ^(a)	14.15-14.45	0.05	0.06
Density (g/l)	3A-C	805.0-816.0	1.83	1.88
Kinematic Viscosity (mm ² /s)	4A-C	4.80-5.70	0.120	0.147
Net Heat of Combustion (MJ/l)	5A-C	34.40-34.90	0.064	0.066
Net Heat of Combustion (MJ/kg)	— ^(a)	43.25-43.45	0.039	0.044
Sulfur (ppm) by SCD	— ^(a)	0.5-25	5.3	4.9
Sustained boiling temp. (°C)	— ^(a)	204.5-212.5	1.40	1.38
Vapor rise temp. (°C)	— ^(a)	207-214	1.41	1.43

Table 3.4. Major contributing compounds identified in LRVs for density (positive) per Figure 3.3B.

#	Compound Identification	¹ t _r (min)	² t _r (s)	MV	Compound Class
1	Undecane, 2,6-dimethyl- (C13H28)	13.42	2.32	880	alkanes
2	Tridecane, 7-methyl- (C14H30)	14.75	2.38	818	alkanes
3	Dodecane, 2,6,10-trimethyl- (C15H32)	17.50	2.24	882	alkanes
4	Cyclohexane, 1-ethyl-2-propyl- (C11H22)	11.92	1.72	835	cycloalkanes
5	Cyclohexane, 1-methyl-4-(1-methylethyl)-, cis- (C10H20)	17.08	1.72	831	cycloalkanes
6	Heptylcyclohexane (C13H26)	18.83	1.66	872	cycloalkanes
7	Naphthalene, decahydro-2-methyl- (C11H20)	14.08	1.44	913	di-&tri-cycloalkanes
8	Adamantane, dimethyl- (C12H20)	14.75	1.42	888	di-&tri- cycloalkanes
9	Decalin, syn-1-methyl-, cis- (C11H20)	16.67	1.32	819	di-&tri- cycloalkanes
10	no significant peak found				mono-aromatics
11	no significant peak found				di-aromatics

Table 3.5. Major contributing compounds identified in LRVs for density (negative) per Figure 3.3C.

#	Compound Identification	¹ t _r (min)	² t _r (s)	MV	Compound Class
1	Decane (C ₁₀ H ₂₂)	8.25	1.98	953	alkanes
2	Undecane (C ₁₁ H ₂₄)	10.92	2.12	842	alkanes
3	Dodecane (C ₁₂ H ₂₆)	10.92	2.32	864	alkanes
4	Cyclohexane, diethyl- (C ₁₀ H ₂₀) isomer	7.67	1.62	873	cycloalkanes
5	Cyclohexane, 1,2-diethyl-, cis- (C ₁₀ H ₂₀) isomer	8.33	1.64	854	cycloalkanes
6	C ₁₀ H ₁₈	11.17	1.4	800	di-&tri- cycloalkanes
7	Xylene (C ₈ H ₁₀) isomer	12.25	0.82	913	mono-aromatics
8	Naphthalene,decahydro-2-methyl- (C ₁₁ H ₂₀)	12.33	1.48	855	di-&tri- cycloalkanes
9	Xylene (C ₈ H ₁₀) isomer	13.50	0.82	957	mono-aromatics
10	no significant peak found				di-aromatics

3.5 Supplemental Material

In PLS, a scree plot provides statistical information, specifically the root mean square error of cross validation (RMSECV), as a function of the number of latent variables (LV). These results are used in selecting the appropriate LV number and to avoid over-factoring. Consider a PLS model with n LV and the RMSECV is at a minimum, the LV selection algorithm in PLS_Toolbox compares the RMSECV values between n LV and $(n-1)$ LV. When the differences between RMSECV values are statistically insignificant then $(n-1)$ LV is selected as more appropriate. The algorithm continues iteratively decreasing n by 1 until the corresponding RMSECV value of the n LV is significantly smaller than $(n-1)$ LV, (or when n is equal to 1). The scree plots of the highlighted PLS models are shown in Figure 3.6.

For Reference, the LRVs (positive and negative values) of the PLS models for n -alkanes, cycloalkanes, and di-cycloalkanes are shown in Figure 3.7. The LRV of the n -alkanes PLS model show major positive contributions from several large peaks (as expected). The LRVs of the cycloalkanes PLS model and the di-cycloalkanes PLS model confirm that the cycloalkanes and di-cycloalkanes are strongly correlated to each other since their respective LRVs are extremely similar.

Included herein, are the tables of identified compounds of interest for the highlighted PLS models for kinematic viscosity and net heat of combustion (MJ/l) provided in Tables 3.6-3.9. As with the tables of identified compounds of interest for the PLS model for density, the compounds were considered 'of interest' based on their greater magnitude in the LRVs.

The following PLS models for iso-alkanes, tri-cycloalkanes, aromatics, hydrogen content, net heat of combustion (MJ/kg), sulfur content, sustained boil, vapor rise and all discussion on

their respective results were originally omitted for reasons of brevity and/or poor LOOCV results and are included under supplementary material to offer the opportunity to observe situations where PLS may struggle to produce a robust model. This is not to say these PLS models are useless; though it is not how PLS is primarily used, one may analyze the PLS model introspectively (i.e. use PLS in a manner similar to PCA). When viewing LRVs of PLS models introspectively, especially when LOOCV results are unsatisfactory, results should be taken with a moderate amount of skepticism, meaning any useful external information (such as the identification of the compounds as demonstrated earlier in this chapter) should be taken into consideration to help distinguish the relevant information from mathematical artifacts (i.e. covariance and error).

The results for the PLS model (5 LV), which was constructed for *iso*-alkanes, are summarized in Table 3.3 in addition to a plot showing predicted values relative to the measured values, the scree plot, and the LRV are provided in Figure 3.8. The RP-1 samples have similar *iso*-alkanes content, making accurate prediction challenging, the cross validation results are shown in Figure 3.8A. For reference, the respective scree plot is shown in Figure 3.8B. The LRV for the *iso*-alkanes model is shown in Figure 3.8C and 3.8D. The LRV contains some positive values in the alkanes region, while the negative contribution is more pronounced and some correspond to peaks with positive values in the LRV for the *n*-alkanes PLS model. Many peaks with negative values in the LRV for *iso*-alkanes were previously identified as branched alkanes, suggesting the source of error is caused due to covariance of these compounds between RP-1 samples.

The results for the PLS model (3 LV) constructed for the measured content of tri-cycloalkanes are summarized in Table 3.3, and the predicted values plotted relative to the

measured values, the scree plot, and the LRV are provided in Figure 3.9. The predicted values shown in Figure 3.9A appear to somewhat agree with the measured values. For reference, the respective scree plot is shown in Figure 3.9B. The LRV for the tri-cycloalkanes is shown in Figure 3.9C and 3.9D, where one can see positive values look very similar to those in the LRVs of the PLS models for cycloalkanes and di-cycloalkanes – only the spread of positively contributing compounds is narrower, and most of the negative contribution also appears similar to the LRV of the di-cycloalkanes model except the spread is wider. The LRV for the tri-cycloalkanes PLS model suggests, as with cycloalkanes and di-cycloalkanes PLS models, there is a strong covariance between all three compound classes.

The results for the PLS model (4 LV) for aromatics are provided in Table 3.3, and the predicted values plotted relative to the measured values, the scree, and the LRV are provided in Figure 3.10. Even though the fuels analyzed exhibited a bi-modal distribution of aromatic content where most samples contained $\sim 0.2\%$, with two samples containing about 3.6% and one sample containing 3.7% . The PLS model (for aromatics) provides good prediction of mass% relative to the measured mass% values. The LRV for the model, seen in Figure 3.10C and 3.10D, shows that the positive values correspond to aromatic compounds while the negative values are almost nonexistent; there are some peaks around 1.1 sec in the second column dimension that corresponds to tri-cycloalkanes found in the region used in analysis.

The results of the PLS (3 LV) model constructed for hydrogen content are summarized in Tables 3.3, 3.10 and 3.11, and the predicted values plotted relative to the measured values, the scree plot, and the LRV are provided in Figure 3.11. PLS struggled in the cross validation results as the plots in Figures 3.11A and 3.11B show, meaning this PLS model did not work well in predicting the correct value of samples outside the calibration as other PLS models. Figure

3.11C shows the positive contributions to the LRV, while the negative contributions are shown in Figure 3.11D. In the LRV, the aromatics (mono- and di-) have a negative scores; suggesting they are correlated with decreasing the hydrogen content. The alkane group has the earlier eluting peaks (5 to 15 min) predominantly contributing to increasing the hydrogen content, while most of the alkanes eluting from 15 to 25 min decrease the predicted hydrogen content. The negative contribution of alkanes in this model may be explained by the inadvertent covariance between aromatics and alkanes. With respect to the cycloalkanes and di- & tri- cycloalkanes classes, negative contributions for peaks are observed between 10 and 20 min with compounds outside this range having a mild positive contribution. Identified compounds of interest (regions in the LRVs with dominating positive/negative values) are provided in Tables 3.10 and 3.11.

The results for the PLS model (4 LV) constructed for mass based net heat of combustion (MJ/kg) are provided in Tables 3.3, 3.8, and 3.9, and the predicted values as a function of the measured values, the scree, and the LRV are provided in Figure 3.12. The data ranges from 43.29 to 43.45 MJ/kg, meaning the measured values are very similar, which may be the reason for PLS having relatively poor RMSECV. The LRV is shown in Figure 3.12C and Figure 3.12D, and it is significantly different than the LRV for net heat of combustion (MJ/l). For the positive values in the LRV, there appears to be a huge contribution from earlier eluting alkanes along with some of the later eluting peaks in the cycloalkanes group. For the negative values in the LRV there are earlier eluting cycloalkanes, later eluting alkanes and the aromatics. Certain alkane peaks and aromatic peaks are repeatedly covaryin LRVs across multiple PLS models, which further confirms they are indeed covarying.

The results for the PLS model (2 LV) constructed for sulfur content (ppm) are provided in Table 3.3, and the predicted values as a function of the measured values, the scree, and the

LRV are provided in Figure 3.13. In the LRV shown in Figure 3.13C and Figure 3.13D, the compounds that appear to contribute positively appear in all major compound groups.

Compounds that appear to contribute negatively appear to be the earlier eluting alkanes and cycloalkanes. This is another case where there appears to be a lot of covariance in the LRV, especially considering the fact that the sulfur content ranges between 0.5 and 24.1 ppm.

Therefore if covarying compounds were not an issue, it would be reasonable to expect several low intensity peaks of sulfur containing compounds (in the chromatograms) to have relatively high positive intensity in the LRV located in the aromatics region (see Appendix for more a detailed investigation).

The results for the PLS model (4 LV) constructed for sustained boiling are summarized in Table 3.3, and the predicted values as a function of the measured values, the scree, and the LRV are provided in Figure 3.14. The LRV is shown in Figure 3.14C and 3.14D and bares similarity to the LRV for the model of kinematic viscosity in that the heavier later eluting compounds (including aromatics) contribute positively while the light earlier eluting compounds contribute negatively.

The results for the PLS model (4 LV) constructed for vapor rising are provided in Table 3.3, and the predicted values as a function of the measured values, the scree, and the LRV are provided in Figure 3.15. The LRV shown in Figure 3.15C and 3.15D is extremely similar to the aforementioned LRV for the PLS model of sustained boiling, though there are subtle differences in the intensities (magnitude of contribution to the LRV) of some peaks.

Supplemental Figure Captions

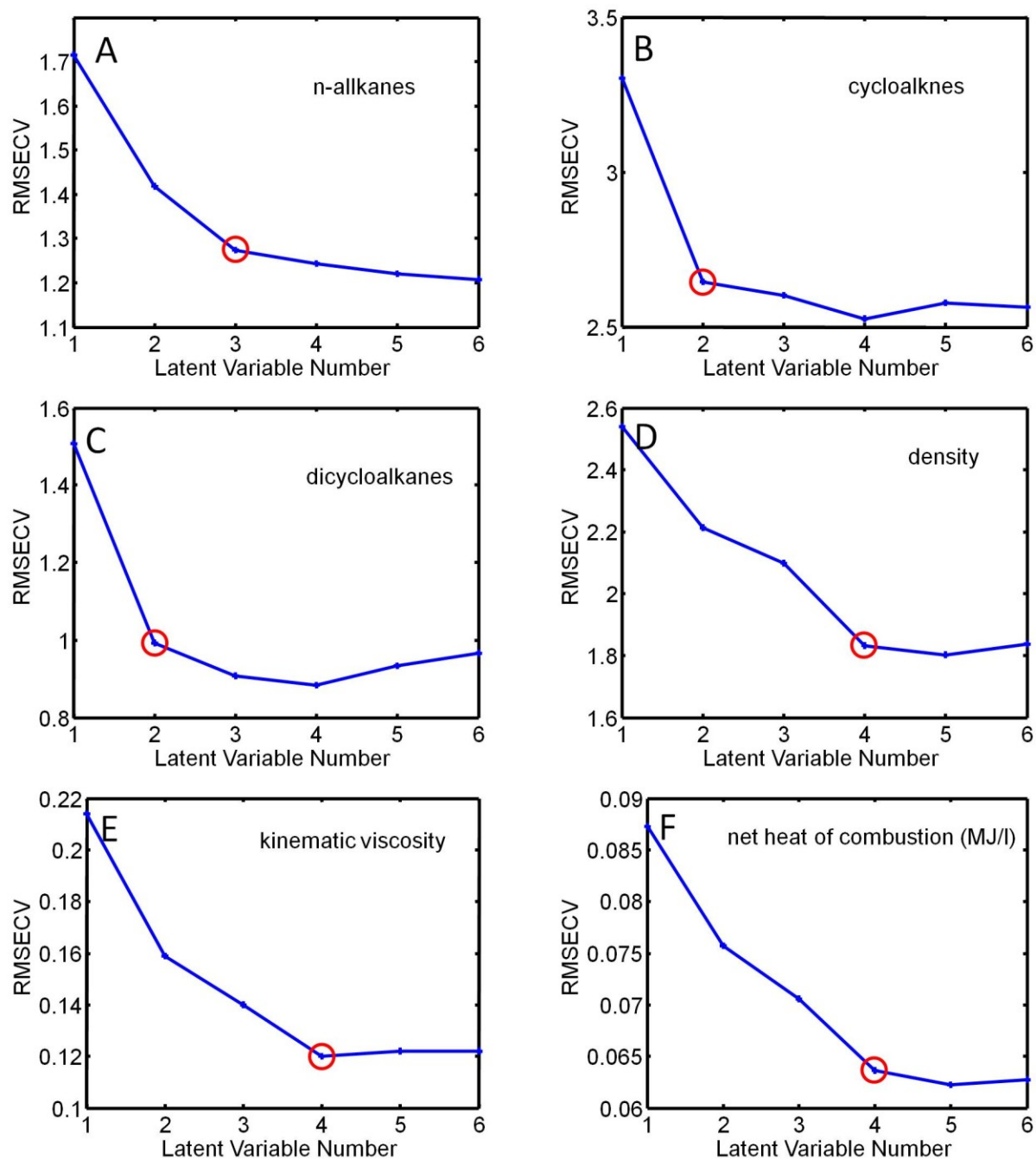


Figure 3.6. Scree plots. (A) *n*-alkanes. (B) cycloalkanes. (C) di-cycloalkanes. (D) density. (E) kinematic viscosity. (F) net heat of combustion (MJ/l).

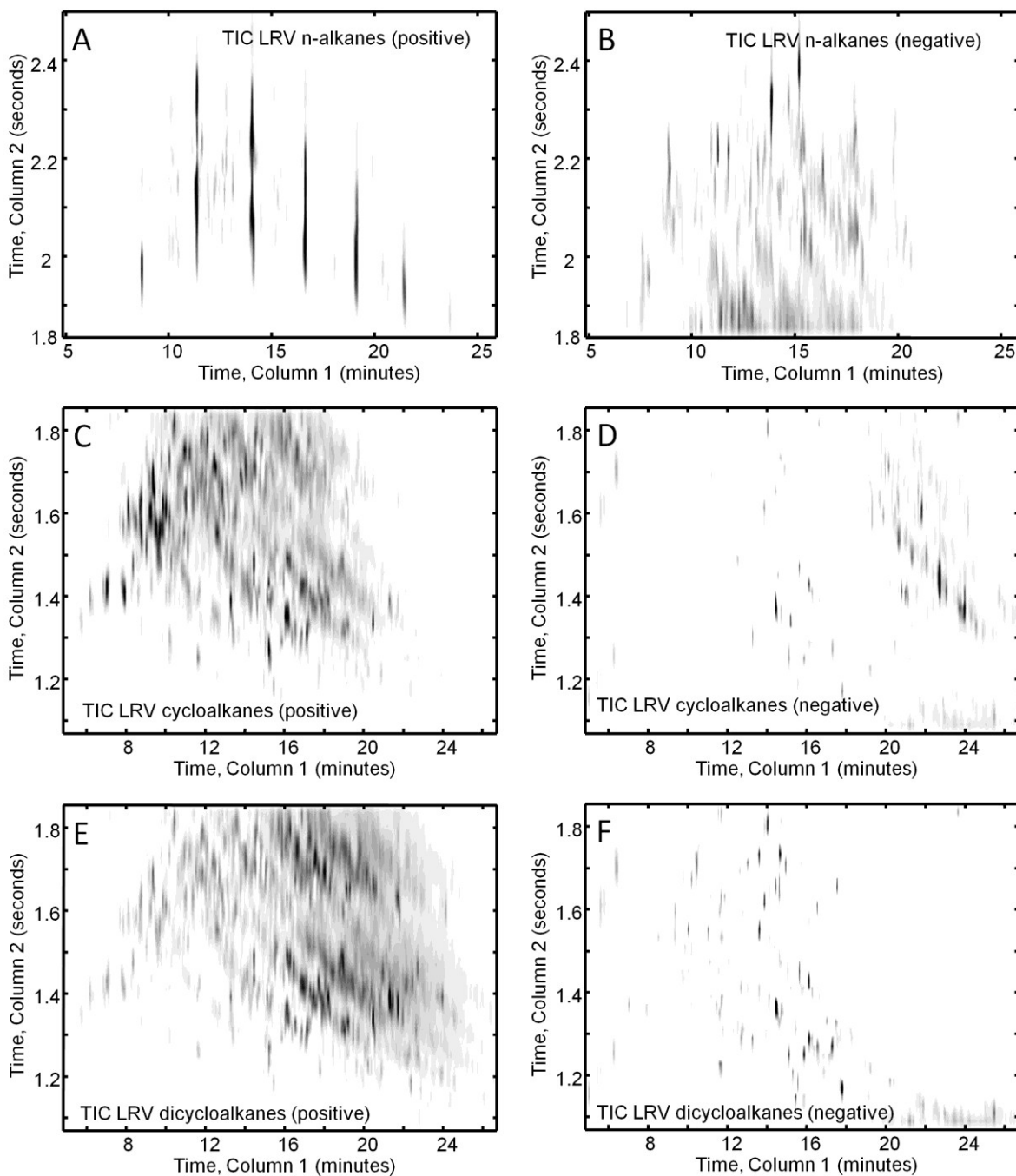


Figure 3.7. LRVs of highlighted PLS models of compositional properties. (A) The positive values in the LRV for *n*-alkanes. (B) The negative values in the LRV for *n*-alkanes. (C) The positive values in the LRV for cycloalkanes. (D) The negative values in the LRV for cycloalkanes. (E) The positive values in the LRV for di-cycloalkanes. (F) The negative values in the LRV for di-cycloalkanes.

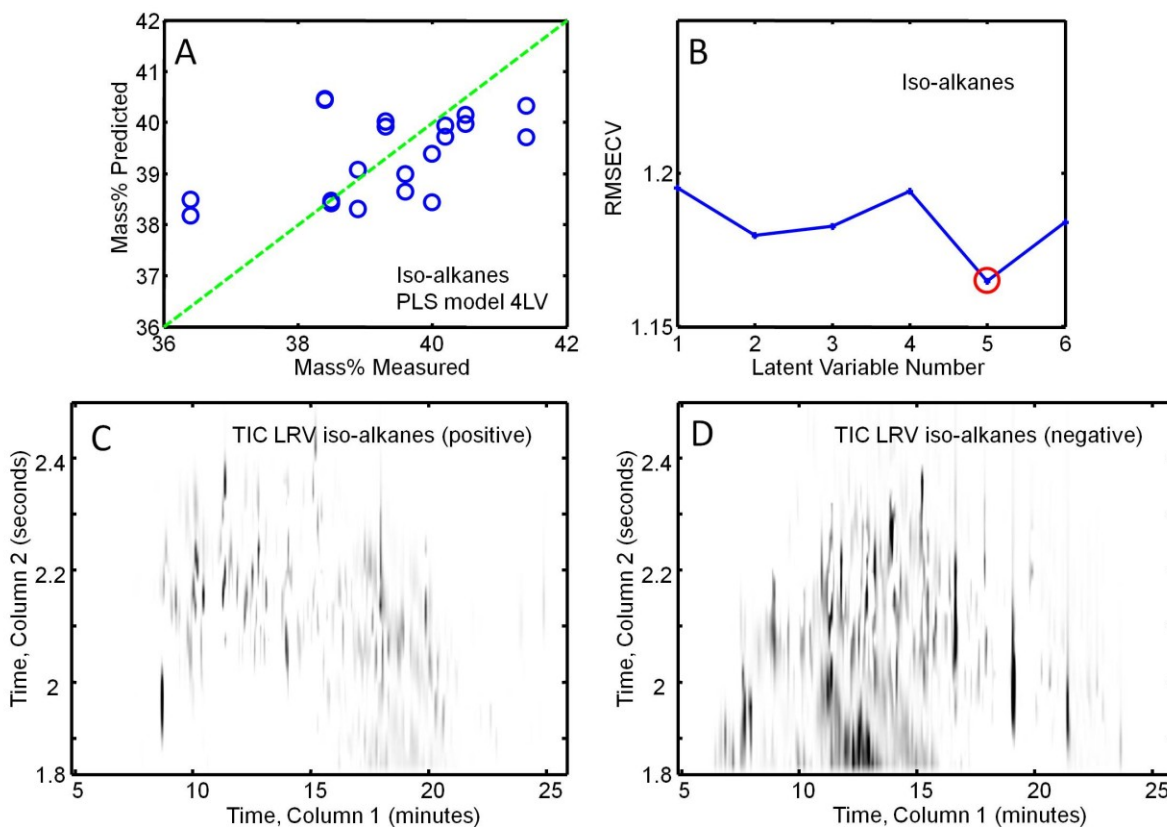


Figure 3.8. Results of the 3 LV PLS model for iso-alkanes. (A) Comparison between the predicted mass% values derived from the GC \times GC – TOFMS with PLS model (5 LV) and the measured mass% values for *iso*-alkanes (ASTM D2425), same as with n-alkanes in Figure 3.3. (B) Scree plot for PLS model for *iso*-alkanes. (C) The positive values in the LRV for *iso*-alkanes. (D) The negative values in the LRV for *iso*-alkanes.

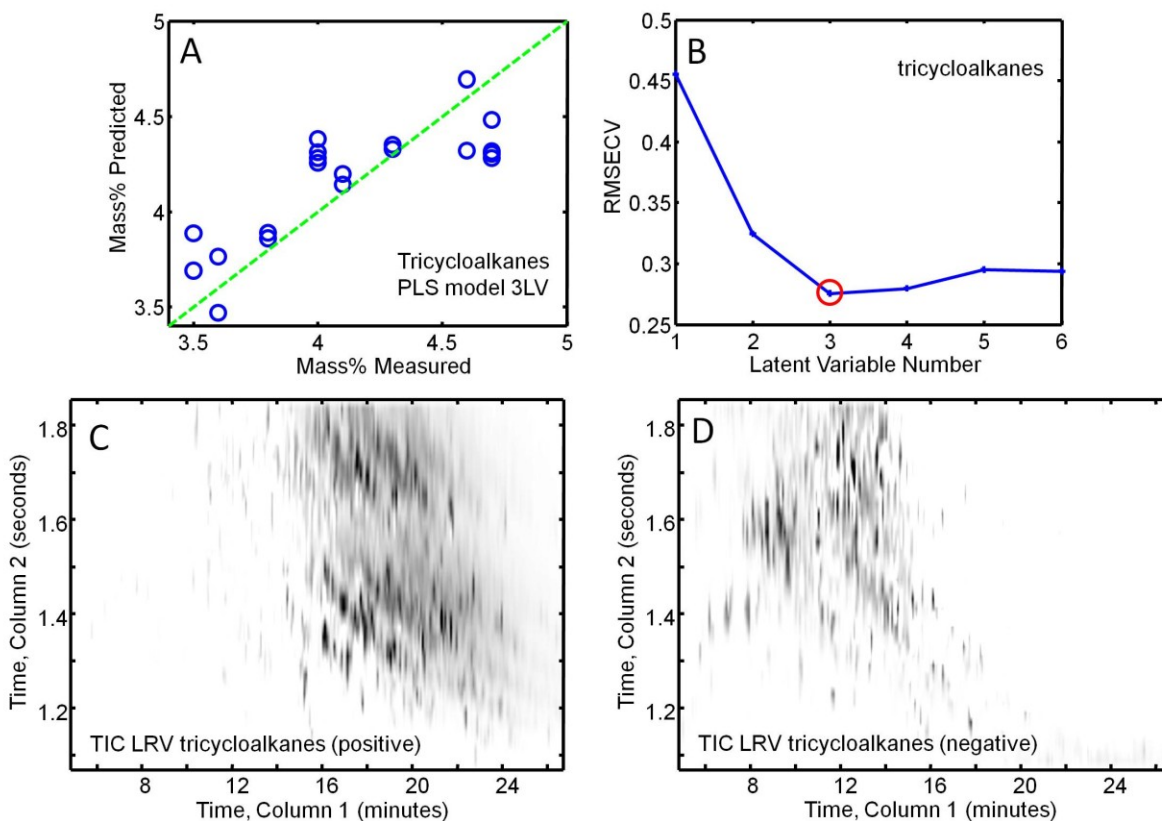


Figure 3.9. Results of the 3 LV PLS model for tri-cycloalkanes. (A) Comparison between the predicted mass% values derived from the GC \times GC – TOFMS with PLS model (3 LV) and the measured mass% values for tri-cycloalkanes (ASTM D2425). (B) Scree plot for PLS model for tri-cycloalkanes. (C) The positive values in the LRV for tri-cycloalkanes. (D) The negative values in the LRV for tri-cycloalkanes.

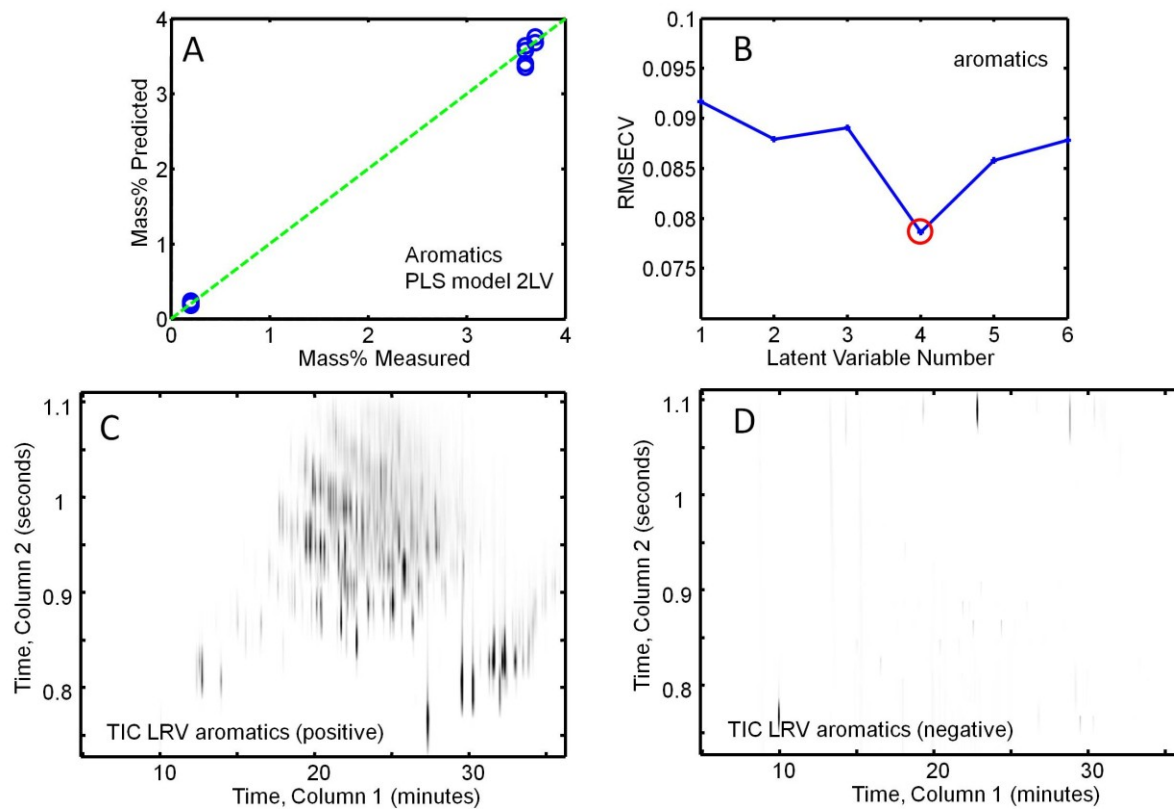


Figure 3.10. Results of the 3 LV PLS model for aromatics. (A) Comparison between the predicted mass% values derived from the GC \times GC – TOFMS with PLS model (3 LV) and the measured mass% values for aromatics (ASTM D6379). (B) Scree plot for PLS model for aromatics. (C) The positive values in the LRV for aromatics. (D) The negative values in the LRV for aromatics.

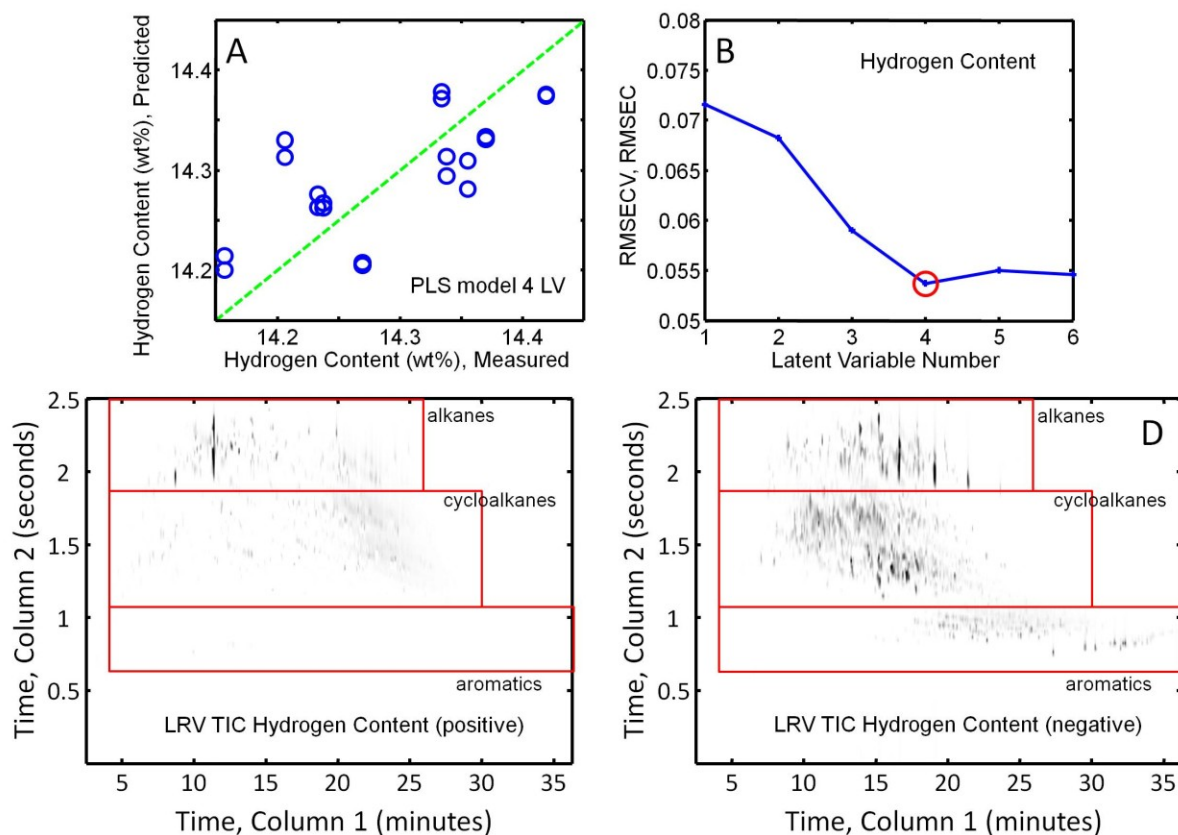


Figure 3.11. Results of the 3 LV PLS model for hydrogen content. (A) Hydrogen content (wt%) predicted values plotted relative to the measured values (Perkin Elmer Elemental Analyzer, Model EA2400). The dashed line represents an ideal agreement between the predicted and measured values and the solid line represents the linear regression best fit line. (B) Scree plot for PLS model for hydrogen content. (C) Hydrogen content LRV, positive values only. (D) Hydrogen content LRV, negative values only.

