# Comprehensive Two-Dimensional Gas Chromatography and Chemometrics for the Analysis of Complex Mixtures

By

Carlos Gerardo Fraga

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

Doctor of Philosophy

University of Washington

2000

Program Authorized to Offer Degree: Chemistry

DISTRIBUTION STATEMENT A
Approved for Public Release
Distribution Unlimited

University of Washington

Abstract

Comprehensive Two-Dimensional Gas Chromatography and Chemometrics for the Analysis of Complex Mixtures

By Carlos Gerardo Fraga

Chairperson of Supervisory Committee:
Associate Professor Robert E. Synovec
Department of Chemistry

The main goal of this dissertation is to enhance the analysis capabilities of comprehensive two-dimensional (2-D) gas chromatography (GC x GC) by applying chemometrics to extract component signals in the presence of significant interference, noise, or both. This is accomplished through three projects.

The first project quantifies the theoretical analysis enhancement provided by the use of the generalized rank annihilation method (GRAM) for the analysis of component signals, i.e., peaks, that are unresolved in GC x GC separations. Monte Carlo simulations, modeled after real GC x GC data, are used to determine the conditions where the use of GRAM results in the successful analysis of unresolved peaks. This information is then used on Monte Carlo simulations of 2-D chromatograms. Ultimately, it is determined that the use of GRAM increases the average number of analyzable peaks by a factor of two for 2-D chromatograms that are 67 percent occupied by randomly distributed peaks. The use of GRAM should increase the number of analyzable peaks for all forms of comprehensive 2-D separations.

The second project extends the use of GRAM analysis to the GC x GC separation of real world complex mixtures by the use of the standard addition method and bilinear data alignment. Standard addition and bilinear data alignment are used to correct peak width variations and retention time shifts that would otherwise negate the utility of GRAM. The use of GRAM, in part, reduces the separation time of three isomers in jet fuel by a factor of five compared to a single column GC separation. The

20010525 057

use of bilinear data alignment improves GRAM quantification accuracy and precision by a factor of four. In addition, 2-D bilinear data alignment is introduced.

The last project demonstrates the signal enhancement provided by GRAM. It substantially improves the quantitative precision and accuracy of GC x GC compared to peak integration. In the case of a 2.7 part per million by mass propylbenzene sample, GRAM analysis is 2.6 times more precise and 4.2 times more accurate than integration. The GC x GC limit of detection is also lowered by a factor of three.

University of Washington

Graduate School

This is to certify that I have examined this copy of a doctoral dissertation by

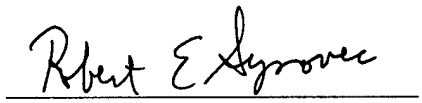
Carlos Gerardo Fraga

and have found that it is complete and satisfactory in all respects,

and that any and all revisions required by the final

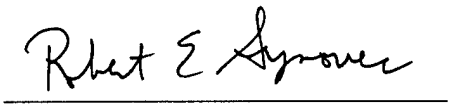
examining committee have been made.

Chair of Supervisory Committee:

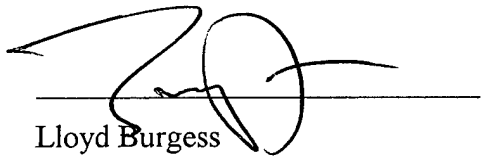


Robert E. Synovec

Reading Committee:



Robert E. Synovec



Lloyd Burgess

Younan Xia

Date: August 15, 2000

Doctoral Dissertation

In presenting this dissertation in partial fulfillment of the requirements for the Doctoral degree at the University of Washington, I agree that the Library shall make its copies freely available for inspection. I further agree that extensive copying of this dissertation is allowable only for scholarly purposes, consistent with "fair use" as prescribed in the U.S. Copyright Law. Requests for copying or reproduction of this dissertation may be referred to Bell and Howell Information and Learning, 300 North Zeeb Road, Ann Arbor, MI 48106-1346, to whom the author has granted "the right to reproduce and sell (a) copies of the manuscript in microform and/or (b) printed copies of the manuscript made from microform."

Signature 

Date 15 AUG 2006

# TABLE OF CONTENTS

# List of Figures

# List of Tables

# List of Abbreviations

| | |
|---|---|
| **1-D** | One Dimensional |
| **2-D** | Two Dimensional |
| **CE** | Capillary Electrophoresis |
| **FID** | Flame Ionization Detector |
| **GC** | Gas Chromatography |
| **GC x GC** | Comprehensive Two Dimensional Chromatography |
| **GRAM** | Generalized Rank Annihilation Method |
| **IPB** | Isopropylbenzene |
| **ITP** | Isotachophoresis |
| **IEC** | Ion Exchange Chromatography |
| **LC** | Liquid Chromatography |
| **LOD** | Limit of Detection |
| **NA** | Not Applicable |
| **OET** | *Ortho*-ethyltoluene |
| **PB** | Propylbenzene |
| **RSD** | Relative Standard Deviation |
| **RPLC** | Reverse Phase Liquid Chromatography |
| **SEC** | Size Exclusion Chromatography |
| **SFC** | Supercritical Fluid Chromatography |
| **SMO** | Statistical Model of Overlap |
| **SVD** | Singular Value Decomposition |
| **S/N** | Signal-to-Noise |
| **TBB** | *Tert*-butylbenzene |
| **TMB** | Trimethylbenzene |

# Acknowledgements

The author wishes to express sincere gratitude to Professor Synovec for his guidance during the author's graduate work. The author likes to thank the members of the Synovec Research group, past and present, for their support. He also thanks the United Sates Air Force for giving him the chance to fulfill a dream of obtaining a Ph.D. The author is extremely grateful to his parents who provided a nourishing and loving environment during his childhood years. Lastly, the author thanks his wife, Trish, for her support and love.

# Chapter 1: Introduction

Comprehensive two-dimensional (2-D) column techniques are perfect for the separation and analysis of complex mixtures. One-dimensional (1-D) column techniques such as single column gas chromatography (GC) lack the separation power for isolating complex mixture components. They produce a 1-D retention space or line that does not have the needed capacity to sufficiently separate many components. Hence, numerous components remain mixed and are not analyzed. In contrast, comprehensive 2-D techniques are specifically designed to produce a planar separation space that has more room to distribute components. A comprehensive 2-D separation is capable of separating more components of a complex mixture in less time than a 1-D separation. However, not all components are adequately separated or at high enough concentration for reliable quantitative analysis. In this dissertation, the analysis capabilities of comprehensive 2-D GC is enhanced by applying chemometrics to extract component signals in the presence of significant interference, noise, or both.

## 1.1 The Separation of Complex Mixtures

A mixture is complex if it contains a multitude of chemical components from different chemical families. Many chemical mixtures of interest are complex. For instance, biological and environmental samples are well known for containing a multitude of different constituents at widely different concentrations [Khaledi, 1998]. Many food and fragrance products are highly complex [James, 1995]. Petroleum products are known to contain hundreds of thousands of components [Bertsch, 1999]. The chemical analysis of complex mixtures requires the separation of complex mixtures into individual constituents or groups. Separation column techniques such as GC and capillary electrophoresis (CE) are used to separate complex mixtures [Khaledi, 1998; Adlard, 1995]. The comprehensive 2-D versions of separation column techniques are discussed in this dissertation.

### 1.1.1 1-D Peak Overlap

The word 'peak' in this dissertation refers to the measured signal of a single component. This stems from the fact that a single component eluting from a separation column into a univariate detector produces a peak shaped signal. For a mixture subjected to a single column separation, peaks are distributed along a line. This line of peaks is referred to as a chromatogram for chromatographic separations. Regardless of the separation mechanism, not all peaks are appreciably separated from one another in complex 1-D separations. Users of single column separations are aware of peak overlap, in part, because mathematical statistics has been used to reveal its prevalence.

Several authors have used a statistical approach to theoretically quantify peak overlap in 1-D chromatographic separations [Rosenthal, 1982; Nagels et al., 1983; Felinger, 1998]. The statistical model of overlap (SMO) developed by Davis and Giddings is one of the most notable theories describing peak overlap in multicomponent chromatograms [Davis and Giddings, 1983]. SMO is based on the assumption that peaks are spaced randomly in complex chromatograms. According to SMO, more separation space is needed than one would intuitively expect necessary. For instance, for 50 peaks randomly distributed on a line capable of fitting 100 peaks side-by-side, SMO predicts that approximately 18 peaks are adequately separated, i.e., they each have a peak resolution of at least one. In fact, a line capable of fitting 1000 peaks would be necessary to provide a 90 % probability that a given peak from a group of 50 will be adequately separated. Achieving a line capable of holding 1000 peaks with a 1-D separation would dramatically increase separation time. In addition, the later eluting peaks would be so broaden by the separation process that peak detection would be diminished. SMO has been confirmed using real and realistically simulated chromatograms [Delinger and Davis, 1990; Samuel and Davis, 1999]. Later versions of SMO have been modified to take into account peak density and height variations found in real 1-D separations [Davis, 1996; Davis, 1994]. Whatever version is used,

SMO supports the notion that 1-D separations lack the separation space needed for complex mixture separations.

### 1.1.2 Planar Separations

Planar separations overcome the lack of separation space associated with 1-D separations of complex mixtures. Planar separations literally have a planar space where peaks can distribute themselves. Obviously, more peaks can fit into a plane then on any of the lines bordering the plane. For that reason, planar separation techniques such as 2-D gel electrophoresis are widely used for separating complex biological samples [Dunbar, 1987; Young and Tracy, 1995]. In planar methods, the sequential application of two different separation mechanisms takes place on or within a support structure having a rectangular configuration [Giddings, 1991]. First, a sample spot is placed on one corner and a force is applied that drives the sample as it separates. Following this a second driving force is applied perpendicular to the initial sample flow. This drives the sample components away from the support's edge into the body of the support where separation is further enhanced. In 2-D gel electrophoresis, isoelectric focusing is used along one direction and gel electrophoresis along the second. Unlike planar techniques, column techniques are inherently 1-D. They must be deliberately configured to produce a 2-D separation space. While such a task is not necessarily straightforward, the faster analysis times and better quantification provided by column techniques makes the endeavor well worth it [Bushey and Jorgenson, 1990a; Khaledi, 1998].

### 1.1.3 Comprehensive 2-D Separation Design

Producing a 2-D separation space with separation columns requires coupling two columns of different selectivities such that all components of a sample are subject to two different separations without either separation nullifying the other [Giddings, 1984; Giddings, 1987]. Figure 1.1 is a schematic for a comprehensive 2-D GC or GC x GC analyzer. The GC x GC analyzer is designed to produce a 2-D separation space

using two GC columns. It provides a good example for all other automated comprehensive 2-D column systems. Figure 1.1 essentially depicts two GC columns coupled by a diaphragm valve. The first column is usually a large bore, non-polar column that separates components primarily based on volatility. The second column is a high-speed, narrower-bore polar column that separates components based on polarity. Its outlet is connected to a flame ionization detector (FID). The diaphragm valve repeatability diverts small portions or "plugs" of the first column eluent to the second column. Most GC x GC experiments use thermal focusing, either heating or cooling, to produce sample plugs [Lee et al., 2000]. In either case, each plug sent to the second column generates a high-speed, secondary chromatogram. Each secondary separation needs to be fast enough so that at least four secondary chromatograms are generated across the width of a peak eluting from the first column [Murphy et al., 1998]. Anything less than 4 might completely nullify the separation achieved by the first column. A series of secondary chromatograms are generated across the span of a first column separation. Arranging the secondary chromatograms side-by-side results in a matrix of data that is a 2-D chromatogram with peaks dispersed over a 2-D space.

Figures 1.2A, B, and C provide a visual example of the data processing required to generate 2-D chromatograms. Figure 1.2A depicts a series of secondary chromatograms generated from two coeluting, first column peaks. In this example, the first column successfully separated the two peaks from others but not from themselves. However, as depicted in Figure 1.2B, each secondary separation is able to fully separate each peak. In Figure 1.2C, the secondary chromatograms are aligned side-by-side producing a matrix of data visualized as two "3-D" peaks. When depicted in a contour plot, the peaks appear as two elliptical zones.

When a GC x GC run is performed on an entire complex mixture, a 2-D chromatogram is generated that can reveal components otherwise hidden in a 1-D separation. Figures 1.3A and 1.3B depict the first column chromatogram and the 2-D chromatogram for a jet fuel sample. The comprehensive 2-D separation of jet fuel (Figure 1.3B) easily draws out polar compounds like aromatics and napthalenes from the long horizontal band of alkane and alkene compounds [Frysinger and Gaines, 1999;

Frysinger et al., 1999]. In addition, the polar compounds cluster into distinct chemical classes [Beens et al., 2000].

### 1.1.4 Sample Dimensionality

The distribution of jet fuel components in Figure 1.3B depends largely on the jet fuel sample dimensionality. Sample dimensionality, s, is the intrinsic property of an analytical sample that determines its amenability to a multidimensional separation. The parameter s is defined as the number of independent variables that must be specified to identify the components of a sample [Giddings, 1995]. It is assumed that the properties of the components (e.g., chromatographic distribution coefficient, K ) vary in a systematic manner with the s variables. Generally, the variables are structural factors that determine molecular identity. For instance, if a sample is composed entirely of straight chain alkanes, then the sample components can be fully characterized in terms of one s variable, which can be the carbon number or molecular weight [Giddings, 1995]. A comprehensive 2-D separation of such a sample would not provide any additional separation of the sample components than a 1-D separation. The sample components would essentially fall on a line. However, if s is increased to two by including mono-chlorinated straight chain alkanes into the sample, then the sample components would disperse into the 2-D space. The sample dimensionality correlates with the number of compound families [Felinger, 1998]. In order to utilize the extra separation space of a comprehensive 2-D separation, a sample needs to have a dimensionality of 2 or more. The larger the parameter s, the more peaks will be dispersed throughout the 2-D separation space.

### 1.1.5 Separation Orthogonality

The separation mechanism on one column or dimension of a comprehensive 2-D separation needs to be completely independent or orthogonal from the separation mechanism on the other column to effectively use the 2-D separation space. Correlation of peak retention in two dimensions reduces the available retention space to

a select region in the 2-D space [Liu et al., 1995]. Hence, correlation inhibits the use of the expanded space provided by a comprehensive 2-D separation. In fact, complete correlation will have all peaks distributed along a diagonal [Venkatramani et al., 1996]. In isothermal GC x GC, correlation between the two GC columns is present but not entirely detrimental. For instance, GC x GC peaks from a chemical family distribute themselves on a diagonal, but are adequately separated from other families. Correlation helps to distinguish different compound classes. However, the 2-D space has empty and unusable areas. Temperature program GC x GC has been used to remove the correlation of peak retention in two dimensions [Liu et al., 1995]. As a result the 2-D separation space is efficiently utilized and analysis times are shortened.

### 1.1.6 GC x GC Attributes and Uses

GC x GC has two main attributes over traditional single-column GC. First, its 2-D separation space enhances peak capacity, which represents the maximum number of peaks that can fit in a separation space. The peak capacity of a GC x GC system is theoretically the product of the peak capacity of each GC column [Liu et al., 1995; Bertsch, 1990; Giddings, 1984; Giddings, 1987]. The enhanced peak capacity of a properly tuned GC x GC system can be used to separate complex samples in less time than if either of the GC x GC columns were used individually. Secondly, GC x GC provides more analyte identification capability than obtained with a single column GC. GC retention time is affected by a compound's polarity and volatility. Identifying unknown compounds is difficult with single-column GC because many combinations of volatility and polarity lead to the same retention time. GC x GC analysis results in a map of the sample in which each compound is described by two retention times. In many cases, this results in a unique signal that indicates the compounds' relative polarity and volatility. Along those same lines, a GC x GC separation groups sample components into chemical classes forming distinct patterns that simplify characterization and class quantification. The location of an analyte peak in a GC x

GC chromatogram can be a powerful, selective identification tool. The above attributes make GC x GC ideal for the analysis of complex mixtures.

GC x GC has been used to analyze complex mixtures such as petrochemical products [Blomberg et al., 1997; Beens et al., 1998; Ledford Jr. et al., 1996; Kinghorn and Marriott, 1999; Beens et al., 2000], essential oils [Dimandja et al., 2000], pesticides extracted from human serum [Liu et al., 1994], polynuclear aromatic hydrocarbons in used engine oil [Ledford Jr. et al., 1996], and aromatics and oxygenates in gasoline [Frysinger et al., 1999; Frysinger and Gaines, 2000].

### 1.1.7 Other Comprehensive 2-D Column Techniques

GC x GC is the most widely used comprehensive 2-D column technique but by no means the only one. Comprehensive 2-D liquid chromatography (LC x LC) of protein mixtures has be demonstrated by coupling two LC techniques such as ion-exchange chromatography (IEC) with size exclusion chromatography (SEC) [Bushey and Jorgenson, 1990a], IEC with reversed phase liquid chromatography (RPLC)[Opiteck et al., 1997], and SEC with RPLC [Opiteck et al., 1998b]. Comprehensive LC-GC (LC x GC) has also been demonstrated for the analysis of volatile organic compounds in water [Quigley et al., 2000]. Comprehensive supercritical fluid chromatography-GC (SFC x GC) has been used for the analysis of polycyclic aromatic hydrocarbons [Liu et al., 1993]. Comprehensive isotachophoresis-CE (ITP x CE) has been used on a mixture of angiotensins [Chen and Lee, 2000]. Comprehensive LC-CE (LC x CE) has been demonstrated for the analysis of complex biological mixtures by coupling SEC with CE and RPLC with CE [Bushey and Jorgenson, 1990b; Hooker and Jorgenson, 1997; Moore and Jorgenson, 1995; Larmann et al., 1993; Lemmo and Jorgenson, 1993].

## 1.2 Chemometrics for Comprehensive 2-D Data

Chemometrics is the use of statistical and mathematical methods to analyze chemical data [Malinowski, 1991; Beebe et al., 1998]. Chemists employ chemometrics

to get useful information from chemical data. Rank annihilation and bilinear data alignment are chemometric methods specifically designed for the type of data produced by comprehensive 2-D separations. They are thoroughly studied in this dissertation to determine the benefits and limits of applying chemometrics to comprehensive 2-D data. Both methods are discussed in the succeeding sections.

### 1.2.1 Bilinear Data

Ideally, the noise-free portion of a peak produced by a comprehensive 2-D system is bilinear. This means the peak, in the form of a data matrix **M**, can be decomposed into the product of three terms:

$$\mathbf{M} = \mathbf{x}c\mathbf{y}^T + \mathbf{E} \qquad (1.1)$$

where $\mathbf{x}$ and $\mathbf{y}^T$ are vectors representing the peak's pure elution profiles on each column time axis, c is a scalar proportional to the analyte concentration, and **E** is a matrix of noise values. The superscript 'T' denotes transpose. From here on, variables representing matrices are uppercase bold letters, variables representing vectors are lowercase bold letters and variables representing scalers are lowercase plain letters. Figure 1.4 provides a pictorial representation of equation 1.1. For a multiple number of overlapped peaks, p, **M** can be expressed as the linear sum of each peak's bilinear signal plus noise:

$$\mathbf{M} = \left( \sum_{i=1}^{p} \mathbf{x}_i \, c_i \, \mathbf{y}_i^T \right) + \mathbf{E}, \text{ or}$$

$$\mathbf{M} = \mathbf{X}\mathbf{C}\mathbf{Y}^T + \mathbf{E} \qquad (1.2)$$

The matrix **X** has p columns where each column is the pure elution profile **x** of a peak, $\mathbf{Y}^T$ has p rows where each row is the pure elution profile **y** of a peak, and the diagonal matrix **C** has each of its p diagonal elements equal to a peak's c term.

### 1.2.2 Rank Annihilation

Rank annihilation is a powerful chemometric method for the quantitative analysis of unresolved bilinear signals. Rank annihilation was originally developed by Ho and co-workers [Ho et al., 1978; Ho et al., 1980; Ho et al., 1981] for the quantification of single components in multicomponent mixtures. The concept of rank annihilation is as follows.

Rank is a property common to all numerical matrices. The rank of a matrix is an indication of correlation among data values in a matrix. For instance, a rank of one indicates that the data in a matrix is highly correlated and can be reduced to two vectors. The product of the two vectors completely reproduces the matrix. On those same lines a calibration matrix, $\mathbf{N}_k$, containing the bilinear signal of a single component, k, has a rank of 1 in the absence of noise. While a sample matrix, **M**, containing a total of p bilinear signals, including the bilinear signal of a given amount of component k, has a rank of p in the absence of noise. If the correct amount of $\mathbf{N}_k$ is subtracted from **M**, the rank of the resultant matrix should be one less than the original rank of **M**, Rank $(\mathbf{M}-d\mathbf{N}_k) = p-1$. The d term corresponds to the relative concentration of component k in **M**.

The original form of rank annihilation is computationally intensive because it involves an iterative process to determine the d term. Lorber modified rank annihilation into a non-iterative procedure where the d term is found by solving a generalized eigenvalue problem [Lorber, 1984]. Sánchez and Kowalski later developed the generalized rank annihilation method (GRAM) which extended Lorber's rank annihilation method to include situations where the calibration matrix, **N**, consists of more than one component [Sánchez and Kowalski, 1986]. In those situations, several d terms are generated. Hence, GRAM provides the resolved bilinear signal for

each component common to **M** and **N** in order to identify the d term corresponding to a specific component. Figure 1.5 provides a visual representation of GRAM analysis.

In Figure 1.5, both **M** and **N** contain the overlapped bilinear signals for three components. The bilinear signals are modeled after those produced by a comprehensive 2-D system. GRAM looks for the bilinear data that is common between **M** and **N**. Hence, it produces the matrices **X** and $\mathbf{Y}^T$, which contain the pure elution profiles at unit concentration for each component. In addition, GRAM generates $\mathbf{C}_M$ and $\mathbf{C}_N$. Dividing the diagonal elements of $\mathbf{C}_M$ by those of $\mathbf{C}_N$ produces the relative concentration of each component or d term. The resolved, "noise-free" signal for a component in **M** or **N** can be reconstructed by multiplying its corresponding elution profiles and diagonal element. A key attribute of GRAM over other analysis methods is that **M** can contain overlapped components not present in **N**.

### 1.2.3 GRAM Algorithm

Several algorithms for GRAM exist [Sánchez and Kowalski, 1986; Faber et al., 1994c; Li et al., 1992; Wilson et al., 1989]. A GRAM algorithm based on the standard eigenvalue method is explained below [Faber et al., 1994c; Bruckner, 1998].

The first step involves singular value decomposition (SVD) of the addition matrix, which is the sum of the sample matrix **M** and the standard matrix **N**.

$$(\mathbf{M} + \mathbf{N}) = \mathbf{USV}^T \qquad (1.3)$$

In this decomposition **U** and **V** are orthogonal matrices and **S** is a diagonal matrix. SVD is an accepted method for decomposing a matrix [Dongarra et al., 1979]. Using the addition matrix ensures that all components present in **M** and **N** are modeled. In addition, the signal-to-noise (S/N) for the components present in both **M** and **N** is improved.

In the second step, **U**, **S**, and **V** are truncated according to the number of significant factors, or pseudorank, of the addition matrix. Several methods exist for

determining the pseudorank of a matrix [Faber and Kowalski, 1997; Faber et al., 1994a; Faber et al., 1994b; Malinowski, 1991]. Ideally, the number of chemical components equals the pseudorank. Hence, if the addition matrix has four chemical components then only the first four columns of **U** and **V** and the first four diagonal elements of **S** are kept. These significant factors accurately reconstruct the combined signal of all chemical components. Each component's unique signal is determined in the fourth step of the GRAM algorithm. The remaining factors describe measurement error, or noise.

The third step involves solving the eigenvalue problem that contains **N** and the truncated SVD components from the addition matrix, **M** + **N** [Bruckner, 1998].

$$\left(\overline{\mathbf{S}}^{-1}\overline{\mathbf{U}}^{T}\mathbf{N}\overline{\mathbf{V}}\right)\mathbf{T} = \mathbf{T}\Pi \qquad (1.4)$$

In eq 1.4, **T** is the resulting matrix of eigenvectors, and $\Pi$ is the associated diagonal matrix containing the eigenvalues. An overbar denotes truncation and '-1' denotes inverse. Eq 1.4 can be viewed as a procedure for "dividing" the unique signals of each chemical component in **N** by the unique component signals in **M** + **N** [Prazen, 1998]. The mathematical relationship shown below supports the pervious statement.

$$\mathbf{N}/(\mathbf{M}+\mathbf{N}) = \mathbf{N}/\mathbf{U}\mathbf{S}\mathbf{V}^{T} = (\mathbf{S}^{-1}\mathbf{U}^{T}\mathbf{N}\mathbf{V}) \qquad (1.5)$$

Dividing the signals in **N** by those in **M** + **N** is equivalent to performing a least squares fit of the unique signals in **N** by those in **M** + **N** [Mathworks, 1998]. In such a case, the concentration of each component in **N** relative to its concentrations in **M** + **N** can be determined. The significance of this observation is addressed in section 1.3.3.

In the fourth step, the analyte concentrations in the sample, $\mathbf{c}_M$, relative to those in the standard, $\mathbf{c}_N$, are determined by solving eq 1.6 for $\mathbf{c}_N/\mathbf{c}_M$.

$$\text{Diagonal}\ (\Pi) = \mathbf{c}_N/(\mathbf{c}_M + \mathbf{c}_N) \qquad (1.6)$$

The pure elution profiles, contained in the matrices **X** for one column and **Y** for the other column, are determined for all components common to both the sample **M** and standard **N** using the eigenvectors and the decomposed components from SVD.

$$\mathbf{X} = \overline{\mathbf{U}}\overline{\mathbf{S}}\mathbf{T} \tag{1.7}$$

$$\mathbf{Y} = \overline{\mathbf{V}}\left(\mathbf{T}^{-1}\right)^{\mathrm{T}} \tag{1.8}$$

The matrix of eigenvectors **T** is really a transformation matrix that rotates the abstract SVD vectors of the addition matrix **M** + **N** into physically meaningful information, i.e., elution profiles.

### 1.2.4 GRAM Applications and Requirements

GRAM has been used to resolve and quantify overlapped signals from a variety of instruments producing bilinear data [Sánchez et al., 1987; Sánchez and Kowalski, 1986; Antalek and Windig, 1996; Poe and Rutan, 1993; Bijlsma et al., 1999; Prazen et al., 1998; Prazen et al., 1999a; Prazen et al., 1999b; Windig and Antalek, 1999]. In terms of comprehensive 2-D techniques, GRAM has been used to successfully quantify GC x GC peaks in modified white gas that were so overlapped quantification by peak height or integration could not even be attempted [Bruckner et al., 1998]. In all the above applications, the success of GRAM depends on adherence to certain prerequisites.

First, the sample data, **M**, and standard data, **N**, must each be bilinear. In terms of GC x GC, this is achieved as long as an analyte's peak shape, width, and retention time in every secondary chromatogram are reproducible during a GC x GC run (see Figure 1.2). Both **M** and **N** are obtained from individual GC x GC runs. Secondly, each analyte's signal in **M** and **N** must be linearly independent on both dimensions. Hence, overlapped GC x GC peaks must each have a unique retention time on each GC column. Thirdly, analytes cannot perfectly covary in concentration from **M** to **N**.

Concentration covariance occurs when the ratio of analyte concentrations in **M** is identical to the ratio of analyte concentrations in **N**. Concentration covariance will cause GRAM to produce inaccurate bilinear profiles, e.g. odd looking elution profiles. However, concentration covariance will not affect quantification accuracy [Sánchez and Kowalski, 1990]. Finally, **M** and **N** together must be considered trilinear. That is to say, the bilinear signals for analytes in common between **M** and **N** must be the same in both **M** and **N** except in terms of signal intensity. In GC x GC, this is achieved when a GC x GC peak's retention time on both column dimensions remains constant from one GC x GC run to the next. Variation in retention times among runs is usually cited for causing problems in the GRAM analysis of bilinear chromatographic data [Ramos et al., 1987; Poe and Rutan, 1993; Prazen et al., 1998]. However, an objective retention time alignment algorithm can be used to correct run-to-run retention time shifts as described next.

### 1.2.5 Bilinear Data Alignment

A bilinear data alignment algorithm developed by Prazen et. al. is designed to align two bilinear data matrices along one dimension prior to GRAM analysis [Prazen et al., 1998]. It has been used to correct retention time shifts along the chromatographic axis of data produced by chromatographic-spectrometric methods such as LC with UV-vis detection [Prazen et al., 1998] and GC with mass spectrometric detection [Prazen et al., 1999b]. It has also been applied to GC x GC data to correct retention time shifts on the first column even though the retention times of the GC x GC data were reproducible enough that retention time alignment was not needed prior to GRAM analysis [Prazen et al., 1999b]. However, in the GC x GC separation of real world complex mixtures, retention time alignment along first column axis is needed prior to GRAM analysis as addressed in section 1.3.2. The theory behind the alignment method is now discussed.

Figure 1.6A depicts the simulated bilinear data matrices **M** and **N** obtained from a 2-D chromatographic instrument. Each matrix is 60-by-60 data points in size.

Both **M** and **N** contain the signals for the same two components. The matrices are stacked to represent a 120-by-60 matrix produced if **M** and **N** where augmented in such away that the second column axis is made twice as long. The rank of the augmented matrix is equal to the number of different components (i.e., two in Figure 1.6A) when the component signals in **M** and **N** are perfectly aligned along the chromatographic axis. However, the rank of the augmented matrix will be greater than the number of components if the component signals between **M** and **N** are shifted on the first column axis. In this example, the augmented matrix's rank could increase to a maximum value of four. The alignment algorithm shifts the **M** matrix until the rank of the augmented matrix reaches a minimum value of two. The simulated data matrices **M** and **N** are purely bilinear and hence have no noise. In the case of real data, the presence of noise requires a slightly different approach for finding the correct shift.

First, the pseudorank of the **M** matrix, which in a real scenario has noise, is estimated (see section 1.2.3). In the simulated example shown in Figure 1.6A, the pseudorank of **M** equals its rank because **M** is absent of noise.

Second, SVD is performed on the augmented matrix, $\left[\frac{\mathbf{M}}{\mathbf{N}}\right]$, which is created by stacking **M** and **N** so that the second column axis is twice the length of that in **M** or **N**. SVD of $\left[\frac{\mathbf{M}}{\mathbf{N}}\right]$ leads to a vector containing the positive square roots of the eigenvalues, **s**, for the cross product or covariance of the augmented matrix [Prazen et al., 1998].

$$\mathbf{s}^2 = \mathbf{u}^{\mathrm{T}}\left(\left[\frac{\mathbf{M}}{\mathbf{N}}\right]\left[\frac{\mathbf{M}}{\mathbf{N}}\right]\right)^{\mathrm{T}}\mathbf{u} \qquad (1.9)$$

Third, the percent residual variance is calculated by dividing the sum of the eigenvalues beyond the pseudorank of **M** by the sum of all the eigenvalues and then multiplying by 100 and a term related to the degrees of freedom. The previous three steps are performed each time the data matrix **M** is shifted by one data point along the

first column axis. Ultimately, an alignment profile is obtained by plotting the percent residual variance vs. number of shifted data points. The profile's minimum indicates the direction and number of data points **M** is shifted from **N** along the first column axis. Figure 1.6B depicts the percent residual variance plot for **M** and **N** depicted in Figure 1.6A. The signals in **M** are shifted five data points to the "right" of those in **N**. In Figure 1.6B, the percent residual variance equals zero at the profile's minimum. At his point, the pseudorank of $\left\lceil \frac{\mathbf{M}}{\mathbf{N}} \right\rceil$ is two and all the data variance in $\left\lceil \frac{\mathbf{M}}{\mathbf{N}} \right\rceil$ is contained in the first two eigenvalues of **s**. In the presence of noise, such as with real data, the percent residual variance reaches a minimum value determined by the data variance originating from noise.

## 1.3 Hypotheses

As stated in the introductory paragraph, this dissertation claims to enhance the separation and analysis capabilities of comprehensive 2-D GC by applying chemometrics to extract component signals in the presence of either significant interference, noise, or both. The material covered so far provides fundamental information that will be used to build hypotheses that when proven correct, substantiate this dissertation's claim. Several hypotheses are made below. They are proven in the subsequent chapters.

### 1.3.1 Chapter 2 Hypotheses

**GRAM analysis increases the total number of peaks that can be analyzed in comprehensive 2-D separations, such as GC x GC, by quantifying peaks that are not fully resolved for quantification by peak integration or height.**

Section 1.2.4 mentioned the successful use of GRAM to quantify GC x GC peaks that could not be accurately quantified by peak integration or peak height because of significant peak overlap. If the assumption of random peak distribution, used by the SMO in section 1.1.1, is applied to comprehensive 2-D separations of

complex mixtures, then numerous peaks in comprehensive 2-D separations will be unresolved. GRAM should be able to quantify a large fraction of those unresolved peaks and hence increase the total number of peaks that can be analyzed.

**The use of data point interpolation as part of the objective retention time alignment algorithm further increases the number of analyzable peaks by improving the quantification accuracy and precision of GRAM analysis for overlapped peaks.**

The objective bilinear data alignment algorithm described in section 1.2.5 corrects shifts between two data matrices to the nearest data point. Hence, if the shift is less than a data point, then the alignment algorithm inaccurately estimates the shift. Minimizing the error in shift estimation should improve the GRAM analysis of overlapped peaks by more accurately correcting shifts so that adherence to the GRAM prerequisite of stable retention times is more closely met (see Section 1.2.4).

### 1.3.2 Chapter 3 Hypotheses

**The use of standard addition and data alignment prior to GRAM analysis permits the successful deconvolution and quantification of overlapped GC x GC peaks in real world analytical applications.**

The hypotheses in section 1.3.1 and the work performed by Bruckner et. al. [Bruckner et al., 1998] focus on GRAM's ability to analyze overlapped peaks from comprehensive 2-D separations. In order to utilize GRAM for the analysis of overlapped peaks, the prerequisites listed in section 1.2.4 must be met. In the pioneering work performed by Bruckner and co-workers, all prerequisites were obtained without additional procedures such as retention time alignment [Prazen et al., 1999b]. However, in the real world analysis of complex mixtures, the GRAM prerequisite of stable peak profiles between runs is not initially obtained from the instrument output. This is because of matrix effects and significant run-to-run retention time variation. The use of standard addition should eliminate matrix effects.

The use of retention time alignment should correct retention time shifts, making a noticeable improvement in GRAM analysis.

**The objective retention time alignment algorithm can be modified to correct run-to-run retention time shifts on both dimensions of a comprehensive 2-D separation.**

The objective retention time alignment algorithm developed by Prazen and co-workers (see section 1.2.5) is designed to correct retention time shifts on one dimension of bilinear data. Retention time alignment on the other dimension is not needed because of the high signal reproducibility achieved by spectrometric measurements or fast second column separations. While the second column separations of GC x GC are highly reproducible, nearly as much as spectrometric measurements, the need for retention time correction on the second dimension is not normally needed. However, significant run-to-run retention time variations on the second dimension can occur in GC x GC experiments that span longer periods of time. The retention time alignment algorithm can be made to fix retention time shifts on both dimensions using a simple algorithm modification.

### 1.3.3 Chapter 4 Hypothesis

**GRAM analysis improves the precision, accuracy, and limit of detection (LOD) of GC x GC.**

GRAM's ability to quantify unresolved bilinear signal has been extensively demonstrated in literature (see section 1.2.4). However, its ability to filter noise and achieve good quantification accuracy and precision for low-level signals has not been reported in literature. In the area of chromatography, the typical method to quantify signal intensity is through signal integration [Felinger, 1998; Braithwaite and Smith, 1996]. There are three steps in the GRAM algorithm (see section 1.2.3) that give GRAM noise filtering capability and better precision and accuracy than quantification methods based on integration. First, the S/N of the data matrix decomposed by SVD is

improved by using the addition matrix instead of the sample or standard matrix. Second, truncating the matrices decomposed by SVD removes noise. This step alone has been used to improve S/N of chromatographic bilinear data [Lee et al., 1991; Statheropoulos et al., 1999]. Finally, as stated in section 1.2.3, the solution to the eigenvalue problem (eq 1.4) is analogous to performing a least-squares fit of the signal profiles from one matrix onto another. Least-squares fitting, both linear and non-linear, is used by curve-fitting methods to quantify chromatographic peaks by finding the best fit between each peak and a given signal profile [Goldberg, 1971; Anderson et al., 1970a; Anderson et al., 1970b]. Curve fitting has been shown to give better quantification precision and accuracy than integration for resolved GC peaks [Goodman and Brenna, 1994].

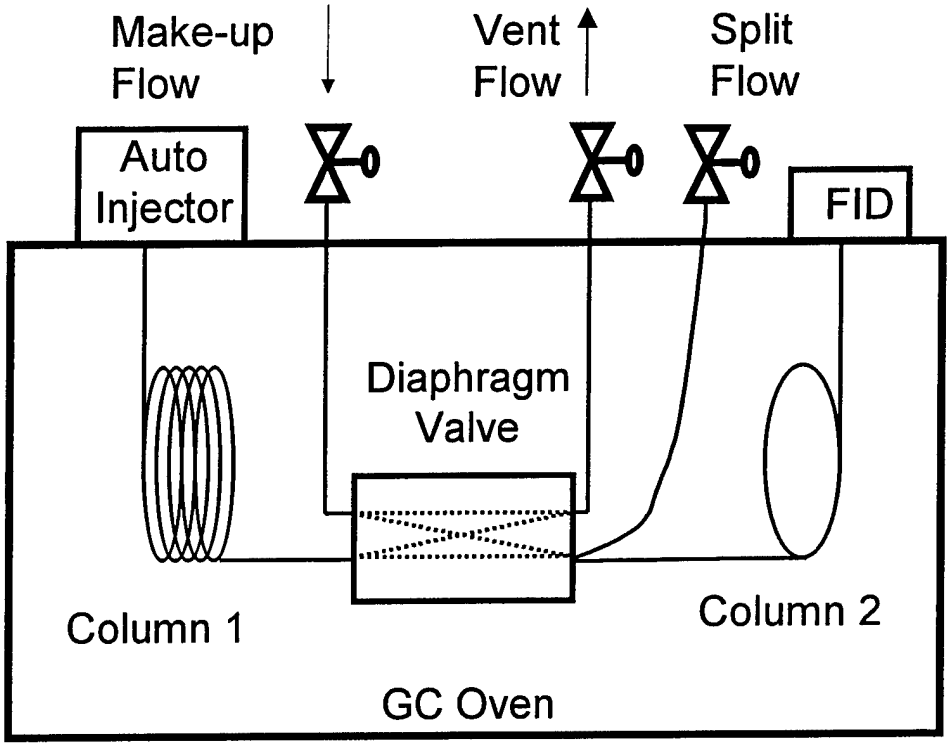



**Figure 1.1** Schematic of a valve-based GC x GC system. The diaphragm valve repeatedly injects segments of the first column eluent into the second column. Each injection into the second column produces a high-speed, secondary chromatogram.

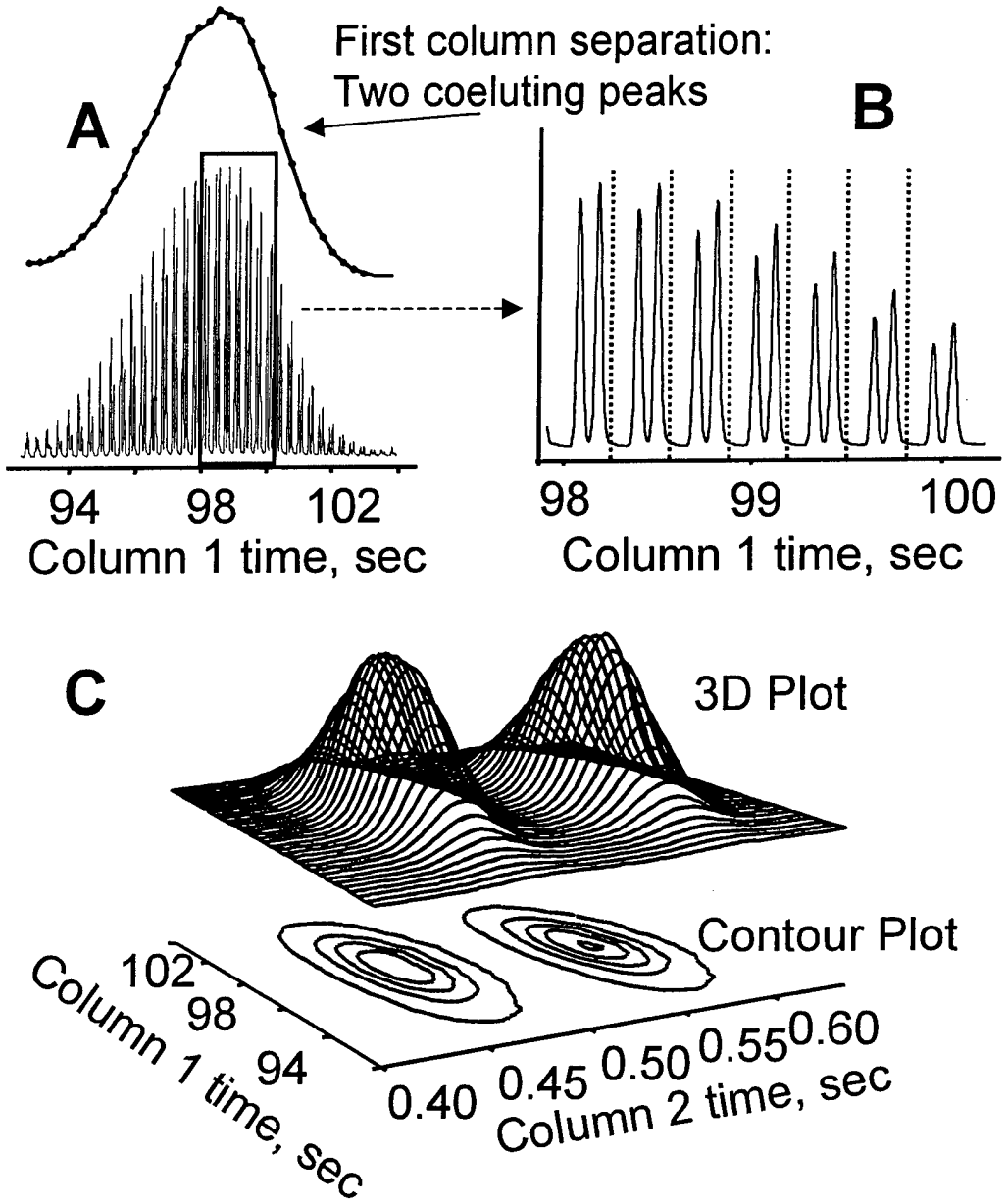



**Figure 1.2** (A) The first column separation of two coeluting peaks as a series of secondary chromatograms. (B) A group of secondary chromatograms from (A). Note the separation of the two peaks. Such a separation is not possible with the first column. (C) Side-by-side alignment of the entire series of secondary chromatograms in (A) produces two comprehensive 2-D peaks.

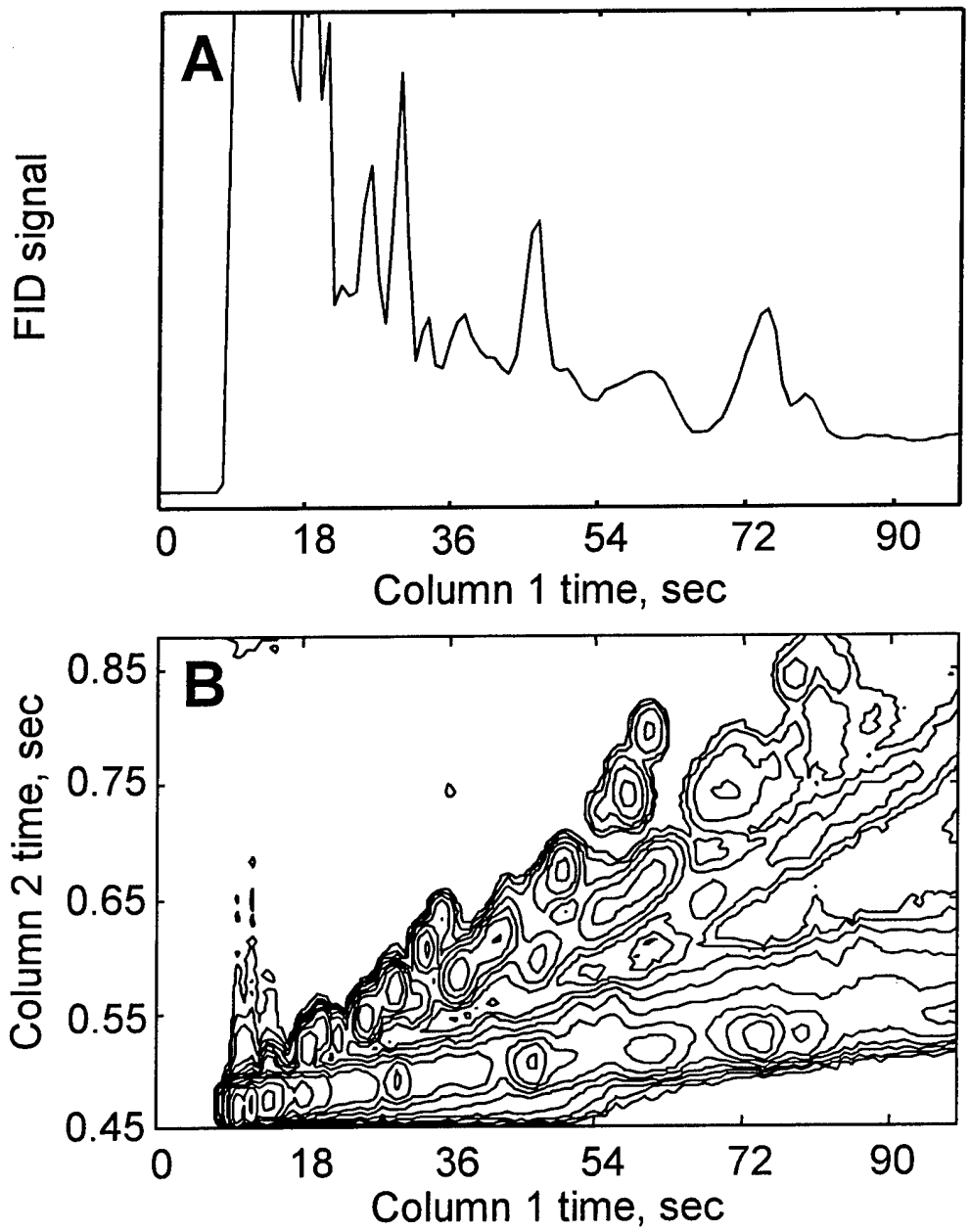



**Figure 1.3** (A) 1-D chromatogram of jet fuel produced by first column of GC x GC. (B) 2-D chromatogram of jet fuel produced by GC x GC. A 2-D chromatogram is able to reveal peaks hidden in a 1-D chromatogram.

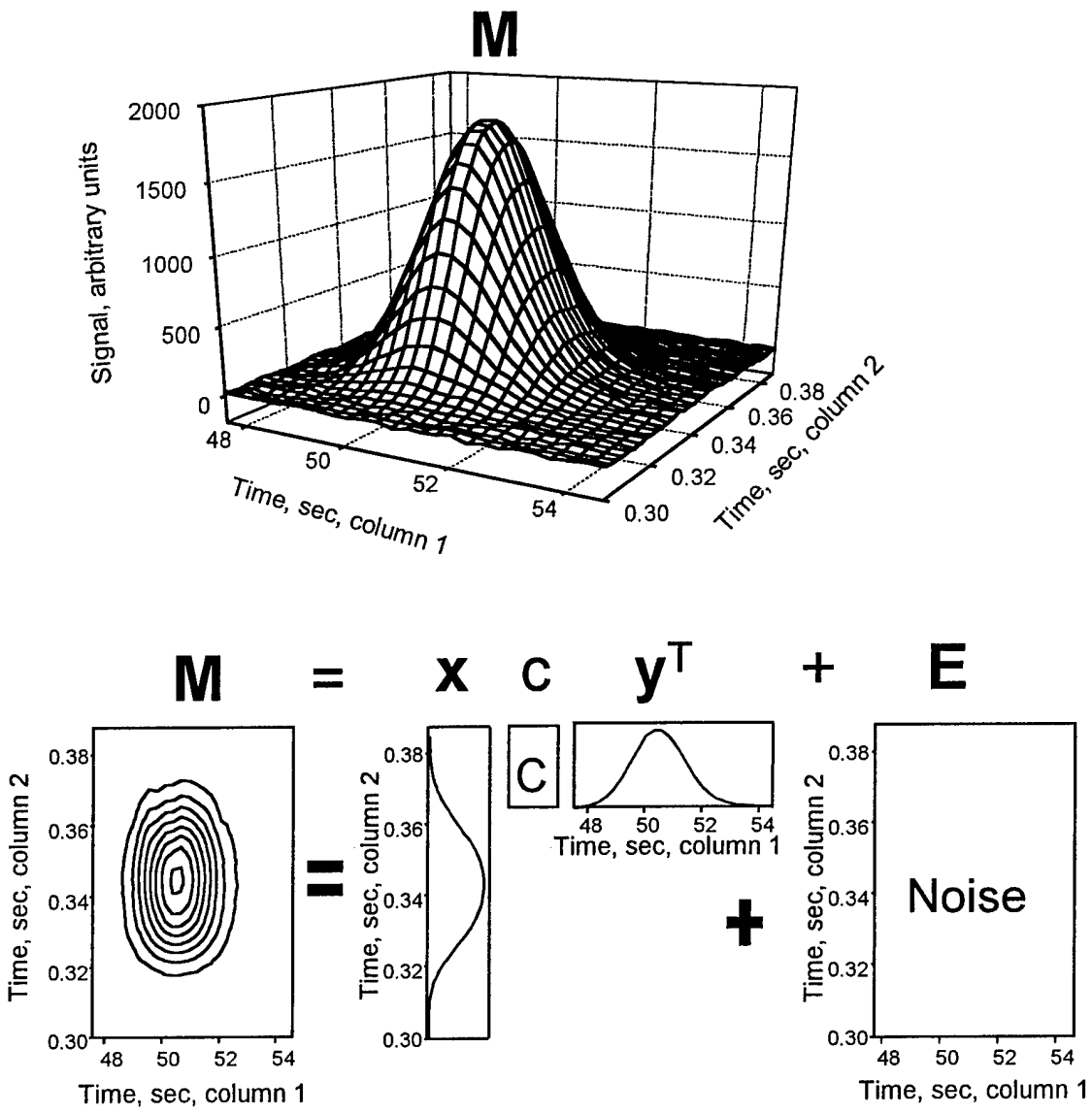



**Figure 1.4** The decomposition of a GC x GC peak, in the form of a data matrix **M**, into its bilinear signal and noise. The peak's bilinear signal represents the peak in the absence of noise. It consists of the peak's pure elution profiles from each GC column, **x** and $\mathbf{y}^T$, and a concentration term, c. The matrix **E** represents noise.