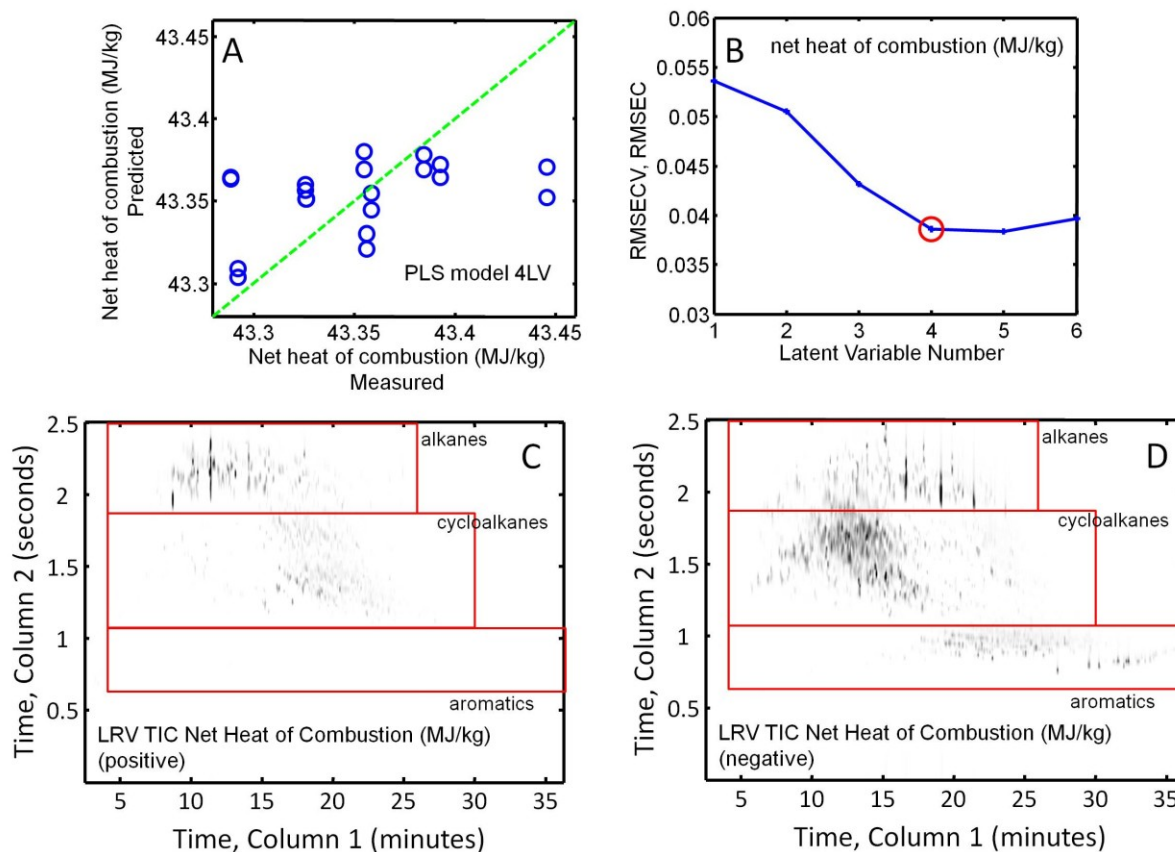


Figure 3.12. Results of the 3 LV PLS model for net heat of combustion (MJ/kg). (A) Net heat of combustion (MJ/kg) predicted plotted relative to the measured values (ASTM D4809). The dashed line represents an ideal agreement between the predicted and measured values and the solid line represents the linear regression best fit line. (B) Scree plot for PLS model for net heat of combustion (MJ/kg). (C) Net heat of combustion (MJ/kg) LRV, positive values only. (D) Net heat of combustion (MJ/kg) LRV, negative values only.

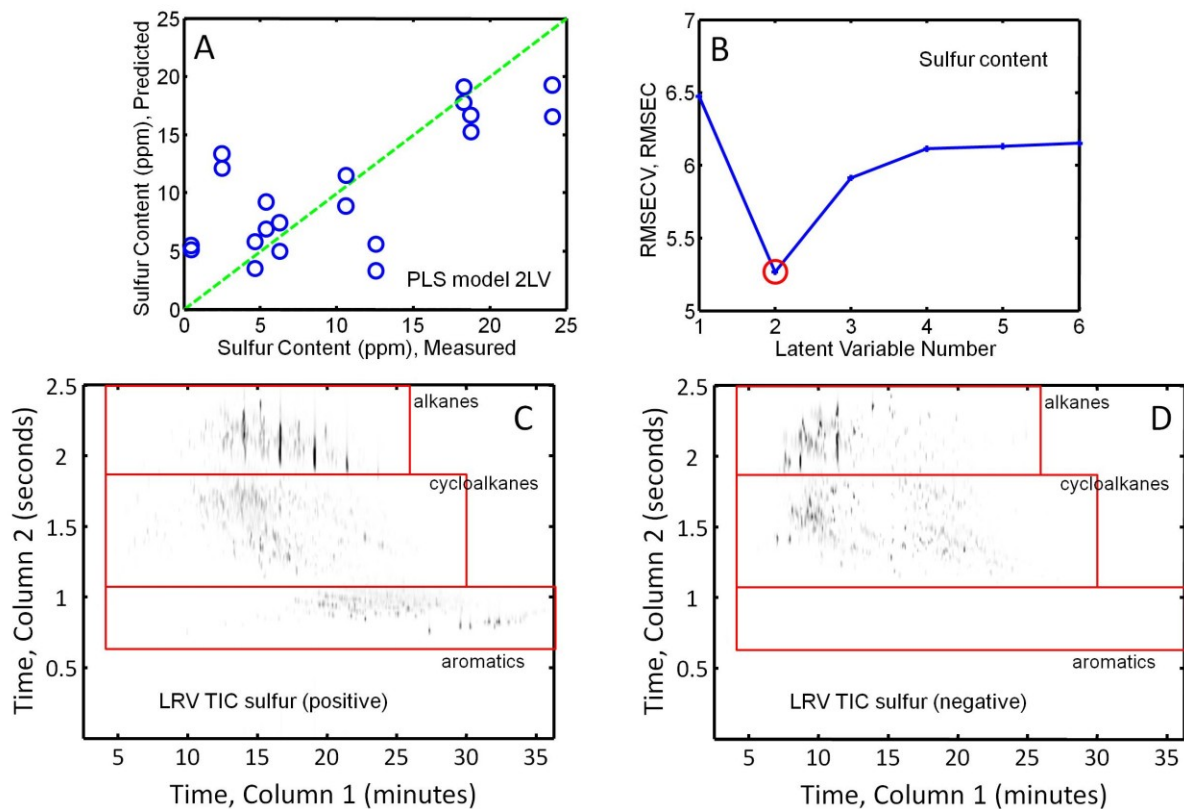


Figure 3.13. Results of the PLS model (3 LV) for sulfur content (ppm). (A) Sulfur content (ppm) predicted values plotted relative to the measured values (using SCD). The dashed line represents an ideal agreement between the predicted and measured values and the solid line represents the linear regression best fit line. (B) Scree plot for PLS model for sulfur content. (C) Sulfur content (ppm) LRV, positive values only. (D) Sulfur content (ppm) LRV, negative values only.

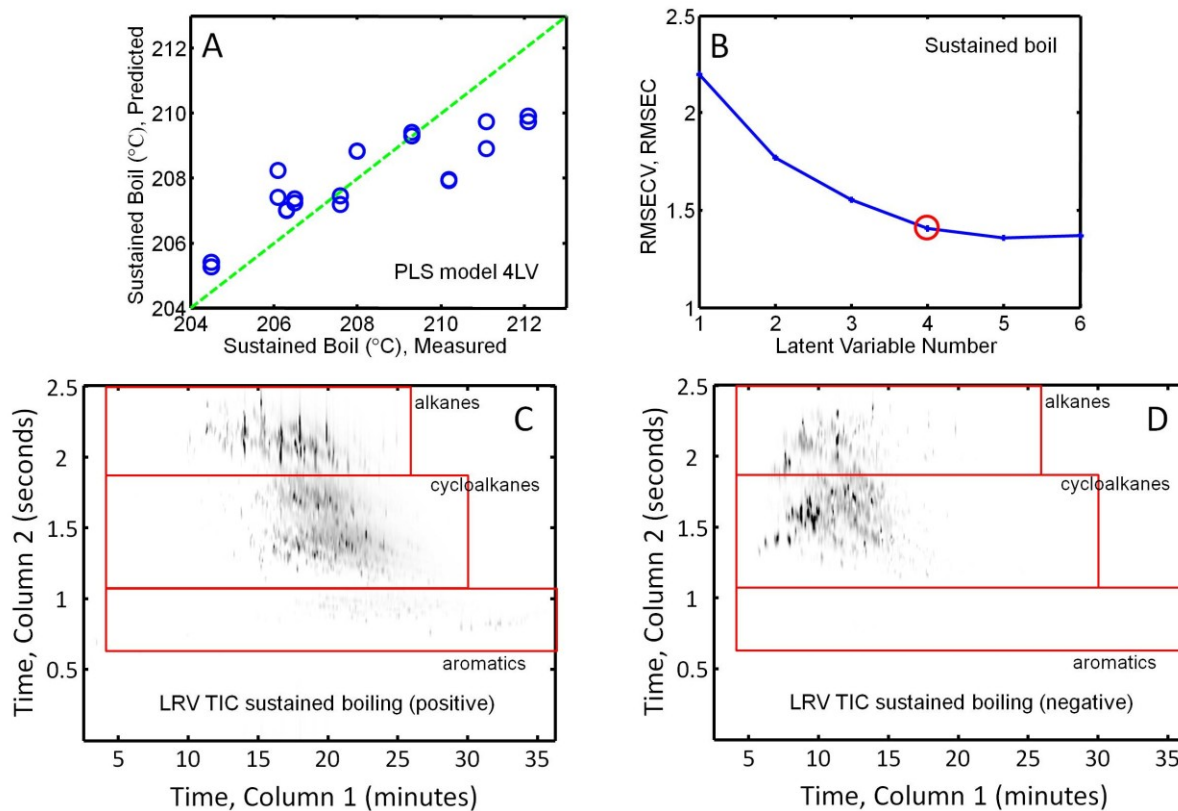


Figure 3.14. Results of the PLS model (3 LV) for sustained boil (°C). (A) Sustained boiling temperature (°C) predicted values plotted relative to the measured values. The dashed line represents an ideal agreement between the predicted and measured values and the solid line represents the linear regression best fit line. (B) Scree plot for PLS model for sustained boil. (C) Sustained boil LRV, positive values only. (D) Sustained boil LRV, negative values only.

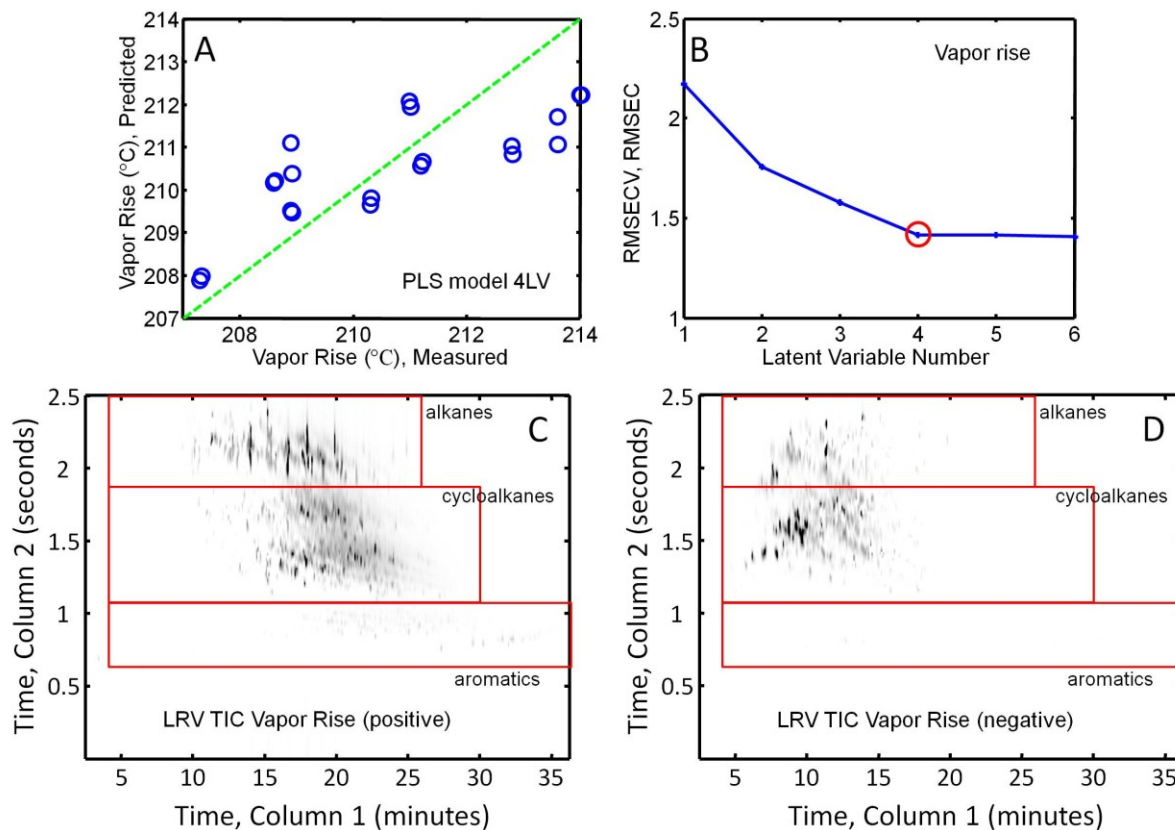


Figure 3.15. Results of the PLS model (3 LV) for vapor rise (°C). (A) Vapor rise temperature (°C) predicted values plotted relative to the measured values. The dashed line represents an ideal agreement between the predicted and measured values and the solid line represents the linear regression best fit line. (B) Scree plot for PLS model for vapor rise. (C) Vapor rise LRV, positive values only. (D) Vapor rise LRV, negative values only.

Supplemental Tables

Table 3.6. Major contributing compounds identified in LRVs for kinematic viscosity (positive) per Figure 3.4B.

#	Compound Identification	¹ t _r (min)	² t _r (s)	MV	Compound Class
1	Tetradecane	19.5	2.01	881	alkanes
2	Tetradecane, 2-methyl-	19.88	2.01	919	alkanes
3	4-Methyltridecane	17.42	2.05	910	alkanes
4	1-Cyclohexylheptane	18.79	1.62	863	cycloalkanes
5	1-Methyl-4-(1-methylbutyl)cyclohexane (C ₁₂ H ₂₄)	19.63	1.68	812	cycloalkanes
6	Octylcyclohexane	21.29	1.6	880	cycloalkanes
7	alkane branched di-cycloalkane (C ₁₅ H ₂₈)	22.21	1.44	825	di- & tri- cycloalkanes
8	1,1'-Bicyclohexyl, 2-methyl-, cis- (C ₁₃ H ₂₄)	20.83	1.36	818	di- & tri- cycloalkanes
9	Bicyclohexane	20	1.31	895	di- & tri- cycloalkanes
10	no significant peak found				mono-aromatics
11	no significant peak found				di-aromatics

Table 3.7. Major contributing compounds identified in LRVs for kinematic viscosity (negative) per Figure 3.4B.

#	Compound Identification	¹ t _r (min)	² t _r (s)	MV	Compound Class
1	Undecane	10.83	1.99	950	alkanes
2	Undecane	10.83	2.23	912	alkanes
3	2-methylnonane	7.17	1.98	927	alkanes
4	1-Methyl-2-propylcyclohexane (C10H20)	8.91	1.71	836	cycloalkanes
5	Cyclohexane, 1-ethyl-1,3-dimethyl-, trans- (C10H20)	7.67	1.52	889	cycloalkanes
6	Cyclohexane, dimethyl-, ethyl- (C10H20)	8.33	1.59	853	cycloalkanes
7	trans-Decahydronaphthalene (C10H18)	14.75	1.24	942	di- & tri- cycloalkanes
8	cis-octahydro-Indene (C9H16)	11.17	1.24	946	di- & tri- cycloalkanes
9	Dimethyladamantane (C12H20)	17.92	1.25	882	di- & tri- cycloalkanes
10	Tetramethylbenzene (C10H14)	21.21	0.87	936	mono-aromatics
11	Tetrahydronaphthalene (C10H12)	22.25	0.81	960	mono-aromatics
12	Xylene	12.25	0.81	969	mono-aromatics
13	Azulene (C10H8)	26.83	0.77	908	di-aromatics
14	Methylnaphthalene	29.04	0.81	920	di-aromatics
15	Methylnaphthalene	29.75	0.80	906	di-aromatics

Table 3.8. Major contributing compounds identified in LRVs for net heat of combustion (MJ/l) (positive) per Figure 3.5B.

#	Compound Identification	¹ t _r (min)	² t _r (s)	MV	Compound Class
1	Tridecane, methyl- (C14H30)	14.75	2.38	807	alkanes
2	2-Methyltridecane (C14H30)	17.5	2.26	925	alkanes
3	Dodecane, trimethyl- (C15H32)	19.42	2.24	856	alkanes
4	Cyclohexane, 1-methyl-4-(1-methylethyl)-, cis- (C10H20)	17.08	1.72	838	cycloalkanes
5	Cyclohexane, 1-ethyl-2-methyl-, cis- (C9H18)	17.58	1.7	817	cycloalkanes
6	Heptylcyclohexane (C13H26)	18.83	1.66	851	cycloalkanes
7	Methyldecahydronaphthalene (C11H20)	16.67	1.32	931	di-&tri- cycloalkanes
8	2-Methyladamantane (C11H18)	17.33	1.22	870	di-&tri- cycloalkanes
9	no significant peak found				mono-aromatics
10	no significant peak found				di-aromatics

Table 3.9. Major contributing compounds identified in LRVs for net heat of combustion (MJ/l) (negative) per Figure 3.5B.

#	Compound Identification	¹ t _r (min)	² t _r (s)	MV	Compound Class
1	Decane isomer (C10H22)	8.25	1.98	944	alkanes
2	Undecane (C11H24)	10.92	2.12	953	alkanes
3	2,6-Dimethyldecane (C12H26)	10.92	2.32	910	alkanes
4	Cyclohexane, 1-ethyl-1,3-dimethyl-, trans- (C10H20)	7.67	1.62	845	cycloalkanes
5	Cyclohexane, 1,2-diethyl-, cis- (C10H20)	8.33	1.64	859	cycloalkanes
6	1-Methyl-4-(1-methylethyl)-cyclohexane (C10H20)	8.92	1.7	882	cycloalkanes
7	Cyclohexane, (2-methyl-1-propenyl)- (C10H18)	11.17	1.4	863	cycloalkanes
8	Naphthalene, decahydro-2-methyl- (C11H20)	14.08	1.42	929	di- & tri- cycloalkanes
9	Xylene (C8H10)	12.25	0.82	962	mono-aromatics
10	Naphthalene, 1,2,3,4-tetrahydro- (C10H12)	22.25	0.86	960	mono-aromatics
11	Naphthalene, 2-methyl- (C11H10)	29.08	0.82	903	di-aromatics
12	2,6-Dimethylnaphthalene (C12H12)	31.83	0.84	920	di-aromatics

Table 3.10. Major contributing compounds identified in LRVs for hydrogen content (positive) per Figure S3.6C.

#	Compound ID	¹ t _r (min)	² t _r (s)	MV	Compound Class
1	Decane (C ₁₀ H ₂₂)	8.25	1.98	875	alkanes
2	Undecane (C ₁₁ H ₂₄)	10.92	2.1	954	alkanes
3	2,6-Dimethyldecane (C ₁₂ H ₂₆)	10.92	2.32	910	alkanes
4	Cyclohexane, 1,2-diethyl-, cis- (C ₁₀ H ₂₀)	8.33	1.64	864	cycloalkanes
5	Cyclohexane, (2-methyl-1-propenyl)- (C ₁₀ H ₁₈)	11.17	1.4	875	cycloalkanes
6	Cyclic (formula unknown)	20	1.68	750+	cycloalkanes
7	Bicyclohexyl (C ₁₂ H ₂₂)	20	1.34	886	di- & tri-
8	Decahydro-pentamethylnaphthalene (C ₁₅ H ₂₈)	23.5	1.4	807	di- & tri-
9	no significant peak found*				mono-aromatics
10	no significant peak found				di-aromatics

* “No significant peak found” indicates no analytes contribute significantly from that compound class.

Table 3.11. Major contributing compounds identified in LRVs for hydrogen content (negative) per Figure S3.6D.

#	Compound ID	¹ t _r (min)	² t _r (s)	MV	Compound Class
1	Tridecane, 7-methyl- (C ₁₄ H ₃₀)	14.75	2.38	805	Alkanes
2	Tridecane (C ₁₃ H ₂₈)	16.17	2.08	946	Alkanes
3	Tetradecane (C ₁₄ H ₃₀)	16.17	2.2	910	Alkanes
4	Isobutyl-Dimethylcyclohexane (C ₁₂ H ₂₄)	13.58	1.74	818	cycloalkanes
5	C ₁₂ H ₂₄ alkylated cycloalkane	14.67	1.72	750+	cycloalkanes
6	Cyclohexane, (3-methylpentyl)- (C ₁₂ H ₂₄)	15.17	1.68	888	cycloalkanes
7	Naphthalene, decahydro-2-methyl- (C ₁₁ H ₂₀)	14.75	1.42	923	di- & tri- cycloalkanes
8	Decalin, syn-1-methyl-, cis- (C ₁₁ H ₂₀)	16.67	1.32	924	di- & tri- cycloalkanes
9	Naphthalene, 1,2,3,4-tetrahydro- (C ₁₀ H ₁₂)	22.25	0.86	957	mono-aromatics
10	Naphthalene, 1,2,3,4-tetrahydro-6-methyl- (C ₁₁ H ₁₄)	24.58	0.9	928	mono-aromatics
11	Benzene, cyclohexyl- (C ₁₂ H ₁₆)	25.33	0.94	928	mono-aromatics
12	Naphthalene (C ₁₀ H ₈)	26.83	0.78	902	di-aromatics
13	Dimethylnaphthalene (C ₁₂ H ₁₂) (isomer)	31.67	0.84	934	di-aromatics
14	Dimethylnaphthalene (C ₁₂ H ₁₂)	31.83	0.84	920	di-aromatics

Chapter 4: Analysis of RP-2 Fuel Using PCA

4.1 Introduction

Along with the RP-1 research discussed in chapter 3, the RP-2 fuels were investigated with respect to variability in their chemical composition. There has been some research on the topic of compositional differences between RP-2 fuels^{77,98}. The purpose of the following study was to provide GC × GC – TOFMS chromatograms of the RP-2 fuels (listed in Table 4.1) using the reversed column configuration⁹ as mentioned in chapter 3, and using principal component analysis (PCA) to highlight the overall compositional variation between the seven analyzed RP-2 fuels. PCA provides not only scores plots for summarizing similarities and differences but also the loadings that can be used to inspect the sources of the variance in the samples, i.e. in the recorded variables. In the loadings, the variables that show a high amount of variance typically correspond to peaks found in the GC × GC – TOFMS chromatograms. A mass spectrum could be extracted and deconvoluted when needed using PARAFAC⁵² and identified using NIST11 V2.0g mass spectral library (National Institute of Standards and Technology, Boulder CO, USA) as previously established⁷⁴.

(98) L.S. Ott, A.B. Hadler, T.J. Bruno, *Ind. Eng. Chem. Res.* 47 (2008) 9225–9233.

4.2 Experimental

The RP-2 samples were run at the same time as the RP-1 fuels using the exact instrument and setup as previously stated in section 3.2 and in a published report⁷⁴.

The samples were obtained from the Air Force Research Laboratory (AFRL) and prepared in the same manner as described in section 3.2⁷⁴. The list of RP-2 fuels analyzed can be found in Table 4.1.

The .peg files were imported into Matlab2009b (Mathworks, Natick MA, USA) and preprocessed using both commercial and in-house software (as in previous reports)^{35-38,74}. The GC × GC – TOFMS chromatograms underwent baseline correction and 2-point binning on both the first and second column dimensions. The GC × GC – TOFMS chromatograms were unfolded and mean centered using PCA algorithm in PLS_Toolbox V6.02 (Eigenvector Inc. Winachee WA, USA).

Compounds of interest found in the loadings of the first two PCs that have high (positive or negative) values and correspond to peaks in the GC × GC – TOFMS chromatograms were identified in the same manner as described in section 3.2 and in a previous report⁷⁴ using ChromaTOF, the in-house non-target PARAFAC, and the NIST mass spectral library^{52,74}.

4.3 Results and Discussion

4.3.1 RP-2 Chromatograms

The RP-2 GC × GC – TOFMS chromatograms initially underwent visual inspection. The separation of compounds showed very similar results to the aforementioned RP-1 GC × GC –

TOFMS chromatograms⁷⁴. Select GC \times GC – TOFMS chromatograms of RP-2 fuels are highlighted in Figure 4.1A-C which are the TIC GC \times GC – TOFMS chromatograms of samples WC0721HW01, YA2921HW10, and LB011011-01, respectively. The alkanes are separated between 5 and 28 minutes on the first column, and between 1.7 and 2.5 seconds on the second column. The cycloalkanes are predominantly in the region between 5 and 28 minutes on the first column, and between 1.4 and 1.7 seconds on the second column. The di- and tri-cycloalkanes can be found in the region between 8 to 31 minutes on the first column, and between 1.0 and 1.5 seconds on the second column. The aromatics can be found in the region between 9 to 31 minutes on the first column and between 0.7 and 1.0 seconds on the second column. The di-aromatics can be found beside the aromatics region between 26 and 37 minutes on the first column and between 0.6 and 1.0 seconds on the second column. The chromatograms shown were chosen due to their visual differences and their PCA scores (further discussed below).

4.3.2 PCA Results

PCA was performed on two sets of data: first the averaged GC \times GC – TOFMS chromatograms, and second on all replicates combined as a single data set. The latter PCA model was done after the original PCA model to check for abnormalities in the replicates that could potentially bias the results of the original PCA model.

According to the PCA results of the averaged GC \times GC – TOFMS chromatograms shown in Figure 4.2, PCs 1 & 2 captured 87.41% of the variance. When PCs 3 & 4 are also included in the model, the total variance captured was 97.94%. WC0721HW01 showed the most difference with respect to PC1, and LB011011-01 shows the most difference with respect to PC2.

YA2921HW10 appears to be representative of the average RP-2 sample as it exhibits scores close to 0 for both PC1 and PC2 and may be considered as a representative average of all GC \times GC – TOFMS chromatograms analyzed. To compliment the scores, the loadings are shown in Figures 4.3 and 4.4. For purposes of clarity, the loadings are broken into two parts; one subfigure showing the positive values and one subfigure showing the negative values. For each PC, regions sharing the same sign are said to be correlated while regions with opposite signs are said to be inversely correlated (i.e. when one is apparent in the sample, the other will be less apparent or not at all). Examples of correlation in PC1 can be seen in the positive score values shown in Figure 4.3A, where some of the dominant peaks that show high values are the alkanes located in the region between 7 and 15 min on the first column dimension, and between 1.7 and 2.5 sec on the second column dimension. An example of correlation in the negative values can be seen in Figure 4.3B where some of the dominant peaks that show high values are the cycloalkanes located in the region between 5 and 10 min on the first column dimension, and between 1.2 and 1.6 sec on the second column dimension. For an example of inverse correlation, one only needs to look at different peaks between positive and negative values in the same PC. A noteworthy feature found in the loadings of PC2 (Figure 4.3C) are the alkane peaks located in the upper right corner, including one big peak that exhibits wrap-around. A complete list of select analytes of interest for PC1 (positive and negative) and PC2 (positive and negative) are shown in Tables 4.2-4.5.

For PCs 3 and 4 of the PCA model made from the averaged GC \times GC – TOFMS chromatograms, the scores are shown as previously stated in Figure 4.2B, and the loadings are shown in Figure 4.4, capturing 6.22% and 4.31% of the variance, respectively. PC3 describes the correlation existing between some of the di- & tri-cycloalkanes and some of the earlier

eluting compounds including some alkanes positive values which are inversely correlated with some other alkanes and cycloalkanes. PC4 describes correlation between some earlier eluting alkanes (including the wrap-around peak) against many later eluting alkanes and cycloalkanes (both cycloalkanes and di- & tri-cycloalkanes). Due to the low variance captured in these two PCs, their contribution to the variance is minor compared to PC1 and PC2.

The scores for the PCA models constructed from all replicates of GC \times GC – TOFMS chromatograms of RP-2 samples are shown in Figure 4.5. The loading was extremely similar to the loadings of the PCA model made from the averaged GC \times GC – TOFMS chromatograms and was omitted for brevity. In PCs 1, 2, and 3, the scores of GC \times GC – TOFMS chromatogram replicates are located very close to each other indicating they are extremely similar. Only in PC 4 do some replicates appear significantly different, however considering this PC captures about 4.45% of the variance the importance of this variance may not be of great significance and could potentially be attributed to noise. Overall, these results suggest the variation between replicates is minor.

4.4 Conclusions

The RP-2 fuels were analyzed using PCA, and as with the RP-1 fuels, the RP-2 fuels appear very similar with respect to chemical composition. There are few noticeable differences between RP-2 samples; a few of the earlier eluting alkanes, one of which exhibits wrap-around on the second column dimension. Also, the compositional difference between RP-2 fuels is more pronounced: some are richer in lighter compounds while others are richer in heavier compounds. The scores and loadings summarize the variance between samples in a few succinct

figures. One PCA model was constructed using the averaged GC \times GC – TOFMS chromatograms and the second PCA model was constructed using all replicates. The second PCA model was constructed to prove the variance between replicates was not significantly biasing the original PCA results.

Figure Captions

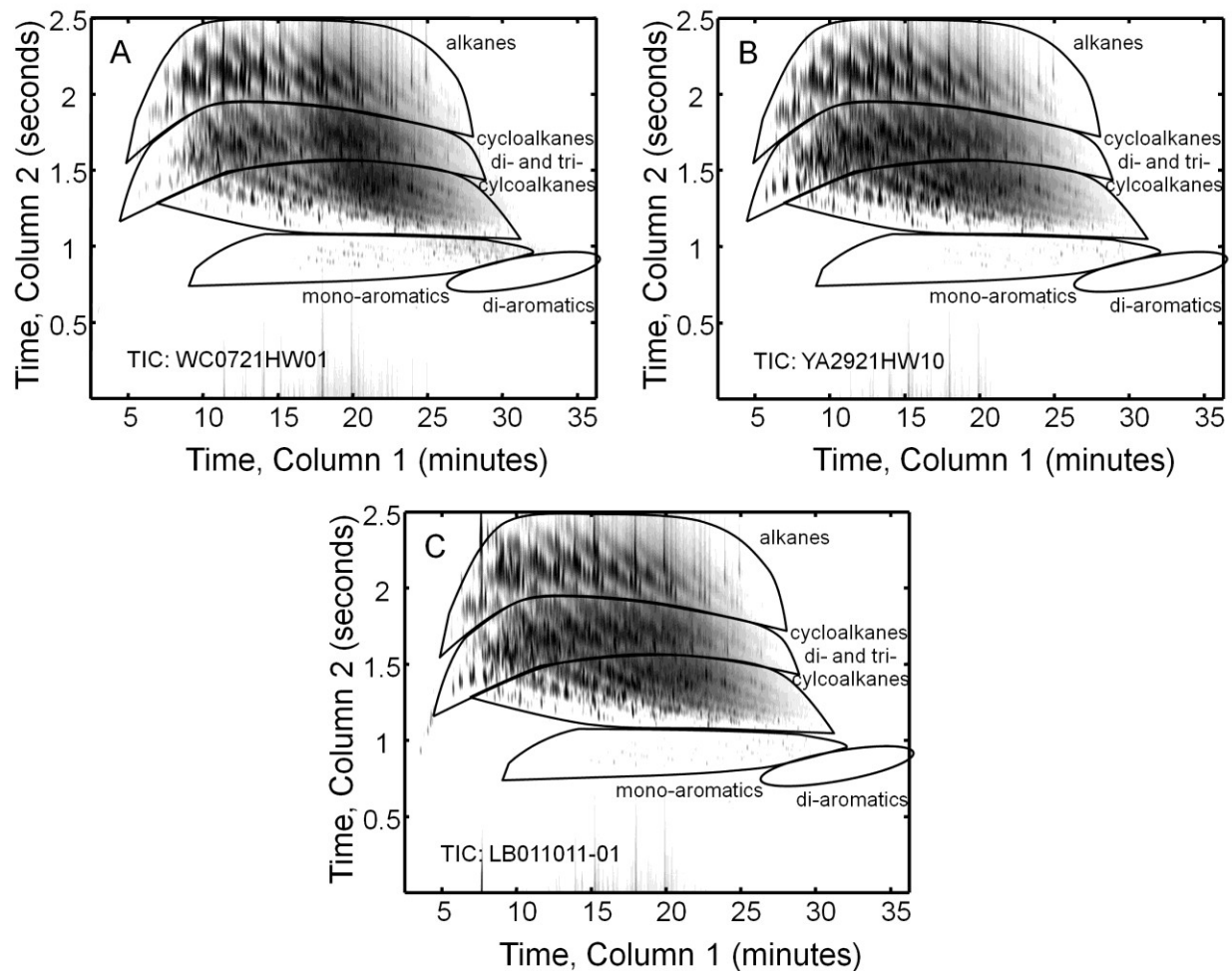


Figure 4.1. RP2 GC \times GC-TOFMS chromatograms, collected by a 30 m Rtx-wax column for the first separation dimension followed by a 1.2 m Rtx-1 column for the second separation dimension at a constant inlet pressure of 35 psig. The x-axis spans \sim 35 min, and the y-axis spans 2.5 s. Compound groups are circled and annotated. (A) TIC chromatogram of WC0721HW01. (B) TIC chromatogram of YA2921HW10. (C) TIC chromatogram of LB011011-01.