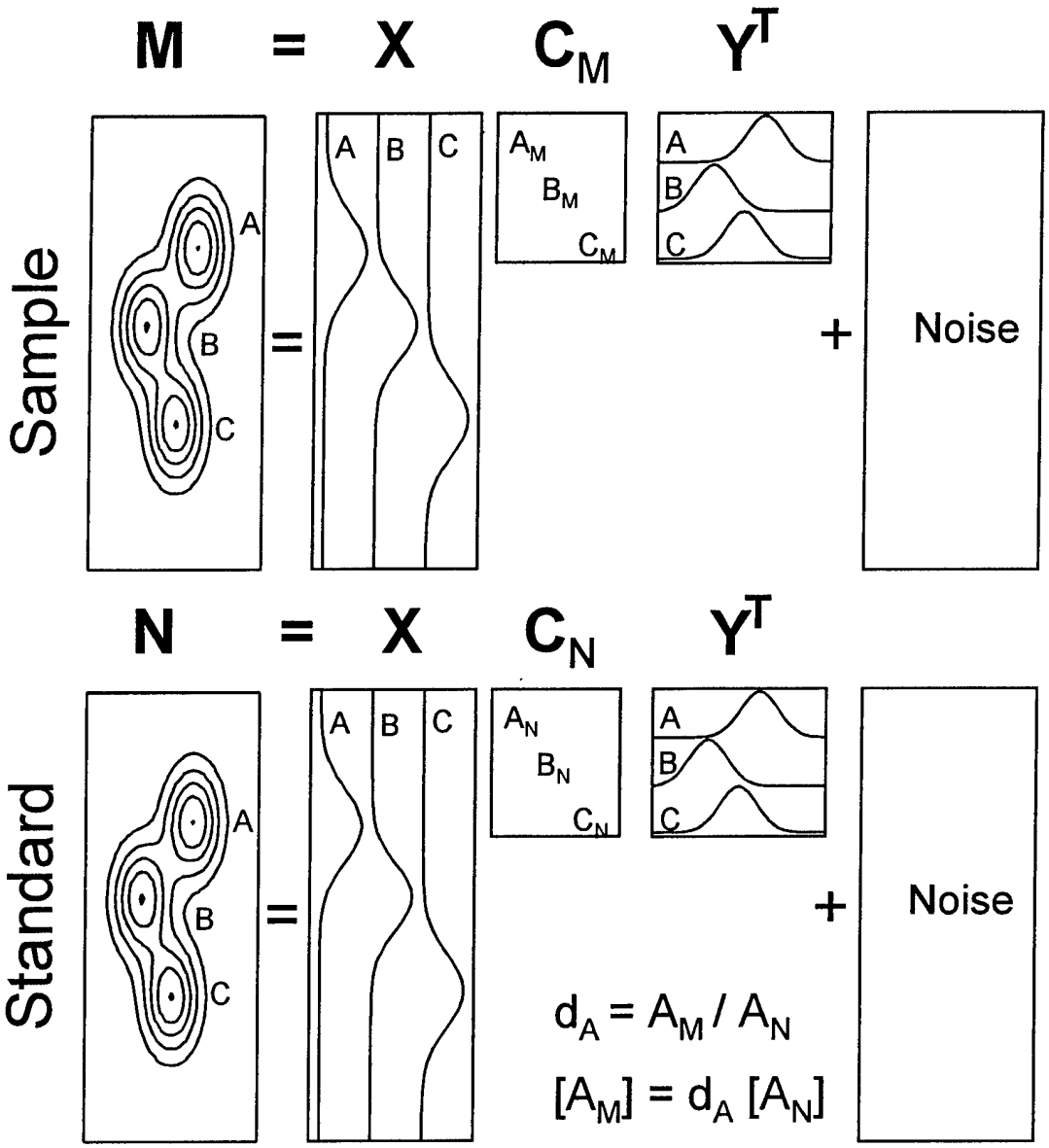



**Figure 1.5** Pictorial representation of GRAM analysis. GRAM finds the pure elution profiles, **X** and **Y**, for each component in common between **M** and **N**. The relative concentration of each component is given by its d term. Multiplying a component's d term with the component's known concentration in **N** produces the component's concentration in **M**.

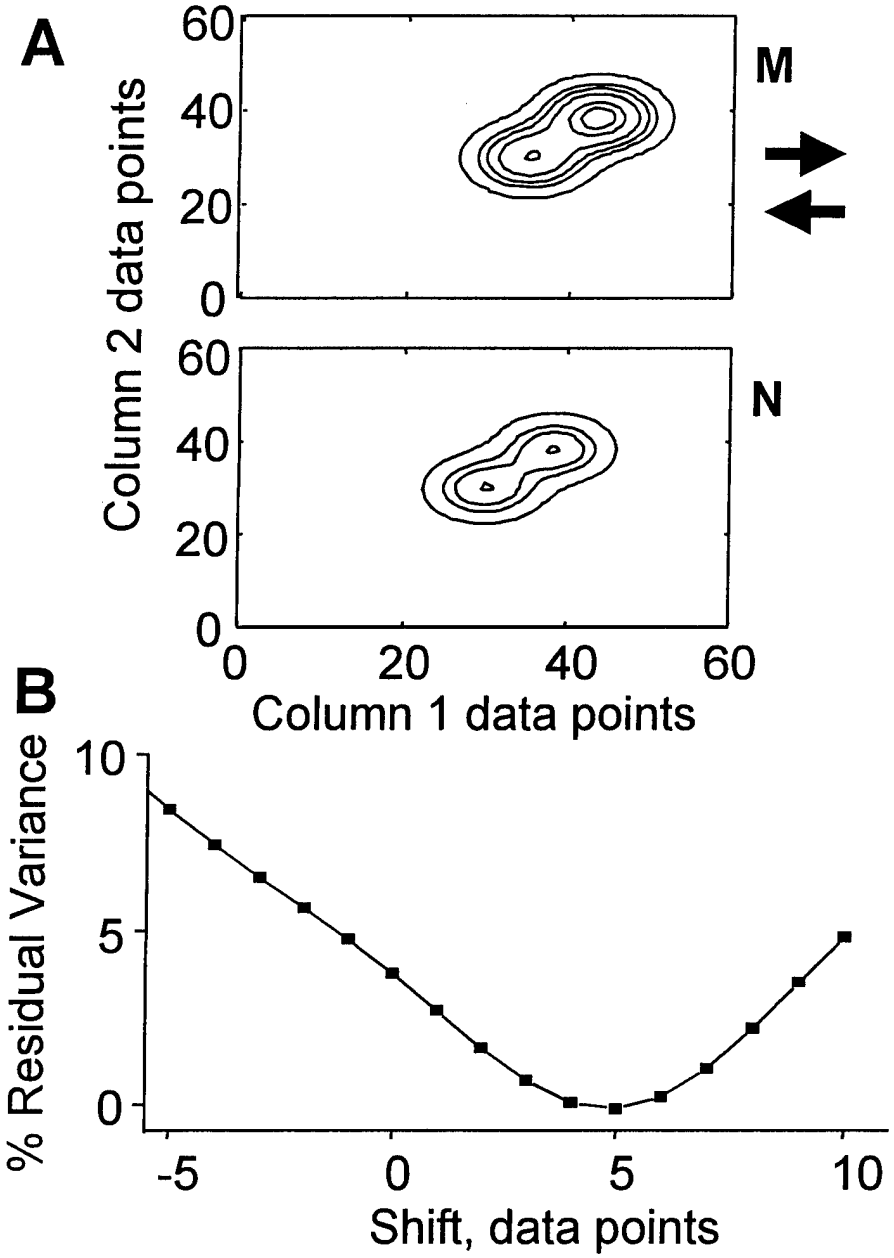



**Figure 1.6** (A) The augmented matrix obtained by stacking the sample data matrix **M** and standard data matrix **N**. The two 2-D peaks in **M** are shifted from those in **N** along the column 1 axis. (B) The percent residual variance of the augmented matrix as a function of data points shifted. The curve in (B) is determined by incrementally shifting **M** with respect to **N** and calculating the percent residual variance of the augmented matrix. The curve's minimum indicates the direction and number of data points **M** is shifted from **N** along the column 1 axis.

# Chapter 2: Enhancing the Number of Analyzable Peaks in Comprehensive 2-D Separations

## 2.1 Introduction

Comprehensive 2-D separations from 2-D column techniques are well suited for the separation and analysis of complex mixtures primarily because of the enhanced peak capacity provided by a 2-D space. Peak capacity is defined as the maximum number of peaks that can fit into a given retention space with a specified resolution [Giddings, 1991]. For a 2-D separation the peak capacity is equal to the product of the peak capacities of each of the two dimensions [Giddings, 1984]. Unfortunately, for moderate to highly complex samples, the peak capacity of a 2-D separation is not enough to ensure the resolution of most components, particularly in a timely manner. This concept was theoretically demonstrated by Davis using the statistical model of peak overlap (SMO) [Davis, 1991]. He showed that for peaks randomly distributed in a reasonably sized 2-D space, the average number of resolved peaks is much lower than the peak capacity. For instance, for a 2-D separation involving 20 peaks and having a peak capacity of 50, the SMO predicts on average that 4 peaks will be resolved. One way of dealing with the likelihood of peak overlap in 2-D separations is through the use of chemometric methods to mathematically resolve and quantify unresolved peaks.

In previous works [Prazen et al., 1998; Prazen, 1998; Prazen et al., 1999a; Prazen et al., 1999b], GRAM has been used to analyze overlapped chromatographic peaks from hyphenated 2-D techniques such as gas chromatography with mass spectrometric detection and liquid chromatography with absorbance detection. In addition, GRAM has been used to successfully deconvolute and quantify overlapped peaks from comprehensive 2-D separations [Bruckner et al., 1998]. Therefore, it is believed that GRAM analysis should enhance the number of analyzable peaks in comprehensive 2-D separations by quantifying peaks not fully resolved for quantification by traditional methods such as peak volume integration or height. This assertion is supported in this chapter by two sets of Monte Carlo simulations modeled

after comprehensive 2-D separations from a GC x GC analyzer. The first set of Monte Carlo simulations are used to generate contour plots that report the quantitative accuracy and precision of GRAM analysis as a function of an unresolved 2-D peak's resolution on each column axis or dimension. These contour plots demonstrate the analysis capabilities of GRAM under a variety of experimental characteristics modeled after real GC x GC data. Characteristics such as retention time variation, signal-to-noise (S/N) levels, and relative peak sizes are modeled. In addition, the beneficial effect of an objective retention time alignment algorithm on GRAM analysis is demonstrated [Prazen et al., 1998]. The added improvement in GRAM quantification through the use of data-point interpolation is also demonstrated [Fraga et al., 2000a]. Ultimately, the contour plots provide the resolution of a 2-D peak needed on each dimension for accurate and precise GRAM quantification. These resolution values are then used to identify the unresolved peaks in simulated comprehensive 2-D chromatograms that can be successfully analyzed by GRAM. Each comprehensive 2-D chromatogram was simulated by randomly distributing peaks into a 2-D space with a peak capacity of 50. Traditionally, the number of analyzable peaks equals the number of resolved peaks. However, with GRAM analysis the number of analyzable peaks becomes the sum of the number of resolved peaks and the number of unresolved peaks analyzable by GRAM within an acceptable quantitative precision and accuracy. After a set of Monte Carlo simulations of comprehensive 2-D chromatograms having randomly distributed peaks, an increase in the number of analyzable peaks in comprehensive 2-D separations is demonstrated through the use of GRAM for a wide range of peaks per unit peak capacity.

## 2.2 Theory

As discussed in sections 1.2.2 - 1.2.4, GRAM permits the quantitative analysis of unresolved bilinear signals. GRAM uses the bilinear data from the 2-D separations of a sample and a calibration standard to resolve and quantify targeted unresolved peaks. The sample contains "unknown" concentrations of the analytes while the

standard contains known amounts. The analyte peak in both the sample and standard can be overlapped with other analytes or interferences. The successful execution of GRAM produces resolved analyte peaks and relative analyte concentrations. In this chapter, GRAM is used to analyze computer simulations of a 2-D analyte peak overlapped with a 2-D interferent peak. The accurate modeling of characteristics present in comprehensive 2-D separations is required for realistic simulations. The characteristics that must be accurately modeled are now discussed.

### 2.2.1 Modeled Experimental Characteristics

Several comprehensive 2-D separations of mixtures produced by a valve-based GC x GC system were studied to determine general experimental characteristics. A GC x GC system is a good model for all other comprehensive 2-D column techniques. The most important characteristic to duplicate is the structure of a 2-D peak. A 2-D peak is made-up of several second dimension peaks whose heights are modulated by the shape of a first dimension peak. Hence, a 2-D peak is reasonably modeled by generating a Gaussian peak on the first dimension and then individually multiplying each of its data points with all the data points of a Gaussian peak generated on the second dimension. The other parameters modeled to mimic comprehensive 2-D separations are now discussed.

Several S/N values are modeled. The S/N is given by the maximum height of a 2-D peak divided by the three times the standard deviation of the detected noise. The detected noise is modeled as pure white noise. Several interferent-to-analyte size ratios are modeled as well as different resolutions on each dimension for the analyte and interferent peak. The reproducibility for repetitive injections of a sample volume onto the second dimension is modeled. In valve-based GC x GC, the valve repeatedly diverts small amounts of the first column eluent to the second column. A relative standard deviation (RSD) of 1.4% for injected volume is typically obtained by a GC x GC diaphragm valve [Bruckner et al., 1998]. Injection volume reproducibility for sample introduction onto the first dimension is not modeled, although it is readily

determined, because unlike other modeled characteristics, it does not directly affect GRAM analysis. This is because injection volume variation on the first column does not affect a peak's profile (i.e., shape, width, or retention time) and only changes in a peak's profile affect the accuracy of peak deconvolution by GRAM. Assuming no peak profile changes between a sample and standard, GRAM can accurately deconvolute each unresolved peak in common between the sample and standard. However, the accuracy of each peak's concentration in the sample relative to that in the standard is subject to the uncertainty associated with injection volume precision on the first column, which directly affect's each peak's intensity in the sample and standard. For the simulations in this paper, if one wants to include the uncertainty associated with automated sample injection onto the first dimension (i.e., 0.7% RSD), then a one-percent error can be propagated into the reported error of GRAM quantification. Run-to-run retention time precision is the final characteristic modeled and is the one having the greatest impact on chemometric analysis. Poor retention time precision, if uncorrected, can adversely effect the quantitative results of GRAM and other chemometric methods [Prazen et al., 1998; Juan et al., 1998]. Fortunately, the issue of run-to-run retention time precision for the GRAM analysis of bilinear chromatographic data, such as from comprehensive 2-D separations, has been successfully addressed [Prazen et al., 1999b; Fraga et al., 2000a].

### 2.2.2 Peak Width-Based Retention Time Precision

Retention time precision is defined as the standard deviation of a peak's retention times, $s_t$, as calculated from multiple chromatographic runs. Three kinds of retention time precision are modeled. Two of them originate because a 2-D peak has two retention times, one for each dimension. The precision values for these retention times determine the retention time shifting occurring on each dimension from one run of a comprehensive 2-D separation to the next. The third type of retention precision time comes from the fact that a 2-D peak is constructed from several second dimension peaks each having its own retention time. This retention time precision determines the

retention time shifting occurring on the second dimension within each run of a comprehensive 2-D separation. When using GRAM or other chemometric methods, retention time precision is better described by the peak width-based retention time precision, $\delta$ [Bahowick and Synovec, 1995]. $\delta$ is a dimensionless quantity defined as $s_t$ divided by the peak's base width, $4\sigma_t$ . The $\delta$ takes into account the effect peak width has on the chemometric analysis of unresolved peaks. For instance, in the case of two different separations each producing data containing two overlapped peaks with a given resolution and constant $s_t$, peak deconvolution results are more accurate and precise for the wider peaks. Wider peaks have a smaller $\delta$. Ideally, the $\delta$ value should be has small as possible to provide the most accurate and precise quantification possible. However, the actual $\delta$ value needed for good quantitative results depends upon resolution, $R_s$. Larger $\delta$ values are acceptable for peaks with larger $R_s$. The relationship between $\delta$ and $R_s$ and their subsequent effect on chemometric analysis can be described by the ratio of $\delta$ and $R_s$ or $\delta/R_s$. $\delta/R_s$ is a dimensionless quantity and is related to chromatographic data by:

$$\delta/R_s = (s_t/4\sigma_t)/(\Delta t_r/4\sigma_t) = s_t/\Delta t_r \qquad (2.1)$$

where $\Delta t_r$ is the time separation between two adjacent peaks of equal width. Bahowick and Synovec determined that a $\delta$ value resulting in a $\delta/R_s$ value of 0.02 or less provides for the accurate and precise quantification of overlapped equal-sized 1-D peaks by classical least squares [Bahowick and Synovec, 1995]. In other words, if the run-to-run variation in retention time relative to the temporal separation of two adjacent peaks is less than or equal to about two percent, then the uncertainty in the quantification of these overlapped peaks will not be adversely affected by the run-to-run retention time variation. Accordingly, if one requires reliable peak identification and sound quantitative precision for peaks with $R_s$ down to 0.4, then eq 2.1 suggests a $\delta$ of 0.008 is needed to achieve this analytical goal. While such a $\delta$ value is specifically valid for

1-D peaks, it can provide general guidance for δ values needed in the chemometric analysis of 2-D peaks. In this chapter, several δ values are simulated with simulated comprehensive 2-D data in order to measure the effect different δ values have on the quantitative results of GRAM analysis as a function of a peak's $R_s$ on each dimension. In real comprehensive 2-D separations, different δ values are anticipated due to differences in experimental parameters between different comprehensive 2-D separation methods. The experimental parameters affecting δ are now discussed.

The δ is affected by the experimental parameters shown in eq 2.2:

$$\delta = s_t/4\sigma_t = (1/4)(s_t/t_r)N^{1/2} \qquad (2.2)$$

where $t_r$ is the mean retention time and N is the plate count [Bahowick and Synovec, 1995]. The value of $s_t/t_r$ is usually instrument dependent, i.e., it is affected by factors controlled by the analytical instrument such as flow rate and temperature [Foley et al., 1989; Goedert and Guiochon, 1970]. The value of N for an optimized system is affected by the separation column properties such as column length and plate height. In this chapter, several δ values are modeled by varying $s_t$ while holding $4\sigma_t$ constant. In terms of eq 2.2, different δ values can be achieved when N is constant and $s_t/t_r$ varies. Such a scenario can occur among several instruments that have the same separation selectivity and efficiency but different $s_t/t_r$ values. For a given instrument under constant experimental conditions such as isothermal GC or isocratic LC, $s_t/t_r$ is usually constant but different from one instrument to the next. Therefore the range of δ values evaluated in this chapter can also be thought to mimic changes in N for a constant $s_t/t_r$ value. Changes in N can occur when column length or plate height is changed. In this chapter, only the δ associated with the first dimension is varied while the two δ values associated with the second dimension are kept constant. It is assumed that the second dimension is fully optimized for fast and reproducible separations. Hence, only changes affecting the first dimension are made. Typically in GC x GC, the δ values from the second dimension are much smaller than those from the first.

This primarily results because a smaller $s_t/t_r$ value on the second dimension is readily achieved experimentally. The $s_t/t_r$ on the second dimension is found to be approximately four times smaller than that on the first dimension. This usually results in $\delta$ values that are much smaller on the second dimension than on the first. The relatively small $\delta$ experimentally achieved with the second dimension, typically about 0.008, permits the use of an objective retention time alignment algorithm, which corrects for retention time shifts on the first dimension [Prazen et al., 1998]. This alignment algorithm is designed for bilinear data where the second dimension has high retention time precision.

## 2.3 Procedures

A computer program written in Matlab 5.2 (The Mathworks Inc, Natick, MA) was used to generate Monte Carlo simulations to determine the effect different experimental characteristics have on the GRAM analysis of unresolved 2-D peaks. Table 2.1 lists the values for the experimental characteristics modeled in five Monte Carlo simulation studies. In each study, sample-standard combinations were generated. For each combination, the sample and standard consist of an analyte and interferent peak in a 2-D space. The 2-D space is a 60-by-60 data-point matrix. Each 2-D peak has a constant base width of 17 points on each matrix dimension. This data density is considered the minimum for accurate peak representation [Braithwaite and Smith, 1996], and it is readily achieved in high speed GC x GC separations. The standard has the same amount of interferent as in the sample but twice the amount of analyte. The standard can be viewed as the original sample with a known amount of added analyte. Such a standard could be produced if the analyte in the sample is quantitatively analyzed using the standard addition method. For each sample and standard, the retention times on each dimension are each randomly chosen from a normal distribution of retention times. The $s_t$ for each retention time distribution is determined by multiplying the $\delta$ value (see Table 2.1) and a given, constant peak width. The mean retention times are chosen to achieve the required separation between the two 2-D

peaks dictated by the given dimensional resolutions, $R_{s1}$ and $R_{s2}$. $R_{s1}$ is a 2-D peak's resolution along the first dimension, and $R_{s2}$ is the 2-D peak's resolution on the second dimension. Both $R_{s1}$ and $R_{s2}$ can be combined to produce a net or 2-D resolution, $R_{2D}$, as shown by eq 2.3 [Giddings, 1987].

$$R_{2D} = [(R_{s1})^2 + (R_{s2})^2]^{1/2} \qquad (2.3)$$

Both sample and standard have the same $R_{s1}$ and $R_{s2}$ values for each combination of $R_{s1}$ and $R_{s2}$. One hundred different combinations of $R_{s1}$ and $R_{s2}$ were generated using resolutions that ranged from 0.1 to 1.0 in increments of 0.1. Each resolution combination has 300 replicate sample-standard combinations having the given dimensional resolutions between peaks. For each resolution combination, GRAM analysis with and without prior retention time alignment is performed on each sample-standard combination and then the mean and RSD of the GRAM sample-to-standard analyte concentrations are calculated. The absolute bias is calculated for the mean GRAM sample-to-standard analyte concentration. Retention time alignment is performed on the first dimension using the objective retention time alignment algorithm [Prazen et al., 1998]. In study A (see Table 2.1), the cubic interpolation function from Matlab is incorporated into the alignment algorithm to interpolate three data points between each original data collected along the first column. This modified algorithm is used in conjunction with the objective alignment algorithm. Ultimately, GRAM quantification accuracy and precision is determined as a function of different experimental characteristics, such as $R_{s1}$ and $R_{s2}$.

Another set of Monte Carlo simulations was performed to determine the additional number of peaks in 2-D chromatograms that are analyzable using GRAM. Each simulated 2-D chromatogram has a peak capacity of 50 as determined by peak capacities of ten and five on the first and second dimension, respectively. For simplicity, the peaks are of equal width and are randomly distributed throughout the 2-D chromatogram. One thousand 2-D chromatograms are generated for each given number of peaks. The number of randomly distributed peaks range from one to 50.

## 2.4 Results and Discussions

### 2.4.1 Comprehensive 2-D Simulations

GRAM analysis using a simulated sample and standard from study A is demonstrated. Figure 2.1A is a sample that represents an analyte and interferent peak that have undergone a comprehensive 2-D separation. A $R_{2D}$ of 0.7, determined from dimensional resolutions $R_{s1}$ and $R_{s2}$ of 0.5 each, makes analyte quantification using peak volume or height inaccurate. However, the analyte peak can be resolved and accurately quantified through the use of standard addition and GRAM. Figure 2.1B is the standard made for GRAM analysis of the selected sample (Figure 2.1A). The standard represents the standard addition of the analyte into the sample such that the amount of analyte is doubled. Standard addition is required for GRAM analysis whenever chemical matrix effects significantly change a peak's retention time, width, or shape in either dimension between a sample and calibration standard [Fraga et al., 2000a]. Chapter 3 demonstrates that such changes occur in the 2-D separation of complex samples, hence, the standard addition method was simulated. Figure 2.1C is the deconvoluted peak of the analyte in the sample obtained by GRAM analysis of the sample (Figure 2.1A) and standard (Figure 2.1B). GRAM analysis is successful as indicated by the appropriate peak shape (i.e., non-negative and unimodal) of the deconvoluted analyte peak and by an accurately predicted analyte concentration of 0.50 times the analyte concentration of the standard (true = 0.5). Such successful results can be reliably achieved with standard addition and GRAM as long as the retention times of the analyte and interferent peak are sufficiently reproducible between the sample and standard.

Run-to-run retention time variations can be detrimental to the successful application of GRAM by causing a given 2-D peak to have a significantly different pair of retention times in the sample than standard. GRAM analyses of replicate runs of a sample and standard usually reveal the detrimental effects of run-to-run retention time variations. Figure 2.2A is an overlay of 30 summed deconvoluted peaks of the analyte

and interferent obtained from the GRAM analysis of 30 randomly generated sample-standard combinations from study A having a resolution of 0.5 on each dimension (e.g., see Figures 2.1A and 2.1B). Each summed deconvoluted peak is the summation of the data matrix containing the resolved peak (e.g., see Figure 2.1C) onto the first dimension. Some of the summed deconvoluted peaks in Figure 2.2A have accurate peak shapes because by chance their corresponding sample and standard had no appreciable retention time shift. However, a sizeable fraction of the peaks depicted in Figure 2.2A have erroneous shapes because their corresponding sample and standard had a significant retention time shift on the first dimension. These peaks also have inaccurate quantitative results compared to those with good shapes. This results in a biased mean concentration for the analyte. In addition to revealing that some deconvoluted peaks have erroneous shapes, Figure 2.2A reveals the poor signal reproducibility caused by significant run-to-run retention time variations. This poor signal reproducibility translates into poor quantitative precision as indicated by a % RSD of 7.0 for the GRAM analyses used in generating Figure 2.2A. However, when retention time alignment is used prior to GRAM analysis, the % RSD drops to 1.8 and the bias of the mean concentration drops from 3% to less than 0.5%. Such improvements in GRAM analysis are depicted in Figure 2.2B by the reproducible and appropriate shapes of the summed deconvoluted peaks. These peaks were obtained from the same 30 sample-standard combinations used to produce Figure 2.2A but with retention time alignment. The above-mentioned improvements in GRAM results closely resemble those obtained when retention time alignment was used prior to the GRAM analysis of real 2-D peaks (see chapter 3). The use of an interpolated retention time alignment algorithm, which is demonstrated later in this chapter, provides an even greater improvement in peak shape and qualitative results than the retention time alignment algorithm used thus far. In either case, retention time alignment is critical for the successful implementation of GRAM analysis in the presence of significant run-to-run retention time shifts.

## 2.4.2 Accuracy and Precision Contour Plots

In addition to precise retention times, an unresolved peak must have some resolution on each dimension of a 2-D separation for the successful application of GRAM. In order to determine what minimal resolutions are needed, 30,000 sample-standard combinations were generated in study A with different combinations of dimensional resolutions between the analyte and interferent peak. Figure 2.3A is a contour plot that depicts the accuracy (i.e., % bias in absolute terms) of GRAM quantification without retention time alignment at different dimensional resolutions. The contour plot was generated by performing study A 10 times and then averaging the absolute bias for all 10 study-A iterations. Figure 2.3B depicts the precision (i.e., % RSD) for GRAM quantification without retention time alignment at different dimensional resolutions. The precision contour plot was generated by averaging the % RSD of analyte quantification for the 10 study-A iterations used to generate Figure 2.3A. Using retention time alignment prior to GRAM analysis resulted in Figures 2.3C and 2.3D, which depict the accuracy and precision contour plots, respectively, for GRAM quantification. Figures 2.3C and 2.3D were obtained from the GRAM analyses of the same sample-standard combinations used to generate Figures 2.3A and 2.3B but with retention time alignment. Figure 2.3D depicts 10 isograms per % RSD. The % RSD values produced from the 10 study-A iterations were not averaged for the purpose of illustration. The distribution of isograms per % RSD is typical and provides a measure of simulation uncertainty, which is helpful when determining if differences in isogram positions are significant. Retention time alignment extends GRAM quantification to more severely overlapped peaks. This is indicated by the observation that the RSD and bias isograms reach further along the first dimension to lower resolutions with retention time alignment than without.

For brevity, only precision contour plots are discussed from this point forth. Focusing only on GRAM quantitative precision is justified considering that it is the limiting factor in the overall quality of the GRAM quantitative results. When comparing a precision contour plot and its corresponding accuracy contour plot, the %

RSD value is always larger than the % bias value for any given $R_{s1}$ and $R_{s2}$ combination.

The dashed curve in Figure 2.3D defines the different combinations of $R_{s1}$ and $R_{s2}$ that result in a $R_{2D}$ of one as defined by eq 2.3. Any peak with dimensional resolutions falling on the dashed curve or to the right is sufficiently resolved for quantitative analysis without GRAM. For a peak with dimensional resolutions to the left of the dashed curve, GRAM analysis works best for peaks whose dimensional resolutions are equal. In addition, Figure 2.3D reveals that if $R_{s1}$ and $R_{s2}$ are not equal, then having $R_{s1}$ larger than $R_{s2}$ results in slightly better GRAM quantitative precision than having $R_{s2}$ larger than $R_{s1}$. This is primarily a consequence of not being able to objectively correct the retention time shifts on the first dimension to the same value of $\delta$ found on the second dimension. The retention time alignment algorithm evaluated thus far can only determine retention time shifts to the nearest data point. That is, it can only determine that portion of the true retention time shift that is an integer multiple of the data-sampling interval. Hence, the retention time imprecision on the first dimension is reduced but not eliminated. A further reduction in retention time imprecision using data-point interpolation is discussed later. A reduction of the $\delta$ value on the first dimension to the same value found on the second dimension would result in a symmetric precision contour plot. In such a case, the values for $R_{s1}$ and $R_{s2}$ could be interchanged and still give the same %RSD. Indeed, this was achieved in study B where each dimension's run-to-run $\delta$ value was the same (not shown for brevity).

### 2.4.3 Interpolated Retention Time Alignment

One approach allowing further corrections of retention time shifts on the first dimension is through interpolating the non-integer data point shift needed to give a better estimate of the true shift between sample and standard. Figure 2.4A depicts the mean isograms for the precision contour plot depicted in Figure 2.3D overlaid with isograms also obtained using retention time alignment but with interpolation. Only the 1% and 4% isograms are shown for clarity. Interpolation significantly extends the

isograms to lower resolutions along the first dimension such that they extend to lower resolutions on the first dimension than along the second dimension. Accordingly, interpolation improves the accuracy and precision of peak deconvolution by GRAM. Figure 2.4B depicts the summed deconvoluted peaks obtained from the GRAM analysis of the same 30 sample-standard combinations used to generate Figure 2.2B but with interpolated retention time alignment. The summed peaks in Figure 2.4B have better signal reproducibility and shapes than those obtained without interpolation in Figure 2.2B. Interpolation artificially increased the data density of the first column allowing a more accurate retention time alignment and ultimately a smaller $s_t/t_r$ and $\delta$ on the first dimension than on the second.

Interpolation is simpler than increasing a peak's data density on the first dimension by either reducing the separation time on the second dimension or increasing a peak's width on the first dimension while simultaneously maintaining the peak's dimensional resolutions. In this chapter, a data density of 17 points per peak on both dimensions is used. Recently, it has been reported that 4 points across the first dimension peak width is sufficient to adequately describe resolution in comprehensive 2-D separations [Murphy et al., 1998]. In those cases or anytime the sampling rate from the first separation dimension to the second is small, retention time alignment with interpolation is critical to obtain a satisfactory estimate of the true retention time shift.

In this chapter, the distribution of retention time shifts between sample and standard is determined by the value of $\delta$. So far a first dimension $\delta$ value of 0.05 has been modeled. This is the largest $\delta$ value measured in recent high-speed GC x GC separations [Bruckner et al., 1998; Fraga et al., 2000b; Fraga et al., 2000a]. However, larger $\delta$ values can be expected with more efficient columns or with first dimension techniques that have poorer retention time reproducibility. When first dimension $\delta$ values greater than 0.5 were simulated in study C, they deteriorated GRAM quantification by stretching the accuracy and precision isograms obtained with a $\delta$ of 0.05 (see Figures 2.3A and 2.3B) to higher dimension 1 resolutions. These results are

not shown for brevity. The larger the δ, the worse the GRAM precision. However, when non-interpolated retention time alignment is used prior to GRAM, δ values between 0.02 and 0.2 are reduced to the same final δ value of approximately 0.01 such that their precision isograms are indistinguishable from those depicted in Figure 2.3D, which where obtained with an original δ value of 0.05. This occurs because peak widths are constant and $s_t$ is reduced to a set value determined by the data density on the first dimension. Retention time alignment is crucial for successful GRAM analysis in cases with large δ values.

### 2.4.4 The Effects of S/N and Interferent Size

S/N and the relative peak sizes of interferent and analyte peak are other factors that affect GRAM quantification to a degree. For instance in study D, a S / N of 30 produced GRAM precision isograms with % RSD values of 2% and greater that were indistinguishable from those obtained with a S/N of 1000 using study A (Figure 2.3D). However, the effect of S/N on GRAM quantification becomes pronounced at very low S/N. Figure 2.5A depicts a representative simulated sample from study D with a S/N of 4.8 and a resolution of 0.5 on each dimension. Figure 2.5B depicts the precision contour plot obtained with GRAM analysis at a S/N of 4.8. Note how all the precision isograms on both dimensions shift to higher resolution compared to the precision contour plot (Figure 2.3D) obtained at an S/N of 1000 in study A. Percent RSD values below 3.8 cannot even be achieved with a S/N of 4.8. However, as shown later in chapter 4, GRAM quantification at this level is still significantly more accurate and precise than quantification by integrating peak volume for resolved GC x GC peaks [Fraga et al., 2000b].

In terms of the interferent-to-analyte size ratio, study E reveals that an interferent peak three times the size of the analyte peak gives precision isograms that were nearly indistinguishable from those obtained with an interferent peak equal in size to the analyte peak (Figure 2.3D). However, for an interferent peak 10 times the size of the analyte peak, GRAM quantification is slightly affected. Figure 2.6A is a

representative simulated sample from study E with the interferent peak ten times the size of the analyte peak and with a resolution of 0.5 on both dimensions. Figure 2.6B is the precision contour plot obtained with an interferent peak ten times the size of the analyte peak in the sample. The isograms in Figure 2.6B are stretched toward higher resolution values compared to those obtained with equal sized analyte and interferent peaks (Figure 2.3D). This is still respectable considering the large size discrepancy between analyte and interferent peak. It should be noted that GRAM quantification can provide respectable results with much larger interferent peaks if retention time shifting is non-existent or corrected. However in study E, the retention time alignment algorithm is unable to properly align samples having an interferent peak 30 times larger than the analyte peak.

### 2.4.5 Analysis Enhancement

Using the information provided by studies A-E, it is possible to quantify the analysis enhancement that is obtained by using GRAM for the quantitative analysis of component mixtures. In this assessment, simulated 2-D chromatograms are generated to determine the number of peaks that are analyzable without GRAM analysis and the number of peaks that are analyzable through the inclusion of GRAM analysis. It is assumed that peaks in a 2-D chromatogram having a $R_{2D}$ of one or greater are sufficiently resolved that quantification by GRAM is not needed. Only peaks having a $R_{2D}$ less than one are considered for GRAM analysis. Of those peaks, only the ones that would result in a GRAM quantitative precision of 4% or less are considered analyzable in this study. Figure 2.7 depicts one of the simulated 2-D chromatograms used in this study. The peaks depicted in the Figure 2.7 are randomly distributed. Random peak distributions have been previously used to simulate 2-D separations for the purpose of quantifying peak overlap [Davis, 1991; Rowe and Davis, 1995; Rowe et al., 1995]. In Figure 2.7, only four peaks are sufficiently resolved to be considered analyzable without needing GRAM. The remaining 11 peaks are considered not analyzable by traditional quantitative methods that employ peak height or integration.

However, GRAM analysis permits the analysis of five of these peaks, which increases the number of analyzable peaks to nine. The five peaks have dimensional resolutions that fall within the 4% dashed isogram in Figure 2.4 obtained using GRAM and non-interpolated retention time alignment. According to the previously mentioned results from studies C, D, and E, the 4% dashed isogram in Figure 2.4A is valid for peaks having a wide range of $\delta$ values, a S/N of at least 30, and an interferent-to-analyte size ratio of three or less. The same is true for the 4% solid isogram in Figure 2.4A obtained using GRAM and interpolated retention time alignment. One thousand simulated 2-D chromatograms each having 15 peaks randomly distributed in a 2-D separation space with a peak capacity of 50 (e.g., see Figure 2.7) were generated. They were used to determine the average number of resolved peaks, which are analyzable without GRAM, and the average number analyzable peaks, which include the use of GRAM. Both 4% isograms depicted in Figure 2.4A were independently used to determine the number of overlapped peaks that are analyzable by GRAM.

Different numbers of peaks, ranging from one through 50, were then randomly distributed into the same 2-D separation space depicted in Figure 2.7. The average number of analyzable peaks without and with GRAM for every given number of peaks were determined by generating one thousand simulated 2-D chromatograms for each given number of peaks. Ultimately Figure 2.8 was produced, which depicts the analysis enhancement achieved with GRAM as a function of the number of peaks per unit peak capacity. The analysis enhancement provided by GRAM is measured by the analysis enhancement factor (AEF), which is the average number of analyzable peaks divided by the average number of resolved peaks. For this calculation, the number of analyzable peaks is equal to the number of resolved peaks plus those unresolved peaks amenable to precise GRAM analysis, i.e., a 4% RSD or better. GRAM increases the number of analyzable peaks, especially for more crowded 2-D chromatograms. For instance, at a number of peaks per unit peak capacity of 0.67, where the AEF using interpolated alignment is approximately two, the number of analyzable peaks is doubled using GRAM analysis. This analysis enhancement is very impressive. However, it is interesting to note that the fraction of peaks that are analyzable is low.

For instance at a peak per peak capacity of 0.67, the average number of peaks that are analyzable out of 100 is 20. Without GRAM only 10 out of 100 are analyzable. These low numbers are an unfortunate consequence of randomly distributed peaks [Davis, 1991; Schure, 1991].

The random distribution of peaks used to generate the 2-D chromatograms does mimic the apparent random scatter of peaks seen in a 2-D chromatoelectropherogram of a complex biological mixture [Hooker and Jorgenson, 1997]. A random distribution of peaks can be expected for multicomponent mixtures that have a large sample dimensionality [Giddings, 1995], or many different chemical families, and are separated by comprehensive 2-D separations that are truly orthogonal, i.e., each column separation process is completely independent [Slonecker et al., 1996].

## 2.5 Conclusions

Using Monte Carlo simulations, GRAM analysis clearly enhances the analysis capabilities of comprehensive 2-D separations. As long as chemical standards exist for the peaks of interest, the use of GRAM permits the analysis of peaks that would otherwise not be analyzed because of inadequate peak resolution. The analysis enhancement demonstrated in this chapter is probably the minimum achievable with GRAM. In cases where peak distributions are ordered, GRAM should be able to analyze more overlapped peaks. For instance, in some GC x GC separations peaks fall along diagonals dispersed orderly throughout a 2-D chromatogram [Beens et al., 2000; Gaines et al., 1999]. In such cases, the peaks are in an optimal configuration for GRAM analysis. GRAM analysis is certainly beneficial and should be seriously considered as an integral component of any comprehensive 2-D analysis package.

**Table 2.1. Experimental characteristics used in Monte Carlo studies**

| Study | Signal-to-noise[a] of analyte in sample | Interferent-to-analyte size in sample | Dimension 1 run-to-run $\delta^b$ | Dimension 2 run-to-run $\delta^b$ | Dimension 2 within-run $\delta^b$ | Dimension 2 injection volume reproducibility[c] |
|---|---|---|---|---|---|---|
| A | 1000 | 1 | 0.05 | 0.008 | 0.006 | 1.4 |
| B | 1000 | 1 | 0.008 | 0.008 | 0.006 | 1.4 |
| C | 1000 | 1 | 0.02, 0.08, 0.1, 0.2 | 0.008 | 0.006 | 1.4 |
| D | 4.8, 6.25, 12.5, 30, 60 | 1 | 0.05 | 0.008 | 0.006 | 1.4 |
| E | 1000 | 3, 5, 10, 30 | 0.05 | 0.008 | 0.006 | 1.4 |

[a] Defined as the maximum signal of a 2-D peak divided by three times the standard deviation of the baseline noise
[b] Peak width-based retention time precision; standard deviation of a peak's retention time divided by the peak's base width
[c] Reproducibility measured as the percent relative standard deviation of amount of sample injected from replicate trials

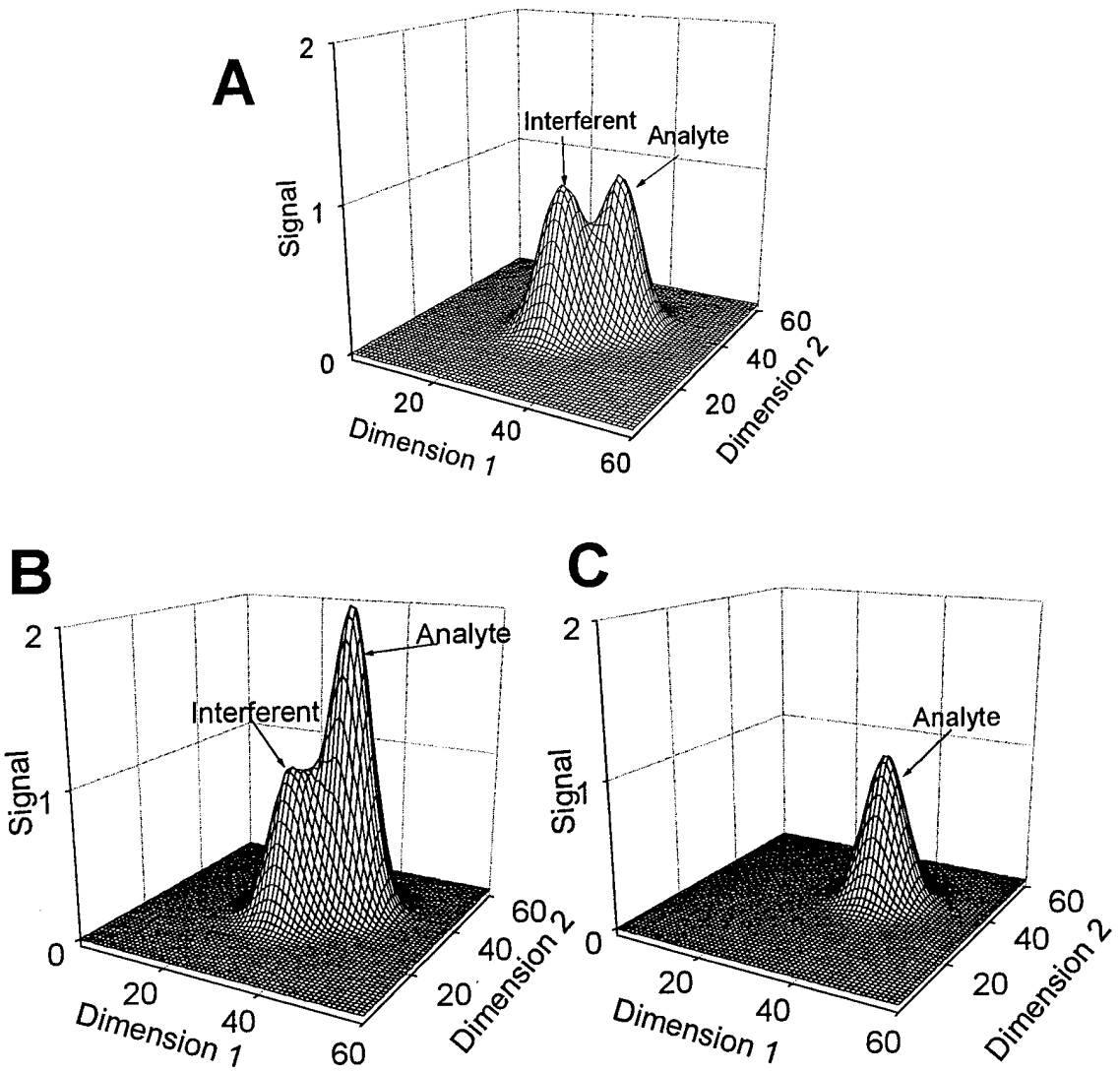



**Figure 2.1.** (A) Computer simulated sample consisting of an analyte and interferent peak that have undergone a comprehensive 2-D separation. Both peaks have a resolution of 0.5 on both dimensions and are modeled according to the parameters listed for study A in Table 2.1. (B) Computer simulation of the calibration standard obtained by the standard addition of analyte to the sample, (A), resulting in a 2-to-1-peak size ratio of analyte to interferent. (C) Deconvoluted peak of the analyte in the sample obtained by GRAM analysis of data from the sample, (A), and standard, (B). The deconvoluted peak of the interferent is not shown for brevity.

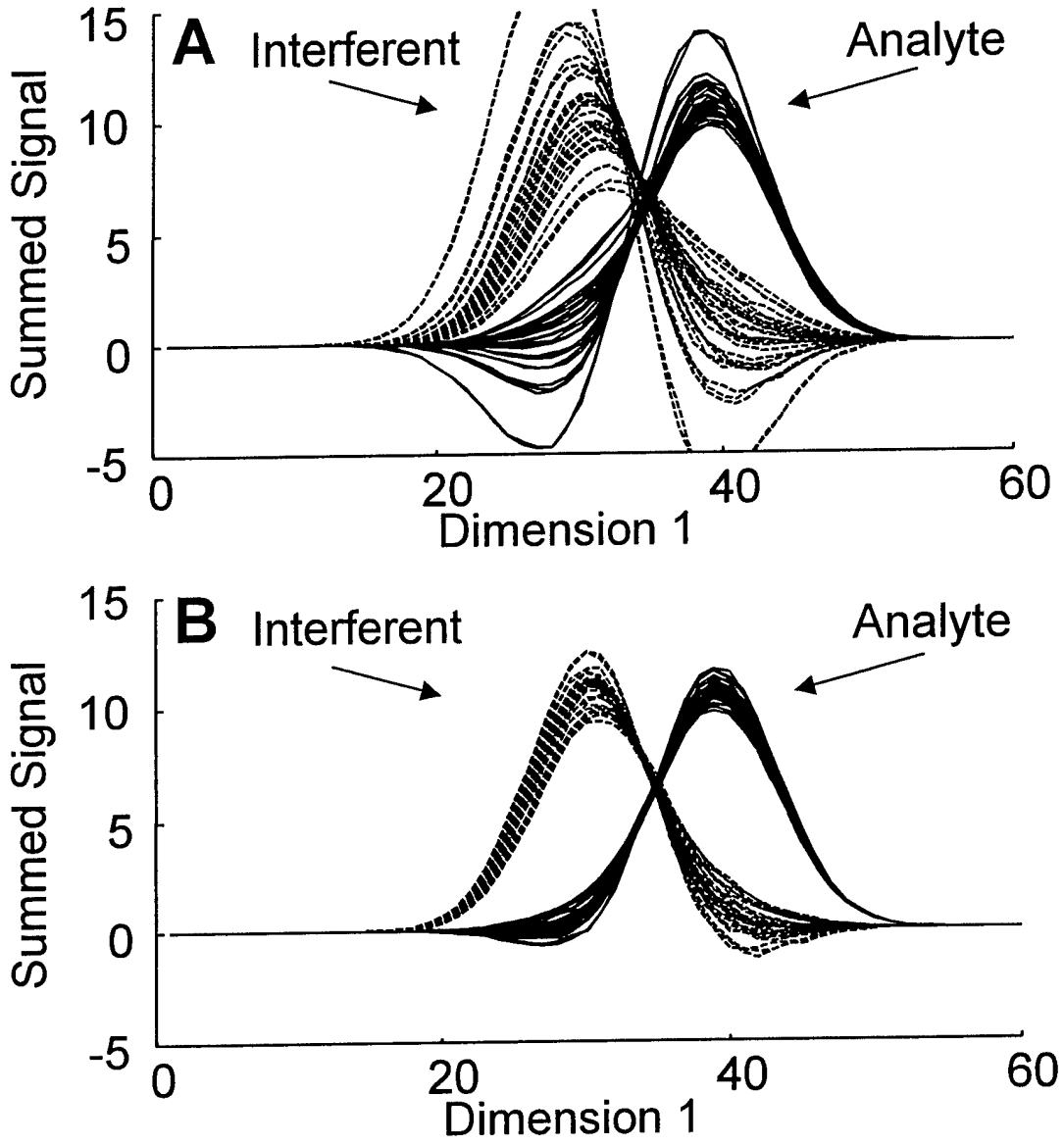



**Figure 2.2.** (A) Overlay of 30 summed deconvoluted peaks of the analyte and interferent obtained from the GRAM analysis of 30 randomly generated sample-standard combinations from study A having a resolution of 0.5 on each dimension (e.g., see Figures 2.1A and 2.1B) without objective retention time alignment. Each deconvoluted peak is the summation of the entire data matrix containing the resolved peak (e.g., see Figure 2.1C) onto the first dimension. (B) Overlay of 30 summed deconvoluted peaks obtained by GRAM analysis after the objective retention time alignment [Prazen et al., 1998] of each sample/standard combinations used in (A).