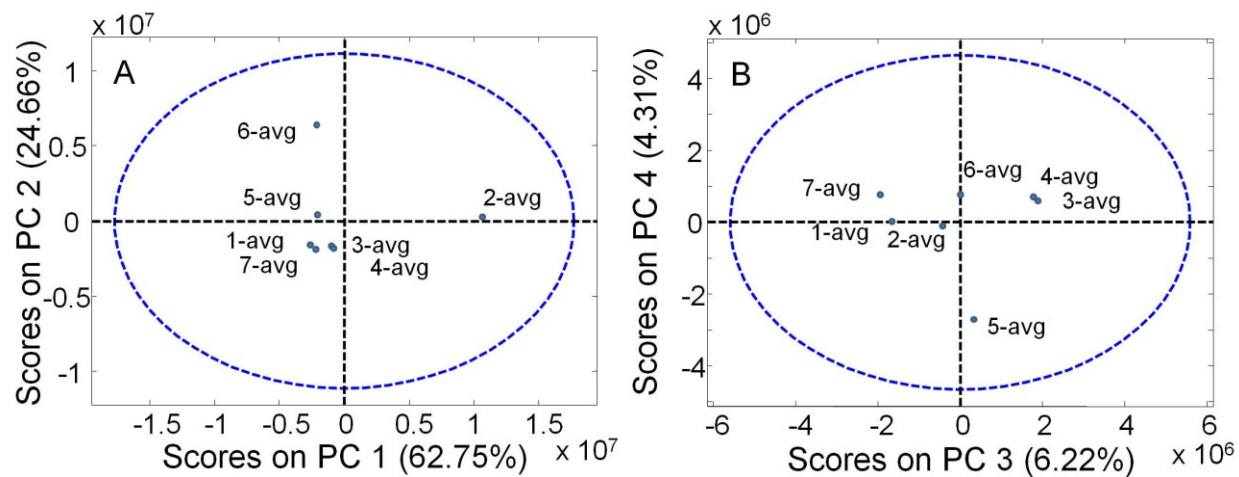


Figure 4.2. PCA scores plots using the averaged chromatograms. (A) Scores plot of PC 1 and PC 2. (B) Scores plot of PC 3 and PC 4.

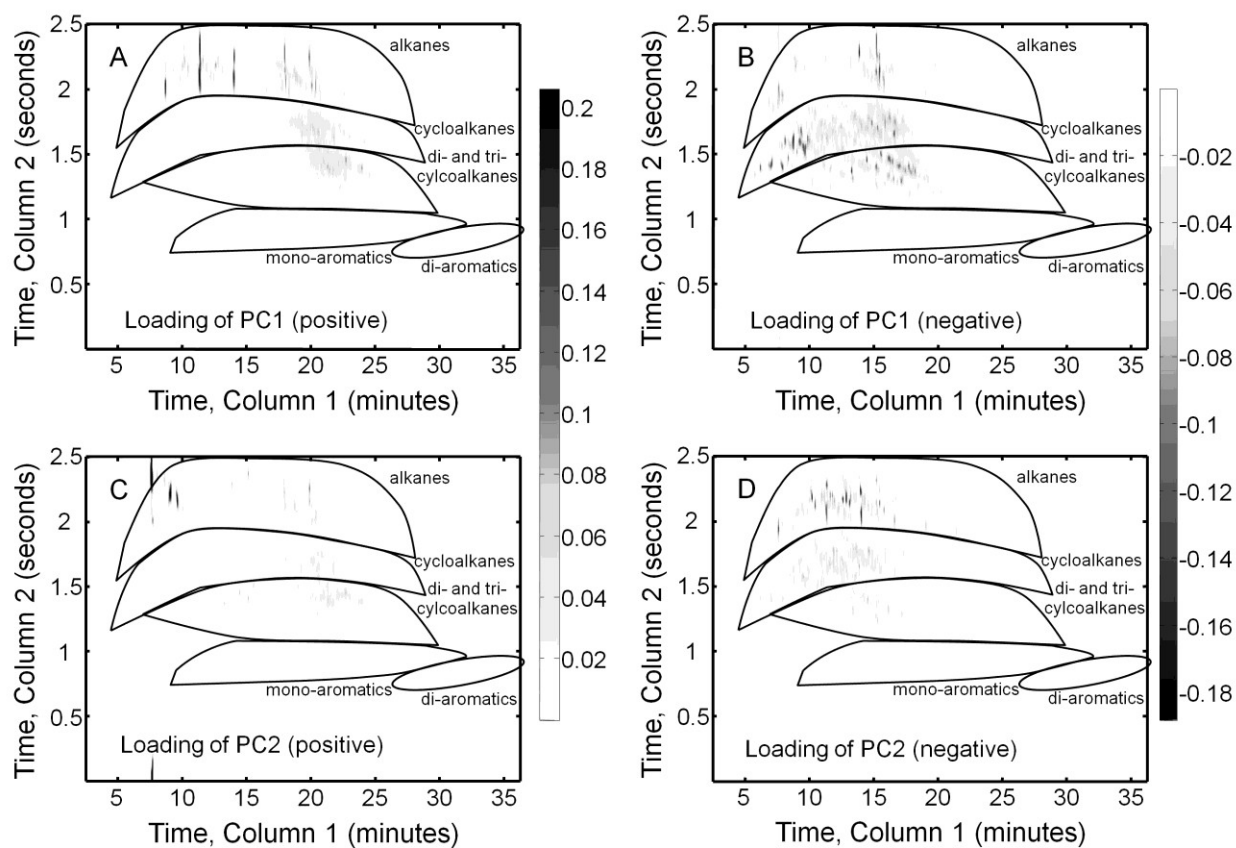


Figure 4.3. PCA (unfolded) loadings plots (TIC) using the averaged chromatograms, the positive and negative values of each loadings were separated for clarity. (A) Loadings (TIC) of PC 1 (positive values only). (B) Loadings (TIC) of PC 1 (negative values only). (C) Loadings (TIC) of PC 2 (positive values only). (D) Loadings (TIC) of PC 2 (negative values only).

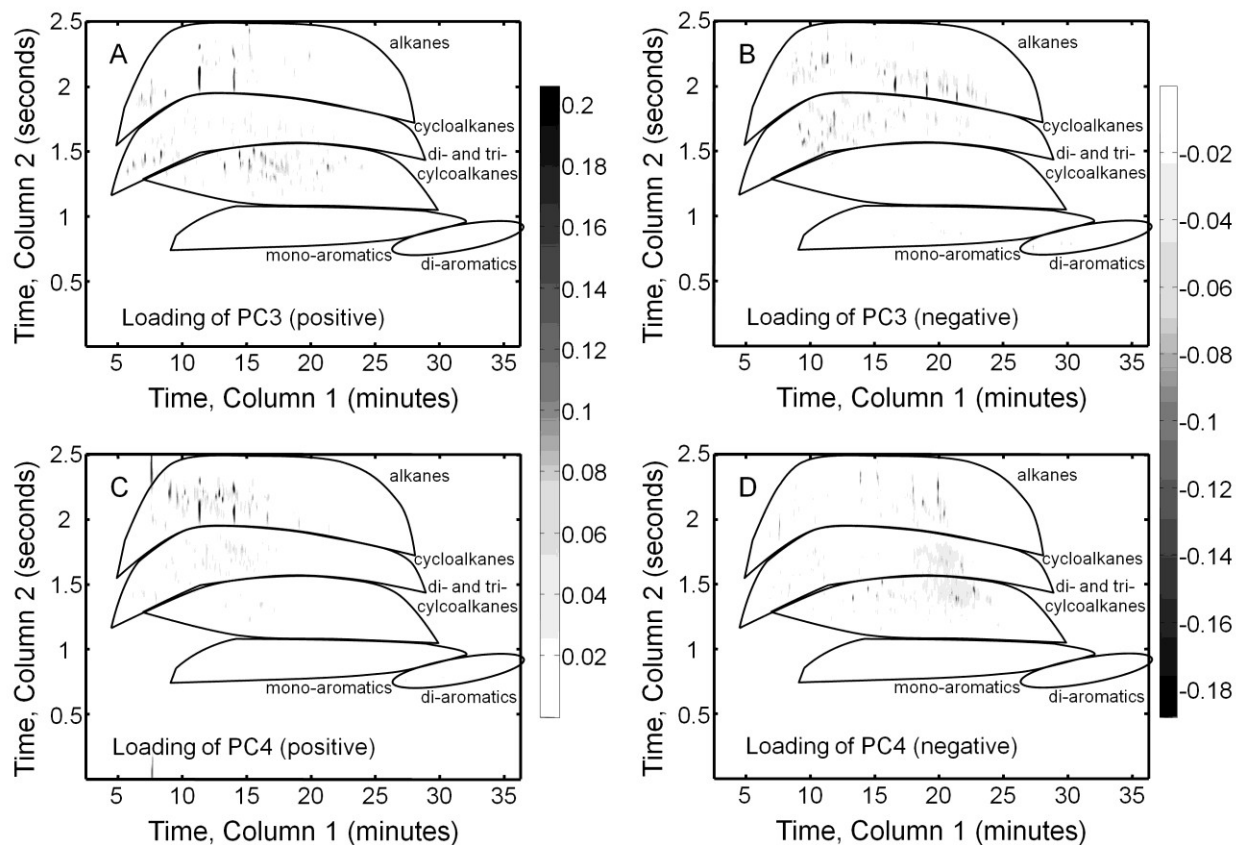


Figure 4.4. PCA (unfolded) loadings plots (TIC) using the averaged chromatograms, the positive and negative values of each loadings were separated for clarity. (A) Loadings (TIC) of PC 3 (positive values only). (B) Loadings (TIC) of PC 3 (negative values only). (C) Loadings (TIC) of PC 4 (positive values only). (D) Loadings (TIC) of PC 4 (negative values only).

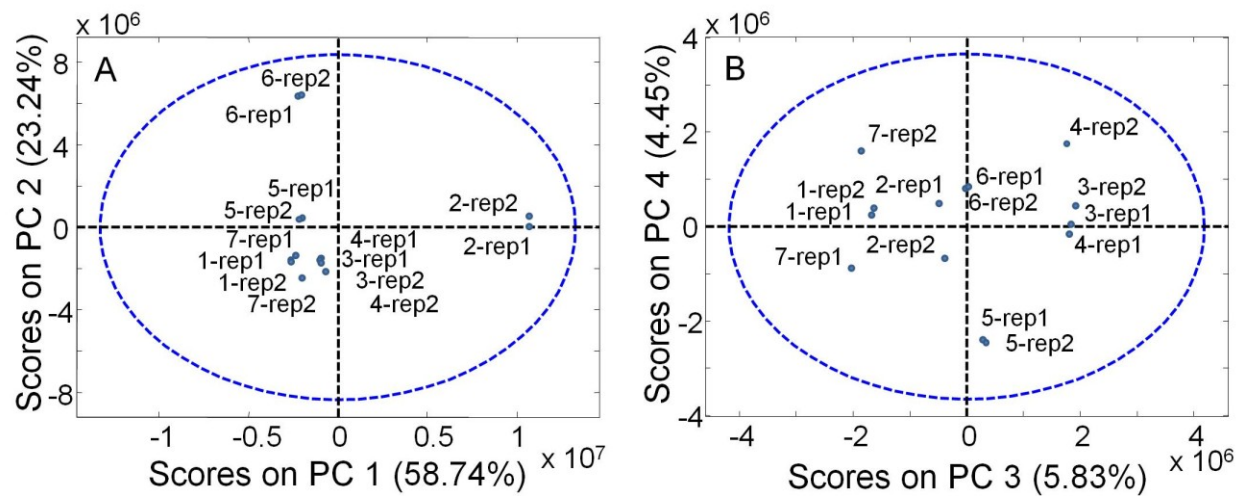


Figure 4.5. PCA scores plots using all chromatograms replicates. (A) Scores plot of PC 1 and PC 2. (B) Scores plot of PC 3 and PC 4.

Tables

Table 4.1. RP-2 Fuel Set

RP-2 Sample	UW Sample No. (RP-2 Set)	AFRL Designation	Fuel Type
1	1	011404/SA1421LS03	2004 UL-RP-1
2	2 (19 RP-1 Set)	WC0721HW01	2008 RP-2
3	3 (5 RP-1 Set)	XK1621HW10	2009 RP-2
4	4 (20 RP-1 Set)	YA2921HW10	2010 RP-2
5	5	2011 UL-Replicate	2011 RP-2
6	6	LB011011-01/Option A	2011 RP-2
7	7	RG3021LS05	2003 TS-5-RP-1

Table 4.2. Compounds of interest based on the PCA loadings of RP-2 Fuels of PC1 (positive values).

#	Compound Identification	¹ t _r (min)	² t _r (s)	MV	Compound Class
1	Decane	8.25	2.04	951	alkanes
2	Undecane	10.92	2.3	954	alkanes
3	Dodecane	13.58	2.26	941	alkanes
4	Heptylcyclohexane (C ₁₃ H ₂₆)	18.83	1.66	847	cycloalkanes
5	C ₁₄ H ₂₈ alkylated Cycloalkane	19.5	1.7	803	cycloalkanes
6	C ₁₀ H ₂₀ alkylated Cycloalkane	19.67	1.72	804	cycloalkanes
7	1,1'-Bicyclohexyl, 2-methyl-, cis- (C ₁₃ H ₂₄)	20.83	1.42	829	di- & tri- cycloalkanes
8	Hexamethyloctahydro-1H-indene (C ₁₅ H ₂₈)	22.25	1.46	824	di- & tri- cycloalkanes
9	Hexamethyloctahydro-1H-indene (C ₁₅ H ₂₈)	23.5	1.4	789	di- & tri- cycloalkanes
10	No significant peak found				mono-aromatics
11	No significant peak found				di-aromatics

Table 4.3. Compounds of interest based on the PCA loadings of RP-2 Fuels of PC1 (negative values).

#	Compound Identification	¹ t _r (min)	² t _r (s)	MV	Compound Class
1	2,6-Dimethylundecane (C13H28)	13.42	2.28	938	alkanes
2	branched alkane (C14H20)	14.75	2.4	777	alkanes
3	Undecane, 2,6-dimethyl- (C13H28)	14.75	2.14	824	alkanes
4	2-Methyldodecane (C13H28)	15	2.14	924	alkanes
5	1-Ethyl-1,3-dimethylcyclohexane (C10H20)	7.67	1.62	866	cycloalkanes
6	1-Ethyl-2,3-dimethylcyclohexane (C10H20)	8.33	1.64	870	cycloalkanes
7	1-Methyl-3-propylcyclohexane (C10H20)	8.92	1.7	855	cycloalkanes
8	decahydro-naphthalene (C10H18)	12.83	1.38	928	di- & tri-cycloalkanes
9	2-Methyldecahydronaphthalene (C11H20)	14	1.46	927	di- & tri-cycloalkanes
10	2-Ethyldecahydronaphthalene (C12H22)	17.25	1.42	805	di- & tri-cycloalkanes
11	No significant peak found				mono-aromatics
12	No significant peak found				di-aromatics

Table 4.4. Compounds of interest based on the PCA loadings of RP-2 Fuels of PC2 (positive values).

#	Compound Identification	¹ t _r (min)	² t _r (s)	MV	Compound Class
1	2,2,4,6,6-Pentamethylheptane (C12H26)	7.17	2.5	902	alkanes
2	2-Methylnonane (C10H22)	7.25	2.02	890	alkanes
3	2,2,4,4-Tetramethyloctane (C12H26)	8.58	2.22	879	alkanes
4	branched alkane (C9H20)	9.17	2.18	824	alkanes
5	Octylcyclohexane (C14H28)	21.25	1.64	884	cycloalkanes
6	trans-Decahydronaphthalene (C10H18)	12.83	1.38	921	di- & tri-cycloalkanes
7	Methyldecahydronaphthalene (C11H20)	14	1.46	921	di- & tri-cycloalkanes
8	2-Methyldecahydronaphthalene (C11H20)	14.67	1.42	926	di- & tri-cycloalkanes
9	1,1'-Bicyclohexyl, 2-methyl-, cis- (C13H24)	20.83	1.4	835	di- & tri-cycloalkanes
10	No significant peak found				mono-aromatics
11	No significant peak found				di-aromatics

Table 4.5. Compounds of interest based on the PCA loadings of RP-2 Fuels of PC2 (negative values).

#	Compound Identification	¹ t _r (min)	² t _r (s)	MV	Compound Class
1	4-Methylnonane (C ₁₀ H ₂₂)	7.17	1.96	882	alkanes
2	dimethyl-undecane (C ₁₃ H ₂₈)	13.5	2.24	871	alkanes
3	Duodecane (C ₁₂ H ₂₆)	13.58	2.06	941	alkanes
4	methyl(methylethyl)cyclohexane (C ₁₀ H ₂₀)	11.67	1.78	803	cycloalkanes
5	pentyl-cyclohexane (C ₁₁ H ₂₂)	13.42	1.66	857	cycloalkanes
6	C ₁₂ H ₂₄ alkylated Cycloalkane	14.17	1.78	875	cycloalkanes
7	2-Methyldecahydronaphthalene (C ₁₁ H ₂₀)	14.08	1.42	914	di- & tri- cycloalkanes
8	2-Methyldecahydronaphthalene (C ₁₁ H ₂₀)	14.75	1.38	930	di- & tri- cycloalkanes
9	2-Methyldecahydronaphthalene (C ₁₁ H ₂₀)	15.67	1.34	920	di- & tri- cycloalkanes
10	Tetramethylbenzene (C ₁₀ H ₁₄)	19.67	0.9	932	mono-aromatics
11	Tetramethylbenzene (C ₁₀ H ₁₄)	19.83	0.88	925	mono-aromatics
12	3,4-Dimethyl-1-ethylbenzene (C ₁₀ H ₁₄)	21.25	0.88	919	mono-aromatics
13	Naphthalene	26.83	0.78	937	di-aromatics
14	1-Methylnaphthalene	29	0.8	928	di-aromatics
15	1-Methylnaphthalene	29.75	0.8	941	di-aromatics

Chapter 5: Analysis of RP-1 Fuel and Advanced Distillation Curves using PLS Regression*

5.1 Introduction

The chemical composition of a kerosene fuel, though complex, holds a key to understanding and altering the physical properties and performance of the fuels^{79–84,87}. Achieving fine control over the chemical composition can be a difficult task. It has become increasingly important to achieve further insight into fuel composition, as well as the sources of variation in the fuel composition to both maintain and control fuel performance, as well as to assess the performance of “field” fuels^{79–83}. Fuel performance is inextricably tied to characterization, and the advanced distillation curve (ADC) method has demonstrated itself as a well suited approach for the analysis and characterization of complex fuels^{99–101}. The ADC method is a state-of-the-art approach to very accurately and precisely analyze the boiling curve of complex liquids. Samples (i.e., distillation fractions) may be obtained during the distillation, and can be further analyzed both qualitatively and quantitatively.

The ADC method was pioneered by Bruno and co-workers^{77,80,83,98–109}. Briefly, the

* Large portions of this chapter are intended for a future publication.

(99) T.J. Bruno, *Ind. Eng. Chem. Res.* 45 (2006) 4371–4380.

(100) T.J. Bruno, B.L. Smith, *Ind. Eng. Chem. Res.* 45 (2006) 4381–4388.

(101) T.J. Bruno, L.S. Ott, B.L. Smith, T.M. Lovestead, *Anal. Chem.* 82 (2010) 777–783.

(102) B.L. Smith, T.J. Bruno, *Ind. Eng. Chem. Res.* 46 (2007) 310–320.

(103) L.S. Ott, B.L. Smith, T.J. Bruno, *Fuel*. 87 (2008) 3055–3064.

(104) T.M. Lovestead, T.J. Bruno, *Energy Fuels*. 23 (2009) 3637–3644.

(105) B.C. Windom, T.J. Bruno, *Energy Fuels*. 25 (2011) 5200–5214.

apparatus for the ADC method utilizes a round-bottom flask connected to an air cooled condenser, a receiver adapter and a calibrated volume receiver. In more detail, the flask is encased with a heater in an aluminum metal jacket. Inside the flask are two thermocouples (suspended with the use of a centering adapter): one thermocouple measures the temperature of the liquid analyzed, and the other thermocouple measures the temperature in the headspace above the liquid being distilled. A stir bar and some glass shards are introduced to promote even heat distribution and prevent “bumping” of the liquid being distilled. Three bore scope ports (5 mm diameter) are strategically located to inspect both the liquid and the thermocouples inside the apparatus; their small size minimizes heat loss. The flask is connected to an air cooled condenser wherein the distillate condenses. The condenser, in turn, is connected to a special adapter where the drops of distillate fall into a small 0.05 mL “hammock.” With the use of a syringe, the distillate may be sampled from the hammock for further analysis including but not limited to gas chromatography (GC)^{98,100–105}, infrared spectroscopy¹⁰³, and measurements of enthalpy of combustion¹⁰⁴. After the adapter, the distillate reaches the calibrated volume receiver. More recently, a variation of the ADC method apparatus was implemented that controls the internal pressure, preventing sample degradation due to reactions that may potentially occur at high temperatures when analyzing samples containing low-volatility compounds¹⁰⁷. This feature was achieved by sealing every connection between parts of the

(106) B.L. Smith, T.J. Bruno, Improvements in the Measurement of Distillation Curves. 3. Application to Gasoline and Gasoline + Methanol Mixtures, *Ind. Eng. Chem. Res.* 46 (2007) 297–309.

(107) B.C. Windom, T.J. Bruno, Improvements in the Measurement of Distillation Curves. 5. Reduced Pressure Advanced Distillation Curve Method, *Ind. Eng. Chem. Res.* 50 (2011) 1115–1126.

(108) B.C. Windom, T.J. Bruno, Application of Pressure-Controlled Advanced Distillation Curve Analysis: Virgin and Waste Oils, *Ind. Eng. Chem. Res.* 52 (2013) 327–337.

(109) T.J. Bruno, A. Wolk, A. Naydich, Stabilization of Biodiesel Fuel at Elevated Temperature with Hydrogen Donors: Evaluation with the Advanced Distillation Curve Method, *Energy Fuels.* 23 (2009) 1015–1023.

apparatus and using a commercial pressure controller. Sampling is performed with a reduced pressure balance syringe.

The ADC method has been instrumental in the study of a variety of complex liquid samples including, but not limited to, crude oil¹⁰³, gasoline¹⁰⁶, biodiesel fuel^{107,109}, jet fuel^{83,101-102}, motor oil¹⁰⁸, and rocket propellant (RP)^{77,80,98,100-101,104-105}. The ADC method can be used not only to provide information regarding sample composition, but also to study the thermodynamic and physical properties such as corrosive effects¹⁰³, enthalpy of combustion (through the use of each distillate fraction to determine the overall enthalpy of combustion and fuel performance)^{83,101-102,105-106}, and the influence of thermal stress on fuels¹⁰⁵. Furthermore, the variability in fuel composition and its impact on physical properties have also been investigated⁷⁷.

In conjunction with implementing the ADC method, it has become apparent that additional chemical composition information should be evaluated to strengthen the information gained from ADC data. For this purpose, in this report we applied the powerful chemical analysis tool known as comprehensive two-dimensional gas chromatography combined with time-of-flight mass spectrometry (GC × GC – TOFMS), using a reverse column GC × GC configuration (i.e., polar primary dimension column coupled with a non-polar secondary dimension column)⁹ building from our previous study⁷⁴, to improve the separation of the various compound classes (eg. alkanes, cycloalkanes, aromatics, etc), and to facilitate extraction of chemical information from a set of ten RP-1 fuel samples. Using chemometrics, we then explored the connection between chemical composition via GC × GC – TOFMS chromatographic data and the ADC data from the RP-1 fuels. Indeed, GC × GC – TOFMS is ideally suited for use in fuels analysis^{9,24-25,48,51-52,74,88-90,92}.

To help glean useful information, multivariate “chemometric” data analysis methods have been developed. Chemometrics have been shown to be able to take advantage of the three-way data provided by the GC \times GC – TOFMS instrumental platform, to help reveal similarities and/or differences between chromatograms^{23,25,48,74,88}. Partial least-squares (PLS) analysis can be used to associate variance in fuel composition to measured physical properties⁷⁴. Detailed information on the theory of PLS can be found elsewhere^{19,93–95}. In short, PLS was used to model the GC \times GC – TOFMS chromatograms (the X-block) for a set of ten RP-1 fuels, in conjunction with the ADC data (the Y-block) for the same fuels, allowing to use GC \times GC – TOFMS chromatograms to predict the ADC (without directly measuring a given ADC). This analysis is accomplished by selecting an appropriate number of latent variables (LVs) that are used to calculate loadings that capture the variance (i.e. chemical information) in the GC \times GC – TOFMS chromatograms that have the maximum covariance (implying a linear relationship) with corresponding information in the ADC data set. This experimental procedure has the ability to provide direct insight into the chemical composition changes as a function of % distilled (and distillation temperature during the ADC experiment).

In this study, GC \times GC – TOFMS chromatographic data of RP-1 fuels and their respective ADC data are analyzed using PLS to provide useful information on chemical compounds that significantly influence the RP-1 fuel properties via inspection of the linear regression vector (LRV) of each PLS model. Our goals are to demonstrate and validate the use of PLS modeling, and to relate chemical information obtained from the GC \times GC – TOFMS chromatograms to the corresponding ADC for each RP-1 fuel, and ultimately to predict the ADC temperatures of a given RP-1 fuel, without directly making those measurements⁸⁰. Eventually,

this chemical analysis approach will provide insight, and to aid, in the process of optimization of fuel performance.

5.2 Experimental

The full details on the GC \times GC – TOFMS instrumental platform and methodology can be found in our previous report⁷⁴. Ten RP-1 fuel samples were obtained from Air Force Research Laboratories (AFRL); the full list of RP-1 fuels used can be found in Table 3.1. The ADC data were obtained from an earlier report⁸⁰. The GC \times GC – TOFMS instrument used was an Agilent 6890A GC with a 7683B auto-injector (Agilent Technologies, Palo Alto, CA, USA) coupled to a LECO Pegasus-III TOFMS (LECO, St. Joseph, MI, USA). Isobaric mode was used with an inlet pressure of 35 psig (241 kPa). The auto-injector was set to 1 μ L injection, a 200:1 split injection with helium carrier gas was used, and acetone was used as the solvent rinse. The first GC \times GC separation dimension (primary column) used a RTX-wax (polar) stationary phase, of 30 m in length, 250 μ m i.d., and a 0.5 μ m film. The modulation period was set to 2.5 s. The second separation dimension (secondary column) used a 1.2 m RTX-1, of 100 μ m i.d., and a 0.18 μ m film. The GC oven was initially set to 40 $^{\circ}$ C for 2 min and ramped to 225 $^{\circ}$ C at a rate of 6 $^{\circ}$ C/min; the final temperature was maintained for 3 min. The GC inlet was set to 225 $^{\circ}$ C and the transfer line temperature was 235 $^{\circ}$ C. The thermal modulator offset was 20 $^{\circ}$ C, with a hot pulse time of 0.59 s and a 0.35 s cool time. The secondary column oven temperature control was not used while still achieving a suitable GC \times GC separation, and the secondary oven (housed in the primary oven) was left open and set at the same nominal temperature as the primary oven. The TOFMS data acquisition parameters were set with a 120 s acquisition delay, a mass channel

(m/z) scan range of 35-334 amu, with a 100 Hz acquisition rate.

The computer used for analysis was an Intel Core i-3-2120 @3.3 GHz, with 16.0 GB of RAM, and included a 60 GB SSD drive used for the purpose of a page disc (“fast” virtual RAM). Two replicate sets of RP-1 GC \times GC – TOFMS chromatograms were analyzed as separate sets of PLS models as described below, and the results for both replicates are provided herein, overlaid in figures, similar to previous reports^{74,92}. Chromatographic runs were imported to MATLAB2009b (MathWorks, Natick MA) using the ‘peg2mat’ function^{35,38,74}.

The GC \times GC – TOFMS data underwent baseline correction using in-house software as reported previously⁷⁴, and to help save memory and computation time, the data also underwent a condensing procedure^{74,96} that included the following operations. First, the chromatographic data were binned (for 2 points in each chromatographic dimension, resulting in GC \times GC – TOFMS chromatograms that are 25% of their original size). The binning also addressed any minor run-to-run misalignment in the data⁹⁶. Second, in the TOFMS domain, omitting m/z channels that were unselective and m/z channels that do not exhibit signal greater than five times the standard deviation of baseline corrected noise (these m/z are: 35-37,43-47, 51, 58-62, 73-76, 87-90, 101-103, 115-118, 133, 207, 214-334). Third, the signal for uninformative temporal regions was set to 0, specifically, GC \times GC regions dominated by column bleed or with no analyte compound signal (these regions were initially inspected while taking chromatogram variability into consideration to prevent the chance of removing compositional variation). The chromatographic and mass spectral dimensions of the GC \times GC – TOFMS data for each RP-1 fuel was vectorized (from 10 fuels \times 125 secondary column data points \times 405 primary column data points \times 148 mass channels to 10 fuels \times 7492500 unfolded data points) prior to PLS analysis along with the ADC (in vector form) for each RP-1 fuel. PLS analysis was performed

using PLS Toolbox 6.7 (Eigenvector Research Inc., Wenatchee WA), with mean centering of the GC \times GC – TOFMS data and auto scaling (subtracting the mean and dividing by the standard deviation) for the ADC temperature values.

Using the ADC method for a RP-1 fuel analysis, the temperature is recorded at the moment a specific percentage of the fuel has been distilled (% distilled point)⁸⁰. For this study, temperatures for the ADC method were measured at nineteen % distilled points: 0.025, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, and 90⁸⁰. Rather than construct a single PLS model for the entire ADC data set (simultaneously on all nineteen measurement points along the % distilled axis of the ADC for all fuels in the sample set), a series of 19 PLS models (a PLS model at each % distilled point) were produced. Performing the PLS analysis using a series of 19 models offered several key advantages. First, this approach lessened the restrictions on PLS when constructing the model(s). Second, this approach offered the ability to change the number of LVs at different % distilled points in the ADC (if necessary). Different numbers of LVs can be expected because the composition of a fuel is known to change over the course of the distillation, i.e. the GC \times GC – TOFMS chromatographic data represents the initial chemical composition of a given fuel, however the composition at a given % distilled is a subset of this composition, with possibly different relative concentrations for the various compounds present. A third important advantage for constructing a series of 19 PLS models was to save computation time. Consider modeling the entire ADC data set (10 fuels \times 19 % distilled points) coupled with the unfolded GC \times GC – TOFMS chromatograms (as stated previously, 7,492,500 unfolded data points per fuel), PLS would require a considerable amount of computer memory (about 13 GB), and the computation time would be prohibitively long, and on some computer systems this computational exercise would fail due to memory constraints. In contrast, applying PLS on the

unfolded GC \times GC – TOFMS chromatograms at one % distilled point at a time, required less LVs and significantly less memory (around 6.5GB), and requires less than a minute to compute per PLS model.

The PLS modeling was validated using leave-one-out-cross-validation (LOOCV). Briefly, LOOCV involves, a series of PLS models from ($n-1$) samples from the original n sample data set, and using the n th sample to predict values from the constructed ($n-1$) model. After all combinations are analyzed the root-mean-square of error of cross-validation of the residuals of the PLS models (RMSECV) was calculated in Eq. 3.1². Moreover, RMSECV results were also used to help determine the most appropriate number of LVs to use for the PLS models.

At each step in the analytical procedure, the LRVs of the PLS models were inspected to qualitatively verify that the connections the PLS models made between the chromatographic information (GC \times GC – TOFMS data) and physical measurements (ADC data) were both logical, and that the LRVs from consecutive models appear continuous. Using information provided by the LRVs, identification of compounds of interest in the GC \times GC – TOFMS data was performed via ChromaTOF V.3.32 (LECO Corporation, St. Joseph, MI, USA), and in-house software for nontarget PARAFAC for well resolved and unresolved peaks, respectively⁵². The NIST11 V2.0g mass spectral library (National Institute of Standards and Technology, Boulder CO, USA) was used for mass spectral identification.