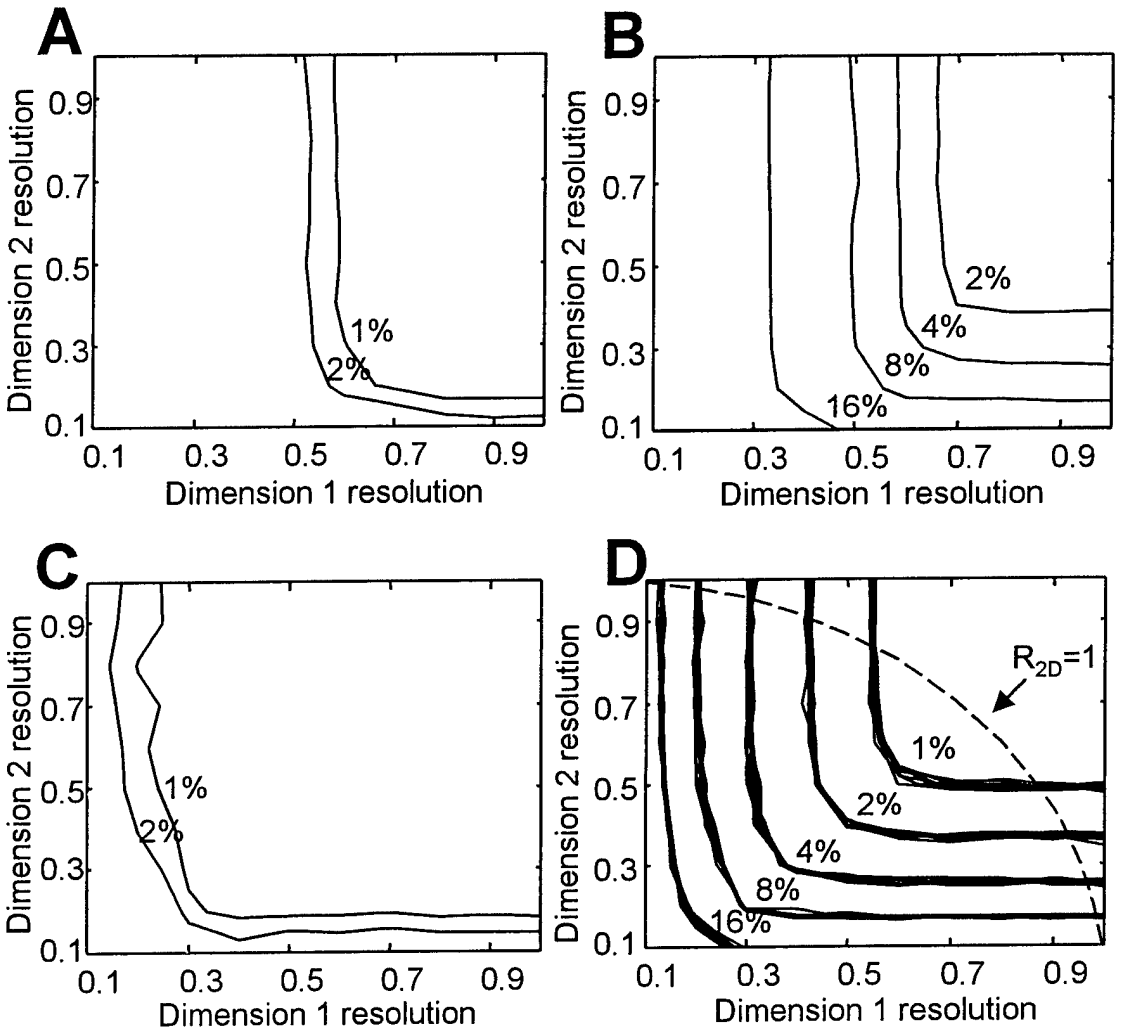



**Figure 2.3.** (A) Contour plot depicting the % absolute bias for GRAM quantification without retention time alignment at different dimensional resolutions from study A. (B) Corresponding contour plot of % RSD of GRAM analyses used to produce (A). (C) Contour plot of % absolute bias obtained with objective retention time alignment prior to GRAM analysis. Exactly the same sample/standard combinations used to generate the previous contour plots were used. (D) Corresponding contour plot of % RSD of GRAM analyses used to produce (C). Ten isograms per % RSD is depicted. The dashed line illustrate a $R_{2D}$ equal to one as defined by eq 2.3.

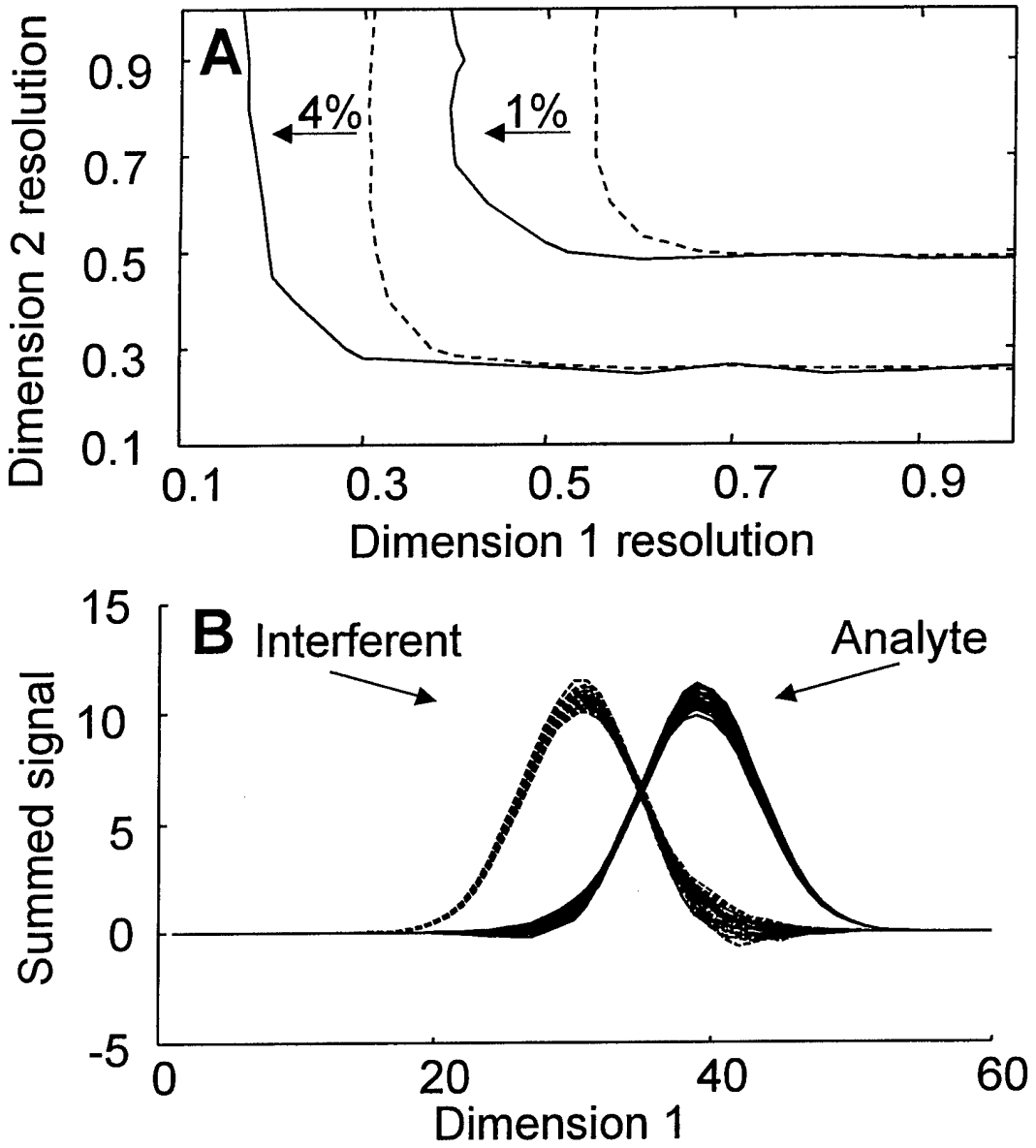



**Figure 2.4.** (A) Overlay of isograms for GRAM quantification precision obtained by applying interpolated retention time alignment (solid lines) and alignment without interpolation (dashed lines). Interpolation extends the isograms to lower resolutions along the first dimension because it more accurately estimates the true retention time shift. (B) Overlay of 30 summed deconvoluted peaks obtained by GRAM analysis after the interpolated retention time alignment of each sample/standard combinations used in Figure 2.2B. Peak shape accuracy and precision are better with interpolation than without.

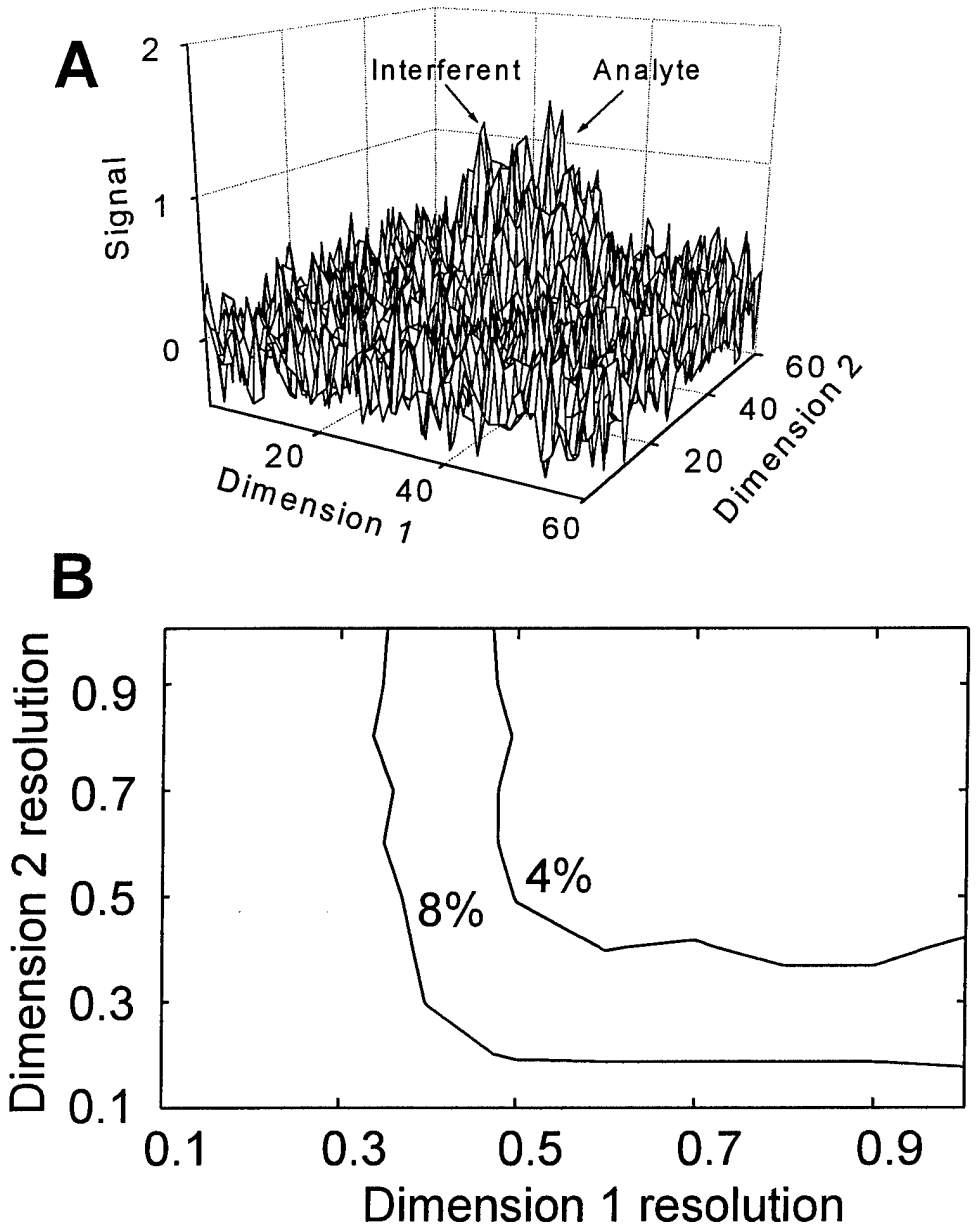



**Figure 2.5.** Evaluating the effect of low signal-to-noise on the GRAM analysis of unresolved peaks. (A) Representative simulation of a sample from study D having a S/N of 4.8 and a peak resolution of 0.5 on both dimensions. (B) Contour plot of % RSD obtained from study D for the retention time alignment and GRAM analysis at a S/N of 4.8. Note how the % RSD isograms shift to higher resolutions on both dimensions compared to the contour plot obtained with a S/N of 1000 (Figure 2.3D).

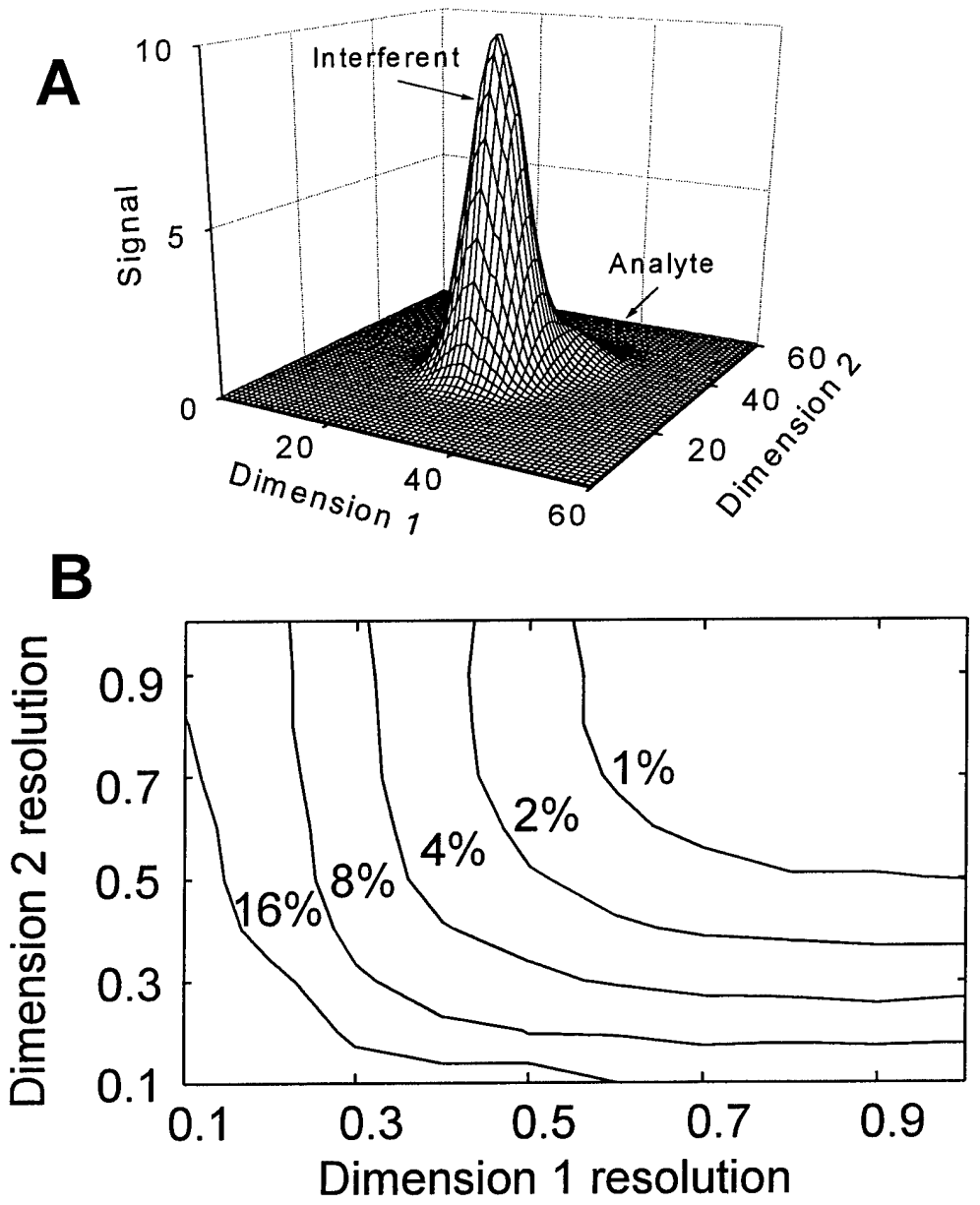



**Figure 2.6.** The effect on GRAM analysis of an interferent peak much larger than the analyte peak. (A) Representative simulation of a sample from study E having an interferent-to-analyte size ratio of 10-to-1, an analyte S/N of 1000 and peak resolution of 0.5 on both dimensions. (B) Contour plot of % RSD obtained from retention time alignment and GRAM analysis of sample/standard combinations from study E with a size ratio of 10-to-1. Larger but respectable % RSD values are obtained than when the interferent and analyte peak are of equal size (see Figure 2.3D).

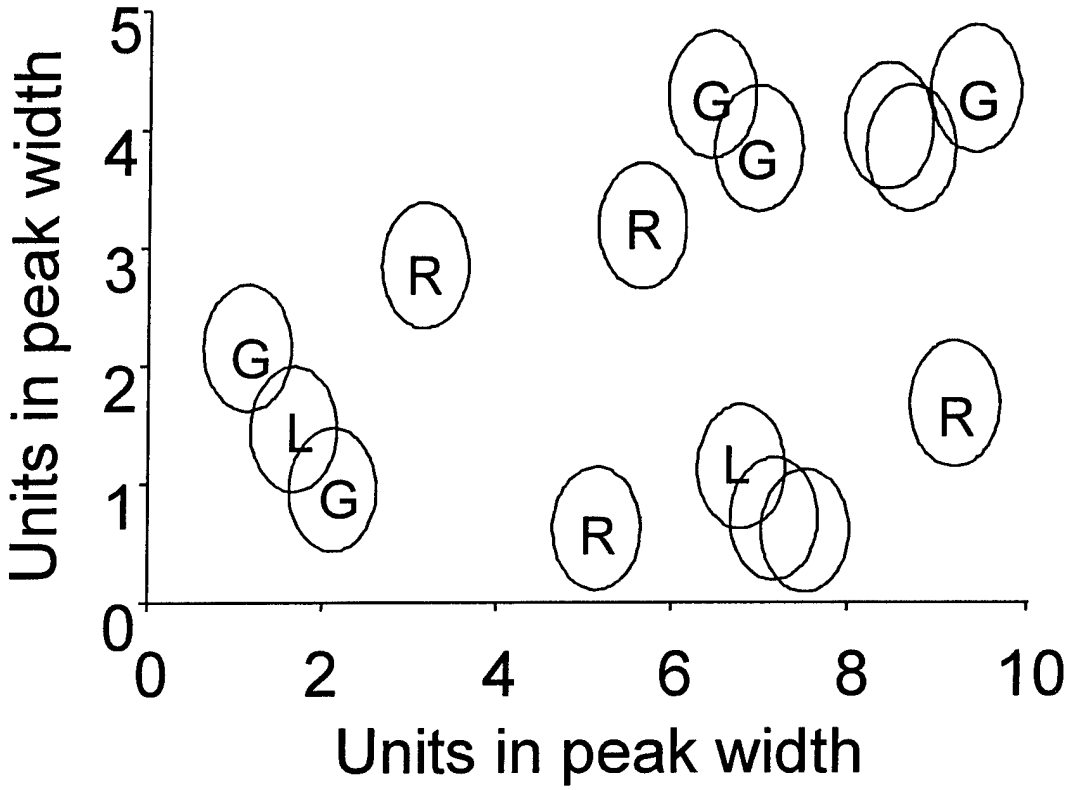



**Figure 2.7.** The boundaries of 15 randomly distributed peaks in a 2-D chromatogram with a peak capacity of 50. Only four peaks, labeled R, are considered resolved ($R_{2D} \geq 1$) for traditional quantitative analysis. Five additional peaks, labeled G, can be accurately analyzed by GRAM with a quantitative precision better than or equal to 4 %. These peaks have dimensional resolution that fall within the 4% dashed isogram in Figure 2.4A. The full benefit of GRAM is underestimated because two more peaks, labeled L, are likely analyzable by GRAM but are not considered for GRAM analysis in the current study. These peaks are overlapped with more than one peak and are therefore beyond the scope of the contour plot in Figure 2.4A.

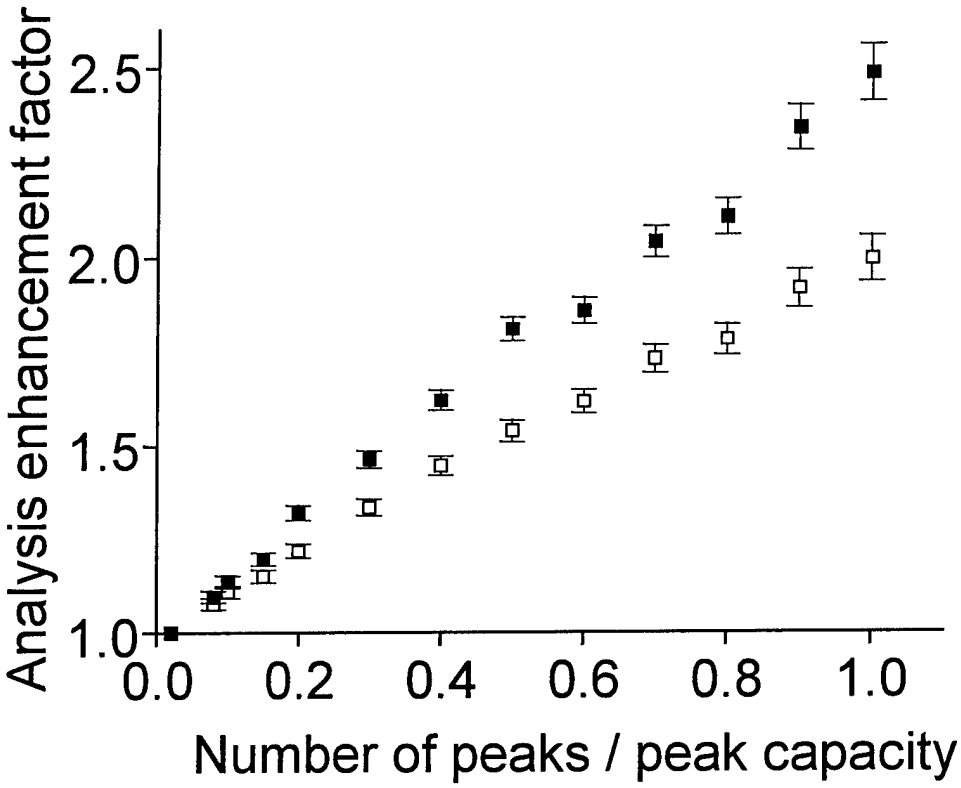



**Figure 2.8.** Analysis enhancement factor (AEF) as a function of peaks per peak capacity. The AEF quantifies the increase in the number of analyzable peaks achieved with GRAM analysis. An AEF of 2 means GRAM analysis doubles the number of analyzable peaks. Retention time alignment with (■) and without (□) interpolation is used prior to GRAM.

# Chapter 3: GRAM Analysis with Standard Addition and Retention Time Alignment

## 3.1 Introduction

In previous work, it was demonstrated that under favorable conditions GRAM can mathematically resolve overlapped GC x GC signals without any preprocessing to align data sets [Prazen et al., 1999b]. The reproducible retention times observed in this study were due to two factors. First, the strong similarity between the chemical matrices of the calibration standard and sample suppressed retention time shifts caused by matrix effects. Second, very short GC x GC run times minimized the run-to-run retention time variations caused by random fluctuations in instrumental conditions such as flow rate and temperature. Such favorable conditions are not always present in the analysis of complex samples encountered in real world applications.

In this chapter, two methods that address retention time and peak width variations that can occur in the GC x GC analysis of complex samples are described. The combination of these methods corrects the retention time variations occurring on the first column axis of GC x GC data making GC x GC data applicable to GRAM and other chemometric methods that compare multiple bilinear data sets. The first technique is the standard addition method. If standards for the analytes of interest are measured in a different chemical matrix than the sample, then analyte retention times and peak widths can differ significantly between a sample and standard [Dietz, 1996]. Standard addition overcomes this problem. The second method described is a retention time alignment algorithm that corrects for the small, unavoidable run-to-run retention time variations caused by fluctuating instrumental parameters such as flow rate and temperature. In brief, this chapter demonstrates that the standard addition method, a retention time alignment algorithm, and chemometric quantification can extend the applicability of GC x GC to the high-speed analyses of complex samples. In a broader sense, the methodology described herein should be of substantial benefit to other comprehensive separation methods such as LC x LC, LC x CE and LC x GC [Bushey and Jorgenson, 1990a; Bushey and Jorgenson, 1990b; Opiteck et al., 1997; Opiteck et

al., 1998a; Opiteck et al., 1998b; Liu and Sweedler, 1996; Hooker and Jorgenson, 1997; Quigley et al., 2000].

JP-TS jet fuel serves as an example of a complex sample to demonstrate the importance of the standard addition method and retention time alignment to the chemometric analysis of high-speed GC x GC data. Petroleum products like jet fuel have been analyzed extensively by GC x GC analyzers because of their environmental and industrial importance. GC x GC is well suited for the analysis of aromatics in petroleum products because of its selectivity for this class of compounds. Frysinger, et.al, recently demonstrated this attribute in the analysis of BTEX and total aromatic compounds in gasoline [Frysinger et al., 1999]. In this chapter, three aromatic compounds found in most petroleum fuels are identified and quantified in the GC x GC analysis of jet fuel. In order to reduce analysis times, a fast GC x GC analysis was done in part by not fully resolving the aromatic peaks of interest. GRAM was used after data collection to mathematically resolve and quantify the unresolved analytes. The GC x GC run time was at least five times faster than current GC methods developed for the analysis of aromatics in fuels. The analysis of JP-TS jet fuel also demonstrated the GRAM advantage of resolving and quantifying signals in the presence of interfering signals whose identity and quantity are not determined.

## 3.2 Theory

In this chapter, GRAM is used as a means to resolve incompletely separated GC x GC signals (see sections 1.2.1 - 1.2.3). GRAM compares GC x GC data obtained from a sample and a calibration standard to determine the pure signals for analytes present in both the sample and standard. In addition, GRAM determines the ratio of the sample and calibration standard GC x GC signals. With standard addition, the GRAM ratio is the concentration of the analyte in the sample divided by the concentration of the analyte plus standard in the spiked sample. Assuming a linear response to the analyte, the analyte concentration in the sample is calculated from the

GRAM ratio and the known quantity of analyte added or spiked into the original solution [Harris, 1995].

The requirements discussed in section 1.2.4 must be met for the successful application of GRAM. The most difficult of these is that the signal profile (i.e., peak shape, width, and retention time) of analytes must be consistent between the sample and the standard data sets. In chromatography run-to-run retention time shifts can reduce the applicability of GRAM and other multivariate techniques. In this chapter two approaches are described that reduce run-to-run retention time variation of GC x GC data sets and thus permit the application of GRAM.