5.3 Results and Discussion

A representative GC \times GC – TOFMS chromatogram of an RP-1 fuel is provided in Figure 5.1A. In this figure the total ion current (TIC) signal is plotted for the GC \times GC separation of the RP-1 fuel LB073009-08. To further demonstrate the separation power for complex samples such as RP-1, in Figure 5.1B, C and D, are provided specific regions of the GC \times GC separation with a representative alkane, cycloalkane, and aromatic compound indicated, respectively. Each of the representative compounds indicated also are key compounds identified in the PLS modeling that will be presented herein. In Figure 5.1B is provided a region of Figure 5.1A at the selective mass channel m/z 57 (between 5 and 12 min along the primary dimension, and between 1.7 and 2.5 s along the secondary dimension); the highlighted peak (located at 8.75 min and 1.94 s on the primary and secondary dimensions, respectively) has been identified as decane. In Figure 5.1C is provided a region of Figure 5.1A at the selective mass channel m/z 136 (between 13 and 19 min, and 1.0 and 1.8 s along the primary and secondary dimensions, respectively); the highlighted peak (located at 15.00 min and 1.17 s on the primary and secondary dimensions, respectively) has been identified as the adamantane. Finally, in Figure 5.1D is provided a region of Figure 5.1A at the selective mass channel m/z 105 (between 18 and 24 min, and between 0.8 and 1.2 s along the primary and secondary dimensions, respectively); the highlighted peak (located at 19.29 min and 0.95 s on the primary and secondary dimensions, respectively) has been identified as methylbutylbenzene.

The previously measured ADC data for all ten RP-1 fuels are provided in Figure 5.2A⁸⁰. The measured ADC data were obtained at a % distilled range from 0.025% to 90%, representing the ADC data set that we subsequently modeled using PLS with GC \times GC – TOFMS data. The

recorded temperatures for the ADC data set range from 207.2°C to 213.5°C at 0.025% distilled, to 235.9°C to 258.1°C at 90% distilled. At various % distilled values the ADC for several fuel pairs cross one another, which is indicative of the rich information that the ADC method provides, but may also potentially make the PLS modeling of ADC data more challenging. For clarity, in Figure 5.2B two representative ADCs are provided that approximately span the range of temperatures at each % distilled. The upper ADC in Figure 5.2B is from RP-1 fuel LB073009-06, having some of the highest measured temperatures for the ten fuels, The lower ADC in Figure 5.2B is from RP-1 fuel LB073009-02, which exhibited some of the lowest recorded temperatures.

For comparison to Figure 5.2B, the ADCs for RP-1 fuels LB073009-06 and LB073009-02 predicted using PLS during the LOOCV procedure are provided in Figure 5.2C. Figures. 5.2B and 5.2C are qualitatively very similar, but in order to obtain a more quantitative evaluation of the accuracy of the PLS modeling, residuals for each ADC were calculated at each % distilled value. The ADC residuals were obtained by subtracting a given measured ADC from the ADC predicted using PLS. The residuals imply an accuracy of the PLS modeling to within +/- 2.5 °C range.

Examination of the LRVs of the PLS models provide additional information, complementary to the ADCs predicted from the PLS models. In Figure 5.3A-C, three of the nineteen LRVs are provided (one for each PLS model constructed, other LRVs omitted for brevity): one LRV from the beginning (0.025% distilled), middle (45% distilled), and end (90% distilled) of the ADC. Through inspection of the positive LRV values, the corresponding peaks tend to be analyte compounds eluting after ~10 min for alkanes, after ~15 min for the cycloalkanes, and di- and tri-cycloalkanes, and after ~17 min for aromatic groups to a lesser

extent. These results in the LRVs display a general pattern that the less volatile compounds contribute positively to an ADC, suggesting less volatile compounds increase the overall predicted temperature of the ADC at a given % distilled point. As the % distilled approaches 90%, the intensities of the positively contributing peaks in the LRVs shift to the right to less volatile compounds, suggesting these compounds may contribute more with respect to the predicted ADC temperature. An interesting observation is that some regions (and peaks therein) in the LRVs change sign as the distillation runs toward completion; a good example is a cluster of peaks located ~13 and 17 min in the primary separation dimension and ~1.2 and 1.5 s in the secondary separation dimension. Although the peaks in the LRVs in this separation region are generally positive at 0.025% distilled, as the distillation progresses the magnitude of many peaks diminish until their contribution is zero, then as the distillation progresses further the signs of these peaks change to negative with a corresponding increase in magnitude. This suggests that early in the distillation, analyte compounds corresponding to peaks in the LRV that are changing from positive to negative during the distillation would contribute to increasing the predicted ADC temperature, but approaching the end of the distillation these compounds would contribute to decreasing the predicted ADC temperature. These compounds seem to act analogous to a chemical buffer in that as buffers moderate changes in pH, these compounds moderate the temperature range of the distillation, i.e. the more of these compounds present the narrower the temperature range over which the distillation will occur.

An interesting phenomenon is observed at the higher % distilled values, as shown in Figure 5.3C. There are several unexpected, slightly positive peaks in the LRV region between 5 and 15 min. At 90% distilled the chemical composition of the fuels is actually a subset of the fuel composition that is analyzed by the GC × GC – TOFMS instrument, since at 90% distilled

the more volatile compounds will have mostly boiled off, and there likely have been some significant changes in the relative compositions of the various compounds in the fuels. Thus, the positive values in the LRVs in the region between 5 and 15 min may be attributed to covariance between peaks that are more volatile and peaks that are less volatile in the PLS models, and not necessarily because these LRV peaks are chemically meaningful; this may be the source of the higher error in PLS models at higher % distilled values.

Inspection of the negative regions (and peaks therein) in all of the LRVs, analyte compounds between 5 and 15 min generally have negative values, suggesting the earlier and more volatile compounds lower the overall temperature of the ADC at a given % distilled. As with the positive LRV values, as the distillation progresses from 0.025% to 90% distilled, the intensity shifts from left to right. As the temperature rises, the more volatile compounds preferentially evaporate, so their decreased presence reduces their influence on the overall temperature at higher % distilled values, while the heavier, less volatile compounds contribute more. A list of representative, yet key, analyte compounds of interest, indicated by large magnitudes of their peaks in the LRVs were identified and summarized in Tables 5.2, 5.3, and 5.4. For example, methylbutylbenzene (identified in Figure 5.1D) is listed as compound #10 in Table 5.1, and is one of the major positively contributing compounds to the LRV. Decane (identified in Figure 5.1B), is listed as compound #2 in Table 5.2, and is one of the major negatively contributing compounds to the LRV. Adamantane (identified in Figure 5.1C), is listed as compound #7 in Table 5.3, and is one of the significant compounds that change sign with respect to their contribution as the ADC nears completion. Identification of compounds that impact the ADC can play an important role in understanding the information provided by the

ADC experiment, and ultimately could play a key role in improving fuel formulation and performance.

Finally, we present the LOOCV summary using the RMSECV calculation defined in Eq. (1) as a function of the % distilled value. The LOOCV procedure for the PLS modeling was performed using both sets of GC \times GC – TOFMS data with the ADC data set. The most appropriate number of latent variables (LVs) was determined to be 4 LVs, based upon the analysis of scree plots⁷⁴. The LOOCV summary in Figure 5.4A provides an assessment of the accuracy of the PLS models and select scree plots (0.025%, 0.45%, 0.90%) are overlaid and highlighted in Figure 5.4B for purpose of reference. The residuals (Figure 5.2D) of many of the RP-1 fuels cross at 80% distilled along with a sharp increase in the RMSECV in Figure 5.4 (at 85% and 90% distilled). These changes are probably linked to the changes in fuel composition as more fuel is distilled and the resulting covariance between compounds of different volatility that appear in the chromatograms. In principle, distillate of RP-1 fuels could be collected at each % distilled and analyzed with the GC \times GC – TOFMS, and the resulting chromatograms could be used to construct the PLS models using their respective temperatures on the ADC data. However this approach is more laborious and impractical, requiring a prohibitively large set of samples, eg., 190 samples, from 10 fuels \times 19 ADC % distilled points (instead of only 10 fuel samples directly analyzed herein in order to demonstrate the methodology principles). The primary benefit of doing so would be to reduce the apparent covariance, thus making the RMSECV values (in Figure 5.4) consistently smaller across the ADC. Another way to think about this source of the error while approaching the end of the distillation, is that PLS is using the chromatograms of un-distilled RP-1 fuels to “predict the future” ADC values. It is likely that better PLS models could be constructed from chromatograms generated from the RP-1 fuels

sampled at each % distilled. Using a respective chromatogram of a fuel at each distillation point would have more representative of the fuel and would have help minimize the error of the PLS models, however obtaining said RP-1 samples at various stages of distillation poses a more laborious proposition.

5.4 Conclusions

In this report we have demonstrated the use of PLS on GC \times GC – TOFMS chromatograms of RP-1 fuels, and their respective ADCs. The PLS modeling provides insight into how the chemical composition weighs differently in determining the temperature for a given % distilled value across the ADC. Compounds were discovered that correlate with narrowing the temperature range of which the distillation occurs. The predictive power of the PLS modeling was assessed using LOOCV, yielding RMSECV with low values, typically below 2.0 °C, at each % distilled measurement point during the ADC analysis.

Acknowledgement

The work at the University of Washington (UW) was performed under subcontract to ERC, Incorporated, Air Force Research Laboratory, Edwards AFB, CA. The fuels were provided by the Air Force Research Laboratory/RQRC, Edwards AFB, CA. Certain commercial equipment, instruments or materials are identified in this paper in order to adequately specify the experimental procedure. Such identification does not imply recommendation or endorsement by the University of Washington, the United States Air Force, or the National Institute of Standards

and Technology, nor does it imply the materials or equipment identified are necessarily the best available for that purpose.

Figure Captions

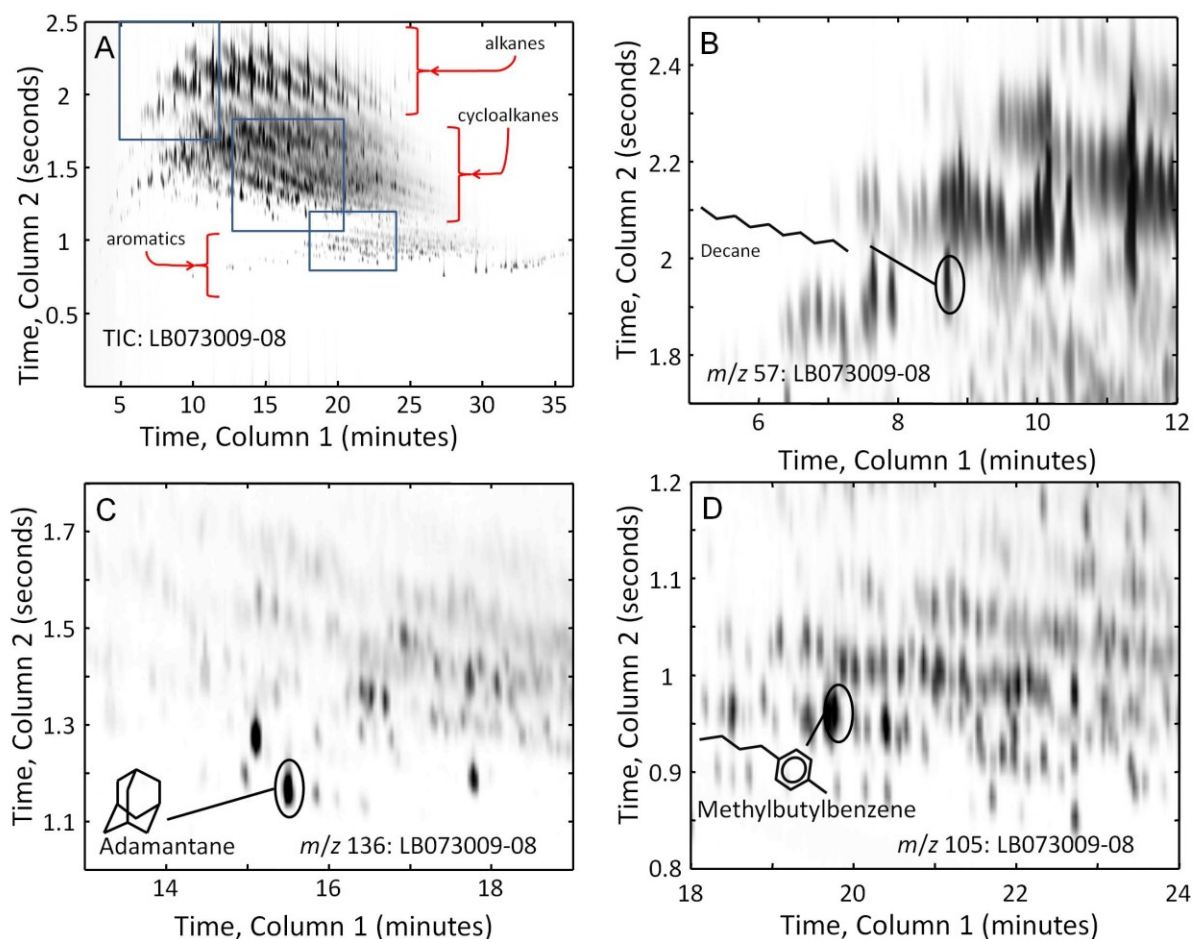


Figure 5.1. (A) Total ion current (TIC) chromatogram of the RP-1 fuel LB073009-08, collected using GC \times GC – TOFMS. Compound classes are indicated. (B) Region between 5 min and 12 min in the primary dimension and 1.7 s and 2.5 s in the secondary dimension at m/z 57, the upper left box in (A), with n-decane identified. (C) Region between 13 min and 19 min in the primary dimension and 1.0 s and 1.8 s in the secondary dimension at m/z 136, the middle box in (A), with adamantane identified. (D) Region between 18 min and 24 min in the primary dimension and 0.8 s and 1.2 s in the secondary dimension, the lower right box in (A), with methylbutylbenzene identified.

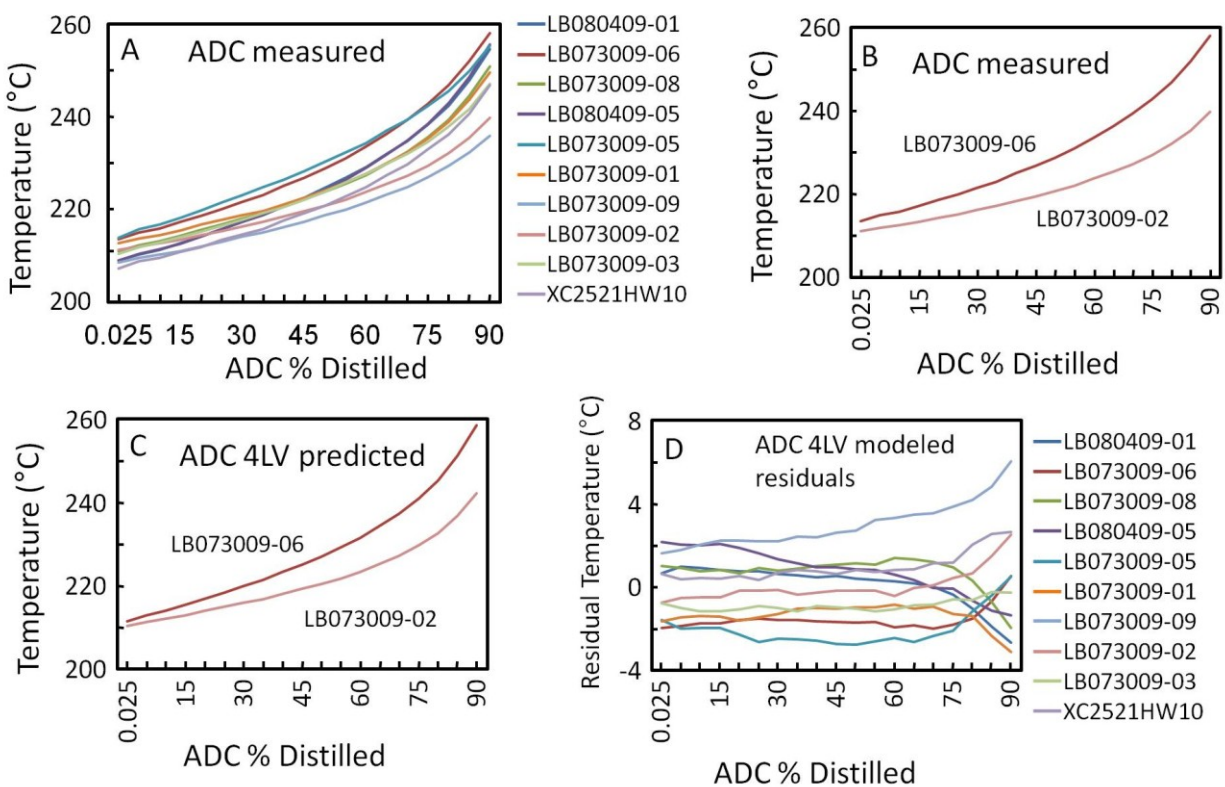


Figure 5.2. (A) Measured ADC data for the ten RP-1 fuels (listed in Table 3.1) are provided. (B) The ADC of two RP-1 fuels are provided that span the approximate range of the ADC data set: top LB073009-06, bottom LB073009-02. (C) The PLS modeled ADC for the two fuels in part (B) are provided: top LB073009-06, bottom LB073009-02. (D) The ADC residuals for all ten of the RP-1 fuels, calculated as the predicted ADC obtained from the cross validation predicted PLS models minus the measured ADC.

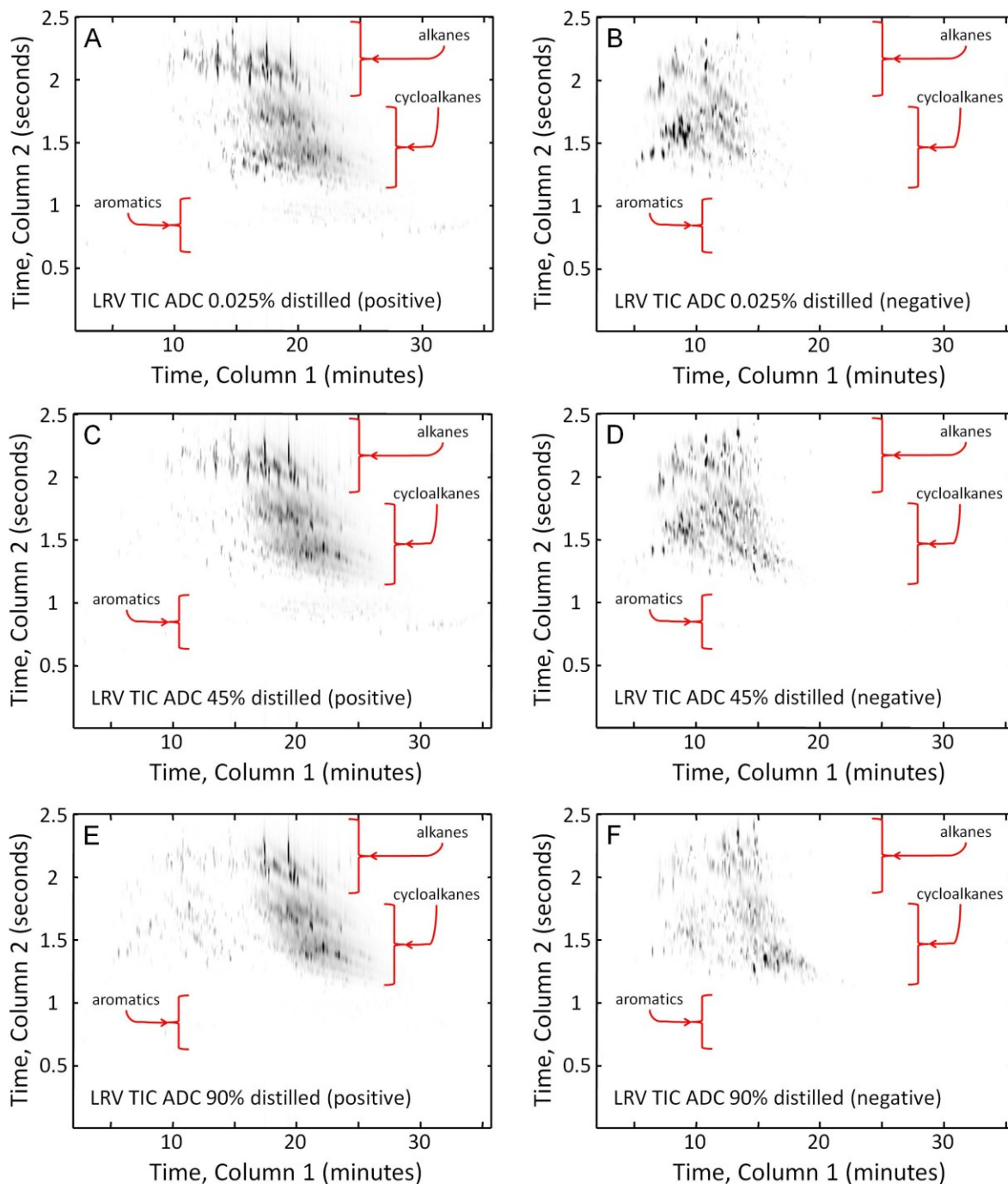


Figure 5.3. (A) Linear regression vector (LRV) of a 4 LV PLS model at 0.025% distilled of the ADC (positive contribution only). (B) Linear regression vector (LRV) of a 4 LV PLS model at 0.025% distilled of the ADC (negative contribution only). (C) LRV of a 4 LV PLS model at 45% distilled (the middle) of the ADC (positive contribution only). (D) LRV of a 4 LV PLS model at 45% distilled (the middle) of the ADC (negative contribution only). (E) LRV of a 4 LV PLS model at 90% distilled (the end) of the ADC (positive contribution only). (F) LRV of a 4 LV PLS model at 90% distilled (the end) of the ADC (negative contribution only).

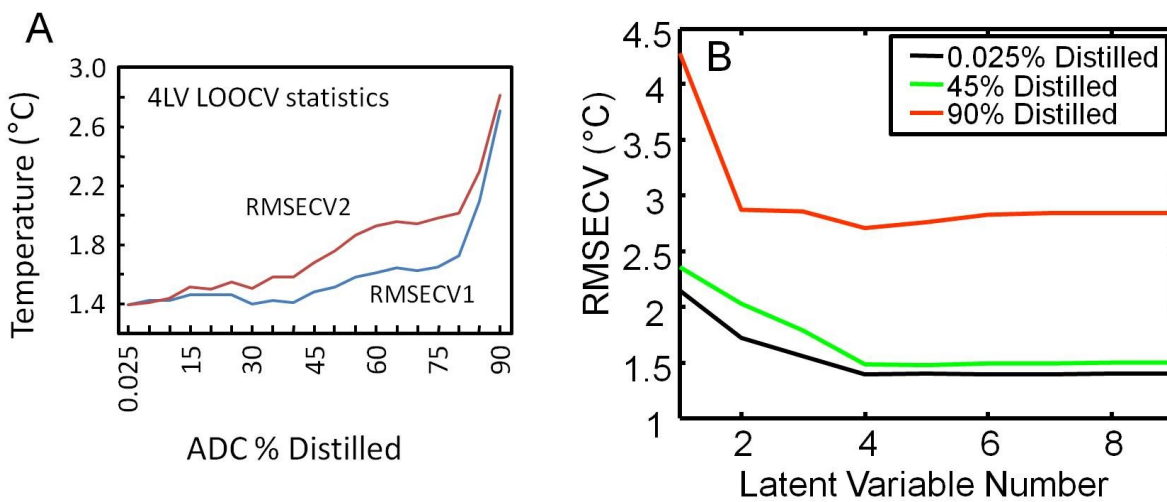


Figure 5.4. Validation results are provided for the PLS models of the ADCs for the ten RP-1 fuels in Table 3.1 using LOOCV. The RMSECV values for PLS modeling of both sets of GC \times GC – TOFMS data are indicated as a function of % distilled.

Tables

Table 5.1. Major contributing compounds identified in the LRVs that contribute positively, per the blue features in Figure 5.3A, B, C. The retention time on primary is labeled 1t_R , and on the secondary column as 2t_R . The mass spectral match value is labeled MV.

#	Compound Identification	1t_R (min)	2t_R (s)	MV	Compound Class
1	Trimethyldodecane (C15H32)	17.42	2.23	924	alkanes
2	3-Methyltridecane (C14H30)	19.88	2.00	910	alkanes
3	3-Methyltetradecane (C15H32)	20.17	1.96	922	alkanes
4	Heptylcyclohexane (C13H26)	18.75	1.62	889	cycloalkanes
5	Octylcyclohexane (C14H28)	21.25	1.60	909	cycloalkanes
6	Nonylcyclohexane (C15H30)	23.63	1.60	929	cycloalkanes
7	Methyl-bicyclohexyl (C13H24)	20.79	1.36	841	di- & tri- cycloalkanes
8	Hexamethyloctahydro-1H-indene (C15H28)	22.21	1.43	832	di- & tri- cycloalkanes
9	Bicyclohexane (C15H28)	20.00	1.38	907	di- & tri- cycloalkanes
10	Methylbutylbenzene (C11H16)	19.29	0.95	908	mono-aromatics
11	Azulene (C10H18)	26.83	0.76	919	di-aromatics

Table 5.2. Major contributing compounds identified in the LRVs that contribute negatively per the red features in Figure 5.3A, B, C.

#	Compound Identification	1t_R (min)	2t_R (s)	MV	Compound Class
1	Methylnonane (C10H22)	7.46	1.94	937	alkanes
2	Decane (C10H22)	8.21	1.94	960	alkanes
3	Dimethylnonane (C11H24)	8.42	2.18	931	alkanes
4	Trimethylcyclohexane (C9H18)	7.42	1.40	943	cycloalkanes
5	Methylpropylcyclohexane (C10H20)	8.92	1.63	873	cycloalkanes
6	Ethylmethylcyclohexane (C10H20)	9.21	1.53	864	cycloalkanes
7	cis-Octahydro-1H-indene (C9H16)	11.17	1.24	948	di- & tri- cycloalkanes
8	Dimethylbicyclo[3.2.1]octane (C10H18)	11.96	1.34	890	di- & tri- cycloalkanes
9	Not found at significant level				mono-aromatics
10	Not found at significant level				di-aromatics

Table 5.3. Compounds of interest identified in the LRVs that exhibit a sign change across the ADC (from positive to negative), per Figure 5.3.

#	Compound Identification	¹ t _R (min)	² t _R (s)	MV	Compound Class
1	Trimethyldecane (C13H28)	14.83	2.24	898	alkanes
2	Methyldodecane isomer (C13H28)	14.67	2.10	926	alkanes
3	Methyldodecane isomer (C13H28)	15.00	2.10	940	alkanes
4	Not found at significant level				cycloalkanes
5	Not found at significant level				cycloalkanes
6	trans-decahydronaphthalene(C10H18)	12.83	1.35	930	di- & tri- cycloalkanes
7	Adamantane (C10H16)	15.00	1.17	959	di- & tri- cycloalkanes
8	Methyldecahydronaphthalene (C11H20)	14.00	1.42	940	di- & tri- cycloalkanes
9	Not found at significant level				mono-aromatics
10	Not found at significant level				di-aromatics

Chapter 6: Analysis of RP-1 Fuel and Physical Properties over a Temperature Range

6.1 Introduction

Complex samples such as Kerosene fuel and RP-1 fuel contain many compounds, and further insight about their composition is crucial in gaining better understanding of their physical properties and performance^{79,84,110}. However, obtaining control over the fuel composition can be difficult considering the stock material may exhibit considerable variation. Achieving further insight into fuel composition and the sources of variation has become increasingly important to help evaluate, control, and or maintain the performance of “field” fuels^{79,83}. The list of RP-1 fuels analyzed is shown in Table 3.1.

GC × GC – TOFMS as stated and demonstrated previously is a workhorse with respect to analysis of complex samples such as fuels^{9,24,25,48,52,74,88}. Multivariate data analysis methods (such as chemometrics) have been developed to help glean useful information from the rich chemical data provided by GC × GC – TOFMS chromatograms^{23,25,48,88}. Partial least-squares (PLS) specifically can be used to associate compositional information in GC × GC – TOFMS chromatograms (X) to physical property measurements of RP fuels (Y)⁷⁴. In depth discussion regarding PLS is available in section 1.3.3 and elsewhere^{93,95}.

(110) J.H. Christensen, G. Tomasi, A.B. Hansen, Environ. Sci. Technol. 39 (2005) 255–260.

A series of PLS models were constructed, one set of PLS models for each measured physical property at every data point measurement over the range of temperatures; this was performed for the same reasons as discussed previously in chapter 5. To recap, the first reason was for lessening the restrictions on PLS, and if needed, to have the ability of changing the number of latent variables at different vectors in Y (physical properties). The second reason was to make the process more computationally affordable: X is as a 10×7492500 element matrix and the sets of Y were matrices of 10×14 and 10×23 . Analysis using a Y for a physical property would require a considerable amount of computer memory (from 13 GB to over 16GB) and would take hours (if successful) to complete. By applying PLS on 10 element vectors from a given Y and X, the computation time was significantly shorter, required less memory (around 7GB), and had less latent variables.

Testing the overall accuracy of the PLS models in predicting the specific measured values is imperative; therefore the validation process was performed as discussed previously in chapters 3 and 5 by using leave one out cross validation (LOOCV). The implementation of LOOCV was mentioned previously in chapters 3 and 5. Briefly, it involves calculating the residuals of every combination of PLS models constructed with $(n-1)$ and testing with the n th sample left out, and then testing the overall accuracy of the PLS model using the n th sample. An important figure of merit for LOOCV is the root mean squared error of cross validation (RMSECV). LOOCV is computationally demanding and performing it on a series of PLS models as presented drops the memory requirements to around 7 GB, making it more feasible.

This chapter is the continuation of the study of GC \times GC – TOFMS chromatograms of RP-1 fuels and relatively large sets of measured physical properties (across considerable range of temperatures) using PLS. The following measured fuel properties were modeled: density, speed

of sound, and kinematic viscosity all measured over a broad range of temperatures⁹⁷. The inspection of the linear regression vectors (LRV) of the PLS models provide useful information on compounds of interest in the RP-1 GC \times GC – TOFMS chromatograms with respect to a given physical property, and further demonstrate the potential for using chemical information from GC \times GC – TOFMS chromatograms to predict physical properties.

6.2 Experimental

The full details on the instrumental setup can be found in a previous report⁷⁴ and in section 3.2. The instrument used in the separation of analytes in the fuels was an Agilent 6890A GC coupled to a LECO Pegasus-III TOFMS and a 7683B auto-injector. The column configuration used was an RTX-wax (polar) column (30 m length, 250 μ m i.d., and 0.5 μ m film thickness), and an RTX-1 (1.2 m long, 100 μ m i.d., and 0.18 μ m film thickness) for the first and second column dimension, respectively. The modulation period was 2.5 sec. The secondary oven was left open and set nominally at the same temperature as the primary oven. The data acquisition parameters had a 120 sec acquisition delay, m/z scan range of 35-334, and data acquisition rate of 100 Hz.

The RP fuels were obtained from Air Force Research Laboratories (AFRL), the full list of samples used can be seen in Table 3.1. For more details, see section 3.2.1.

The measured physical properties (as a function of temperature) for the RP fuels can be found in a previous report⁹⁷. The complete list of physical properties measured over temperature ranges are the following: Density (g/mL), Speed of Sound (m/s), Kinematic Viscosity (mm^2/s),

Adiabatic Compressibility (TPa^{-1}), and Dynamic Viscosity ($\text{mPa}\cdot\text{s}$).

The computer used was the same as described in section 5.2: an Intel Core i-3 @3.3GHz, with 16.0GB of RAM. GC \times GC–TOFMS chromatograms were imported to MATLAB2009b (MathWorks, Natick MA) using the ‘peg2mat’ function^{35,37,74}, and underwent preprocessing (baseline correction, data reduction) as stated in a previous report⁷⁴. Chromatographic and mass spectral dimensions were unfolded and mean centered. PLS Toolbox 6.7 (Eigenvector Research Inc., Wenatchee WA) was used for PLS. The aforementioned measured values underwent autoscaling in PLS Toolbox prior to analysis. Separate PLS models were constructed for each physical property at each set of values. The LRVs from consecutive PLS models were inspected to ensure they appeared continuous. Validation of the PLS models was performed using LOOCV as stated earlier⁷⁴. ChromaTOF V.3.32 (LECO Corporation, St. Joseph, MI, USA), and the NIST11 V2.0g mass spectral library (National Institute of Standards and Technology, Boulder CO, USA), and an in-house software for non-target PARAFAC⁵² were used in identifying compounds.