Retention time fluctuations caused by typical instrument instability such as variations in column flow and temperature are corrected by an objective retention time alignment algorithm designed for bilinear chromatographic methods such as liquid chromatography with UV-vis detection [Prazen et al., 1998] and gas chromatography with mass spectrometric detection [Prazen et al., 1999b]. This method is based on using the bilinear structure of data and the relatively high repeatability of spectrometric measurements to correct for shifts along the chromatographic axis. The retention time alignment method is an iterative technique that finds the retention time difference or shift between a sample and standard data set containing analytes in common. The retention time shift is indicated by the minimum found in the percent residual variance versus retention time shift. The sample data set is then shifted along the chromatographic axis by the amount determined to align the sample and standard. This retention time alignment algorithm requires an estimation of the pseudorank of the sample data set and the expected maximum retention time shift. A detailed description of the algorithm is given in section 1.2.5. While this algorithm was originally designed for chromatographic methods coupled with spectrometric detection, it is successful on GC x GC data because the second GC column has the needed signal precision to behave like a spectrometric detector [Prazen et al., 1999b; Bruckner et al., 1998]. Hence, the retention time alignment algorithm can correct retention time shifts along the first column. Originally, retention times were aligned to the nearest data point on

the first column axis using this method. However in chapter 2, the use of data point interpolation allows the alignment of simulated data having retention times that are in part less than a data point. This chapter demonstrates the use of interpolated retention time alignment on real data. In addition, a modified retention time alignment algorithm for alignment along both chromatographic axes is introduced for the first time.

## 3.3 EXPERIMENTAL

### 3.3.1 Jet Fuel Sample

A neat sample of JP-TS jet fuel (U.S. Air Force Aerospace Fuel Laboratory, Mukilteo, WA) was analyzed. A standard addition sample of the JP-TS jet fuel was made by spiking the compounds isopropylbenzene, propylbenzene and 1,3,5-trimethylbenzene (Aldrich Co., Milwaukee, WI) to 60 mL of the neat jet fuel such that the added concentration (w/w) of each compound was 0.756%, 0.351% and 0.742%, respectively. Generally when using standard addition with GRAM, enough analyte should be added to the sample such that the relative peak height ratios between the analyte and interfering components are close to one or slightly higher. A standard not containing jet fuel was made by adding isopropylbenzene, propylbenzene and 1,3,5-trimethylbenzene to heptane such that the concentration (w/w) of each compound was 0.27 %. Replicate GC x GC runs were made of the samples and standard.

### 3.3.2 GC x GC Method

The diaphragm valve GC x GC system that was used in these experiments is described in section 1.1. The GC x GC system was set up on a Varian 3600cx gas chromatograph (Varian, Sugar Land, TX) with flame ionization detection (FID). An autosampler (Hewlett-Packard, Wilmington, DE) was used to inject 1-μL samples to a split/splitless injector at 250°C. Direct injection [Grant, 1996] of each sample was accomplished by operating in splitless mode during an entire 4-min GC x GC run. The column oven temperature was held at 75°C for the entire GC x GC run. In order to

clean the GC x GC system for the next run, the oven temperature was then increased at 40°C / min to 175 °C where it was held constant for 0.5 min. The detector was set at 250°C. The column 1 head pressure was 17.8 psi. Helium was the carrier gas. The first column of the GC x GC system was a 9.2-m x 530-μm i.d. capillary column with a 3-μm poly(dimethylsiloxane) film (SPB-1, Supelco, Bellefonte, PA). The second column was a 0.89-m x 180-μm i.d. column with a 0.15-μm poly(ethylene glycol) stationary phase (Carbowax, Quadrex Corp, New Haven, CT). The diaphragm valve (DV-12, 6 port, Valco Instrument Co. Inc., Houston, TX) was actuated for 15 ms every 800 ms during a 4-min GC x GC run. Thus, the valve repeatedly diverted a small portion of the first column effluent towards the second column. The effluent from the first column was split after the diaphragm valve between the second column and 0.5 m of 180-μm i.d. silica tubing.

### 3.3.3 Data Collection and Processing

The FID signal from the Varian was measured at a rate of 20,000 points/s by a data acquisition board (model AT-MIO-16XE-50, National Instruments, Austin, TX) connected to a PC running LabView 5.0 (National Instruments). The raw data was then boxcar averaged to 500 points/s and transferred into Matlab 5.2 (The Mathworks Inc, Natick, MA) where it was converted into a matrix of data (see Figure 3.1) such that each row of the matrix represented a fixed time on the second GC column and each column of the matrix represented a fixed time on the first GC column.

Two smaller sub-matrices were selected from each data matrix for GRAM analysis. The two sub-matrices are depicted in Figure 3.1 by the data within the two dashed rectangles. The first matrix was 136.0 to 185.6 s on the first column axis and 0.96 to 1.27 s on the second column axis which corresponded to 63- by 156-points in size. The first matrix contained the signals for propylbenzene, 1,3,5-trimethylbenzene and at least two interfering components. The second matrix was 114.4 to 133.6 s on the first column axis and 0.88 to 1.02 s on the second column axis which corresponded

to 24- by 70-points in size. The second matrix contained the signal for isopropylbenzene which had a portion of its signal overlapped with the trailing edge of a much larger peak.

Prior to GRAM analysis, each neat jet fuel matrix was retention time aligned to its corresponding matrix from the spiked jet fuel sample using the retention time alignment algorithm discussed in the theory section and in section 1.2.5. In the alignment algorithm, the cubic interpolation function from Matlab was used to interpolate 4 data points between each data point collected along the first column axis. This increase in the data density allowed for a more accurate determination of the retention time shift between the neat and spiked jet fuel matrices than alignment without interpolation.

Two data matrices were analyzed by GRAM for each of the neat and spiked jet fuel samples. One contained the signals for isopropylbenzene and the other contained the signals of propylbenzene and 1,3,5-trimethylbenzene. Each GRAM analysis involved one neat jet fuel matrix as the sample and one spiked jet fuel matrix as the calibration standard. The GRAM resolved signals and sample-to-standard signal ratios for each analyte of interest were obtained. The concentrations of isopropylbenzene, propylbenzene and 1,3,5-trimethylbenzene in the jet fuel sample were obtained by solving eq. (2) for $[X]_{sample}$ given $G_x$, D and $[X]_{added}$.

$$G_x = \frac{\text{Signal}\quad X_{sample}}{\text{Signal}\quad X_{standard}} = \frac{[X]_{sample}}{[X]_{added} + D \times [X]_{sample}} \qquad (3.1)$$

$G_x$ is the GRAM signal ratio for analyte X, [X] is the concentration of X, and D is the factor by which [X] in the sample is diluted by the addition of X.

Data preprocessing and GRAM analysis in Matlab took less than 45 sec for each sample-standard pair. The retention time alignment with data point interpolation accounted for 99 % of the data analysis time. Future software versions can be written to be more time efficient.

### 3.3.4 Reference GC Method

A reference single column GC method was used to validate GC x GC quantification. The neat and spiked jet fuel samples were analyzed by a HP 6890 gas chromatograph (Hewlett-Packard) equipped with FID. An autosampler was used to inject 1-μL samples to a split/splitless injector at 250 °C with a 100 : 1 split ratio. Helium was used as the carrier gas at a constant flow of 0.6 ml / min. The column oven temperature was held at 70 °C for 2 min, then increased at 1.0 °C / min to 85 °C and then increased at 35 °C / min to 145 °C where it held constant for 15 min. The detector was set at 250 °C. The GC column was a 60 m x 250-μm i.d. x 0.44-μm TCEP capillary column (Supelco, Bellefonte, PA). The GC column for the single column reference method is a relatively long, polar column specifically designed for the separation of aromatics in mineral spirits as described in Supelco application note No. 23 (Supelco). Propylbenzene and 1,3,5-trimethylbenzene were adequately resolved (see Figure 3.2) and their concentrations in the JP-TS jet fuel sample were determined by the standard addition method. The neat-to-spike signal ratio needed for standard addition calculations was determined by the ratio of the analyte's mean peak area in the neat and spiked jet fuel samples. Four replicate analyses of each sample were made. A PC running HP Chemstation rev. A.06.03 (Hewlett-Packard) was used to measure peak areas.

## 3.4 Results and Discussions

The goal of this chapter is to demonstrate that standard addition and an objective retention time alignment program can extend the applicability of GC x GC to high-speed quantitative chemometric analyses of complex sample matrices. First, the standard addition method is demonstrated as a means to reduce the retention time and peak width variability caused by matrix effects. Second, an objective retention time correction algorithm is demonstrated as a means to reduce retention time variations

caused by run-to-run instrumental instability and thus improve the qualitative and quantitative information obtained.

### 3.4.1 GC x GC and GC Comparison

The three aromatic isomers in the JP-TS jet fuel sample used in this study were analyzed by single column gas chromatography to demonstrate the high-speed advantage of GC x GC with GRAM analysis, and to serve as a reference method to verify the accuracy of GC x GC with GRAM analysis. Figure 3.2 depicts the single GC column chromatogram for JP-TS jet fuel. The signals for the aromatic isomers isopropylbenzene, propylbenzene and 1,3,5-trimethylbenzene in JP-TS jet fuel are labeled in Figure 3.2.

The high-speed advantage of GC x GC can be seen by comparing the separation time for the three aromatic isomers in Figure 3.1 and Figure 3.2. A separation time of 2.8 min is needed to elute the three aromatic compounds with GC x GC (see Figure 3.1) while 14.4 min is needed using the reference GC method (see Figure 3.2). Note that only two of the three aromatic isomers are adequately resolved for quantification using the reference GC method while all three isomers in the GC x GC separation can be quantified by chemometric analysis as is subsequently demonstrated. Other GC methods that fully resolve the three aromatic isomers in gasoline, which has a similar chemical composition as jet fuel, fare even worse in terms of separation time. For example, a separation time of 17.4 minutes was needed using a single column GC-FID method described in Chrompack GC application note 31 (Chrompack International, Middelburg, The Netherlands ) and 35 min using a thermal modulation GC x GC [Frysinger et al., 1999]. Note that the thermal modulation GC x GC system was probably not optimized to demonstrate speed, but serves as a good example of the current state-of-the-art.

### 3.4.2 Standard Addition Method

Significant differences between the chemical matrix of samples can cause analytes to have considerably different peak shapes, widths and retention times on the first column of GC x GC data. These matrix effects can occur in the GC x GC analysis of complex samples such as jet fuel where the matrices of the sample and calibration standard can differ considerably. Figure 3.3 displays two-overlaid GC x GC chromatograms obtained by running a JP-TS jet fuel sample and a calibration standard containing isopropylbenzene, propylbenzene and 1,3,5-trimethylbenzene in heptane. The chemical matrices of the jet fuel and calibration standard differ enough that the three aromatic analytes experience a significantly different separation process with each chemical matrix. For instance, the difference in retention times seen between the aromatic analytes in jet fuel and those in heptane (see Figure 3.3) is due primarily to the presence of the large unresolved band of components found below the aromatic analytes in the GC x GC chromatogram of jet fuel (see Figure 3.1). This unresolved band of components traveled with the three aromatic analytes through the first GC column altering the column's chromatographic properties and changing retention times [Dietz, 1996]. Analysis by the standard addition method would eliminate this matrix effect. The presence of the heptane solvent in the calibration standard can alter the peak shapes and/or reduce the peak widths of the three aromatic analytes by solvent trapping in the first column inlet [Grob, 1991]. Indeed, the three aromatic analytes in heptane (see Figure 3.3) had peak widths on the first column that where 50 to 70 % of their peak widths in jet fuel determined by GRAM analysis. No obvious changes in peak shapes were observed. The wider peaks in jet fuel indicate that some separation efficiency is lost by direct injection. A split injection would have improved separation efficiency at the expense of losing sample through the injector's split vent. Thus, the separation efficiency of the first GC column needed to be somewhat reduced in order to maximize detection sensitivity of the GC x GC system.

Matrix effects do not affect peaks along the second column dimension because most matrix effects are removed by the first column separation. The sharp injections

provided by the diaphragm valve also mitigate matrix effects by diverting a small slice of the first column components onto the second column. Any significant change in peak shape, width, or retention time violates the GRAM requirement for a constant signal profile for each analyte common to a sample and calibration standard [Bruckner et al., 1998; Fraga et al., 2000b]. Such changes in peak characteristics reduce the applicability of GRAM and other similar bilinear chemometric techniques. The use of standard addition would eliminate matrix effects that cause changes in an analyte's signal profile.

Figures 3.4A and 3.4B depict representative GC x GC data sets for the sample and standard in the standard addition experiment of JP-TS jet fuel. Each data set is a subset of a GC x GC run containing the signals of propylbenzene and 1,3,5-trimethylbenzene in JP-TS jet fuel. Although not immediately evident in these figures, the constancy of signal profiles has been achieved through standard addition. The run-to-run retention time differences for the standard addition data are three times less than those obtained without standard addition (see Figure 3.3). Major differences in peak shape and width are also absent even though differences cannot be initially observed in Figures 3.4A and 3.4B because the individual chemical components are not resolved. Improved run-to-run retention time and signal profile precision achieved by standard addition makes the successful application of GRAM more likely, as will now be demonstrated.

### 3.4.3 Retention Time Alignment

GRAM analysis of the jet fuel data (Figure 3.4A) with the spiked jet fuel data as the calibration standard (Figure 3.4B) resulted in the resolved signals of propylbenzene and 1,3,5-trimethylbenzene in jet fuel depicted in Figure 3.4C along with their concentrations in jet fuel. In an analysis like this one, where quantitative information must be obtained from signals that are highly overlapped with multiple interfering signals, run-to-run retention time precision becomes an important factor in determining the quality of the quantitative information. Prior to GRAM analysis the

two data matrices from Figure 3.4A and 3.4B were aligned using the retention time alignment algorithm [Prazen et al., 1998; Prazen et al., 1999b]. Figure 3.5 depicts the alignment profile used to align the matrices shown in Figure 3.4A and 3.4B. All other alignment profiles obtained in this work closely resembled the profile depicted in Figure 3.5. Application of the retention time alignment algorithm serves as a second order correction following use of the standard addition method. In Figure 3.5 the data point at the profile minimum indicates that the jet fuel signals are shifted 1.92 s towards an earlier retention time relative to the spiked jet fuel matrix in this example. The retention time alignment algorithm included a 4-point cubic interpolation of data along the first column axis. Interpolation allows for a more accurate determination of the actual retention time difference between the two jet fuel data sets than has been previously achieved. In the example shown in Figure 3.5, shifting the sample matrix 1.92 s in the appropriate direction along the first column axis aligned the data sets. If the alignment algorithm was implemented without interpolation a shift of 1.6 s would have been the most accurate alignment possible.

The retention time alignment of jet fuel data dramatically improved GRAM signal deconvolution. Figures 3.6A, 3.6B, 3.6C and 3.6D depict 25 overlaid GRAM resolved chromatographic profiles per analyte for propylbenzene and 1,3,5-trimethylbenzene. These were obtained from GRAM analyses on 25 different combinations of five neat and five spiked jet fuel data sets. For clarity, the GRAM calculated signals have been reduced from a matrix of data, like that shown in Figure 3.4C, to a vector by integrating each analyte signal onto each column axis. Figures 3.6A and 3.6B were calculated without retention time alignment and Figures 3.6C and 3.6D were calculated with retention time alignment. There is a pronounced improvement in chromatographic signal shape and in the reproducibility of the appropriate peak shape when the retention time alignment program is used prior to GRAM. The shape of a resolved chromatographic signal should be unimodal (i.e., one peak) and non-negative (i.e., always above the baseline). Only when these requirements are met can the user feel confident that GRAM has accurately identified

the true signal. Figures 3.6A and 3.6B clearly depict peaks that do not have proper shapes. Many of the peaks are neither unimodal nor non-negative. However, after retention time alignment (see Figures 3.6C and 3.6D), the peak shapes are sufficiently unimodal and nonnegative suggesting that the true signal was accurately identified by GRAM. Nearly the same signal is found for each of the 25 replicate analyses. Thus, the 25-GRAM signals have reproducible signal profiles and appropriate peak shapes. The remaining irreproducibility in the GRAM signals is due to experimental uncertainty not corrected by retention time alignment such as injection volume irreproducibility and retention time variations among spiked jet fuel replicate runs. Only retention time shifts between neat and spiked jet fuels were corrected in this study.

The largest improvement in signal shape is seen for 1,3,5-trimethylbenzene. This is probably because it was shifted further away from its signal in the spiked jet fuel sample than propylbenzene. This occurs because for isothermal GC x GC runs the degree of retention time shift along the first column, as measured by the retention time standard deviation, increases almost proportionally with a peak's mean retention time. However, peak width increases by the square root of the retention time [Miller, 1988]. Hence, the later eluting 1,3,5-trimethylbenzene peak had approximately the same peak width as propylbenzene but a larger retention time shift. Propylbenzene and 1,3,5-trimethylbenzene were analyzed simultaneously (i.e., one data matrix), because both are contained in a single unresolved group of peaks in the GC x GC separation (see Figure 3.1). Thus, only one retention time correction value was determined by the retention time alignment algorithm. Quantitative results indicate that the retention time correction obtained through the single matrix lead to quantitative results that are within the experimental uncertainty.

A study was done to quantify the accuracy of the second-column GRAM profiles (i.e., peak shape, width and retention time) determined with and without retention time alignment. To quantify the accuracy of GRAM profiles, part of the signal that is exclusively due to the analyte propylbenzene was used as the true signal

profile for propylbenzene. The pure elution slice used was at 144.8 s on the first column time axis. Because data from the diaphragm valve GC x GC system is bilinear, the true signal for propylbenzene has the same signal profile at all the second column slices that span the propylbenzene signal. Figure 3.7A depicts the true signal for propylbenzene from the JP-TS jet fuel data set depicted in Figure 3.4A and the GRAM signals for propylbenzene determined with and without retention time alignment. The GRAM signal for propylbenzene with alignment is the mean signal on the second column dimension determined by GRAM with retention time alignment (see Figure 3.6D). The GRAM signal for propylbenzene without alignment is the mean signal determined by GRAM without retention time alignment (see Figure 3.6B). In Figure 3.7A, the profile for the true signal is compared to the profiles of the GRAM signals with and without retention time alignment in order to determine the accuracy of the GRAM signal profiles. Signal heights were normalized to facilitate the comparisons of signal profiles. The mean signal profile produced by GRAM with retention time alignment has an appropriate peak shape and closely resembles the true signal profile. This is confirmed by measuring the correlation coefficient (r) between the true signal profile and each mean signal profile produced by GRAM. As the match between the true signal profile and the GRAM profile becomes more accurate, r approaches one [Felinger, 1998]. The mean GRAM profile with alignment had an r of 0.9972 and without alignment an r of 0.9780. Retention time alignment permits GRAM to obtain a more accurate signal profile, which is critical to properly identify the analyte peak(s) of interest.

The relationship between retention time alignment and GRAM profile accuracy is shown in Figure 3.7B. Correlation coefficients were calculated between the true propylbenzene signal profile and each of the 25 GRAM calculated profiles from the combination of five sample and five standard data sets. Figure 3.7B depicts the accuracy of the GRAM profiles compared with the absolute retention time shift as determined by the retention time alignment algorithm. A linear decline in signal profile accuracy as the retention time shift increases is seen for the analyses in which

retention times are not aligned. This trend has been previously noted using simulated bilinear chromatographic data [Juan et al., 1998]. Correcting the retention time shift is shown to improve the signal profile accuracy for all the combinations especially for those with large retention time shifts between the sample and the sample spiked with the standard addition.

Quantification accuracy and precision improvements resulting from the retention time alignment algorithm were also studied. Table 3.1 depicts the quantification results achieved with and without retention time alignment for isopropylbenzene, propylbenzene and 1,3,5-trimethylbenzene in JP-TS jet fuel. The reported GRAM concentration for each compound is the mean concentration of 25-GRAM analyses originating from five replicate runs each of the neat and spiked jet fuel samples. Each reference concentration was determined from the single GC column method described in the experimental section (see Figure 3.2). The single column GC method was unable to quantify isopropylbenzene in JP-TS jet fuel because of significant signal overlap. The retention time alignment algorithm clearly improved the accuracy of GRAM quantification for propylbenzene and 1,3,5-trimethylbenzene. This is to be expected because a more accurate representation of the true signal profile is achieved when using GRAM with retention time alignment (see Figure 3.7A). The precision was also dramatically improved for 1,3,5-trimethylbenzene as indicated by the four-fold reduction in the relative standard deviation when using alignment.

### 3.4.4 2-D Retention Time Alignment

The sample-standard data sets used in this chapter have very reproducible run-to-run retention times on the second column. Only the first column has poor signal reproducibility. This is not always the case. Relatively poor run-to-run signal precision on the second column has been reported with LC x CE data [Bushey and Jorgenson, 1990b]. Bushey and Jorgenson note that within one LC x CE run, the RSD for CE migration times is insignificant. Indeed, the LC x CE signals look bilinear because of the good within-run migration time precision of the CE separation.

However, the run-to-run CE migration times have a significant RSD of 1.18%. The GC x GC data in this chapter, have an RSD of 0.1% for the run-to-run retention times on the second column. However with LC x CE, larger RSD values are anticipated on the second dimension, i.e., CE column. In order to correct retention time shifts on two dimensions, the retention time algorithm is modified.

The retention time alignment algorithm (see section 1.2.5) can be modified to correct shifts on two dimensions because in terms of rank analysis, signal shifts between **M** and **N** along one dimension are irrelevant to shifts on the other. For instance, in Figure 1.6A the rank of the augmented matrix $\begin{bmatrix} \mathbf{M} \\ \mathbf{N} \end{bmatrix}$ is two when the signals in **M** and **N** are align on the chromatographic dimension even if they are not align on the other chromatographic dimension. Hence, the modified alignment algorithm first corrects retention time shifts on the first column using the augmented matrix $\begin{bmatrix} \mathbf{M} \\ \mathbf{N} \end{bmatrix}$ and then on the second using the augmented matrix [**M** | **N**]. The modified algorithm is discussed below.

Given a sample data matrix, **M**, and a standard data matrix, **N**, of equal dimensions, the modified algorithm first augments the two matrices producing a $\begin{bmatrix} \mathbf{M} \\ \mathbf{N} \end{bmatrix}$ matrix that has its second column axis twice as long as it is in **M**. Next, the residual variance of the $\begin{bmatrix} \mathbf{M} \\ \mathbf{N} \end{bmatrix}$ matrix as a function of shifted first-column data points is determined. This portion of the algorithm is no different from the original in section 1.2.5. However after the first column shift is determined from the percent residual variance plot of $\begin{bmatrix} \mathbf{M} \\ \mathbf{N} \end{bmatrix}$, then **M** and **N** are augmented such that the first column axis is made twice as long as it is in **M**. The resulting [**M** | **N**] matrix is then subjected to the original algorithm whereby the percent residual variance of the [**M** | **N**] matrix as a function of shifted second-column data points is determined. The profile's minimum

from the percent residual variance plot dictates the needed shift correction along the second column axis.

Applying the modified retention time algorithm to the GC x GC data in this chapter made no impact, as one would expect with reproducible second column retention times. However in order to demonstrate the modified alignment algorithm, the data from the five replicate GC x GC runs of the neat jet fuel sample were artificially shifted on both column axes. Figure 3.8A is a contour plot depicting five overlaid GC x GC signals each containing the overlapped signals of the four isomers depicted in Figure 3.4A. Figure 3.8B depicts the five signals after retention time alignment with the original algorithm. Figure 3.8C depicts the five signals after retention time alignment on both dimensions using the modified alignment algorithm. This modified algorithm should expand the use of GRAM to more comprehensive 2-D analyses, particularly in LC x CE.

## 3.5 Conclusions

Retention time precision has been a major obstacle in applying chemometrics to all types of chromatographic data. The use of the standard addition method and the retention time alignment algorithm make it feasible to use GRAM and other bilinear chemometric methods on GC x GC data from complex samples. Standard addition method ensures that major changes in peak shapes, widths and retention times caused by matrix effects are corrected before bilinear chemometric analyses. The objective retention time alignment program corrects for the slight retention time shifts caused by instrumental variations. The time associated with standard addition and chemometric analysis (i.e., data preprocessing, alignment and GRAM) does lessen the dramatic impact of short GC x GC separation times. However, having to make one standard addition sample for each sample to be analyzed can be avoided in cases where samples with similar chemical matrices are to be analyzed such as in industrial and environmental applications. In these cases, after standard addition and chemometrics is

used to analyze the first sample, the first sample's GC x GC data can be used as the calibration standard for subsequent sample analyses. This first sample would not need another GC x GC run until deemed necessary to recalibrate the GC x GC instrument. Faster computers and efficient computer algorithms are consistently reducing chemometric analysis time. Coupling GC x GC and chemometrics has the potential for the routine high-speed analysis of complex samples.