6.3 Results and Discussion

6.3.1 Modeling Density as a Function of Temperature

The physical property of density was originally modeled using PLS in an earlier study⁷⁴ at a temperature of 288.15K, in compliance with the ASTM D 4052 method¹¹¹. In the study

(111) D02 Committee, Test Method for Density, Relative Density, and API Gravity of Liquids by Digital Density Meter, ASTM International, 2011.

presented herein, the density of RP-1 fuels, measured over a range from 278.15 K to 343.15 K⁹⁷, were modeled using PLS and results are summarized in Figure 6.1 and Figure 6.2. Figure 6.1A shows the measured values and a zoom in of select values for both clarity and emphasizing the small difference that exists between them. The measured density values appear as parallel lines between samples. The observed range was between 812.28 kg/m³ and 822.9 kg/m³ at 278.15 K and between 765 kg/m³ and 775.88 kg/m³ at 343.15 K, respectively. The RMSECV values recorded over the temperature range is shown in Figure 6.1B; the RMSECV values are just below 1.87 kg/m³, suggesting that the expected error in prediction from these PLS models for an RP fuel not used in the model calibration is at most around 0.25%.

Inspecting the LRVs, there is no visible chemical shift or sign changes across the temperature range, only a slight shift in the magnitude of the values in the LRVs. Due to the fact that the LRVs are almost identical to the one shown previously, tables containing lists of compounds on interest were omitted for brevity but can be found in a previous report⁷⁴ and in Tables 3.4 and 3.5. In more detail, the LRVs have negative values before 15 min for alkanes and before 10 min for the rest of the compound classes; positive values in the LRVs appear after 15 min. This intuitively suggests that the later eluting analytes are correlated with increases in density while earlier eluting analytes are correlated with decreases in density. The LRVs appear to have only subtle differences over the range of temperatures.

6.3.2 Modeling Speed of Sound as a Function of Temperature

The results for the PLS models for speed of sound are summarized in Figure 6.3 and Figure 6.4. The measured values for speed of sound are shown in Figure 6.3A; the temperature

range is from 278.15 K to 343.15 K⁹⁷. The range in measured values for speed of sound was from 1382.7 m/s to 1391.3 m/s at 278.15 K and from 1136.5 m/s to 1145.5 m/s at 343.15 K, respectively. The trend in the measured values for speed of sound as a function of temperature is similar to that for density. Considering the relationship between speed of sound and density, these results appear consistent. A zoom of the values is shown for clarity and to demonstrate that as with density there is little difference between the values. The LOOCV results are shown in Figure 6.3B, which shows the RMSECV are within 1.93 m/s suggesting the expected error is within 0.17% of the measured values.

Since the speed of sound is directly related to density, it comes as no surprise that the LRVs (shown in Figure 6.4) for the speed of sound models appear similar to the LRVs for density. As the temperature increases there is a subtle shift in intensity from the earlier eluting compounds to the later eluting compounds. Compounds of interest based on the LRVs are shown in Tables 6.1 and 6.2.

6.3.3 Kinematic Viscosity as a Function of Temperature

The results for the PLS models for kinematic viscosity are summarized in Figure 6.5 and Figure 6.6. The measured values are shown in Figure 6.5A. As with density, kinematic viscosity was investigated in an earlier study⁷⁴ at 263.15 K in accordance with ASTM D 445¹¹² (see section 3.3.3 and Figure 3.11). Herein, kinematic viscosity measurements span a range of temperatures, from 263.15 K to 373.15 K, shown in Figure 6.5A with a zoom of select values for

(112) D02 Committee, Test Method for Kinematic Viscosity of Transparent and Opaque Liquids (and Calculation of Dynamic Viscosity), ASTM International, 2012.

clarity. The ranges of measured values are from 4.863 mm²/s to 5.624 mm²/s at 263.15 K, and from 0.8024 mm²/s to 0.8542 mm²/s at 373.15 K. Tables containing lists of compounds of interest were omitted for brevity as they can be seen in a previous report⁷⁴ and in Tables 3.6 and 3.7.

For the LOOCV results shown in Figure 6.6B, one can see the RMSECV is below 0.155 mm²/s (or 3.3%). Despite this relatively high error, about halfway through the measurements (after 318.15K) the RMSECV values drop almost a whole order of magnitude: RMSECV values of PLS models for measured kinematic viscosity at 338.15K are 0.1520 mm²/s (or 3.19%) and at 373.15 K are 0.0088 mm²/s (1.09%).

Select LRVs of kinematic viscosity are shown in Figure 6.6; there were no major chemical shifts or sign changes across the temperature range, only a minute change were visible in the magnitude of the LRVs. In detail, the LRVs appear similar to density and speed of sound; the earlier eluting compounds (between 5 and 16 min) correlate with lower overall kinematic viscosities while the later eluting compounds (most are after 16 min) correlate with higher overall kinematic viscosities. Interestingly, aromatics appear to have slightly negative scores; this could be due to covariance with alkanes. Overall, there appears to be a minor shift in both the negative and positive scores towards the later eluting compounds as the temperature increases.

6.3.4 Dynamic Viscosity as a Function of Temperature

The results of the PLS models for dynamic viscosity are summarized in Figure 6.7 and Figure 6.8. The measured values for dynamic viscosity are shown in Figure 6.7A and the

temperatures range from 263.15K to 373.15K. The range of measured values was between 4.029 mPa·s and 4.663 mPa·s, at 263.15K, and between 0.6001 mPa·s and 0.6401 mPa·s at 373.15K. A zoom of the data is shown for clarity and to demonstrate that as with density, the values run parallel to each other. The LOOCV results are shown in Figure 6.7B. The RMSECV values in particular are 0.1304 mPa·s (3.237%) at 263.15 K and 0.0077 mPa·s (1.283%) at 373.15 K.

The LRVs, as with previous analysis of properties discussed above, show little difference over the temperature range and are shown in Figure 6.8. Negative values for alkanes are between 5 and 15 min, and between 5 and 12 min for the other compound classes. Regions with positive values appear after 15 min for alkanes and after 12 min for the other compound classes (aromatics and diaromatics). Selected lists of identified compounds of interest based on the results of LRVs are shown in tables 6.3 and 6.4.

6.3.5 Adiabatic Compressibility as a Function of Temperature

The results of the PLS models for adiabatic compressibility are summarized in Figure 6.9 and Figure 6.10. The measured values for adiabatic compressibility are shown in Figure 6.9A. The measured values range from 627.8 TPa⁻¹ to 643.9 TPa⁻¹ at 278.15 K and from 983 TPa⁻¹ to 1011.0 TPa⁻¹ at 343.15 K. Also shown in Figure 6.9A is a zoom of a small region of the data that is shown for clarity and to emphasize the small differences between values. The LOOCV results are shown in Figure 6.9B; the RMSECV values in particular are 2.997 TPa⁻¹ or 0.477% of the measured values at 278.15 K and 5.644 TPa⁻¹ or 0.574% at 343.15 K.

There appears to be only small differences in the LRVs over the temperature range shown in Figure 6.10 and they bear a resemblance to the LRVs from the speed of sound PLS models, only the signs appear reversed. Selected lists of identified compounds of interest based on the results of LRVs are shown in tables 6.5 and 6.6.

6.4 Conclusions

RP-1 GC \times GC–TOFMS chromatograms and large matrices of various measured physical properties (density, speed of sound, kinematic viscosity, dynamic viscosity, and adiabatic compressibility) over a range of temperatures were analyzed using PLS. By breaking the physical properties matrices into a series of vectors to be analyzed as consecutive PLS models, computer memory was conserved during analysis. The PLS models were validated via LOOCV. The LRVs of the PLS models as indicated herein, were examined to ensure continuity and further demonstrated the suitability of PLS on RP-1 chromatograms and large sets of measurements over a range of temperatures.

Figure Captions

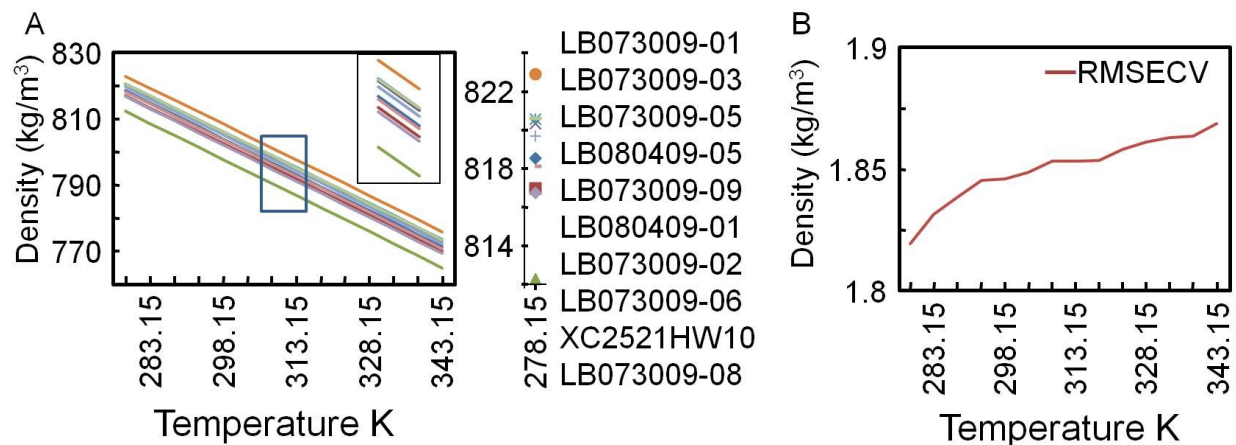


Figure 6.1. PLS 5 LV model results for density as a function of temperature. (A) Measured values, a zoomed-in region highlight the differences and show the density values run parallel to each other. (B) The root mean squared error of cross validation (RMSECV) values for the density PLS models.

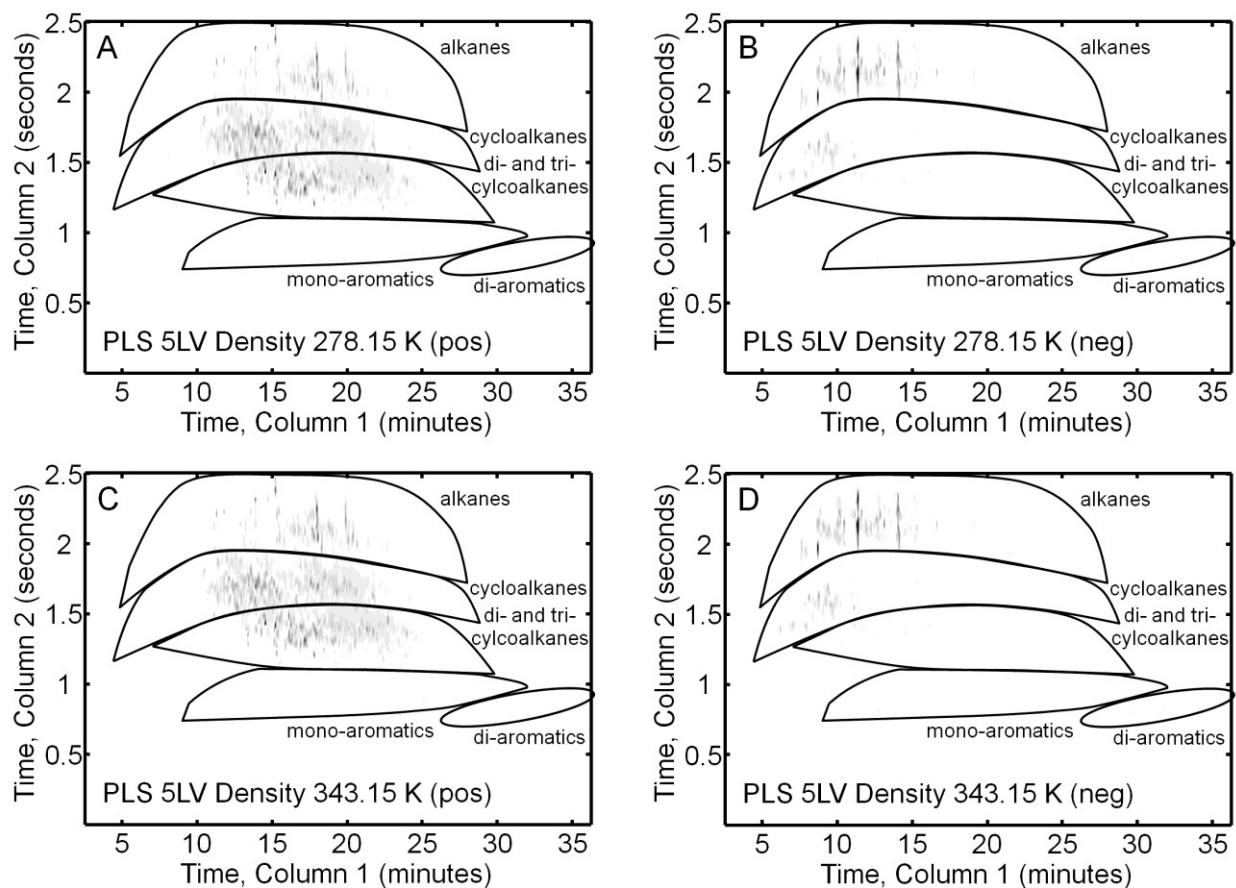


Figure 6.2. Select LRVs of PLS (5 LV) models for density. (A) Positive values of the LRV at 278.15 K. (B) Negative values of the LRV at 278.15 K. (C) Positive values of the LRV at 343.15 K. (D) Negative

values of the LRV at 343.15 K.

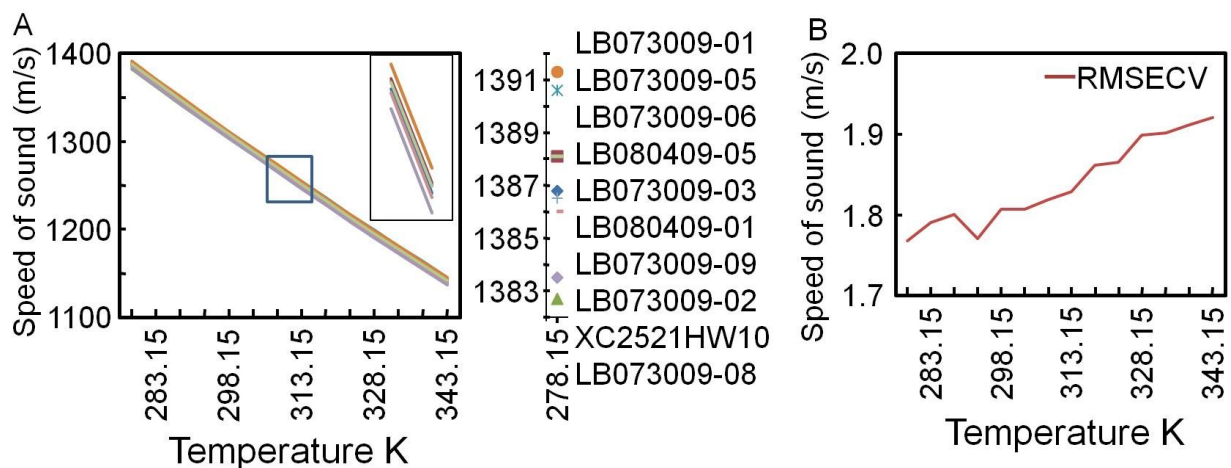


Figure 6.3. PLS 5 LV model results for speed of sound as a function of temperature. (A) Measured values, the zoomed in region highlight the differences. The speed of sound values run parallel to each other. (B) The RMSECV values for speed of sound PLS models.

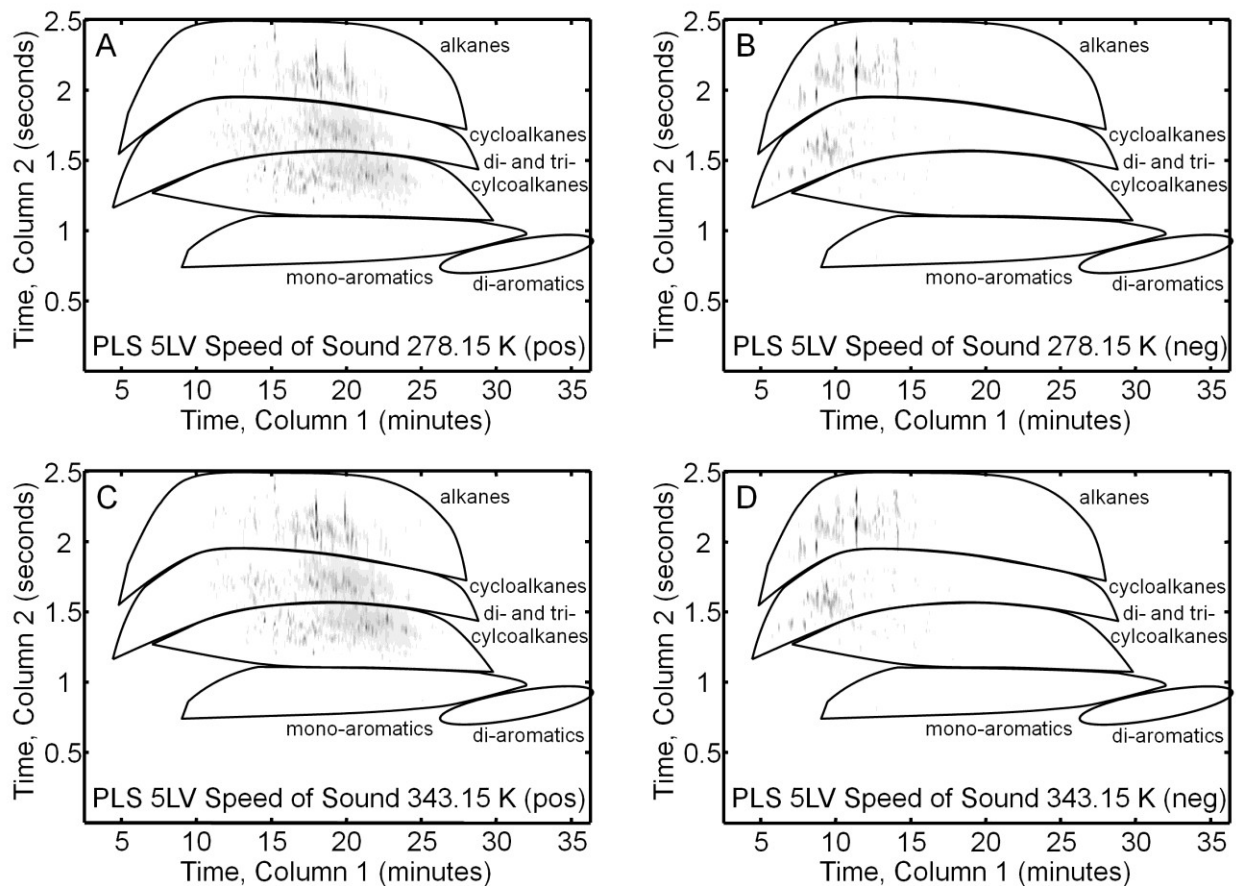


Figure 6.4. Select LRVs of PLS (5 LV) models for speed of sound. (A) Positive values of the LRV at 278.15 K. (B) Negative values of the LRV at 278.15 K. (C) Positive values of the LRV at 343.15 K. (D)

Negative values of the LRV at 343.15 K.

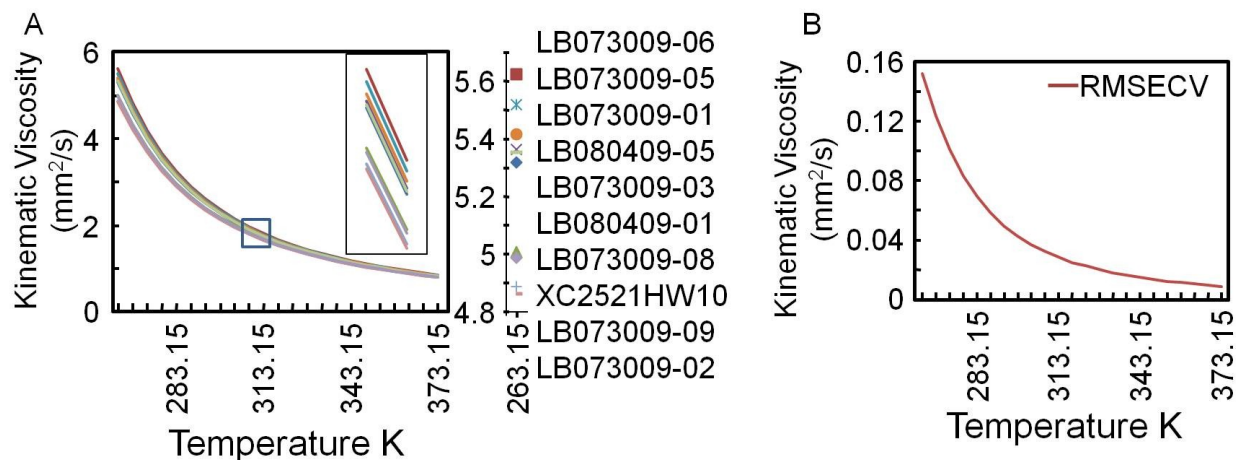


Figure 6.5. PLS 5 LV model results for kinematic viscosity as a function of temperature. (A) Measured values, the zoomed in region highlight the differences. The kinematic viscosity values do not cross each other. (B) The RMSECV values for kinematic viscosity PLS models.

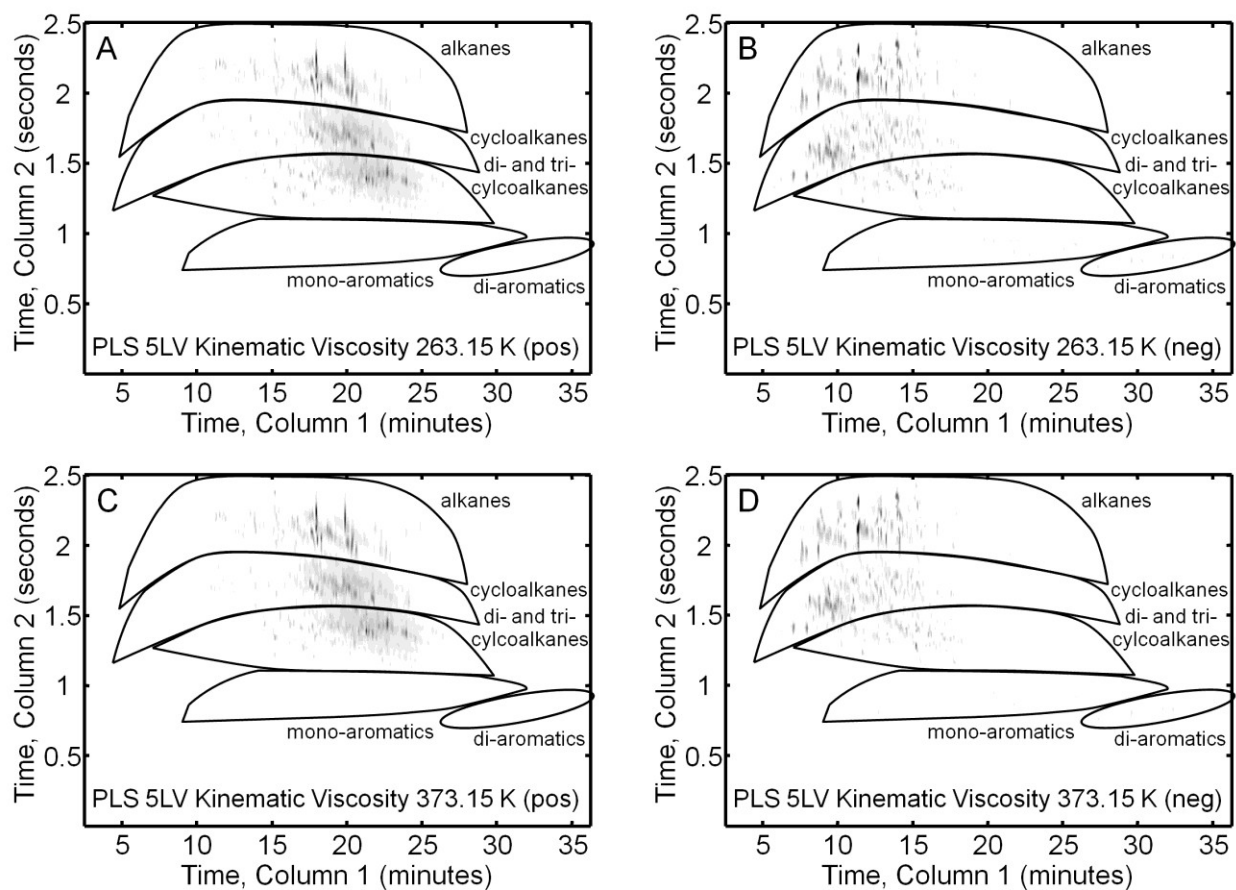


Figure 6.6. Select LRVs of PLS (5 LV) models for kinematic viscosity. (A) Positive values of the LRV at 263.15 K. (B) Negative values of the LRV at 263.15 K. (C) Positive values of the LRV at 373.15 K. (D) Negative values of the LRV at 373.15 K.

(D) Negative values of the LRV at 373.15 K.

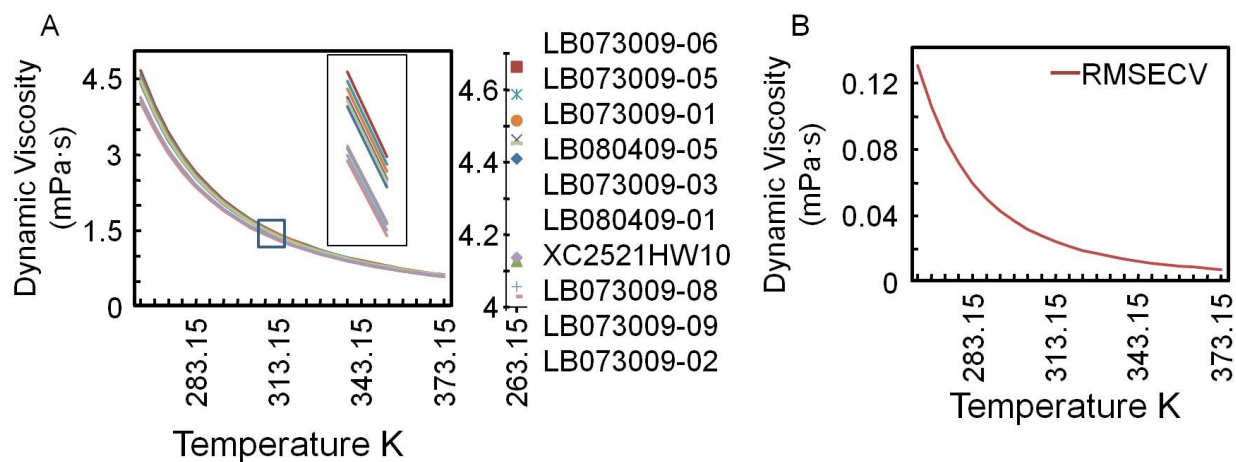


Figure 6.7. PLS 5 LV model results for dynamic viscosity as a function of temperature. (A) Measured values, the zoomed in region highlight the small difference between measured physical properties of samples. (B) The RMSECV values for dynamic viscosity PLS models.

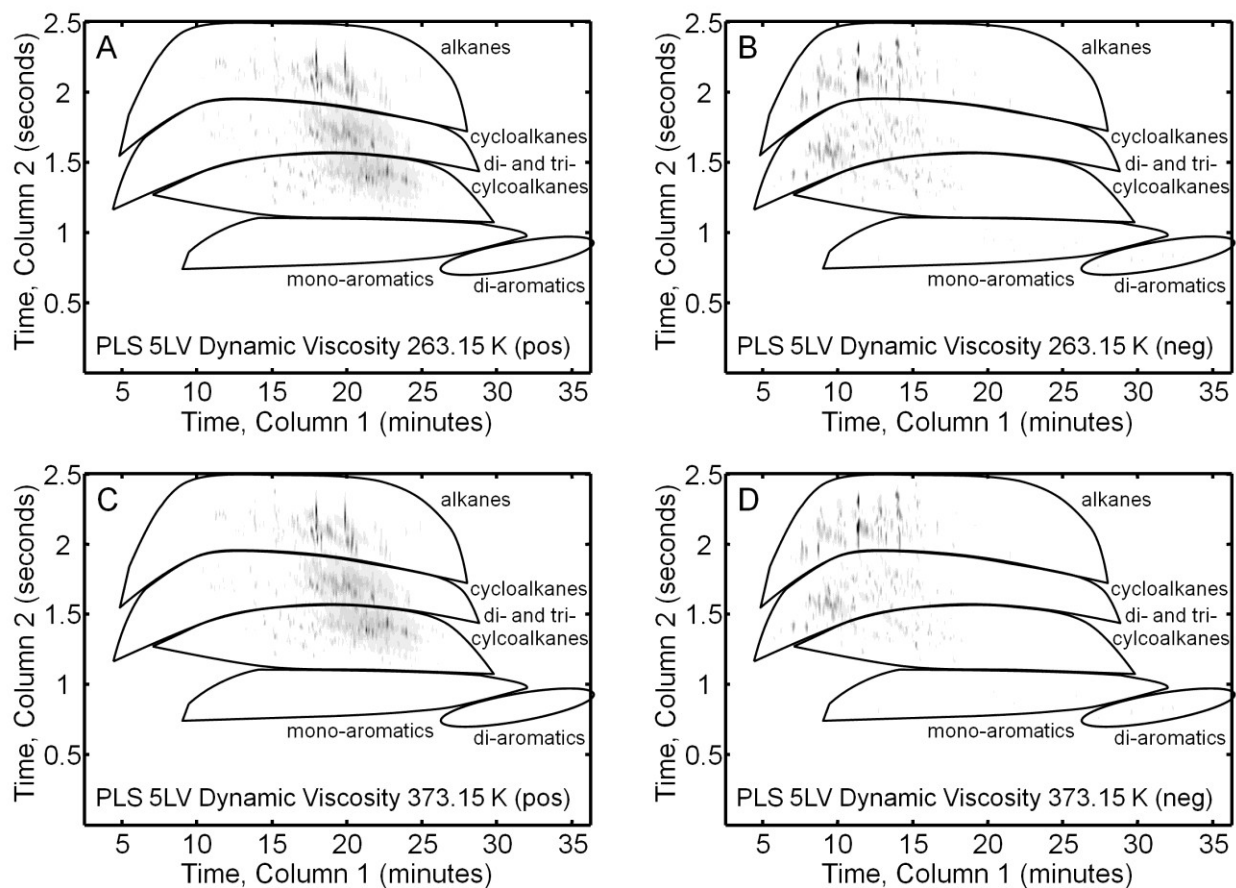


Figure 6.8. Select LRVs of PLS (5 LV) models for dynamic viscosity. (A) Positive values of the LRV at 263.15 K. (B) Negative values of the LRV at 263.15 K. (C) Positive values of the LRV at 373.15 K. (D)

Negative values of the LRV at 373.15 K.

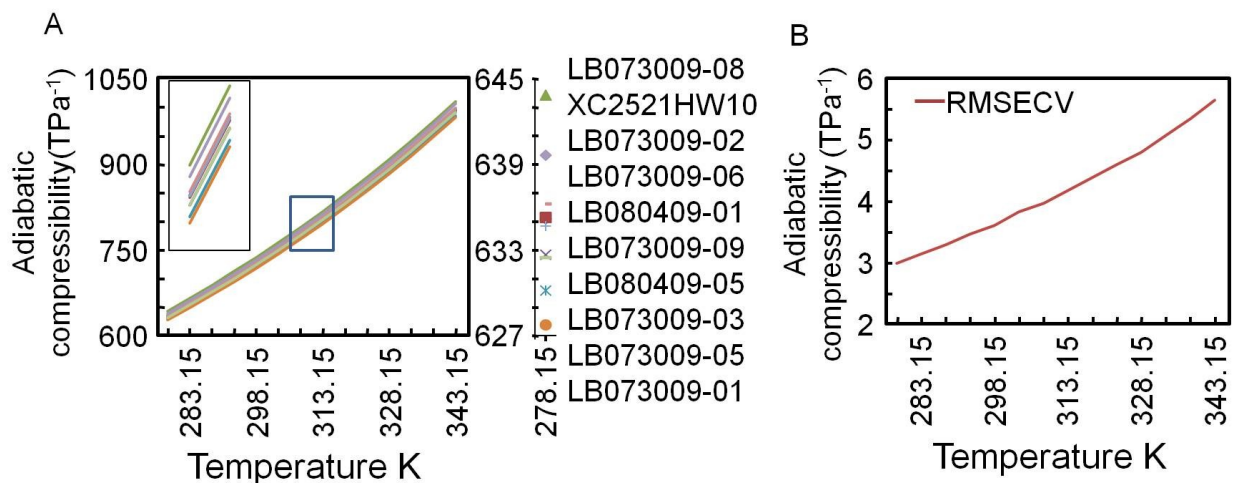


Figure 6.9. PLS 5 LV model results for adiabatic compressibility as a function of temperature. (A) Measured values, the zoomed in region highlight the little differences between values. (B) The RMSECV values for adiabatic compressibility PLS models.

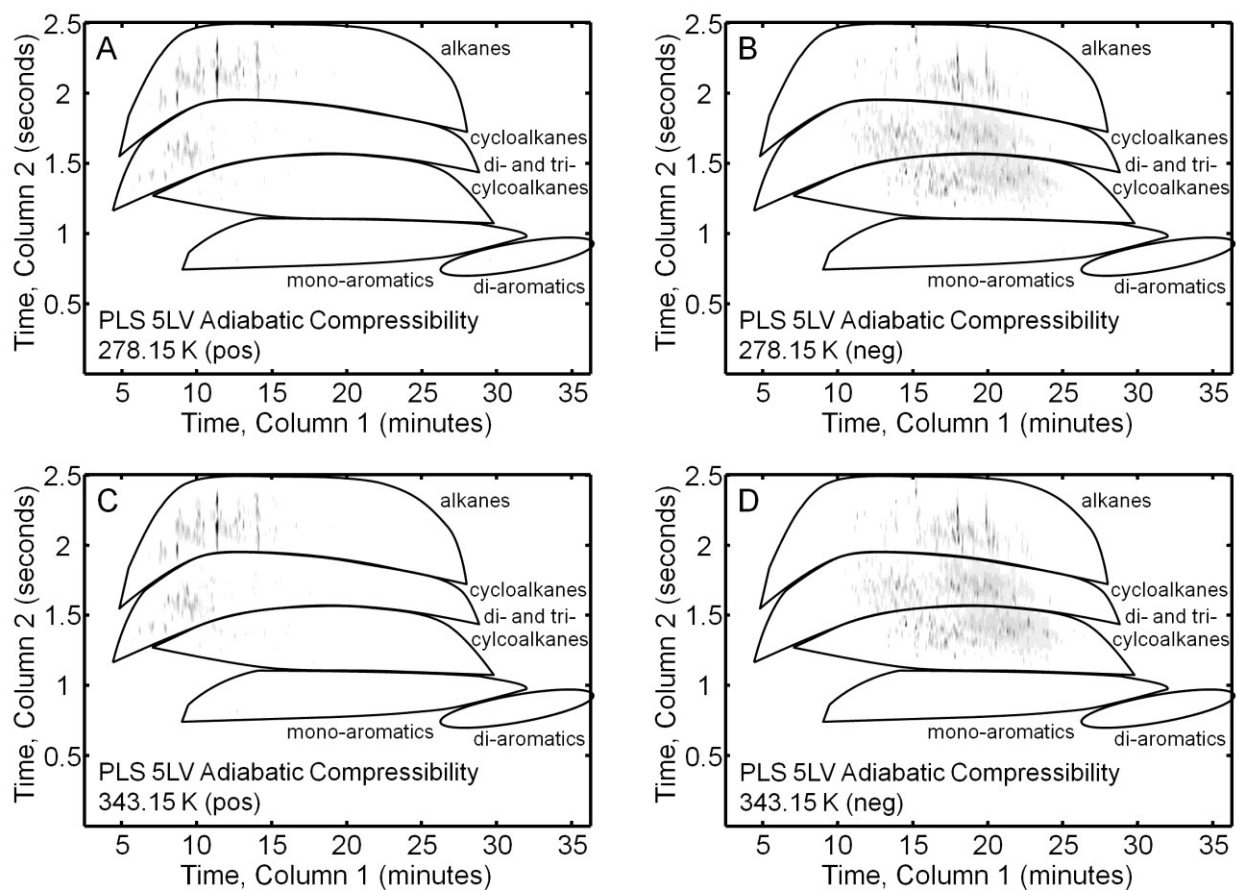


Figure 6.10. Select LRVs of PLS (5 LV) models for adiabatic compressibility. (A) Positive values of the

LRV at 278.15 K. (B) Negative values of the LRV at 278.15 K. (C) Positive values of the LRV at 343.15 K. (D) Negative values of the LRV at 343.15 K.

Tables

Table 6.1. Major contributing compounds identified in the speed of sound LRV that contribute positively, per Figures 6.4A & 6.4C.

#	Compound Identification	¹ t _R (min)	² t _R (s)	MV	Compound Class
1	Trimethyldodecane	17.42	2.23	922	alkanes
2	2-Methyltridecane	17.5	2.04	920	alkanes
3	3-Methyltridecane	17.79	2.04	910	alkanes
4	Heptylcyclohexane (C13H26)	18.75	1.62	863	cycloalkanes
5	Octylcyclohexane (C14H28)	21.29	1.6	863	cycloalkanes
6	Methyldecahydronaphthalene	14	1.46	919	di- & tri- cycloalkanes
7	Methyldecahydronaphthalene	14.71	1.41	928	di- & tri- cycloalkanes
8	2-Methyladamantane	17.29	1.18	866	di- & tri- cycloalkanes
9	Not found at significant level				mono-aromatics
10	Not found at significant level				di-aromatics

Table 6.2. Major contributing compounds identified in the speed of sound LRV that contribute negatively, per Figures 6.4B & 6.4D.

#	Compound Identification	¹ t _R (min)	² t _R (s)	MV	Compound Class
1	Decane	8.25	1.94	953	alkanes
2	Undecane	10.83	2.07	948	alkanes
3	Dimethyldecane (C12H26)	10.83	2.23	920	alkanes
4	Trimethylcyclohexane	5.75	1.39	921	cycloalkanes
5	Ethyl-dimethylcyclohexane (C10H20)	8.29	1.6	862	cycloalkanes
6	Methyl-isopropylcyclohexane(C10H20)	8.92	1.63	864	cycloalkanes
7	Decahydroazulene (C10H18)	10.71	1.41	880	di- & tri- cycloalkanes
8	Octahydro-1H-indene	11.21	1.24	927	di- & tri- cycloalkanes
9	Not found at significant level				mono-aromatics
10	Not found at significant level				di-aromatics

Table 6.3. Major contributing compounds identified in the dynamic viscosity LRV that contribute positively, per Figures 6.8A & 6.8C.

#	Compound Identification	¹ t _R (min)	² t _R (s)	MV	Compound Class
1	Trimethyldodecane	17.42	2.23	922	alkanes
2	2-Methyltridecane	17.5	2.04	920	alkanes
3	3-Methyltridecane	17.79	2.04	910	alkanes
4	Methyl-pentylcyclohexane (C ₁₂ H ₂₄)	14.38	1.72	876	cycloalkanes
5	Methyl-pentylcyclohexane (C ₁₂ H ₂₄)	14.79	1.68	820	cycloalkanes
6	Octylcyclohexane (C ₁₄ H ₂₈)	21.25	1.6	881	cycloalkanes
7	Methyl-bicyclohexyl (C ₁₃ H ₂₄)	20.83	1.37	824	di- & tri- cycloalkanes
8	Dicyclohexylmethane (C ₁₃ H ₂₄)	21.21	1.36	842	di- & tri- cycloalkanes
9	Hexamethyloctahydro-1H-indene (C ₁₅ H ₂₈)	22.21	1.43	821	di- & tri- cycloalkanes
10	Not found at significant level				mono-aromatics
11	Not found at significant level				di-aromatics

Table 6.4. Major contributing compounds identified in the dynamic viscosity LRV that contribute negatively, per Figures 6.8B & 6.8D.

#	Compound Identification	¹ t _R (min)	² t _R (s)	MV	Compound Class
1	Decane	8.25	1.94	953	alkanes
2	Undecane	10.83	2.07	948	alkanes
3	Dimethyldecane (C ₁₂ H ₂₆)	10.83	2.23	920	alkanes
4	1,2,4-Trimethylcyclohexane (C ₉ H ₁₈)	6.58	1.4	916	cycloalkanes
5	1-Methyl-3-ethylcyclohexane	6.67	1.44	932	cycloalkanes
6	1,2,3-Trimethylcyclohexane (C ₉ H ₁₈)	7.46	1.39	945	cycloalkanes
7	Octahydro-1H-indene (C ₉ H ₁₆)	11.17	1.24	926	di- & tri- cycloalkanes
8	cyclic (C ₁₁ H ₁₈)	14.63	1.27	863	di- & tri- cycloalkanes
9	Not found at significant level				mono-aromatics
10	Not found at significant level				di-aromatics

Table 6.5. Compounds of interest identified in the adiabatic compressibility LRV that contributed positively, per Figures 6.10A & 6.10C.

#	Compound Identification	¹ t _R (min)	² t _R (s)	MV	Compound Class
1	Undecane	10.83	2.07	948	alkanes
2	Dimethyldecane (C ₁₂ H ₂₆)	10.83	2.23	920	alkanes
3	Decane	8.25	1.94	953	alkanes
4	1,2,3-Trimethylcyclohexane (C ₉ H ₁₈)	7.46	1.39	945	cycloalkanes
5	1-Ethyl-2,3-dimethylcyclohexane	8.29	1.6	861	cycloalkanes
6	1-Methyl-4-isopropylcyclohexane	8.92	1.65	864	cycloalkanes
7	n-Butylcyclohexane	10.54	1.58	907	cycloalkanes
8	Not found at significant level				di- & tri- cycloalkanes
9	Not found at significant level				mono-aromatics
10	Not found at significant level				di-aromatics

Table 6.6. Compounds of interest identified in the adiabatic compressibility LRV that contributed negatively, per Figures 6.10B & 6.10D.

#	Compound Identification	¹ t _R (min)	² t _R (s)	MV	Compound Class
1	Trimethyldodecane	17.42	2.23	922	alkanes
2	2-Methyltridecane	17.5	2.04	920	alkanes
3	3-Methyltridecane	17.79	2.04	910	alkanes
4	Methyl-pentylcyclohexane (C ₁₂ H ₂₄)	14.375	1.72	876	cycloalkanes
5	Methyl-pentylcyclohexane (C ₁₂ H ₂₄)	14.79	1.68	820	cycloalkanes
6	Octylcyclohexane (C ₁₄ H ₂₈)	21.25	1.6	881	cycloalkanes
7	Methyldecahydronaphthalene	14	1.45	925	di- & tri- cycloalkanes
8	2-Methyldecahydronaphthalene	14.67	1.37	932	di- & tri- cycloalkanes
9	Methyldecahydronaphthalene	16.67	1.28	928	di- & tri- cycloalkanes
10	Not found at significant level				mono-aromatics
11	Not found at significant level				di-aromatics

Chapter 7: Conclusions and Future Directions

7.1 Conclusions

Studies of rocket propellant fuels using the GC \times GC – TOFMS instrument and the reverse column configuration have been discussed. The use of GC \times GC – TOFMS has demonstrated the improved separation of compounds such as alkanes and aromatics when compared to GC-MS. The reverse column configuration for GC \times GC – TOFMS offers excellent separation of analytes (especially various alkanes) in complex samples such as RP fuels.

Interpretation of data is an inseparable part of research. Preprocessing techniques such as baseline correction or peak alignment, when used appropriately prior to analysis, can significantly improve results. The ability of TMSRA to single out pure m/z of a targeted analyte has been demonstrated by utilizing the ratios between m/z . Chemometric techniques such as PCA, PLS and PARAFAC have become increasingly important in data analysis, and their use in studying RP fuels has been demonstrated. The sensitivity of PCA with respect to identifying differences between samples has been applied to RP-2 fuels, and PCA highlighted peaks which were sources of variation in the pool of samples analyzed. The ability of PLS to both correlate the variance in one set of measurements (X) to the variance in a different set of measurements (Y) has been applied on RP-1 fuels and their respective compositional and physical properties. The importance of validating PLS models has been discussed in detail and LOOCV has been applied to all PLS models presented. After the results from the PLS models and their respective LRVs were reviewed, analytes of interest were located and identified. When separation of compounds was less than ideal, i.e. when some analytes appear to have significant overlap,

instead of redoing the separation under different conditions to achieve better results, PARAFAC was applied to deconvolute the covarying signals of these analytes from each other. PARAFAC was helpful in purifying the mass spectra of interfered peaks for better match values in the mass spectral library.

7.2 Future Directions

The natural future direction of multi-dimensional chromatography (including but not limited to GC \times GC – TOFMS) is the push for improved separation, an example of which is the recent study of the reverse column configuration^{9,74}. The drive to improve separation may progress further via reducing the separation time, as highlighted in previous reports^{24,25}. As multi-dimensional GC becomes more widespread, more advancements and studies will follow. New means of increasing peak capacity and increasing sample throughput needs to be explored, such as the high-speed cryo-focusing injection (HSCFI)⁵.

In general, the use of chemometrics and the successful interpretation of their results often require a good conceptual understanding, and usually require optimizing various parameters. Though chemometric tools generally are not automated, there are situations where automation was achieved²³. Automation of chemometric methods such as PLS and summarizing the results into more easily interpretable answers with the option to view the results in greater detail should be considered.

With respect to RP fuels, immediate future research would be the continuation of current research with respect to correlating physical properties and chemical composition. The eventual future direction would be the investigation on the interrelation between compositional

information and figures of merit with respect to fuel performance. An interesting proposition would be to use the gathered information regarding the influence of chemical composition on fuel properties for testing the possibility of modifying a more readily available fuel to make an adequate approximation of an RP fuel. Another interesting but perhaps more challenging proposition would be approximating an RP fuel using a hydrocarbon mixture containing a dozen or so compounds.

Appendix: Further investigation into the sulfur content of RP-1 fuels.

The results from the PLS model for sulfur content of RP-1 fuels, which was originally discussed in supplementary material in chapter 3 and as shown in Figure 3.13, exhibit some covariance with other compound classes. The mono-aromatics and di-aromatics compound classes specifically exhibit a particularly strong covariance with the sulfur containing compounds. The purpose of the following investigation was to identify some of the sulfur containing compounds and attempt to construct a PLS model that accurately predicts sulfur content while minimizing the covariance exhibited between the sulfur containing compounds and other compound classes. Though the latter goal was not achieved, the investigation serves as a good example for explaining what can happen when a PLS model is overfactored.

Since the concentration of sulfur containing compounds in RP-1 fuels is relatively low, it was expected that analytes of interest would appear as low intensity peaks. To make these analytes as visible as possible, it would be best to inspect the GC \times GC – TOFMS chromatogram with the greatest concentration of sulfur content (LB073009-02 with 24.1 ppm), and inspect both the TIC and selective m/z . To further aid in locating these analytes of interest, the LRV of the PLS model for sulfur content was inspected for peaks with positive values where aromatics are located as they could potentially be sulfur containing. Several sulfur containing compounds were successfully identified and their peaks can be seen in the (binned) positive portion of the LRV shown in Figure A.1, and a list summarizing identified analytes is shown in Table A.1. Of the four analytes listed, only three were positively identified, and their match values are lower than previously identified compounds. The reason for the lower match values is likely due to low signal-to-noise. Dimethyl benzothiophene listed in Table A.1 has ‘suspect’ listed as its

match value since the possible matches were relatively low and the mass spectrum for dimethyl benzothiophene had a low signal-to-noise and may have also been interfered. In the LRV TIC shown in Figure A.1A, the arrows point to the location of identified sulfur containing peaks. The positive contribution of the sulfur containing compounds is overwhelmed by the signal from covarying compounds, making the values of the identified sulfur containing compounds difficult to see. Inspecting the LRV (positive portion only) at selective m/z 147, 148, and 162 (shown in Figures A.1B, A.1C, and A.1D, respectively), the sulfur containing compounds are more easily visible (pointed to by the included arrows). Note that in Figure A.1D the intensity for the two left most peaks (methyl benzothiophene isomers) have relatively low signal in m/z 162, barely over the noise level, while ethyl benzothiophene is more visible.

In an attempt to break the covariance between sulfur containing compounds from other compounds (specifically the aromatics), a series of PLS models were constructed, each using smaller sub regions from the original $GC \times GC - TOFMS$ chromatograms, as shown in Figure A.2 (denoted as large, medium and small). A set of three PLS models (3, 4, and 5 LV, respectively) were constructed for each sub region analyzed.

When constructing PLS models, the root mean square error of calibration (RMSEC) values generally decrease as the number of LVs increase. Caution must be taken to avoid overfactoring; although PLS models with higher LVs appear to make better predictions with X samples introspectively, they may not predict values as well when applied to samples not used in the calibration. Overfactoring can be avoided by performing cross validation and visually comparing the RMSECV values as a function of LVs; a model can be considered overfactored when the RMSECV value increases when compared to the previous model. As stated previously, inspecting the LRVs can help gain more insight. The LRVs should display

chemically meaningful information: if the LRV shows significant amount of artifacts or appear to be almost random, it is a strong possibility the model is overfactored. LRVs that exhibit little chemically meaningful information help explain why overfactored models do have higher RMSECV values (i.e. noise is factored into the model).

For the large sub region, the results of RMSEC and RMSECV are shown in table A.2, and the LRV for the 5 LV PLS model is shown in Figure A.3. The LRVs for the other PLS models are visually very similar; only the intensities in the alkanes and cyclic regions increase with increasing number of LV. Based on the results it would appear 4 LV is the best PLS model to use, however there is still significant covariance with other non sulfur containing compounds.

For the medium sub region, the results of RMSEC and RMSECV are shown in table A.2, and the highlighted LRVs (4 and 5) are shown in Figure A.4: only the positive portion of the LRVs are shown for brevity. Based on Table A.1, the di-aromatics region contains several sulfur containing compounds, therefore there should be peaks with significantly positive value – this is featured in the LRV for the 4 LV model but not in the 5 LV model, suggesting the 5 LV model is overfactored. Using the medium sub region however there still appears to be covarying di-aromatic compounds.

For the small sub region, the results of RMSEC and RMSECV are shown in Table A.2, and the highlighted LRVs (3, 4, and 5) are shown in Figure A.5. Note only the positive values of the three LRVs are shown for brevity. The PLS predictions of sulfur content fared better compared to PLS models using the large or medium sub regions. The 3 LV LRV appears to focus on peaks in the di-aromatics region, indicating it may be the better (yet far from ideal) choice under these circumstances. The 4 LV model has placed more emphasis on the mono-aromatics that appear in the upper right corner of the figure, suggesting the 4 LV PLS

model may be focusing a little more on covarying compounds. Finally, the LRV of the 5 LV model does not appear to contain information that is as chemically meaningful since positive values appear to be spread more randomly. These results suggest 3 LV PLS model is the most appropriate, when using the small sub region, especially after considering the RMSECV values.

In conclusion, several sulfur containing compounds were identified and their covariance with aromatics could not be broken based on results of the PLS models constructed using smaller sub regions. Each subsequent sub region focused on smaller regions of the original GC \times GC – TOFMS chromatograms, and the use of smaller sub regions produced better LOOCV results. Also the effects of overfactoring the PLS models with respect to the LRVs was highlighted.

Figure Captions

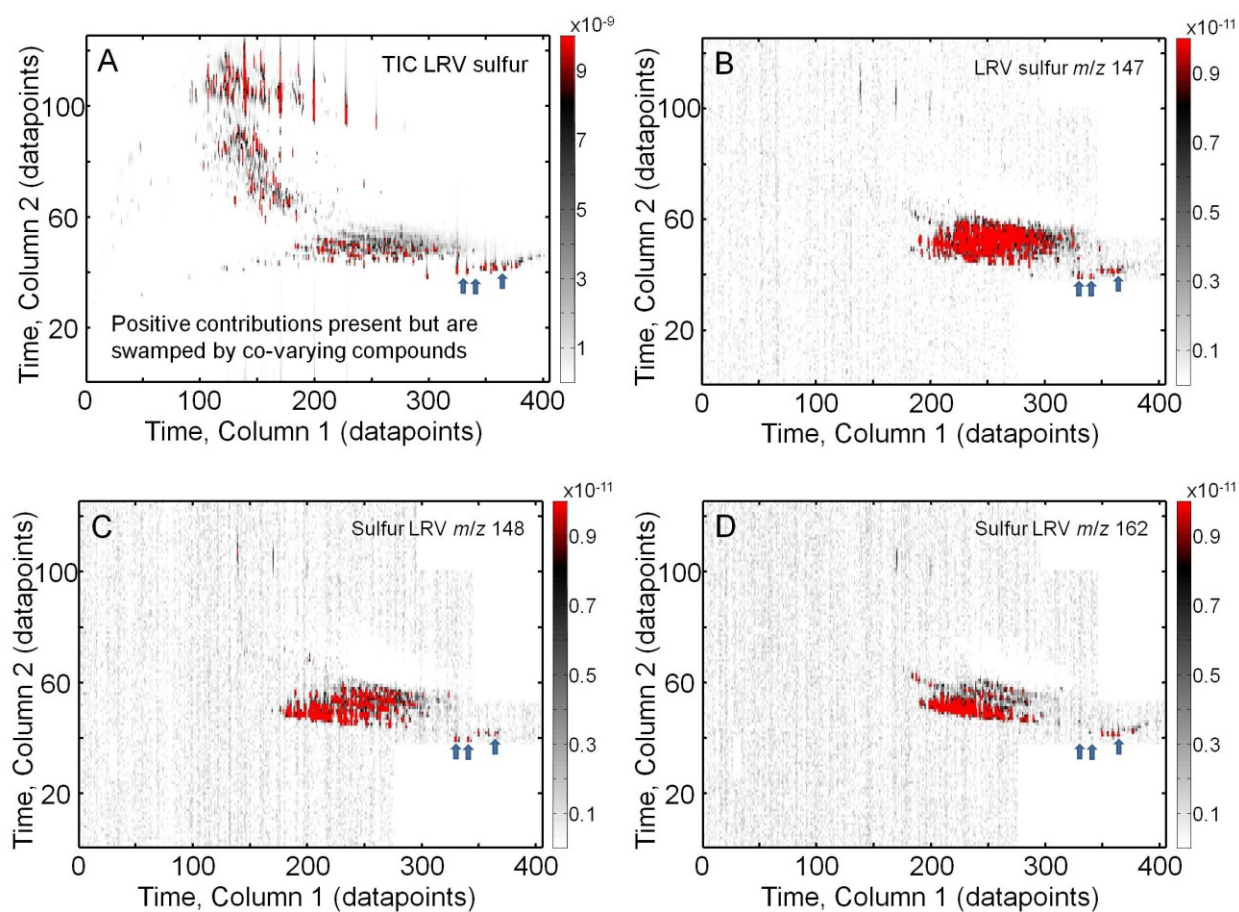


Figure A.1. Positive portion of the LRV for the PLS model for sulfur content (from section 3.3.3). The arrows on the figures indicate the location of the identified compounds (from left to right: two methyl benzothiophene isomers and ethyl benzothiophene). (A) LRV TIC. (B) LRV at m/z 147. (C) LRV at m/z 148. (D) LRV at m/z 162.

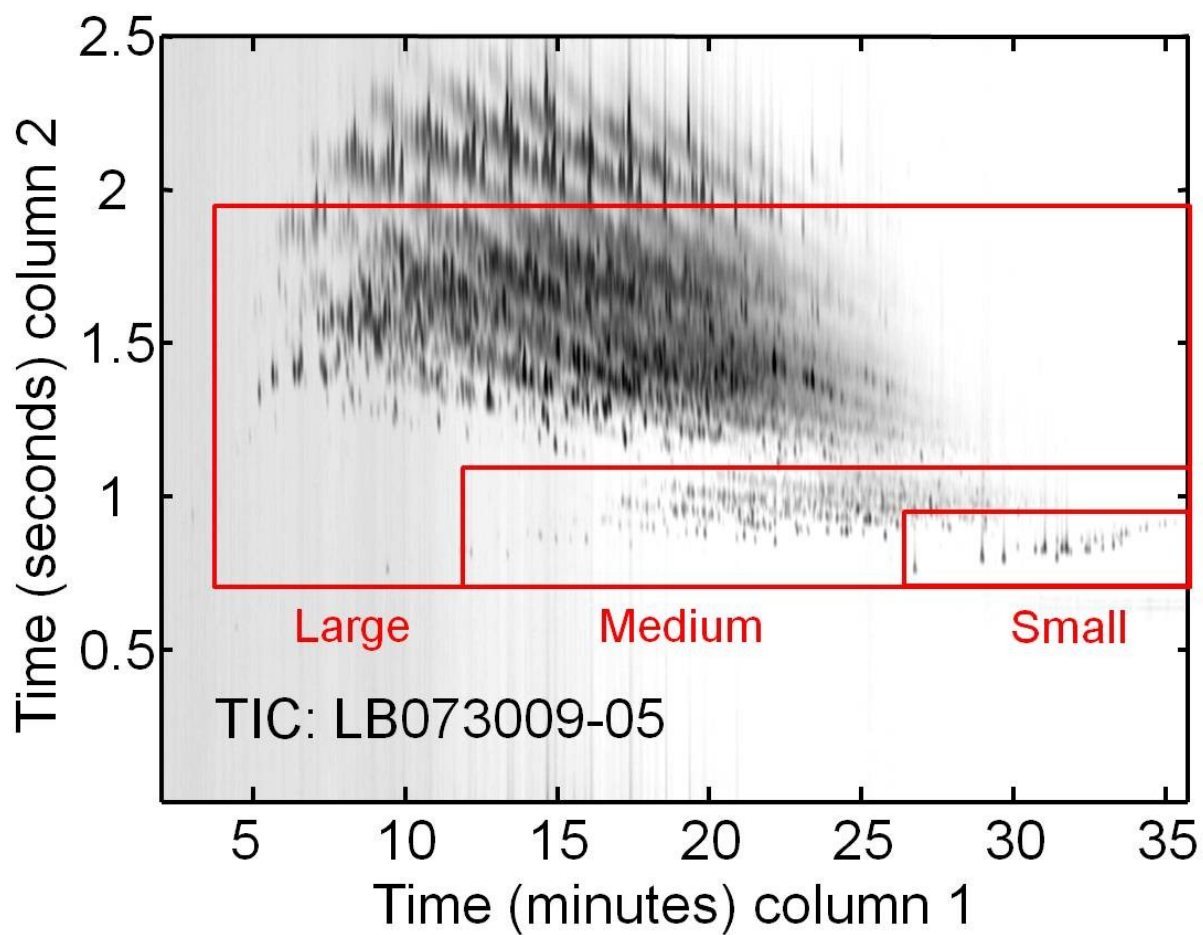


Figure A.2. Noted sub regions used in creating PLS models; ‘Large’ excludes most of the alkanes, ‘Medium’ includes the mono-aromatics and di-aromatics, and ‘Small’ contains mostly the di-aromatics.

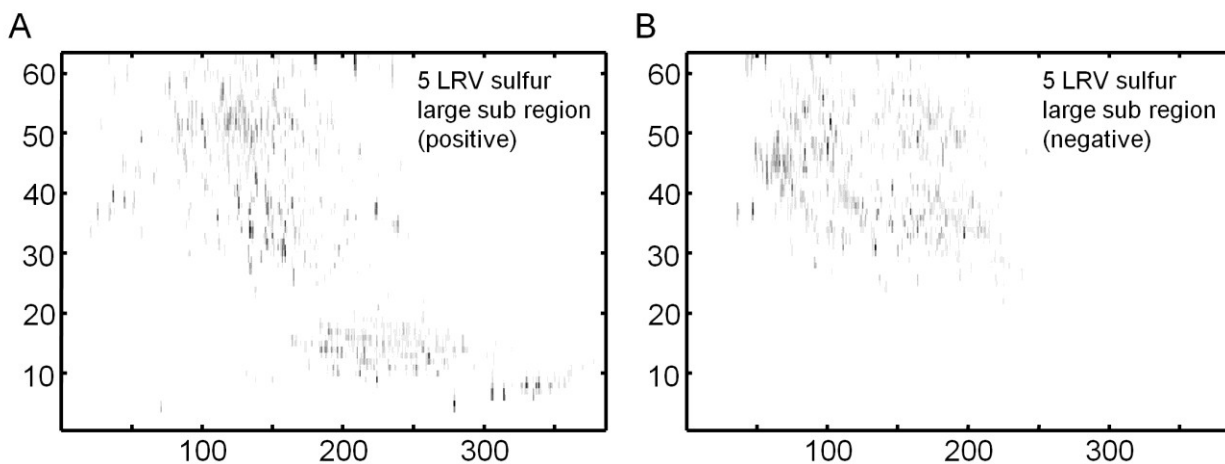


Figure A.3. LRV from 5 LV PLS model using the 'Large' sub region chromatograms. (A) Positive values only. (B) Negative values only.

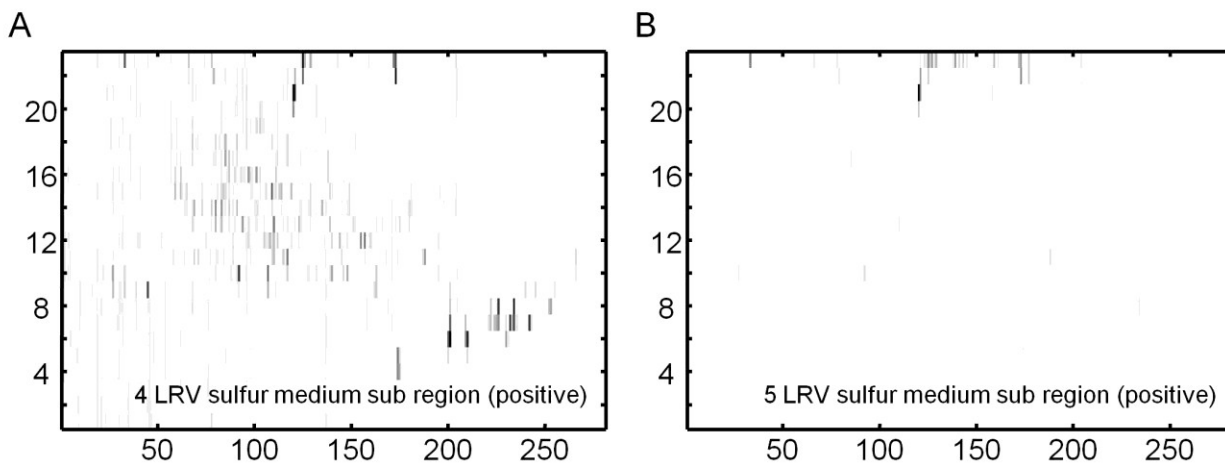


Figure A.4. LRVs from PLS models using the 'Medium' sub region chromatograms. (A) 4 LV LRV, positive values only. (B) 5 LV LRV, positive values only.

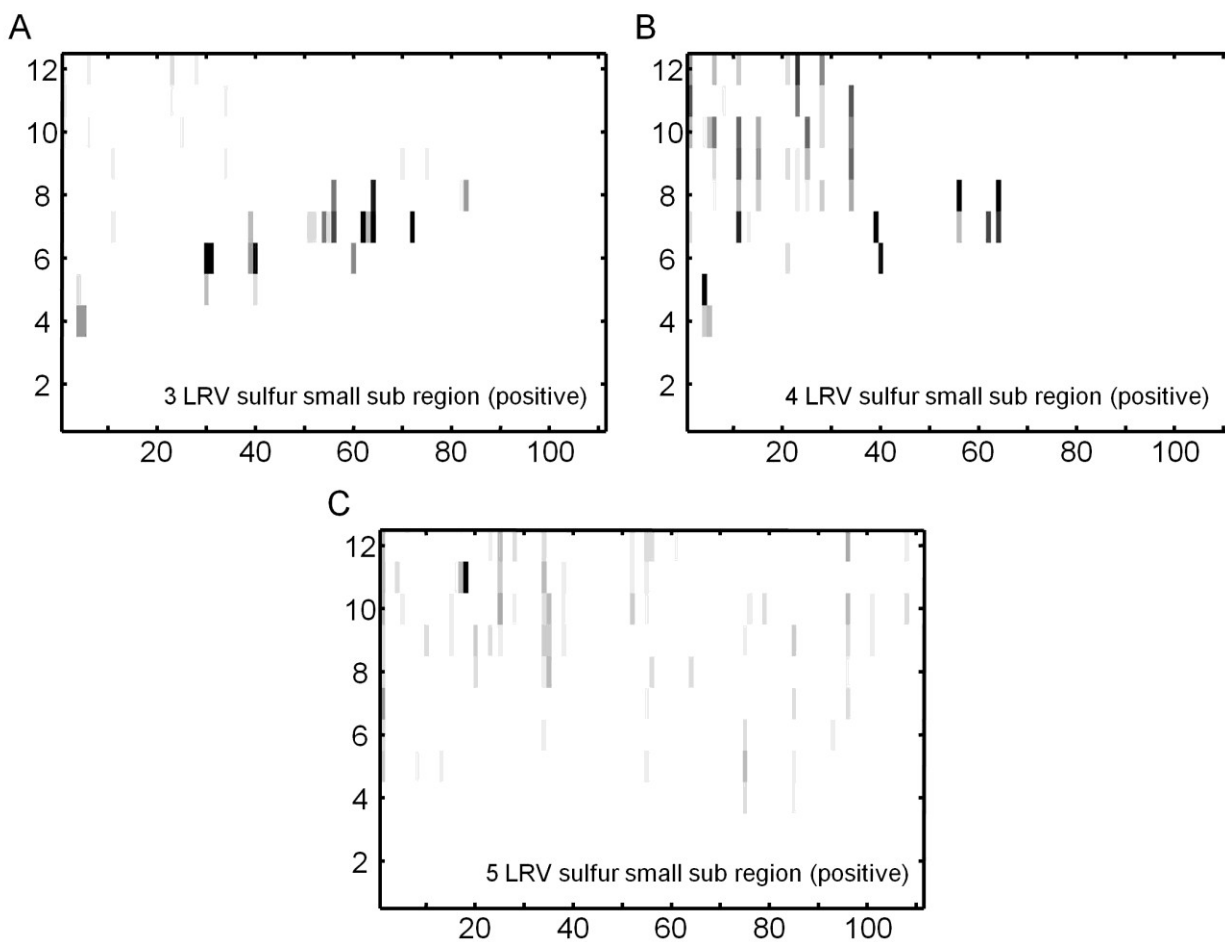


Figure A.5. LRVs from PLS models using the 'Small' sub region chromatograms. (A) 3 LV LRV, positive values only. (B) 4 LV LRV, positive values only. (C) 5 LV LRV, positive values only.