**Table 3.1. Quantifying selected compounds in JP-TS jet fuel with and without retention time alignment**

| Compound | Reference (%, w/w)[a] | GRAM (%, w/w)[b] without alignment [ Bias, % ][c] | GRAM (%, w/w)[b] with alignment [ Bias, % ][c] | GRAM RSD[d] (%) without alignment | GRAM RSD[d] with alignment (%) |
|---|---|---|---|---|---|
| isopropylbenzene | NA | 0.148 [ NA ] | 0.158 [ NA ] | 3.4 | 3.2 |
| propylbenzene | 0.229 | 0.240 [ 4.8 ] | 0.230 [ 0.44 ] | 3.8 | 3.1 |
| 1,3,5-trimethylbenzene | 0.790 | 0.932 [ 18 ] | 0.823 [ 4.2 ] | 13.0 | 3.2 |

[a] Quantifying resolved peaks by peak area using a reference single column GC method and the standard addition method
[b] Quantifying unresolved GC x GC peaks by GRAM and the standard addition method; mean concentration reported
[c] Difference between GRAM and reference concentration divided by reference concentration
[d] RSD is the relative standard deviation using 25 GRAM concentrations
NA : Not applicable; compound could not be quantified by reference GC method because of signal interference

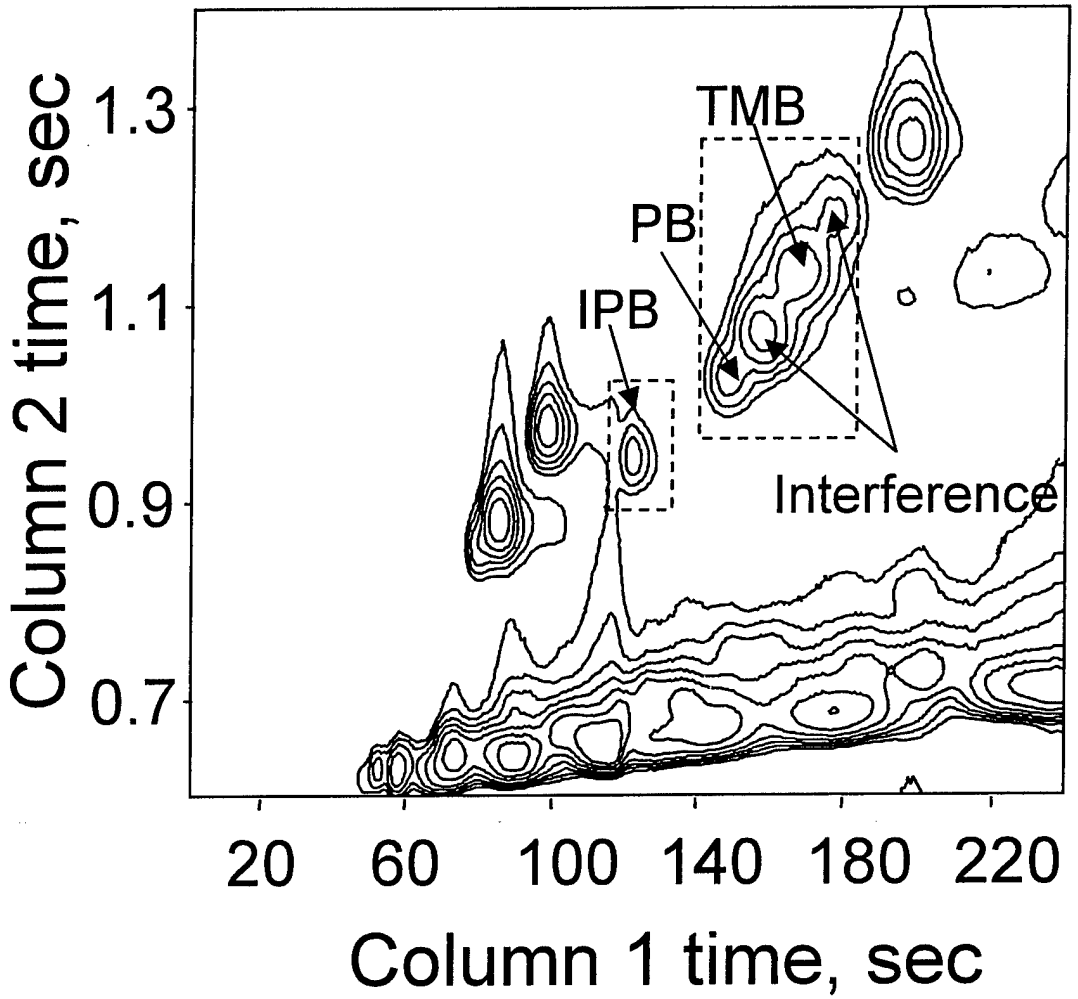



**Figure 3.1.** Contour plot of data from the GC x GC analysis of JP-TS jet fuel. The data within each dashed rectangle contain the unresolved signals of the aromatic analytes isopropylbenzene (IPB), propylbenzene (PB) and 1,3,5-trimethylbenzene (TMB). The unresolved signals for IPB, PB and TMB are used to quantify each analyte using GRAM and the standard addition method. The interference are aromatic compounds whose identities are known [12] but are treated as unknowns for the sake of demonstrating the GRAM analysis.

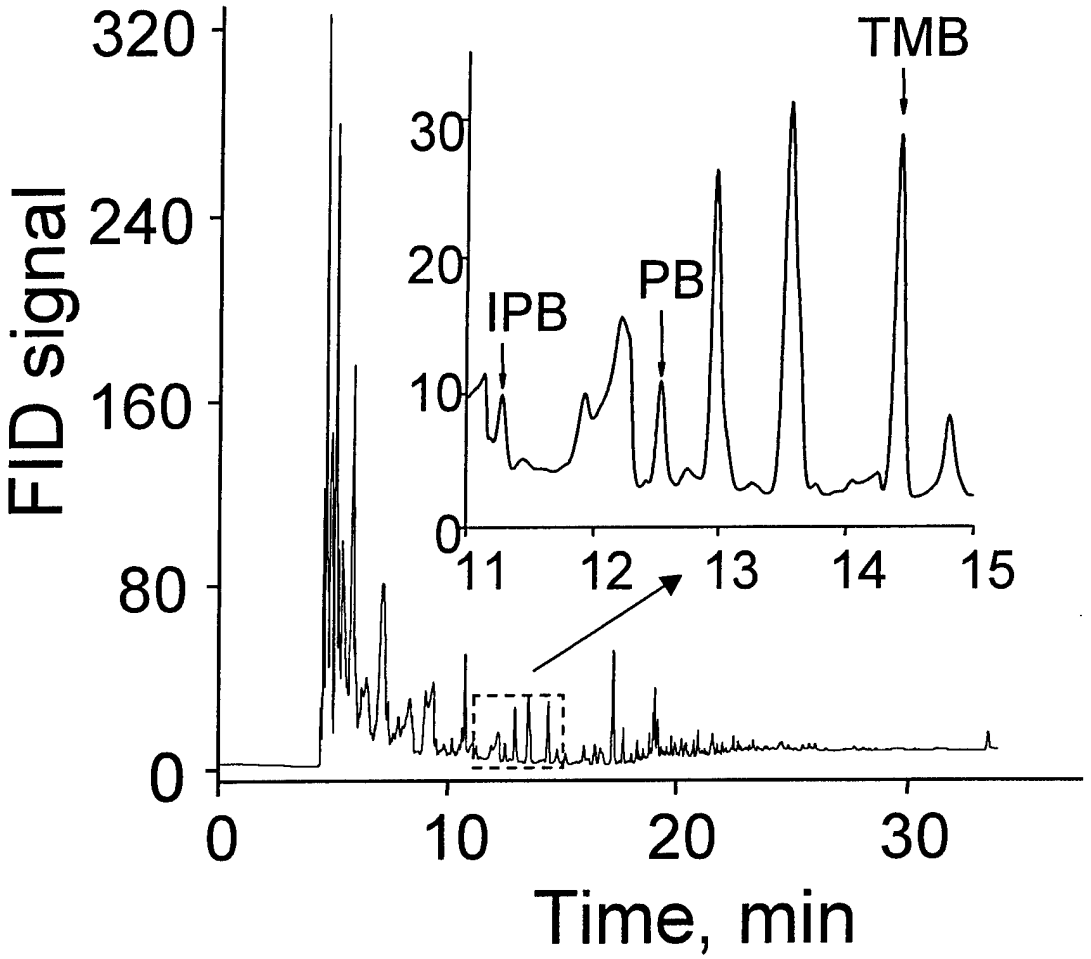



**Figure 3.2.** Single column reference chromatogram of JP-TS jet fuel. The signals for the aromatic analytes IPB, PB and TMB are labeled. The signals for PB and TMB are adequately resolved for quantification using peak area and the standard addition method. The signal for IPB is not adequately resolved for quantification. The column temperature was dramatically increased at 17 minutes to more quickly elute later eluting JP-TS jet fuel components.

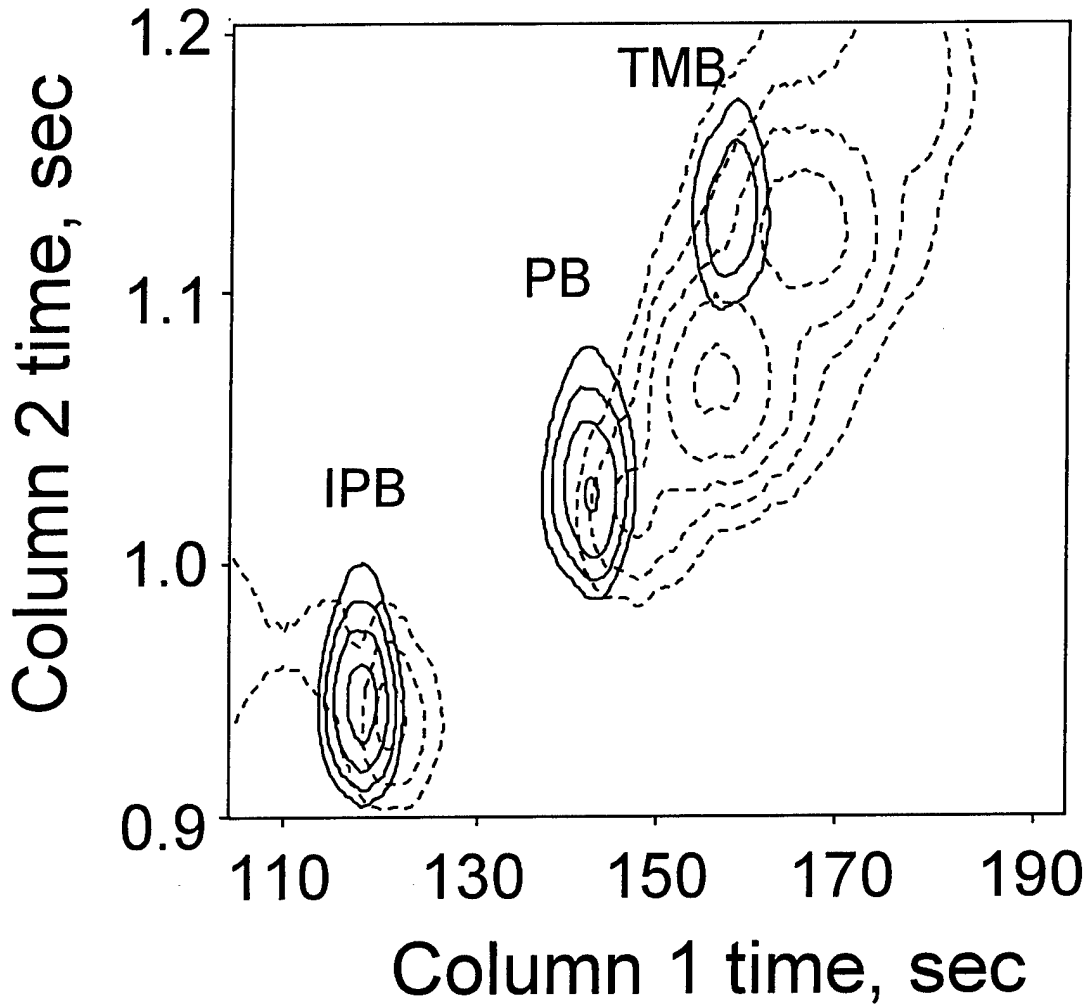



**Figure 3.3.** Contour plot of the GC x GC data from a JP-TS jet fuel sample (dashed lines) overlaid with a standard (solid lines) containing IPB, PB and TMB in heptane. This comparison demonstrates the need for standard addition techniques in order to correct major differences in retention times and peak widths that would inhibit the chemometric analysis of the GC x GC data.

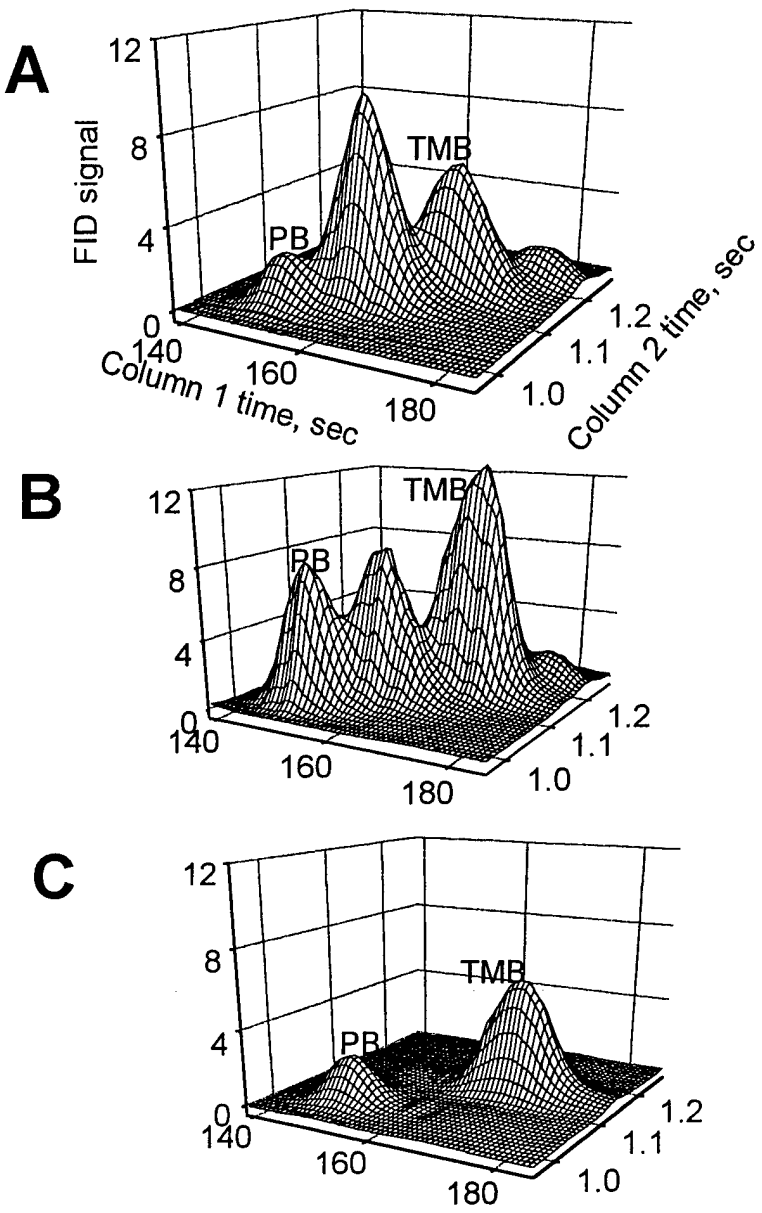



**Figure 3.4.** (A) GC x GC data containing PB and TMB signals from a JP-TS jet fuel sample. (B) GC x GC data containing PB and TMB signals from JP-TS jet fuel with standard addition. (C) GRAM calculated signals for PB and TMB from (A) using the retention time alignment algorithm to correct for the retention time differences between (A) and (B).

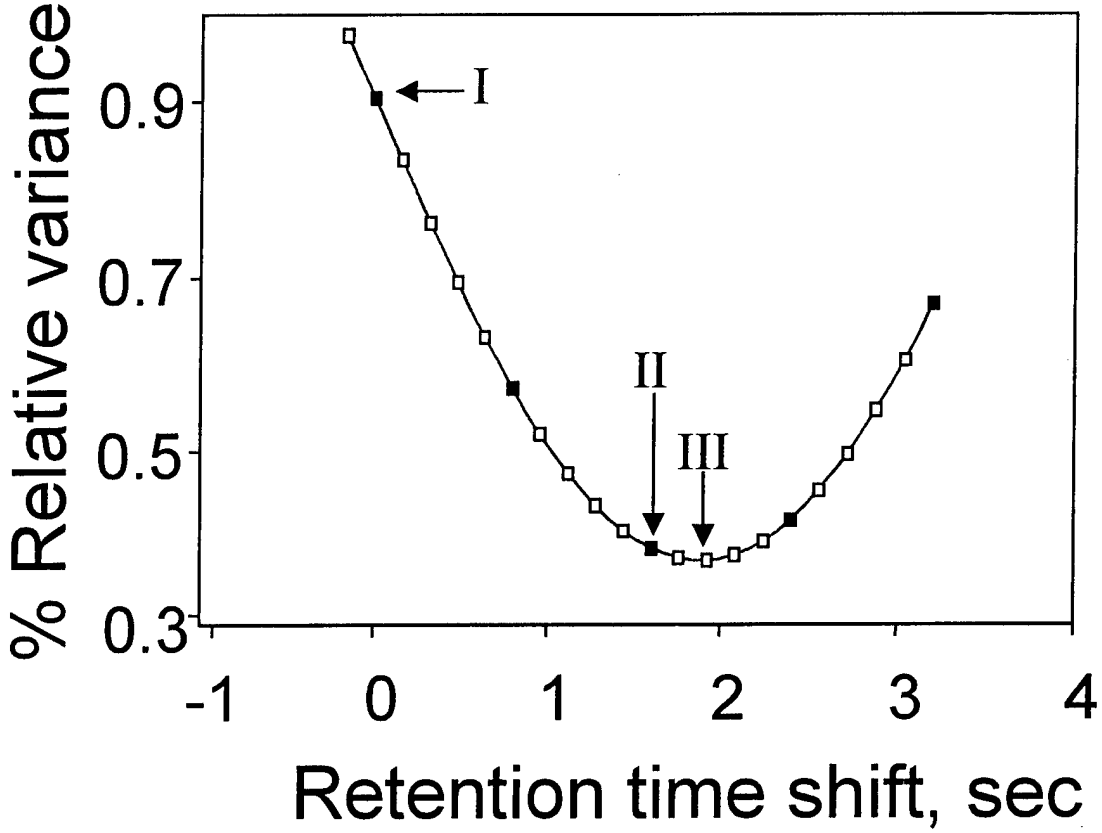



**Figure 3.5.** Bilinear chromatographic alignment profile used for the retention time alignment of the JP-TS jet fuel data in Figure 3.4A to the JP-TS jet fuel with standard addition data in Figure 3.4B. The solid squares represent shifts of data point increments while the open squares represent interpolated points. (I) marks the alignment without applying retention time alignment, (II) marks the alignment without interpolation, and (III) marks the profile's interpolated alignment indicating the required retention time correction. Interpolation allows for a more accurate determination of the actual retention time difference between the two jet fuel data sets.

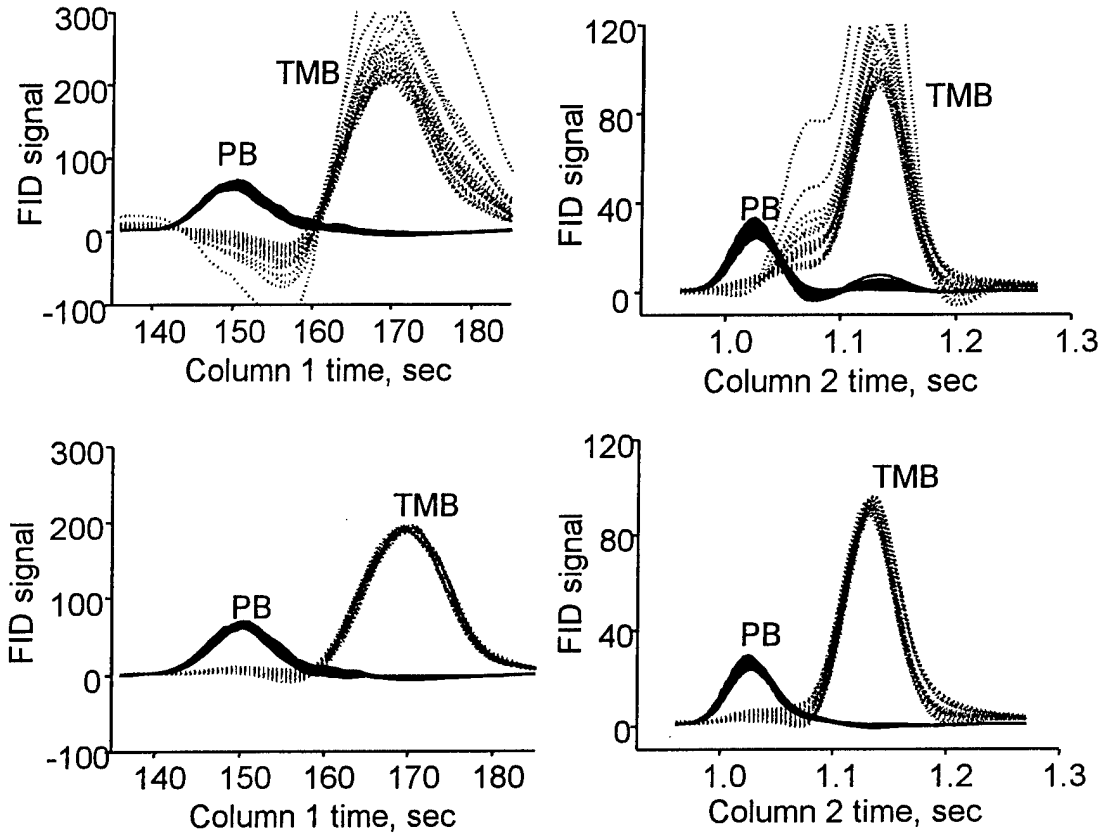



**Figure 3.6.** (A) Overlay of 25 GRAM signals (see Figure 3.4C) per analyte determined without retention time alignment for analytes PB (solid line) and TMB (dotted line) summed onto the first column axis. (B) Overlay of GRAM signals used in (A) summed onto second column axis. (C) Overlay of 25 GRAM signals per analyte determined with retention time alignment for analytes PB (solid line) and TMB (dotted line) summed onto the first column axis. (D) Overlay of GRAM signals used in (C) summed onto the second column axis. Retention time alignment dramatically improves the signal profiles on both chromatographic axes.

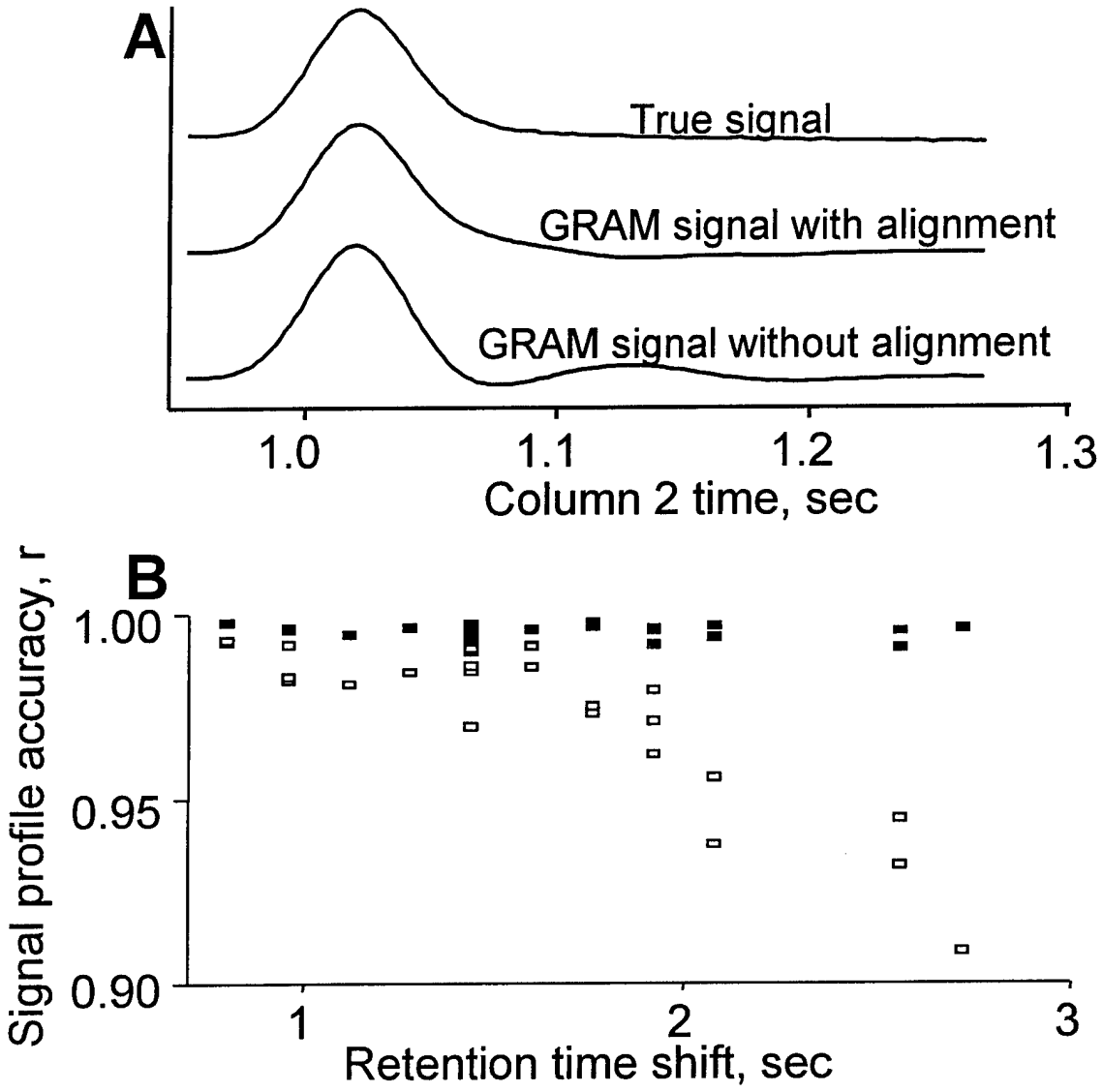



**Figure 3.7.** (A) The true signal for PB and the GRAM signals for PB determined with and without retention time alignment. Retention time alignment allows GRAM to obtain a signal profile having an appropriate shape that closely matches the true profile. (B) The effects of absolute retention time shifts on the accuracy of PB signal profiles from JP-TS jet fuel determined from 25 GRAM analyses each with (■) and without (□) retention time alignment. A correlation coefficient of 1.0 is a perfect signal profile fit between the GRAM and true signal.