Tables

Table A.1. Identified sulfur containing compounds.

Compound Identification	¹ t _r (min)	² t _r (s)	MV	m/z
methyl benzothiophene	29.5	0.78	761	147, 148
methyl benzothiophene	30.42	0.77	660	147, 148
dimethyl benzothiophene	31.13	0.81	suspect	147, 162
ethyl benzothiophene	32.33	0.81	790	147, 162

Table A.2. Results for the PLS models for sulfur content using smaller sub regions (* bold indicates the recommended PLS model for the specific sub region).

Sub region size	# of LVs	RMSEC	RMSECV
Large	1	4.416	7.824
Large	2	2.997	5.915
Large	3	1.951	5.965
Large	4	1.348	5.530
Large	5	0.2953	5.848
Medium	1	3.422	4.143
Medium	2	2.964	4.632
Medium	3	2.622	4.903
Medium	4	1.571	4.407
Medium	5	0.8378	4.524
Small	1	3.405	4.105
Small	2	3.189	4.600
Small	3	3.119	4.177
Small	4	0.8679	4.545
Small	5	0.08048	4.547

Bibliography

- J.C. Giddings, *Unified Separation Science*, John Wiley & Sons, New York, 1991.
- G.D. Christian, *Analytical Chemistry*, 6th ed., John Wiley & Sons, New York, 2004.
- Z. Liu, M. Zhang, J.B. Philips, High speed Chromatographic Analysis of a simulated process stream using on column thermal desorption modulation for sample preconcentration and introduction, *J. Chromatogr. Sci.* 28 (1990) 567-571.
- R.E. Mohler, B.P. Tu, M. Dombek, J.C. Hoggard, E.T. Young, R.E. Synovec, Identification and evaluation of cycling yeast metabolites in two dimensional comprehensive gas chromatography time of flight mass spectrometry data, *J Chromatogr. A.* 1186 (2008) 401-411.
- B.D. Fitz, R.B. Wilson, B.A. Parsons, J.C. Hoggard, R.E. Synovec, Fast high peak capacity separations in comprehensive two dimensional gas chromatography with time of flight mass spectrometry, *J. Chromatogr A.* 1266 (2012) 116-123.
- F.W. McLafferty, F. Tureček, *Interpretation of Mass Spectra*, 4th ed., University Science Books, Sausalito Ca USA, 1993.
- E. de Hoffmann, V. Stroobant, *Mass Spectrometry: Principles and Applications*, 3rd ed., John Wiley & Sons, New York, 2007.
- S. Yang, J.S. Nadeau, E.M. Humston-Fulmer, J.C. Hoggard, M.E. Lidstrom, R.E. Synovec, Gas chromatography–mass spectrometry with chemometric analysis for determining ¹²C and ¹³C labeled contributions in metabolomics and ¹³C flux analysis, *J. Chromatogr. A.* 1240 (2012) 156–164.
- B. Omais, M. Courtiade, N. Charon, D. Thiébaud, A. Quignard, M.-C. Hennion, Investigating comprehensive two-dimensional gas chromatography conditions to optimize the separation of oxygenated compounds in a direct coal liquefaction middle distillate, *J. Chromatogr. A.* 1218 (2011) 3233–3240.
- C. Vendevre, R. Ruiz-Guerrero, F. Bertoncini, L. Duval, D. Thiébaud, M.-C. Hennion, Characterisation of middle-distillates by comprehensive two-dimensional gas chromatography (GC × GC): A powerful alternative for performing various standard analysis of middle-distillates, *J. Chromatogr. A.* 1086 (2005) 21–28.
- M. Adahchour, J. Beens, R.J.J. Vreuls, A.M. Batenburg, U.A.T. Brinkman, Comprehensive two-dimensional gas chromatography of complex samples by using a “reversed-type” column combination: application to food analysis, *J. Chromatogr. A.* 1054 (2004) 47–55.
- K.R. Beebe, R.J. Pell, M.B. Seasholtz, *Chemometrics A practical guide*, John Wiley & Sons, Inc., New York, 1998.
- W.G. Pool, J.W. de Leeuw, B. van de Graaf, A rapid routine to correct for skewing in gas chromatography/mass spectrometry, *J. Mass Spectrom.* 31 (1996) 213–215.
- A. Savitzky, M.J.E. Golay, Smoothing and Differentiation of Data by Simplified Least Squares Procedures., *Anal. Chem.* 36 (1964) 1627–1639.
- J.S. Nadeau, R.B. Wilson, J.C. Hoggard, B.W. Wright, R.E. Synovec, Study of the interdependency of the data sampling ratio with retention time alignment and principal component analysis for gas chromatography, *J. Chromatogr. A.* 1218 (2011) 9091–9101.
- J.S. Nadeau, B.W. Wright, R.E. Synovec, Chemometric analysis of gas chromatography–mass spectrometry data using fast retention time alignment via a total ion current shift function, *Talanta.* 81 (2010) 120–128.

- K.M. Pierce, B.W. Wright, R.E. Synovec, Unsupervised parameter optimization for automated retention time alignment of severely shifted gas chromatographic data using the piecewise alignment algorithm, *J. Chromatogr. A.* 1141 (2007) 106–116.
- R.A. Davis, A.J. Charlton, J. Godward, S.A. Jones, M. Harrison, J.C. Wilson, Adaptive binning: An improved binning method for metabolomics data using the undecimated wavelet transform, *Chemom. Intell. Lab. Syst.* 85 (2007) 144–154.
- P. Geladi, B.R. Kowalski, Partial least-squares regression: a tutorial, *Anal. Chim. Acta.* 185 (1986) 1–17.
- S. Wold, K. Esbensen, P. Geladi, Principal component analysis, *Chemom. Intell. Lab. Syst.* 2 (1987) 37–52.
- A. Daffertshofer, C.J.C. Lamoth, O.G. Meijer, P.J. Beek, PCA in studying coordination and variability: a tutorial, *Clin. Biomech.* 19 (2004) 415–428.
- R. Bro, PARAFAC. Tutorial and applications, *Chemom. Intell. Lab. Syst.* 38 (1997) 149–171.
- J.C. Hoggard, W.C. Siegler, R.E. Synovec, Toward automated peak resolution in complete GC × GC–TOFMS chromatograms by PARAFAC, *J. Chemom.* 23 (2009) 421–431.
- K.M. Pierce, B. Kehimkar, L.C. Marney, J.C. Hoggard, R.E. Synovec, Review of chemometric analysis techniques for comprehensive two dimensional separations data, *J. Chromatogr. A.* 1255 (2012) 3–11.
- K.M. Pierce, J.C. Hoggard, R.E. Mohler, R.E. Synovec, Recent advancements in comprehensive two-dimensional separations with chemometrics, *J. Chromatogr. A.* 1184 (2008) 341–352.
- M.J. Gómez, S. Herrera, D. Solé, E. García-Calvo, A.R. Fernández-Alba, Automatic Searching and Evaluation of Priority and Emerging Contaminants in Wastewater and River Water by Stir Bar Sorptive Extraction followed by Comprehensive Two-Dimensional Gas Chromatography–Time-of-Flight Mass Spectrometry, *Anal. Chem.* 83 (2011) 2638–2647.
- S. Yang, M. Sadilek, R.E. Synovec, M.E. Lidstrom, Liquid chromatography–tandem quadrupole mass spectrometry and comprehensive two-dimensional gas chromatography–time-of-flight mass spectrometry measurement of targeted metabolites of *Methylobacterium extorquens* AM1 grown on two different carbon sources, *J. Chromatogr. A.* 1216 (2009) 3280–3289.
- E.M. Humston, J.C. Hoggard, R.E. Synovec, Utilizing the Third Order Advantage with Isotope Dilution Mass Spectrometry, *Anal. Chem.* 82 (2010) 41–43.
- J.C. Hoggard, J.H. Wahl, R.E. Synovec, G.M. Mong, C.G. Fraga, Impurity Profiling of a Chemical Weapon Precursor for Possible Forensic Signatures by Comprehensive Two-Dimensional Gas Chromatography/Mass Spectrometry and Chemometrics, *Anal. Chem.* 82 (2010) 689–698.
- E.M. Humston, K.M. Dombek, J.C. Hoggard, E.T. Young, R.E. Synovec, Time-Dependent Profiling of Metabolites from Snf1 Mutant and Wild Type Yeast Cells, *Anal. Chem.* 80 (2008) 8002–8011.
- T. Skov, J.C. Hoggard, R. Bro, R.E. Synovec, Handling within run retention time shifts in two-dimensional chromatography data using shift correction and modeling, *J. Chromatogr. A.* 1216 (2009) 4020–4029.
- J. de Vos, R. Dixon, G. Vermeulen, P. Gorst-Allman, J. Cochran, E. Rohwer, J.-F. Focant, Comprehensive two-dimensional gas chromatography time of flight mass spectrometry (GC × GC–TOFMS) for environmental forensic investigations in developing countries, *Chemosphere.* 82 (2011) 1230–1239.
- T. Sfetsas, C. Michailof, A. Lappas, Q. Li, B. Kneale, Qualitative and quantitative analysis of pyrolysis oil by gas chromatography with flame ionization detection and comprehensive two-dimensional gas chromatography with

time-of-flight mass spectrometry, *J. Chromatogr. A.* 1218 (2011) 3317–3325.

B. Guthery, T. Bassindale, A. Bassindale, C.T. Pillinger, G.H. Morgan, Qualitative drug analysis of hair extracts by comprehensive two-dimensional gas chromatography/time-of-flight mass spectrometry, *J. Chromatogr. A.* 1217 (2010) 4402–4410.

R.E. Mohler, K.M. Dombek, J.C. Hoggard, K.M. Pierce, E.T. Young, R.E. Synovec, Comprehensive analysis of yeast metabolite GC×GC–TOFMS data: combining discovery-mode and deconvolution chemometric software, *Analyst.* 132 (2007) 756–767.

K.M. Pierce, J.C. Hoggard, J.L. Hope, P.M. Rainey, A.N. Hoofnagle, R.M. Jack, B.W. Wright, R.E. Synovec, Fisher Ratio Method Applied to Third-Order Separation Data To Identify Significant Chemical Components of Metabolite Extracts, *Anal. Chem.* 78 (2006) 5068–5075.

J.C. Hoggard, The Synovec Research Group at UW, (<http://depts.washington.edu/synlab/people/dr-jamin-hoggard-2/>).

K.M. Pierce, J.C. Hoggard, Chromatographic data analysis. Part 3.3.4: handling hyphenated data in chromatography, *Anal Methods* (2014), DOI 10.1039/C3AY40965A

M.B. Lucitt, T.S. Price, A. Pizarro, W. Wu, A.K. Yocum, C. Seiler, M.A. Pack, I.A. Blair, G.A. FitzGerald, T. Grosser, Analysis of the Zebrafish Proteome during Embryonic Development, *Mol. Cell. Proteomics.* 7 (2008) 981–994.

T. Gröger, R. Zimmermann, Application of parallel computing to speed up chemometrics for GC × GC–TOFMS based metabolic fingerprinting, *Talanta.* 83 (2011) 1289–1294.

M.P. Pedroso, L.A.F. de Godoy, E.C. Ferreira, R.J. Poppi, F. Augusto, Identification of gasoline adulteration using comprehensive two-dimensional gas chromatography combined to multivariate data processing, *J. Chromatogr. A.* 1201 (2008) 176–182.

L.A.F. de Godoy, E.C. Ferreira, M.P. Pedroso, C.H. de V. Fidélis, F. Augusto, R.J. Poppi, Quantification of Kerosene in Gasoline by Comprehensive Two-Dimensional Gas Chromatography and N-Way Multivariate Analysis, *Anal. Lett.* 41 (2008) 1603–1614.

K.M. Pierce, S.P. Schale, Predicting percent composition of blends of biodiesel and conventional diesel using gas chromatography–mass spectrometry, comprehensive two-dimensional gas chromatography–mass spectrometry, and partial least squares analysis, *Talanta.* 83 (2011) 1254–1259.

E.M. Humston, J.D. Knowles, A. McShea, R.E. Synovec, Quantitative assessment of moisture damage for cacao bean quality using two-dimensional gas chromatography combined with time-of-flight mass spectrometry and chemometrics, *J. Chromatogr. A.* 1217 (2010) 1963–1970.

E.M. Humston, Y. Zhang, G.F. Brabeck, A. McShea, R.E. Synovec, Development of a GC×GC–TOFMS method using SPME to determine volatile compounds in cacao beans, *J. Sep. Sci.* 32 (2009) 2289–2295.

A.C. Beckstrom, E.M. Humston, L.R. Snyder, R.E. Synovec, S.E. Juul, Application of comprehensive two-dimensional gas chromatography with time-of-flight mass spectrometry method to identify potential biomarkers of perinatal asphyxia in a non-human primate model, *J. Chromatogr. A.* 1218 (2011) 1899–1906.

S. Peters, G. Vivó-Truyols, P.J. Marriott, P.J. Schoenmakers, Development of a resolution metric for comprehensive two-dimensional chromatography, *J. Chromatogr. A.* 1146 (2007) 232–241.

J.L. Hope, A.E. Sinha, B.J. Prazen, R.E. Synovec, Evaluation of the DotMap algorithm for locating analytes of interest based on mass spectral similarity in data collected using comprehensive two-dimensional gas chromatography coupled with time-of-flight mass spectrometry, *J. Chromatogr. A.* 1086 (2005) 185–192.

- A.E. Sinha, J.L. Hope, B.J. Prazen, E.J. Nilsson, R.M. Jack, R.E. Synovec, Algorithm for locating analytes of interest based on mass spectral similarity in GC \times GC-TOF-MS data: analysis of metabolites in human infant urine, *J. Chromatogr. A.* 1058 (2004) 209–215.
- H.-Y. Fu, H.-L. Wu, Y.-J. Yu, L.-L. Yu, S.-R. Zhang, J.-F. Nie, S.-F. Li, R.-Q. Yu, A new third-order calibration method with application for analysis of four-way data arrays, *J. Chemom.* 25 (2011) 408–429.
- K.J. Johnson, R.E. Synovec, Pattern recognition of jet fuels: comprehensive GC \times GC with ANOVA-based feature selection and principal component analysis, *Chemom. Intell. Lab. Syst.* 60 (2002) 225–237.
- J.C. Hoggard, R.E. Synovec, Automated Resolution of Nontarget Analyte Signals in GC \times GC-TOFMS Data Using Parallel Factor Analysis, *Anal. Chem.* 80 (2008) 6677–6688.
- B. Kehimkar, J.C. Hoggard, J.S. Nadeau, R.E. Synovec, Targeted mass spectral ratio analysis: A new tool for gas chromatography—mass spectrometry, *Talanta.* 103 (2013) 267–275.
- F.L. Dorman, J.J. Whiting, J.W. Cochran, J. Gardea-Torresdey, Gas Chromatography, *Anal. Chem.* 82 (2010) 4775–4785.
- W. Zhang, P. Wu, C. Li, Study of automated mass spectral deconvolution and identification system (AMDIS) in pesticide residue analysis, *Rapid Commun. Mass Spectrom.* 20 (2006) 1563–1568.
- M.R. Meyer, F.T. Peters, H.H. Maurer, Automated Mass Spectral Deconvolution and Identification System for GC-MS Screening for Drugs, Poisons, and Metabolites in Urine, *Clin. Chem.* 56 (2010) 575–584.
- A. Genovese, R. Dimaggio, M.T. Lisanti, P. Piombino, L. Moio, Aroma Composition of Red Wines by Different Extraction Methods and Gas Chromatography-SIM/Mass Spectrometry Analysis, *Ann. Chim.* 95 (2005) 383–394.
- E. Jellum, O. Stokke, L. Eldjarn, Application of gas chromatography, mass spectrometry, and computer methods in clinical biochemistry, *Anal. Chem.* 45 (1973) 1099–1106.
- J.M. Halket, A. Przyborowska, S.E. Stein, W.G. Mallard, S. Down, R.A. Chalmers, Deconvolution gas chromatography/mass spectrometry of urinary organic acids – potential for pattern recognition and automated identification of metabolic disorders, *Rapid Commun. Mass Spectrom.* 13 (1999) 279–284.
- W.G. Pool, J.W. de Leeuw, B. van de Graaf, Automated Extraction of Pure Mass Spectra from Gas Chromatographic/Mass Spectrometric Data, *J. Mass Spectrom.* 32 (1997) 438–443.
- Y. Koh, K.K. Pasikanti, C.W. Yap, E.C.Y. Chan, Comparative evaluation of software for retention time alignment of gas chromatography/time-of-flight mass spectrometry-based metabonomic data, *J. Chromatogr. A.* 1217 (2010) 8308–8316.
- E. Sanchez, B.R. Kowalski, Tensorial resolution: A direct trilinear decomposition, *J. Chemom.* 4 (1990) 29–45.
- C.G. Fraga, Chemometric approach for the resolution and quantification of unresolved peaks in gas chromatography–selected-ion mass spectrometry data, *J. Chromatogr. A.* 1019 (2003) 31–42.
- M. Daszykowski, B. Walczak, Use and abuse of chemometrics in chromatography, *TrAC Trends Anal. Chem.* 25 (2006) 1081–1096.
- M. Garrido, F.X. Rius, M.S. Larrechi, Multivariate curve resolution–alternating least squares (MCR-ALS) applied to spectroscopic data from monitoring chemical reactions processes, *Anal. Bioanal. Chem.* 390 (2008) 2059–2066.
- Y. Marrero-Ponce, E.R. Martínez-Albelo, G.M. Casañola-Martín, J.A. Castillo-Garit, Y. Echevería-Díaz, V.R. Zaldivar, J. Tygat, J.E.R. Borges, R. García-Domenech, F. Torrens, F. Pérez-Giménez, Bond-based linear indices of

the non-stochastic and stochastic edge-adjacency matrix. 1. Theory and modeling of ChemPhys properties of organic molecules, *Mol. Divers.* 14 (2010) 731–753.

M. Randić, On Characterization of Chemical Structure, *J. Chem. Inf. Comput. Sci.* 37 (1997) 672–687.

E. Estrada, Edge Adjacency Relationships and a Novel Topological Index Related to Molecular Volume, *J. Chem. Inf. Comput. Sci.* 35 (1995) 31–33.

T.I. Dearing, J.S. Nadeau, B.G. Rohrback, L.S. Ramos, R.E. Synovec, Real-time target selection optimization to enhance alignment of gas chromatograms, *Talanta.* 83 (2011) 738–743.

M. Fransson, S. Folestad, Real-time alignment of batch process data using COW for on-line process monitoring, *Chemom. Intell. Lab. Syst.* 84 (2006) 56–61.

M. Daszykowski, B. Walczak, Target selection for alignment of chromatographic signals obtained using monochannel detectors, *J. Chromatogr. A.* 1176 (2007) 1–11.

K.M. Pierce, J.L. Hope, K.J. Johnson, B.W. Wright, R.E. Synovec, Classification of gasoline data obtained by gas chromatography using a piecewise alignment algorithm combined with feature selection and principal component analysis, *J. Chromatogr. A.* 1096 (2005) 101–110.

N.E. Watson, M.M. VanWingerden, K.M. Pierce, B.W. Wright, R.E. Synovec, Classification of high-speed gas chromatography–mass spectrometry data by principal component analysis coupled with piecewise alignment and feature selection, *J. Chromatogr. A.* 1129 (2006) 111–118.

B. Kehimkar, J.C. Hoggard, L.C. Marney, M.C. Billingsley, C.G. Fraga, T.J. Bruno, Synovec RE, Correlation of rocket propulsion fuel properties with chemical composition using comprehensive two-dimensional gas chromatography with time-of-flight mass spectrometry followed by partial least squares regression analysis, *J. Chromatogr.* 1327 (2014) 132–140.

D.J. Cookson, B.E. Smith, Calculation of jet and diesel fuel properties using carbon-13 NMR spectroscopy, *Energy Fuels.* 4 (1990) 152–156.

G. Liu, L. Wang, H. Qu, H. Shen, X. Zhang, S. Zhang, Z. Mi, Artificial neural network approaches on composition–property relationships of jet fuels based on GC–MS, *Fuel.* 86 (2007) 2551–2559.

M.L. Huber, E.W. Lemmon, T.J. Bruno, Effect of RP-1 Compositional Variability on Thermophysical Properties, *Energy Fuels.* 23 (2009) 5550–5555.

M.J. DeWitt, T. Edwards, L. Shafer, D. Brooks, R. Striebich, S.P. Bagley, M.J. Wornat, Effect of Aviation Fuel Type on Pyrolytic Reactivity and Deposition Propensity under Supercritical Conditions, *Ind. Eng. Chem. Res.* 50 (2011) 10434–10451.

M.C. Billingsley, T. Edwards, L.M. Shafer, T.J. Bruno, Extent and Impacts of Hydrocarbon Fuel Compositional Variability for Aerospace Propulsion Systems, *Proc. 46th AIAA/ASME/SAE/ASEE Joint Propulsion Conference*, 2010.

T.M. Lovestead, B.C. Windom, J.R. Riggs, C. Nickell, T.J. Bruno, Assessment of the Compositional Variability of RP-1 and RP-2 with the Advanced Distillation Curve Approach, *Energy Fuels.* 24 (2010) 5611–5623.

R.V. Gough, T.J. Bruno, Composition-Explicit Distillation Curves of Alternative Turbine Fuels, *Energy Fuels.* 27 (2013) 294–302.

P.Y. Hsieh, K.R. Abel, T.J. Bruno, Analysis of Marine Diesel Fuel with the Advanced Distillation Curve Method, *Energy Fuels.* 27 (2013) 804–810.

- J.L. Burger, T.J. Bruno, Application of the Advanced Distillation Curve Method to the Variability of Jet Fuels, *Energy Fuels*. 26 (2012) 3661–3671.
- N.J. Begue, J.A. Cramer, C. Von Bargen, K.M. Myers, K.J. Johnson, R.E. Morris, Automated Method for Determining Hydrocarbon Distributions in Mobility Fuels, *Energy Fuels*. 25 (2011) 1617–1623.
- T.J. Bruno, L.S. Ott, T.M. Lovestead, M.L. Huber, The composition-explicit distillation curve technique: Relating chemical analysis and physical properties of complex fluids, *J. Chromatogr. A*. 1217 (2010) 2703–2715.
- R.B. Wilson, W.C. Siegler, J.C. Hoggard, B.D. Fitz, J.S. Nadeau, R.E. Synovec, Achieving high peak capacity production for gas chromatography and comprehensive two-dimensional gas chromatography by minimizing off-column peak broadening, *J. Chromatogr. A*. 1218 (2011) 3130–3139.
- J.H. Christensen, G. Tomasi, A.B. Hansen, Chemical Fingerprinting of Petroleum Biomarkers Using Time Warping and PCA, *Environ. Sci. Technol.* 39 (2005) 255–260.
- R.E. Mohler, K.M. Dombek, J.C. Hoggard, E.T. Young, R.E. Synovec, Comprehensive Two-Dimensional Gas Chromatography Time-of-Flight Mass Spectrometry Analysis of Metabolites in Fermenting and Respiring Yeast Cells, *Anal. Chem.* 78 (2006) 2700–2709.
- C.G. Fraga, B.J. Prazen, R.E. Synovec, Comprehensive Two-Dimensional Gas Chromatography and Chemometrics for the High-Speed Quantitative Analysis of Aromatic Isomers in a Jet Fuel Using the Standard Addition Method and an Objective Retention Time Alignment Algorithm, *Anal. Chem.* 72 (2000) 4154–4162.
- B.J. Prazen, K.J. Johnson, A. Weber, R.E. Synovec, Two-Dimensional Gas Chromatography and Trilinear Partial Least Squares for the Quantitative Analysis of Aromatic and Naphthene Content in Naphtha, *Anal. Chem.* 73 (2001) 5677–5682.
- K. J. Johnson, R. E. Synovec, *J. Chemom. Intell. Lab. Syst.* 60 (2002) 225.
- K.J. Johnson, B.J. Prazen, D.C. Young, R.E. Synovec, Quantification of naphthalenes in jet fuel with GC×GC/Tri-PLS and windowed rank minimization retention time alignment, *J. Sep. Sci.* 27 (2004) 410–416.
- F. Westad, N.K. Afseth, R. Bro, Finding relevant spectral regions between spectroscopic techniques by use of cross model validation and partial least squares regression, *Anal. Chim. Acta.* 595 (2007) 323–327.
- T. Rajalahti, O.M. Kvalheim, Multivariate data analysis in pharmaceuticals: A tutorial review, *Int. J. Pharm.* 417 (2011) 280–290.
- A.A. Gowen, G. Downey, C. Esquerre, C.P. O'Donnell, Preventing over-fitting in PLS calibration models of near-infrared (NIR) spectroscopy data using regression coefficients, *J. Chemom.* 25 (2011) 375–381.
- L. C. Marney, W. C. Siegler, B. A. Parsons, J. C. Hoggard, B. W. Wright, R. E. Synovec, *Talanta* 115 (2013) 887–895.
- T.J. Fortin, Assessment of Variability in the Thermophysical Properties of Rocket Propellant RP-1, *Energy Fuels*. 26 (2012) 4383–4394.
- L.S. Ott, A.B. Hadler, T.J. Bruno, Variability of the Rocket Propellants RP-1, RP-2, and TS-5: Application of a Composition- and Enthalpy-Explicit Distillation Curve Method, *Ind. Eng. Chem. Res.* 47 (2008) 9225–9233.
- T.J. Bruno, Improvements in the Measurement of Distillation Curves. 1. A Composition-Explicit Approach, *Ind. Eng. Chem. Res.* 45 (2006) 4371–4380.
- T.J. Bruno, B.L. Smith, Improvements in the Measurement of Distillation Curves. 2. Application to

- Aerospace/Aviation Fuels RP-1 and S-8, *Ind. Eng. Chem. Res.* 45 (2006) 4381–4388.
- T.J. Bruno, L.S. Ott, B.L. Smith, T.M. Lovestead, Complex Fluid Analysis with the Advanced Distillation Curve Approach, *Anal. Chem.* 82 (2010) 777–783.
- B.L. Smith, T.J. Bruno, Improvements in the Measurement of Distillation Curves. 4. Application to the Aviation Turbine Fuel Jet-A, *Ind. Eng. Chem. Res.* 46 (2007) 310–320.
- L.S. Ott, B.L. Smith, T.J. Bruno, Advanced distillation curve measurements for corrosive fluids: Application to two crude oils, *Fuel*. 87 (2008) 3055–3064.
- T.M. Lovestead, T.J. Bruno, A Comparison of the Hypersonic Vehicle Fuel JP-7 to the Rocket Propellants RP-1 and RP-2 with the Advanced Distillation Curve Method, *Energy Fuels*. 23 (2009) 3637–3644.
- B.C. Windom, T.J. Bruno, Assessment of the Composition and Distillation Properties of Thermally Stressed RP-1 and RP-2: Application to Fuel Regenerative Cooling, *Energy Fuels*. 25 (2011) 5200–5214.
- B.L. Smith, T.J. Bruno, Improvements in the Measurement of Distillation Curves. 3. Application to Gasoline and Gasoline + Methanol Mixtures, *Ind. Eng. Chem. Res.* 46 (2007) 297–309.
- B.C. Windom, T.J. Bruno, Improvements in the Measurement of Distillation Curves. 5. Reduced Pressure Advanced Distillation Curve Method, *Ind. Eng. Chem. Res.* 50 (2011) 1115–1126.
- B.C. Windom, T.J. Bruno, Application of Pressure-Controlled Advanced Distillation Curve Analysis: Virgin and Waste Oils, *Ind. Eng. Chem. Res.* 52 (2013) 327–337.
- T.J. Bruno, A. Wolk, A. Naydich, Stabilization of Biodiesel Fuel at Elevated Temperature with Hydrogen Donors: Evaluation with the Advanced Distillation Curve Method, *Energy Fuels*. 23 (2009) 1015–1023.
- J.H. Christensen, G. Tomasi, A.B. Hansen, Chemical Fingerprinting of Petroleum Biomarkers Using Time Warping and PCA, *Environ. Sci. Technol.* 39 (2005) 255–260.
- D02 Committee, Test Method for Density, Relative Density, and API Gravity of Liquids by Digital Density Meter, ASTM International, 2011.
- D02 Committee, Test Method for Kinematic Viscosity of Transparent and Opaque Liquids (and Calculation of Dynamic Viscosity), ASTM International, 2012.
- K.E. Swearingen, W.P. Loomis, B. Kehimkar, B.T. Cookson, N.J. Dovichi, Quantification of green fluorescent protein in cellular supernatant by capillary electrophoresis with laser-induced fluorescence detection for measurement of cell death, *Talanta*. 81 (2010) 948–953.
- D. Cohen, J.A. Dickerson, C.D. Whitmore, E.H. Turner, M.M. Palcic, O. Hindsgaul, N.J. Dovichi, Chemical Cytometry: Fluorescence-Based Single-Cell Analysis, *Annu. Rev. Anal. Chem.* 1 (2008) 165–190.
- P. Jandik, G. Bonn, *Capillary Electrophoresis of Small Molecules and Ions*, VCH Publishers Inc., New York, 1993.
- R. Weinberger, *Practical Capillary Electrophoresis*, Academic Press Inc., San Diego CA, 1993.
- M.M.C. Ferreira, W.C. Ferreira, B.R. Kowalski, Rank determination and analysis of non-linear processes by global linearizing transformation, *J. Chemom.* 10 (1996) 11–30.
- S. Wold, H. Antti, F. Lindgren, J. Öhman, Orthogonal signal correction of near-infrared spectra, *Chemom. Intell. Lab. Syst.* 44 (1998) 175–185.

- Z. Wang, T. Dean, B.R. Kowalski, Additive Background Correction in Multivariate Instrument Standardization, *Anal. Chem.* 67 (1995) 2379–2385.
- Y. Wang, D.J. Veltkamp, B.R. Kowalski, Multivariate instrument standardization, *Anal. Chem.* 63 (1991) 2750–2756.
- R.E. Synovec, Refractive index effects in cylindrical detector cell designs for microbore high-performance liquid chromatography, *Anal. Chem.* 59 (1987) 2877–2884.
- P.R. Bevington, *Data Reduction and Error Analysis for the Physical Sciences*, McGraw-Hill, New York, 1969.
- D02 Committee, Test Method for Heat of Combustion of Liquid Hydrocarbon Fuels by Bomb Calorimeter (Precision Method), ASTM International, 2013.
- L.R. Snyder, J.C. Hoggard, T.J. Montine, R.E. Synovec, Development and application of a comprehensive two-dimensional gas chromatography with time-of-flight mass spectrometry method for the analysis of l- β -methylaminoalanine in human tissue, *J. Chromatogr. A.* 1217 (2010) 4639–4647.
- B.M. Wise, N.B. Gallagher, S.W. Butler, D.D. White, G.G. Barna, A comparison of principal component analysis, multiway principal component analysis, trilinear decomposition and parallel factor analysis for fault detection in a semiconductor etch process, *J. Chemom.* 13 (1999) 379–396.
- N.E. Watson, W.C. Siegler, J.C. Hoggard, R.E. Synovec, Comprehensive Three-Dimensional Gas Chromatography with Parallel Factor Analysis, *Anal. Chem.* 79 (2007) 8270–8280.
- J.J. Harynuk, A.D. Rossé, G.B. McGarvey, Study of alkyl phosphates in industrial petroleum mixtures by comprehensive two-dimensional gas chromatography time-of-flight mass spectrometry, *Anal. Bioanal. Chem.* 401 (2011) 2415–2422.
- S. Dasgupta, K. Banerjee, S.H. Patil, M. Ghaste, K.N. Dhumal, P.G. Adsule, Optimization of two-dimensional gas chromatography time-of-flight mass spectrometry for separation and estimation of the residues of 160 pesticides and 25 persistent organic pollutants in grape and wine, *J. Chromatogr. A.* 1217 (2010) 3881–3889.
- M. Mieth, J.K. Schubert, T. Gröger, B. Sabel, S. Kischkel, P. Fuchs, D. Hein, R. Zimmermann, W. Miekisch, Automated Needle Trap Heart-Cut GC/MS and Needle Trap Comprehensive Two-Dimensional GC/TOF-MS for Breath Gas Analysis in the Clinical Environment, *Anal. Chem.* 82 (2010) 2541–2551.
- R. Stepan, P. Cuhra, S. Barsova, Comprehensive two-dimensional gas chromatography with time-of-flight mass spectrometric detection for the determination of anabolic steroids and related compounds in nutritional supplements, *Food Addit. Contam. Part. 25* (2008) 557–565.
- C. von Mühlen, P.J. Marriott, Retention indices in comprehensive two-dimensional gas chromatography, *Anal. Bioanal. Chem.* 401 (2011) 2351–2360.
- D.A. Skoog, F.J. Holler, T.A. Nieman, *Principles of Instrumental Analysis*, 5th ed., Saunderson College Publishing, 1998.
- W.P. Gardiner, *Statistical Analysis Methods For Chemists A Software-based Approach*, The Royal Society of Chemistry, Cambridge, UK, 1997.

H.G. Hecht, *Mathematics in Chemistry- An Introduction to Modern Methods*, Prentice-Hall Inc, Englewood cliffs, New Jersey, 1990.

R. Aroca, *Surface Enhanced Vibrational Spectroscopy*, John Wiley & Sons, Ltd, Chichester, England, 2006.

P.W. Bohn, *Localized Optical Phenomena and the Characterization of Materials Interfaces*, *Annu. Rev. Mater. Sci.* 27 (1997) 469–498.

Y.R. Shen, *The Principles of Nonlinear Optics*, John Wiley & Sons, Inc, New York, 1984.

sea-surface-chemistry, (<http://www.ozean-der-zukunft.de/english/research-areas/ocean-change/sea-surface-chemistry/facts/>).

liposomes, (<http://www.britannica.com/EBchecked/topic-art/342910/92244/Phospholipids-can-be-used-to-form-artificial-structures-called-liposomes>).

J. Liu, J.C. Conboy, *Asymmetric Distribution of Lipids in a Phase Segregated Phospholipid Bilayer Observed by Sum-Frequency Vibrational Spectroscopy*, *J. Phys. Chem. C.* 111 (2007) 8988–8999.

S. Baldelli, C. Schnitzer, D. Simonelli, *Aqueous Solution/Air Interfaces Probed with Sum Frequency Generation Spectroscopy*, *J. Phys. Chem. B.* 106 (2002) 5313–5324.

Z. Chen, Y.R. Shen, G.A. Somorjai, *Studies of Polymer Surfaces by Sum Frequency Generation Vibrational Spectroscopy*, *Annu. Rev. Phys. Chem.* 53 (2002) 437–465.

Y.R. Shen, *A few selected applications of surface nonlinear optical spectroscopy*, *Proc. Natl. Acad. Sci.* 93 (1996) 12104–12111.

D.K. Hore, D.K. Beaman, D.H. Parks, G.L. Richmond, *Whole-Molecule Approach for Determining Orientation at Isotropic Surfaces by Nonlinear Vibrational Spectroscopy*, *J. Phys. Chem. B.* 109 (2005) 16846–16851.

X. Chen, Z. Chen, *SFG studies on interactions between antimicrobial peptides and supported lipid bilayers*, *Biochim. Biophys. Acta BBA - Biomembr.* 1758 (2006) 1257–1273.

J. Liu, J.C. Conboy, *Structure of a Gel Phase Lipid Bilayer Prepared by the Langmuir–Blodgett/Langmuir-Schaefer Method Characterized by Sum-Frequency Vibrational Spectroscopy*, *Langmuir.* 21 (2005) 9091–9097.

S. Ye, K.T. Nguyen, S.V.L. Clair, Z. Chen, *In situ molecular level studies on membrane related peptides and proteins in real time using sum frequency generation vibrational spectroscopy*, *J. Struct. Biol.* 168 (2009) 61–77.

T.C. Anglin, J. Liu, J.C. Conboy, *Facile Lipid Flip-Flop in a Phospholipid Bilayer Induced by Gramicidin A Measured by Sum-Frequency Vibrational Spectroscopy*, *Biophys. J.* 92 (2007) L01–L03.

G.A. Somorjai, A.L. Marsh, *Active sites and states in the heterogeneous catalysis of carbon–hydrogen bonds*, *Philos. Trans. R. Soc. Math. Phys. Eng. Sci.* 363 (2005) 879–900.

J.B. Rollins, B.D. Fitchett, J.C. Conboy, *Structure and Orientation of the Imidazolium Cation at the Room-Temperature Ionic Liquid/SiO₂ Interface Measured by Sum-Frequency Vibrational Spectroscopy*, *J. Phys. Chem. B.*

111 (2007) 4990–4999.

A. Opdahl, S. Hoffer, B. Mailhot, G. a. Somorjai, Polymer surface science, *Chem. Rec.* 1 (2001) 101–122.

M.A. De Jesús, K.S. Giesfeldt, M.J. Sepaniak, Improving the analytical figures of merit of SERS for the analysis of model environmental pollutants, *J. Raman Spectrosc.* 35 (2004) 895–904.

J. Liu, J.C. Conboy, Phase Transition of a Single Lipid Bilayer Measured by Sum-Frequency Vibrational Spectroscopy, *J. Am. Chem. Soc.* 126 (2004) 8894–8895.

P.B. Miranda, Y.R. Shen, Liquid Interfaces: A Study by Sum-Frequency Vibrational Spectroscopy, *J. Phys. Chem. B.* 103 (1999) 3292–3307.

G.A. Somorjai, J.Y. Park, Molecular surface chemistry by metal single crystals and nanoparticles from vacuum to high pressure, *Chem. Soc. Rev.* 37 (2008) 2155–2162.

G. Temporão, S. Tanzilli, H. Zbinden, N. Gisin, T. Aellen, M. Giovannini, J. Faist, Mid-infrared single-photon counting, *Opt. Lett.* 31 (2006) 1094–1096.

G. Niaura, Z. Kuprionis, I. Ignatjev, M. Kazemkaiatè, G. Valincius, Z. Talaikytè, V. Razumas, A. Svendsen, Probing of Lipase Activity at Air/Water Interface by Sum-Frequency Generation Spectroscopy, *J. Phys. Chem. B.* 112 (2008) 4094–4101.

D.K. Hore, D.K. Beaman, G.L. Richmond, Surfactant Headgroup Orientation at the Air/Water Interface, *J. Am. Chem. Soc.* 127 (2005) 9356–9357.

J. Kim, G.A. Somorjai, Molecular Packing of Lysozyme, Fibrinogen, and Bovine Serum Albumin on Hydrophilic and Hydrophobic Surfaces Studied by Infrared–Visible Sum Frequency Generation and Fluorescence Microscopy, *J. Am. Chem. Soc.* 125 (2003) 3150–3158.

C. Aliaga, C.S. Santos, S. Baldelli, Surface chemistry of room-temperature ionic liquids, *Phys. Chem. Chem. Phys.* 9 (2007) 3683–3700.

B.D. Fitchett, J.B. Rollins, J.C. Conboy, Interfacial Tension and Electrocapillary Measurements of the Room Temperature Ionic Liquid/Aqueous Interface, *Langmuir.* 21 (2005) 12179–12186.

A. Volkmer, Vibrational imaging and microspectroscopies based on coherent anti-Stokes Raman scattering microscopy, *J. Phys. Appl. Phys.* 38 (2005) R59–R81.

T.E. Wales, K.E. Fadgen, G.C. Gerhardt, J.R. Engen, High-Speed and High-Resolution UPLC Separation at Zero Degrees Celsius, *Anal. Chem.* 80 (2008) 6815–6820.

A.J. Bard, L.R. Faulkner, *Electrochemical Methods- Fundamentals and Applications*, 2nd ed., John Wiley & Sons, Inc, New York, 2001.

J.J. Stine, C.P. Palmer, Covalent modification of fused silica capillaries with quaternized polyamines to achieve robust and stable anodic electroosmotic flow, *J. Sep. Sci.* 32 (2009) 446–456.

M.M. Harwood, E.S. Christians, M.A. Fazal, N.J. Dovichi, Single-cell protein analysis of a single mouse embryo by

two-dimensional capillary electrophoresis, *J. Chromatogr. A.* 1130 (2006) 190–194.

L.H.H. Silvertand, E. Machtejevas, R. Hendriks, K.K. Unger, W.P. van Bennekom, G.J. de Jong, Selective protein removal and desalting using microchip CE, *J. Chromatogr. B.* 839 (2006) 68–73.

H. Shadpour, S.A. Soper, Two-Dimensional Electrophoretic Separation of Proteins Using Poly(methyl methacrylate) Microchips, *Anal. Chem.* 78 (2006) 3519–3527.

R.S. Foote, J. Khandurina, S.C. Jacobson, J.M. Ramsey, Preconcentration of Proteins on Microfluidic Devices Using Porous Silica Membranes, *Anal. Chem.* 77 (2005) 57–63.

T. Guo, C.S. Lee, W. Wang, D.L. DeVoe, B.M. Balgley, Capillary separations enabling tissue proteomics-based biomarker discovery, *Electrophoresis.* 27 (2006) 3523–3532.

X. Fang, L. Yang, W. Wang, T. Song, C.S. Lee, D.L. DeVoe, B.M. Balgley, Comparison of Electrokinetics-Based Multidimensional Separations Coupled with Electrospray Ionization-Tandem Mass Spectrometry for Characterization of Human Salivary Proteins, *Anal. Chem.* 79 (2007) 5785–5792.

N.M. Homan, C.W. Windt, F.J. Vergeldt, E. Gerkema, H.V. As, 0.7 and 3 T MRI and Sap Flow in Intact Trees: Xylem and Phloem in Action, *Appl. Magn. Reson.* 32 (2007) 157–170.

M. Gopal, E. Mishra, Analytical Method for Estimation of a New Insecticide Flubendiamide and its Safety Evaluation for Usage in Rice Crop, *Bull. Environ. Contam. Toxicol.* 81 (2008) 360–364.

H. Yin, K. Killeen, R. Brennen, D. Sobek, M. Werlich, T. van de Goor, Microfluidic Chip for Peptide Analysis with an Integrated HPLC Column, Sample Enrichment Column, and Nanoelectrospray Tip, *Anal. Chem.* 77 (2005) 527–533.

J.S. Albuquerque, M.F. Pimentel, V.L. Silva, I.M. Raimundo, J.J.R. Rohwedder, C. Pasquini, Silicone Sensing Phase for Detection of Aromatic Hydrocarbons in Water Employing Near-Infrared Spectroscopy, *Anal. Chem.* 77 (2005) 72–77.

M. Arredondo, G.M. LaPorte, J.D. Wilson, T. McConnell, D.K. Shaffer, M. Stam, Analytical Methods Used for the Discrimination of Substances Suspected to be Bar Soap: A Preliminary Study*, *J. Forensic Sci.* 51 (2006) 1334–1343.

J.A. Dowell, D.C. Frost, J. Zhang, L. Li, Comparison of Two-Dimensional Fractionation Techniques for Shotgun Proteomics, *Anal. Chem.* 80 (2008) 6715–6723.

M. Gómez-Romero, D. Arráez-Román, R. Moreno-Torres, P. García-Salas, A. Segura-Carretero, A. Fernández-Gutiérrez, Antioxidant compounds of propolis determined by capillary electrophoresis–mass spectrometry, *J. Sep. Sci.* 30 (2007) 595–603.

A. Yanagida, A. Shoji, Y. Shibusawa, H. Shindo, M. Tagashira, M. Ikeda, Y. Ito, Analytical separation of tea catechins and food-related polyphenols by high-speed counter-current chromatography, *J. Chromatogr. A.* 1112 (2006) 195–201.

F.-K. Liu, M.-H. Tsai, Y.-C. Hsu, T.-C. Chu, Analytical separation of Au/Ag core/shell nanoparticles by capillary electrophoresis, *J. Chromatogr. A.* 1133 (2006) 340–346.

- M.V. Mirkin, A.J. Bard, Simple analysis of quasi-reversible steady-state voltammograms, *Anal. Chem.* 64 (1992) 2293–2302.
- M.W. Kanan, D.G. Nocera, In Situ Formation of an Oxygen-Evolving Catalyst in Neutral Water Containing Phosphate and Co^{2+} , *Science*. 321 (2008) 1072–1075.
- M.E. Smith, J.H. Strange, NMR techniques in materials physics: a review, *Meas. Sci. Technol.* 7 (1996) 449.
- S. Van Bramer, *An Introduction to NMR concepts*, Widener University, Chester, PA, 2000.
- A.W. Overhauser, Polarization of Nuclei in Metals, *Phys. Rev.* 92 (1953) 411–415.
- B.-W. Hu, P. Zhou, I. Noda, G.-Z. Zhao, An NMR Approach Applicable to Biomolecular Structure Characterization, *Anal. Chem.* 77 (2005) 7534–7538.
- B. Cramer, M. Königs, H.-U. Humpf, Identification and in Vitro Cytotoxicity of Ochratoxin A Degradation Products Formed during Coffee Roasting, *J. Agric. Food Chem.* 56 (2008) 5673–5681.
- S. Scharbert, N. Holzmann, T. Hofmann, Identification of the Astringent Taste Compounds in Black Tea Infusions by Combining Instrumental Analysis and Human Bioresponse, *J. Agric. Food Chem.* 52 (2004) 3498–3508.
- C.E. McNamara, L. Larsen, N.B. Perry, J.L. Harper, M.V. Berridge, E.W. Chia, M. Kelly, V.L. Webb, Anti-inflammatory Sesquiterpene-quinones from the New Zealand Sponge *Dysidea cf. cristagalli*, *J. Nat. Prod.* 68 (2005) 1431–1433.
- M.J. Jablonsky, P.L. Jackson, N.R. Krishna, Solution Structure of an Insect-Specific Neurotoxin from the New World Scorpion *Centruroides sculpturatus* Ewing, *Biochemistry (Mosc.)*. 40 (2001) 8273–8282.
- S. Haiber, H. Herzog, P. Burba, B. Gosciniak, J. Lambert, Two-Dimensional NMR Studies of Size Fractionated Suwannee River Fulvic and Humic Acid Reference, *Environ. Sci. Technol.* 35 (2001) 4289–4294.
- G.Q. Ding, M.J. Zheng, W.L. Xu, W.Z. Shen, Fabrication of controllable free-standing ultrathin porous alumina membranes, *Nanotechnology*. 16 (2005) 1285–1289.
- M. Gowtham, L. Eude, C.S. Cojocar, B. Marquardt, H.J. Jeong, P. Legagneux, K.K. Song, D. Pribat, Controlled fabrication of patterned lateral porous alumina membranes, *Nanotechnology*. 19 (2008) 035303–035309.
- T.T. Xu, R.D. Piner, R.S. Ruoff, An Improved Method To Strip Aluminum from Porous Anodic Alumina Films, *Langmuir*. 19 (2003) 1443–1445.
- Z. Su, W. Zhou, Formation Mechanism of Porous Anodic Aluminium and Titanium Oxides, *Adv. Mater.* 20 (2008) 3663–3667.
- H. Asoh, K. Nishio, M. Nakao, T. Tamamura, H. Masuda, Conditions for Fabrication of Ideally Ordered Anodic Porous Alumina Using Pretextured Al, *J. Electrochem. Soc.* 148 (2001) B152–B156.
- W. Chen, J.-S. Wu, X.-H. Xia, Porous Anodic Alumina with Continuously Manipulated Pore/Cell Size, *ACS Nano*. 2 (2008) 959–965.

- A. Rauf, M. Mehmood, M.A. Rasheed, M. Aslam, The effects of electropolishing on the nanochannel ordering of the porous anodic alumina prepared in oxalic acid, *J. Solid State Electrochem.* 13 (2009) 321–332.
- M.A. Kashi, A. Ramazani, The effect of temperature and concentration on the self-organized pore formation in anodic alumina, *J. Phys. Appl. Phys.* 38 (2005) 2396–2399.
- Z. Su, G. Hähner, W. Zhou, Investigation of the pore formation in anodic aluminium oxide, *J. Mater. Chem.* 18 (2008) 5787–5795.
- G.E. Thompson, G.C. Wood, Porous anodic film formation on aluminium, *Nature.* 290 (1981) 230–232.
- G.H. Li, Y. Zhang, Y.C. Wu, L.D. Zhang, Photoluminescence of anodic alumina membranes: pore size dependence, *Appl. Phys.* 81 (2005) 627–629.
- R. Nayak, D.R. Knapp, Effects of Thin-Film Structural Parameters on Laser Desorption/Ionization from Porous Alumina, *Anal. Chem.* 79 (2007) 4950–4956.
- S. Singh, W.R.T. Barden, P. Kruse, Nanopatterning of Transition Metal Surfaces via Electrochemical Dimple Array Formation, *ACS Nano.* 2 (2008) 2453–2464.
- D.F. Hunt, J.R. Yates, J. Shabanowitz, S. Winston, C.R. Hauer, Protein sequencing by tandem mass spectrometry, *Proc. Natl. Acad. Sci.* 83 (1986) 6233–6237.
- E.C. Yi, M. Marelli, H. Lee, S.O. Purvine, R. Aebersold, J.D. Aitchison, D.R. Goodlett, Approaching complete peroxisome characterization by gas-phase fractionation, *ELECTROPHORESIS.* 23 (2002) 3205–3216.
- A. Scherl, S.A. Shaffer, G.K. Taylor, H.D. Kulasekara, S.I. Miller, D.R. Goodlett, Genome-Specific Gas-Phase Fractionation Strategy for Improved Shotgun Proteomic Profiling of Proteotypic Peptides, *Anal. Chem.* 80 (2008) 1182–1191.
- J. Liu, S. Lin, D. Qi, C. Deng, P. Yang, X. Zhang, On-chip enzymatic microreactor using trypsin-immobilized superparamagnetic nanoparticles for highly efficient proteolysis, *J. Chromatogr. A.* 1176 (2007) 169–177.
- R. Garrido-Delgado, S. López-Vidal, L. Arce, M. Valcárcel, Differentiation and identification of white wine varieties by using electropherogram fingerprints obtained with CE, *J. Sep. Sci.* 32 (2009) 3809–3816.
- G. Fiorini, D. Chiu, Disposable microfluidic devices: fabrication, function, and application, *BioTechniques.* 38 (2005) 429–446.
- L.M. Ramsay, J.A. Dickerson, N.J. Dovichi, Attomole protein analysis by CIEF with LIF detection, *Electrophoresis.* 30 (2009) 297–302.
- M. Herrero, E. Ibañez, A. Cifuentes, Capillary electrophoresis-electrospray-mass spectrometry in peptide analysis and peptidomics, *Electrophoresis.* 29 (2008) 2148–2160.
- D. Gonzalez-Gomez, D. Cohen, J.A. Dickerson, X. Chen, F. Cañada-Cañada, N.J. Dovichi, Improvement in protein separation in Barrett's esophagus samples using two-dimensional capillary electrophoresis analysis in presence of cyclodextrins as buffer additives, *Talanta.* 78 (2009) 193–198.

- M. Li, J. Zhou, X. Gu, Y. Wang, X. Huang, C. Yan, Quantitative capillary electrophoresis and its application in analysis of alkaloids in tea, coffee, coca cola, and theophylline tablets, *J. Sep. Sci.* 32 (2009) 267–274.
- W. Berthold, A. Wucher, Electronic Excitation during Sputtering of Silver Atoms, *Phys. Rev. Lett.* 76 (1996) 2181–2184.
- M.W. Evans, C.R. Pelkie, Optical nuclear magnetic resonance: theory, simulation, and animation, *J. Opt. Soc. Am. B.* 9 (1992) 1020–1029.
- C. Cordella, I. Moussa, A.-C. Martel, N. Sbirrazzuoli, L. Lizzani-Cuvelier, Recent Developments in Food Characterization and Adulteration Detection: Technique-Oriented Perspectives, *J. Agric. Food Chem.* 50 (2002) 1751–1764.
- H. Zhu, S.M. Clark, S.C. Benson, H.S. Rye, A.N. Glazer, R.A. Mathies, High-Sensitivity Capillary Electrophoresis of Double-Stranded DNA Fragments Using Monomeric and Dimeric Fluorescent Intercalating Dyes, *Anal. Chem.* 66 (1994) 1941–1948.
- S.M. Clark, R.A. Mathies, Multiplex dsDNA Fragment Sizing Using Dimeric Intercalation Dyes and Capillary Array Electrophoresis: Ionic Effects on the Stability and Electrophoretic Mobility of DNA–Dye Complexes, *Anal. Chem.* 69 (1997) 1355–1363.
- T.E. Wales, K.E. Fadgen, G.C. Gerhardt, J.R. Engen, High-Speed and High-Resolution UPLC Separation at Zero Degrees Celsius, *Anal. Chem.* 80 (2008) 6815–6820.
- C. Pan, L. Zhang, J. Zhu, J. Luo, Z. Cheng, C. Wang, Surface decoration of anodic aluminium oxide in synthesis of Nafion®-115 nanowire arrays, *Nanotechnology.* 18 (2007) 015302-015308.
- I. Riipinen, B. Svenningsson, M. Bilde, A. Gaman, K.E.J. Lehtinen, M. Kulmala, A method for determining thermophysical properties of organic material in aqueous solutions: Succinic acid, *Atmospheric Res.* 82 (2006) 579–590.
- V. Philipsen, J. Bastiaansen, P. Lievens, E. Vandeweert, R.E. Silverans, Resonant electron transfer in the emission of ion-beam sputtered metal atoms studied by double resonance laser ionization, *Vacuum.* 56 (2000) 269–274.
- F.J. Marcie, X-ray fluorescence determination of trace toxic elements in water, *Environ. Sci. Technol.* 1 (1967) 164–166.
- E.R. Williams, S.Y. Loh, F.W. McLafferty, R.B. Cody, Hadamard transform measurement of tandem Fourier-transform mass spectra, *Anal. Chem.* 62 (1990) 698–703.
- D.F. Hunt, J. Shabanowitz, J.R. Yates, R.T. McIver, R.L. Hunter, Tandem-quadrupole Fourier transform mass spectrometry of oligopeptides, *Anal. Chem.* 57 (1985) 2728–2733.
- V.B. Stein, E. Canelli, A.H. Richards, Determination of dissolved Selenium in fresh and estuarine waters by flameless atomic absorption, *At. Spectrosc.* 1 (1980) 61–65.
- E. Vandeweert, V. Philipsen, W. Bouwen, P. Thoen, H. Weidele, R.E. Silverans, P. Lievens, Anomalous Sputtering Behavior Observed by Quantitative Measurements of the Population Partition of Metastable Ni Atoms, *Phys. Rev.*

Lett. 78 (1997) 138–141.

M.A. Kuzyk, L.B. Ohlund, M.H. Elliott, D. Smith, H. Qian, A. Delaney, C.L. Hunter, C.H. Borchers, A comparison of MS/MS-based, stable-isotope-labeled, quantitation performance on ESI-quadrupole TOF and MALDI-TOF/TOF mass spectrometers, *PROTEOMICS*. 9 (2009) 3328–3340.

P. Kubalczyk, E. Bald, Analysis of orange juice for total cysteine and glutathione content by CZE with UV-absorption detection, *Electrophoresis*. 30 (2009) 2280–2283.

S. Vichi, M.R. Aumatell, S. Buxaderas, E. López-Tamames, Assessment of some diterpenoids in commercial distilled gin, *Anal. Chim. Acta*. 628 (2008) 222–229.

B.R. Fonslow, J.R. Yates III, Capillary electrophoresis applied to proteomic analysis, *J. Sep. Sci.* 32 (2009) 1175–1188.

M.C. Billingsley, A.K. Weaton, B.A. Wight, Compositional Variability of RP-1: Chemical Analysis and Impacts on Physical Properties, in: 2010.

C.R. Borges, Concept for Facilitating Analyst-Mediated Interpretation of Qualitative Chromatographic–Mass Spectral Data: An Alternative to Manual Examination of Extracted Ion Chromatograms, *Anal. Chem.* 79 (2007) 4805–4813.

J.M. Büscher, D. Czernik, J.C. Ewald, U. Sauer, N. Zamboni, Cross-Platform Comparison of Methods for Quantitative Metabolomics of Primary Metabolism, *Anal. Chem.* 81 (2009) 2135–2143.

E.H. Turner, K. Lauterbach, H.R. Pugsley, V.R. Palmer, N.J. Dovichi, Detection of Green Fluorescent Protein in a Single Bacterium by Capillary Electrophoresis with Laser-Induced Fluorescence, *Anal. Chem.* 79 (2007) 778–781.

H. Zhang, X.-F. Li, X.C. Le, Differentiation and Detection of PDGF Isomers and Their Receptors by Tunable Aptamer Capillary Electrophoresis, *Anal. Chem.* 81 (2009) 7795–7800.

L.M. Ramsay, J.A. Dickerson, O. Dada, N.J. Dovichi, Femtomolar Concentration Detection Limit and Zeptomole Mass Detection Limit for Protein Separation by Capillary Isoelectric Focusing and Laser-Induced Fluorescence Detection, *Anal. Chem.* 81 (2009) 1741–1746.

E.V. Dose, G.A. Guiochon, High-resolution modeling of capillary zone electrophoresis and isotachopheresis, *Anal. Chem.* 63 (1991) 1063–1072.

Z. Liu, M. Zhang, J.B. Phillips, High-Speed Gas Chromatographic Analysis of a Simulated Process Stream using On-Column Thermal Desorption Modulation for Sample Preconcentration and Introduction, *J. Chromatogr. Sci.* 28 (1990) 567–571.

A.K. Boardman, S.C. McQuaide, C. Zhu, C.D. Whitmore, M.E. Lidstrom, N.J. Dovichi, Interface of an Array of Five Capillaries with an Array of One-Nanoliter Wells for High-Resolution Electrophoretic Analysis as an Approach to High-Throughput Chemical Cytometry, *Anal. Chem.* 80 (2008) 7631–7634.

D. Jacob, *Introduction to Atmospheric Chemistry*, 1st ed., Princeton University Press, 1999.

A.E. Sinha, J.L. Hope, B.J. Prazen, C.G. Fraga, E.J. Nilsson, R.E. Synovec, Multivariate selectivity as a metric for

evaluating comprehensive two-dimensional gas chromatography–time-of-flight mass spectrometry subjected to chemometric peak deconvolution, *J. Chromatogr. A.* 1056 (2004) 145–154.

A. Krishnan, L.J. Williams, A.R. McIntosh, H. Abdi, Partial Least Squares (PLS) methods for neuroimaging: A tutorial and review, *NeuroImage.* 56 (2011) 455–475.

K.E. Swearingen, W.P. Loomis, M. Zheng, B.T. Cookson, N.J. Dovichi, Proteomic Profiling of Lipopolysaccharide-Activated Macrophages by Isotope Coded Affinity Tagging, *J. Proteome Res.* 9 (2010) 2412–2421.

K. Sobhani, D.A. Michels, N.J. Dovichi, Sheath-Flow Cuvette for High-Sensitivity Laser-Induced Fluorescence Detection in Capillary Electrophoresis, *Appl. Spectrosc.* 61 (2007) 777–779.

P.J. Marriott, W. Khummueng, The Nomenclature of Comprehensive Two-Dimensional Gas Chromatography: Defining the Modulation Ratio (MR), *LCGC Eur.* (2009) 38–45.

Upper Critical Values of the F Distribution, (<http://www.itl.nist.gov/div898/handbook/eda/section3/eda3673.htm>).

S. Yang, J.S. Nadeau, E.M. Humstron Fulmer, J.C. Hoggard, M.E. Lidstrom, R.E. Synovec, chromatography mass spectrometry with chemometric analysis for determining ¹²C and ¹³C labeled contributions in metabolomics and ¹³C flux analysis, *J. Chromatogr. A.* 1240 (2012) 156–164.

Y.-K. Tzeng, C.-C. Chang, C.-N. Huang, C.-C. Wu, C.-C. Han, H.-C. Chang, Facile MALDI-MS Analysis of Neutral Glycans in NaOH-Doped Matrixes: Microwave-Assisted Deglycosylation and One-Step Purification with Diamond Nanoparticles, *Anal. Chem.* 80 (2008) 6809–6814.

H. Martens, M. Høy, B.M. Wise, R. Bro, P.B. Brockhoff, Pre-whitening of data by covariance-weighted pre-processing, *J. Chemom.* 17 (2003) 153–165.

Curriculum Vitae

Education

- 2014 Ph.D., Chemistry, University of Washington (Seattle), WA
- 2013 Master of Science, Chemistry, University of Washington (Seattle), WA
- 2008 Bachelor of Science, Chemistry, University of Washington (Seattle), WA
- 2006 Associates in Arts and Science (transfer program), Bellevue College, WA

Research Experience

- Research assistant, Synovec Research Group, University of Washington 2009-current.
- Research assistant, Dovichi labs research Group, University of Washington, 2009-2010.
- Undergraduate researcher, L. Dalton Research Group, University of Washington, 2007-2008.

Teaching

- University of Washington (Seattle): 2007-2013.
- Chemistry Teaching Assistant
- 100 Level: General Chemistry (chem142, chem152, 1 quarter as Lead Teaching Assistant)
- 300 Level: Quantitative analysis Laboratory (chem321)
- 400 Level: Bio-instrumental Analysis (chem428), Separation Techniques (chem429).

Research Summary and interests

In the field of analytical chemistry, analysis of complex chemical samples and the need for better separation of analytes has led to the development of more powerful instrumentation, such as multi-dimensional chromatography. With the introduction of multidimensional separation the size and complexity of the raw data has become more difficult to analyze. With the rise of more powerful computers, chemometrics has become faster and more robust in analyzing complex data such as information rich multidimensional chromatograms and glean useful chemical information.

My earliest (unpublished) work on data analysis and chemometrics was analysis of capillary electrophoresis with laser induced fluorescence (CE-LIF) electropherograms of human tissue samples, specifically analyzing multiple CE-LIF electropherograms for variance between healthy and various stages of cancer using software tools such as peak alignment, baseline correction and chemometric tools such as principle component analysis (PCA) to identify biomarkers for cancer at various stages. The ultimate goal was to quantify the variance of various stages of cancer via PCA and use the models as a more reliable means for prognosis.

Following my early work, my research went towards analyzing chromatograms from gas chromatograph coupled to a quadrupole mass spectrometer (GC-MS), specifically quantifying poorly resolved temporal analyte peaks (even in cases where two peaks are completely unresolved and appear as only a single peak). This led to becoming a Matlab programmer and constructing the temporal mass spectral ratio analysis (TMSRA) algorithm. The principle behind TMSRA was identifying (at least three) pure m/z channels for quantifying a targeted analyte and approximate its pure mass spectrum. This algorithm allows users to use m/z scan mode from the mass spectrometer while enjoying both the benefits from scan mode such as the ability to identify compounds from their mass spectra and the benefits from selective ion monitoring (SIM) such as greater signal to noise of selective ion monitoring without knowing *a priori* which m/z would be selective in the separation.

My recent work has mostly been on analysis of rocket propellant fuel (RP fuels) using chromatograms obtained from a two dimensional gas chromatograph coupled to time-of-flight mass spectrometer (GC \times GC-TOF). Unfortunately, I am unable to disclose any more information concerning these fuels at this time.

In conjunction with my aforementioned graduate research on RP fuels, I taught myself the programming language C, in order to improve my Matlab programming skills and as a programmer in general. Programming skills in my research were extremely helpful in implementing and testing chemometric tools as well as using them with in-house/custom methods, especially for software with poorly documented data formats. I also used my Matlab programming skills for both writing script files (for automation purposes) and developing tools as the need arises, such as calculating peak width at half height, simulating peak overlap, data binning in multiple dimensions or formats (standard or ribbons) and construction of figures. I strive to teach myself more programming languages and skills as they appear needed or useful in the future.

Publications

Primary Author

Targeted mass spectral ratio analysis: a new tool for gas chromatography-mass spectrometry, Benjamin Kehimkar, Jamin C. Hoggard, Jeremy S. Nadeau, Robert E. Synovec, *Talanta* (2013) 103, 267-275

Correlation of rocket propulsion fuel properties with chemical composition using comprehensive two-dimensional gas chromatography with time-of-flight mass spectrometry followed by partial least squares regression analysis, Benjamin Kehimkar, Jamin C. Hoggard, Luke C. Marney, Matthew C. Billingsley, Carlos G. Fraga, Thomas J. Bruno, Robert E. Synovec, J. Chromatogr A, (2014) 1327, 132-140

Secondary Author

Quantification of green fluorescent protein in cellular supernatant by capillary electrophoresis with laser-induced fluorescence detection for measurement of cell death, Kristian E Swearingen, Wendy P. Loomis, Benjamin Kehimkar, Brad T. Cookson, Norman J. Dovichi, Talanta (2010) 81, 948-953

Review of chemometric analysis techniques for comprehensive two dimensional separation data, Karisa M. Pierce , Benjamin Kehimkar, Luke C. Marney, Jamin C. Hoggard, Robert E. Synovec, J. Chromatogr A (2012) 1255, 3-11

Other Publications

Novel Instrumentation in One and Two-Dimensional Gas Chromatography, Brian D. Fitz, Benjamin Kehimkar, Jamin C. Hoggard, Robert E. Synovec ,Center for Process Analysis & Control, University of Washington, Seattle WA, May 6, 2013.