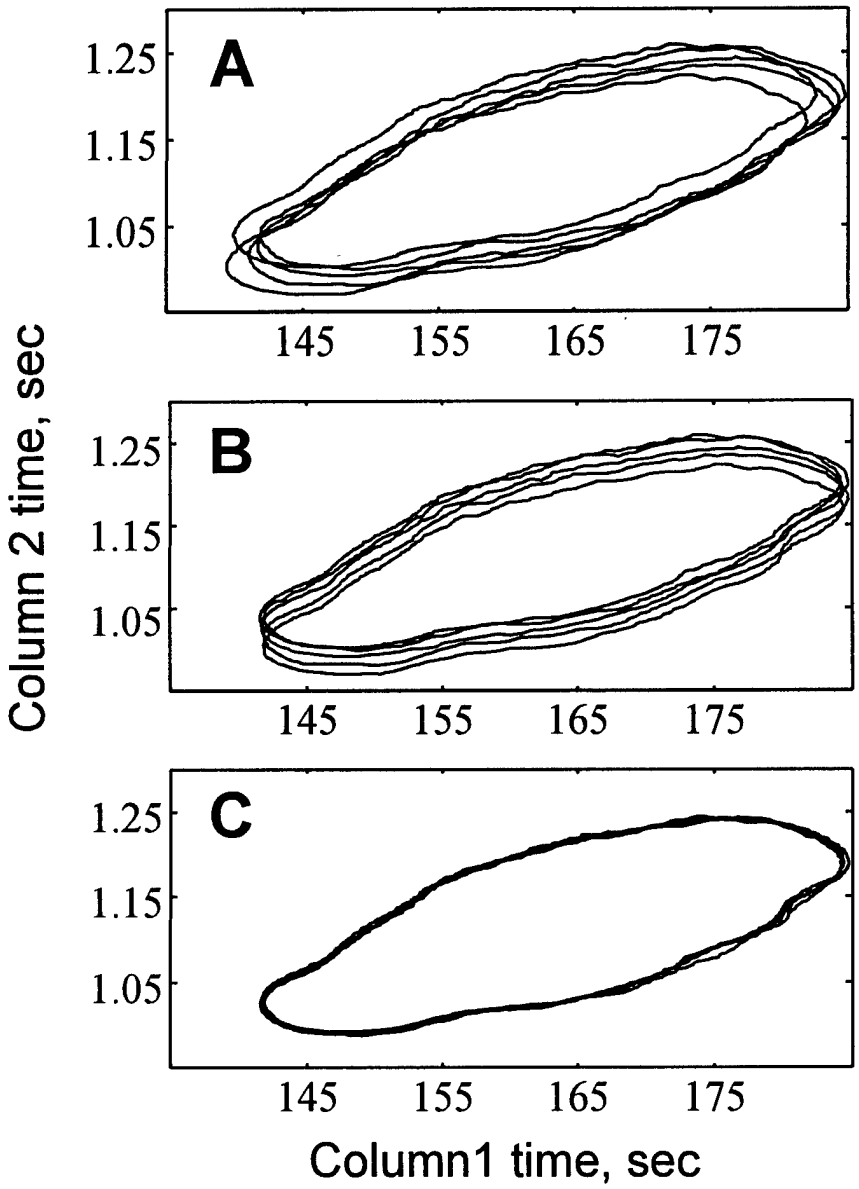



**Figure 3.8.** (A) Contour plot depicting five overlaid GC x GC signals each containing the overlapped signals of the four isomers depicted in Figure 3.4A. The five GC x GC signals were artificially shifted. (B) The five GC x GC signals after retention time alignment on one dimension using the original retention time algorithm. (C) The five GC x GC signals after retention time alignment on both dimensions using the modified alignment algorithm

# Chapter 4: Enhancing the Quantitative Precision, Accuracy, and Limit of Detection of GC x GC

## 4.1 Introduction

The valve-based GC x GC system used in this dissertation produces fast, high peak capacity separations. Excellent data precision and accuracy is achieved, which is ideal for routine analyte quantification, especially chemometric analysis. This chapter will further demonstrate that a diaphragm valve GC x GC system, when coupled to chemometric analysis, is well suited for quantitative analysis of low S/N data. This chapter reports a comparison of traditional approaches of using signal integration for quantification versus using chemometric data analysis.

Integration is commonly used for the quantification of analytes by GC x GC [Gaines et al., 1998; Frysinger et al., 1999]. Integrating a GC x GC analyte peak involves determining the beginning and ending points of the peak on the first and second columns and then, following baseline correction, summing the values of the data points within the selected space. The integration boundaries of a GC x GC peak can be determined by visually selecting the smallest data space that contains the peak [Liu et al., 1994] or by automated peak detection software [Beens et al., 1998]. Selection of accurate peak integration boundaries is difficult when the signal-to-noise ratio (S/N) is low [Li, 1999]. While mathematically simple, successful integration of a peak requires that the peak be resolved, and that the S/N of the peak be well above the limit of detection (LOD). The separation power of GC x GC reduces the chances of having overlapped peaks. However, in complex mixtures the likelihood of many peaks overlapping remains high.

One way of enhancing the separation power of GC x GC is through the use of multivariate chemometric analysis techniques such as GRAM. Multivariate techniques can use the inherent GC x GC data structure to mathematically separate and quantify incompletely resolved signals. Furthermore, multivariate data analysis techniques may enhance the LOD. In this chapter, GRAM is used to quantify GC x GC peaks for analyte signals initially at low S/N. GRAM is critically compared to the traditional

method of peak integration. The advantage of using GRAM to enhance the LOD of GC x GC data is reported. The advantage is shown to be applicable to both resolved and overlapped peaks.

In general, GRAM enables the resolution and quantification of data from hyphenated chemical analyzers. Previously, GRAM was applied for the quantification of overlapped GC x GC signals in modified white gasoline and jet fuel [Bruckner et al., 1998; Fraga et al., 2000a]. To estimate the true signals of individual chemical components, GRAM compares the data of a sample and calibration standard and extracts the individual signals of the chemical components from that of interference and noise. The GRAM extracted signals are used to determine the relative amount of each analyte of interest in the sample. This extraction of individual signals is usually used to separate unresolved peaks. This chapter demonstrates GRAM's ability to filter noise from GC x GC data, thus improving the LOD. While the previous application of GRAM to GC x GC demonstrated the deconvolution of overlapped signals, the effect of GRAM on quantification of small signals has not been rigorously tested because of the high S/N of these previous examples [Bruckner et al., 1998; Prazen et al., 1999b; Fraga et al., 2000a]. The present work is a direct extension of factor analysis [Malinowski, 1991] as a noise filter for data obtained from a hyphenated instrument such as GC x GC. This chapter investigates how the bilinear model used by GRAM improves the quantification of GC x GC signals at low S/N. A rigorous comparison of data analysis by integration and GRAM is performed on resolved and unresolved signals having approximate S/N values from 3 to 800.

## 4.2 Theory

Ideally, GC x GC data sets fit a bilinear structure (see section 1.2.1). Thus, the matrix of data from each chemical component can be described as the product of its first column chromatographic profile, a concentration term, and its second column profile plus a noise matrix. For each chemical component, the shape of one column profile remains constant at each point along the other column profile. Also, the signals

from each analyte peak are independent of each other. Thus, as a peak elutes from the first column the signal profile of the second column modulates in size while retaining the second column profile shape. This data structure is utilized in this chapter to extract analyte signals from noise. The use of this data structure as a noise filter will be referred to as bilinear signal enhancement.

Quantification can be obtained using GRAM in the presence of unknown interference such as when one or more unknown GC x GC peaks overlap with an analyte peak. This capability is termed the second order advantage and does not exist with first order instruments such as one-dimensional gas chromatography with flame ionization detection [Booksh and Kowalski, 1994]. The second order advantage also does not exist for GC x GC data that is effectively reduced to first order by using traditional quantification methods such as integration.

For GRAM to work successfully the measured signal must be linear with concentration and the chromatographic profiles must be independent and constant for each analyte present in the sample and standard. These requirements are attained when the detector and chromatograph behave linearly, when each analyte has some separation on both columns, and when retention times remain constant between the sample and the standard, or are made constant by an objective alignment procedure (see section 1.2.5). For the experiments reported in this chapter, these conditions were met.

The GRAM algorithm used in this study is a commercially available version included with the PLS Toolbox 2.0 [Wise and Gallagher, 1998]. In the first step of the GRAM algorithm the addition matrix (**Q**), which is the sum of sample data and standard data, is decomposed by singular value decomposition (SVD). SVD is done on the addition matrix to ensure all components contained in the sample and standard are included in the model. The SVD routine results in the decomposition of the addition matrix into **U**, **S**, and **V** matrices. **U** and **V** are matrices of eigenvectors and **S** is a diagonal matrix of singular values, such that $\mathbf{U}*\mathbf{S}*\mathbf{V}^T = \mathbf{Q}$. The **U** and **V** matrices are then truncated according to the number of chemical components, or rank, of the addition matrix. The number of components was estimated by inspection of the

singular values [Faber et al., 1994a]. The truncated **U** and **V** matrices are used to project the addition and standard matrices into square matrices containing the bilinear portion of the signal. Also, this truncation step works as a "smart filter." The large correlated signals are retained and the random uncorrelated signals are removed in the truncation [Faber et al., 1997].

In the final step of GRAM, a generalized eigenvalue analysis of the addition and standard matrices leads to the transformation matrix that converts the abstract SVD matrices into the individual chromatographic profiles, i.e., GC x GC signals, of the analytes. The PLS Toolbox version of GRAM uses a generalized eigenvalue routine that is solved with the QZ algorithm [Wilson et al., 1989]. Ratios of the calculated analyte signals of the sample and the standard data matrices are determined by least squares. These ratios are used to obtain the concentrations of the analytes in the sample.

## 4.3 Experimental

### 4.3.1 Sample Preparation

Two sets of experiments were performed in this study. For the first set of experiments, a 450-ppm (w/w) propylbenzene (98% purity, Aldrich Co., Milwaukee, WI), 1060-ppm ethylbenzene (99% Aldrich) and 1630-ppm *sec*-butylbenzene (99% Aldrich) solution was prepared in heptane (HPLC grade, Fisher Scientific, Fair Lawn, NJ). Similar quantitative results were achieved for all three solutes and hence for brevity only the analysis of propylbenzene is reported in detail. Several serial dilutions were made, by weight, with heptane in order to produce eight solutions that ranged from 1.3- to 450-ppm propylbenzene. Six to twelve replicate GC x GC runs were made of each solution.

For the second set of experiments, a sample and a calibration standard each containing isopropylbenzene, propylbenzene, 1,3,5-trimethylbenzene, *o*-ethyltoluene, 1,2,4-trimethylbenzene, *tert*-butylbenzene and *sec*-butylbenzene were prepared in heptane. All seven alkyl-benzene compounds were at least 98 % pure (Aldrich). For

brevity, the quantification of only propylbenzene, 1,3,5-trimethylbenzene, *o*-ethyltoluene, and 1,2,4-trimethylbenzene is reported. The sample solution was diluted with heptane in order to accurately produce analyte concentrations in the sample solution that were 14 ppm (w/w). The standard solution was diluted with heptane to produce concentrations of 11.23 ppm, 22.56 ppm, 34.86 ppm and 58.11 ppm for propylbenzene, 1,3,5-trimethylbenzene, *o*-ethyltoluene, and 1,2,4-trimethylbenzene, respectively. Four replicate GC x GC runs were made of the sample solution and six of the standard solution.

### 4.3.2 GC x GC Method

The GC x GC experiments were performed with a Varian gas chromatograph (model 3600CX, Varian, Sugar Land, TX) with a flame ionization detector (FID). An autosampler was used to inject 1-μL samples to a split injector at 250 °C, which operated in the split-less mode for both experiments. The column temperature was 95 °C for the first experiments and 80 °C for the second. The FID temperature was 250 °C. The first column head pressure was 12.5 psi. The first column of the GC x GC system was a 9.2-m x 530-μm capillary column with a 3-μm poly(dimethylsiloxane) film (SPB-1, Supelco, Bellefonte, PA). The second column was a 0.89-m x 180-μm column with a 0.15-μm poly(ethylene glycol) stationary phase (Carbowax, Quadrex Corp, New Haven, CT). The diaphragm valve (model 11, Applied Automation, Bartlesville, OK) diverted the first column effluent for 15 ms to the second column every 280 ms during an 80-s GC x GC run for the first set of experiments and every 320 ms during a 120-s run for the second set of experiments. The carrier flow from the first column was split after the diaphragm valve between the second column and 0.5 m of 180-μm silica tubing. This split allows the flow rates of each column to be optimized independently.

The diaphragm valve split ratio for the first experiment was a 15ms/280ms or 1:19 split, and for the second experiment a 15ms/320ms or 1:21 split. Note that the sample was then split past the diaphragm valve by 1:7. Thus, the overall split of the

sample injected into the GC by the autoinjector was about 1:140 for the diaphragm valve GC x GC system, since no split is applied prior to the first column. Recently reported split ratios applied to the first column for the thermal modulation GC x GC system are 475:1 [Frysinger et al., 1999] and 500:1 [Beens et al., 1998]. A recently reported split ratio for the cryogenic modulation GC x GC system was 300:1 [Kinghorn and Marriott, 1999]. Thus, thermal and cryogenic modulation GC x GC systems generally have a slightly larger overall split due to the use of narrower bore first columns and interdependent flows between the two columns. The relatively low overall split ratio of the diaphragm valve GC x GC system aids in obtaining low limits of detection and good quantification.

### 4.3.3 Data Processing

The FID signal was measured at a rate of 20,000 points/s, and these data were boxcar averaged to 267 points/s and 250 points/s for the first and second sets of experiments, respectively. Collecting data at a high rate and then boxcar averaging benefits the S/N of the data. The data was converted into a matrix such that each row of the matrix represented a fixed time on the second GC column and each column of the matrix represented a fixed time on the first GC column. All data matrices were preprocessed with background subtraction, boxcar averaging and retention time alignment. Background was subtracted by taking the difference of the matrix and a matrix of the same size containing only baseline noise. Prior to applying GRAM and integration, each sample matrix was retention time aligned to the calibration standard using a previously developed algorithm [Prazen et al., 1998]. The number of chemical components (analytes) used in the GRAM analyses was one for the first set of experiments and four for the second.

For the first set of experiments, 59- by 23-point matrices from 46.35 to 62.59 s on the first column axis and 0.300 to 0.385 s on the second column axis were created, encompassing the propylbenzene peak. This matrix size was used for the analyses shown in Figure 4.1 and Figure 4.2. This matrix size is larger than the propylbenzene

peak so signal can be compared to baseline noise. In order to improve quantification by integration, a smaller sub-matrix was analyzed for the integration results presented in Figure 4.3 and Figure 4.4. These matrices were from 48.01 to 53.61 s on the first axis and 0.300 to 0.385 s on the second. The center data point of each matrix had time coordinates of 50.83 and 0.343 s, which corresponded to the peak maximum of a representative 22-ppm propylbenzene peak. The matrix size corresponded to plus and minus three standard deviations of the 22-ppm propylbenzene peak modeled by a Gaussian profile along each axis. Integration of the signal within the smaller matrix was studied using two methods: with and without a signal threshold. Integration with a signal threshold was accomplished by summing the signal from each data point having a value that was equal to or greater than three times the standard deviation of baseline noise [Mittermayr et al., 1997; Ouchi, 1991]. A section of baseline noise near the propylbenzene peak was used to determine the signal threshold. Integration without a signal threshold was achieved by summing the signal from all data points within the smaller matrix containing the propylbenzene peak. Integrating the signal within the smaller matrix led to better quantitative precision and accuracy than integration methods using peak detection algorithms [Braithwaite and Smith, 1996; Ouchi, 1991]. Such peak detection methods are unreliable at low signal-to-noise due to the difficulty of recognizing a peak's starting and ending points, particularly for those methods that rely on a signal's first derivative [Mittermayr et al., 1997].

For the second set of experiments, 105- by 80- point matrices from 70.72 to 104 s on the first column axis and 0.322 to 0.642 s on the second column axis were created. This matrix size included the signals of propylbenzene, 1,3,5 trimethylbenzene, *o*-ethyltoluene, 1,2,4 trimethylbenzene and *tert*-butylbenzene as depicted in Figure 4.5.A and Figure 4.5.B. GRAM analysis was done on this matrix after zero-filling the matrix points that contained the *tert*-butylbenzene peak. This was done in order to quantify the four analytes of interest simultaneously using a single matrix. Thus, the GRAM analysis of the four analytes together did not require prior knowledge of each peak's exact location, and all four analytes of interest were identified and quantified simultaneously. However, in order to quantify by integration, the exact location and

peak boundaries of all analytes had to be determined. Another drawback to integration was that only propylbenzene and 1,2,4-trimethylbenzne could be quantified by integration because the other two components were overlapped. The integration boundaries were selected visually by using a sample mixture that had the analytes found in the sample but at much higher S/N than those in Figure 4.5. The GC x GC run of the higher concentration mixture was retention time aligned to a calibration standard run. The sub-matrix defining the integration boundaries for the propylbenzene peak was from 72.32 to 80.64 s on the first column and 0.378 to 0.454 s on the second column. The sub-matrix for 1,2,4-trimethylbeznene peaks was from 93.12 to 103.36 s on the first column axis and 0.734 to 0.634 s on the second column. As in the first experiment, these sub-matrices where centered on each peak's maximum and had matrix dimensions that where six standard deviations in length. Integration of the signal within each matrix was accomplished with and without a signal threshold, as previously described. The signal threshold was set at three times the standard deviation of baseline noise.

## 4.4 Results and Discussion

The objective of this chapter is to show that the S/N and quantitative analysis of low concentration samples can be improved with bilinear signal enhancement. Two sets of experiments were performed. The first study demonstrates quantitative precision and accuracy improvements obtained when GRAM is applied on a resolved analyte peak. The second study combines an example of bilinear signal enhancement with mathematical signal resolution to further demonstrate the utility of combining multivariate analysis and GC x GC.

### 4.4.1 Bilinear Signal Enhancement of Resolved Signals

In the first set of experiments, bilinear signal enhancement is demonstrated through the comparison of GRAM and integration for the analysis of low concentration samples of propylbenzene by GC x GC. Figure 4.1.A is a contour plot of a

propylbenzene standard. The uniformity of the propylbenzene signal demonstrates that a diaphragm valve GC x GC system produces bilinear data (see eq. 1), because retention times on the second column are constant throughout for this GC x GC peak. A previous study demonstrated the retention time and injection volume reproducibility of the diaphragm valve GC x GC system produces bilinear data [Bruckner et al., 1998]. Representative examples of standard and sample data obtained in this study are depicted in Figure 4.1.B and Figure 4.1.C. These data sets contain the signal for propylbenzene in the 22-ppm standard and the 2.7-ppm propylbenzene sample, respectively. Figure 4.1.D contains the GRAM calculated signal for the 2.7-ppm sample. The GRAM calculated signal is that portion of the signal that correlates with that of the more concentrated standard. The bilinear signal enhancement is evident from a comparison of Figure 4.1.C and Figure 4.1.D. The signal of the analyte is clearly distinguishable in Figure 4.1.D while it is difficult if not impossible to see in Figure 4.1.C.

In order to allow visual assessment of the S/N and precision of replicate GC x GC runs, each data matrix of the 2.7-ppm sample was reduced to a vector by summing the data matrix onto the first column time axis using only the data within the sub-matrix that contained the analyte peak. Figure 4.2.A shows the signals of seven replicate runs of the 2.7-ppm propylbenzene sample solution. This figure demonstrates the original S/N for this sample. Figure 4.2.B is the seven GRAM computed signals in vector form. These were determined by pairing the seven 2.7-ppm samples with the 22-ppm standard (Figure 4.1.B). Again, the increased S/N and precision resulting from bilinear signal enhancement can be seen by comparing Figure 4.2.A (original data) to Figure 4.2.B (GRAM).

Typically, in chromatography, a single representative run is used to calculate the S/N by taking the ratio of the peak height and the standard deviation of a section of baseline noise [Sun et al., 1994]. This approach cannot be taken on a GRAM signal (e.g., Figure 4.2.B) because the baseline fluctuations are not noise but part of the calculated signal shape. GRAM finds the underlying signal shape for the entire data region (i.e., peak and baseline). A significant portion of the noise is excluded in the

process. Noise that cannot be factored out becomes embedded in the GRAM calculated signal shape. This can be seen in Figure 4.1.D and Figure 4.2.B. Hence, the S/N is determined using the ratio of the mean signal of replicate runs to the signal standard deviation [Hieftje, 1972].

Using the replicate runs depicted in Figure 4.2.A and Figure 4.2.B, the mean and standard deviation of peak heights is used to calculate the S/N values with and without bilinear signal enhancement. The S/N values for the 2.7-ppm sample with and without bilinear signal enhancement are 14 and 4.8, respectively. The improvement in S/N using GRAM is approximately a factor of three, which results in lowering the GC x GC limit of detection (LOD) by the same factor. Defining the LOD as the concentration that gives a S/N of 3, the LOD's without and with bilinear signal enhancement are predicted to be 2 ppm and 0.6 ppm, respectively.

The improvement in S/N when using GRAM is due to two sources. First, GRAM extracts the signal shape common to both the sample and standard. In the first set of experiments, the S/N for the standard is approximately 8 times greater than the 2.7-ppm sample. The higher concentration standard allows GRAM to more effectively estimate the true signal. The second source of improvement is the extraction of signal that follows the bilinear model. The SVD portion of the GRAM algorithm extracts the largest variance from the signal that follows the bilinear model. Because the diaphragm valve GC x GC system generally creates bilinear signals, the remaining signal is considered noise. Thus, GRAM acts as a "smart filter" by extracting signal from the sample which is shaped like signal from the standard.

To demonstrate the precision improvement created by bilinear signal enhancement, GRAM and peak integration were compared for quantification of replicate GC x GC runs of propylbenzene at multiple concentrations. A single 22-ppm propylbenzene solution was used as the calibration standard for all samples and both quantification methods. Figure 4.3 depicts the precision achieved by integration without a threshold and GRAM. The precision for integration without a threshold was nearly the same as with a threshold (which was not shown for brevity). The precision at each concentration was based on at least 48 quantification results from each method.

Each integration result was the ratio of the propylbenzene peak volume in the sample and calibration standard. Each GRAM result was the ratio of the calculated sample and calibration standard signals. The improvement in precision obtained with GRAM is most pronounced at low concentration. The precision for all methods is approximately the same at high concentrations, because the S/N is not the main factor determining the quantitative precision at these concentrations.

The bias in quantification in the propylbenzene study for GRAM, integration without a threshold, and integration with a threshold are depicted in Figure 4.4.A and Figure 4.4.B. The bias was determined as the percent difference between the mean quantification results and those determined by weight (gravimetric determination). The error bars in Figure 4.4.A and Figure 4.4.B are based on the standard deviation of the mean. The bias is due to three sources. The first source affects GRAM and both integration methods to the same degree and is due to the difference in concentration between the sample and standard. The integration methods and GRAM are more biased the further the sample concentration is from the standard. This is to be expected if a one-point calibration is applied. The GRAM bias at low concentrations can be removed by using a standard that is similar in concentration to the sample. For instance, using an 11-ppm standard reduced the GRAM bias to nearly zero. The second source of bias is due to baseline offset. At low concentrations GRAM is significantly less biased than both integration methods. The extra bias found in integration at low concentrations is thought to be caused by a slight baseline offset. GRAM compensates for a baseline, whereas integration incorporates the offset into the integrated signal. In one example, when the baseline of a 1.3-ppm sample data set was artificially moved by one FID signal unit (see Figure 4.2.A for scale) the integration bias increased to over one hundred percent, whereas the GRAM bias remained within the error bars shown in Figure 4.4.A. The third source of bias is introduced when a threshold is used to determine which data points are summed to obtain the total analyte signal. Because the threshold method only includes data point signals that meet or exceed the detection limit, some small signals in the edges of a low-level peak will be lost in the baseline noise [Taraszewski et al., 1984]. Therefore, integrating a low S/N

GC x GC peak results in a computed peak signal that differs significantly from the true peak signal. Figure 4.4.B shows that threshold integration methods can lead to a substantial quantitative negative bias when analyzing low S/N peaks. Thus, going from a threshold integration method to GRAM results in a substantial improvement in the quantitative analysis.

### 4.4.2 GRAM Analysis of Low-Level Unresolved Signals

While the above set of experiments demonstrated the ability of bilinear signal enhancement to improve the precision and accuracy when analyzing a resolved analyte peak, the second set of experiments demonstrates the S/N enhancement coupled with the ability to resolve overlapped signals using multivariate analysis. Figure 4.5.A and Figure 4.5.B contain GC x GC data for the calibration standard and sample used in the second set of experiments. Quantification results obtained using integration and GRAM were also compared for these samples. Only propylbenzene, 1,3,5-trimethylbenzene, *o*-ethyltoluene and 1,2,4 trimethylbenzene where used in the GRAM and integration quantification studies. Chromatographic profiles of the four analytes of interest are depicted in Figures 4.6.A, 4.6.B, 4.6.C, and 4.6.D. Figure 4.6.A is the summed first column profiles without GRAM analysis. Figure 4.6.A depicts the S/N obtained by the first column using traditional integration quantification techniques. Figure 4.6.B is the overlay of summed first column profiles for the four analytes resulting from GRAM analysis. GRAM results in an individual signal for each component, thus resolving the overlapped signals. Figure 4.6.C and Figure 4.6.D are the summed second column profiles without and with GRAM analysis. Three components are severely overlapped on the second column axis, yet GRAM is able to resolve the signals. The bilinear signal enhancement resulting from GRAM analysis is evident in the improved peak shapes seen in Figure 4.6.B and Figure 4.6.D. GRAM is able to provide the qualitative information that is obscured by noise and signal overlap along both column dimensions.

The results for the quantification of the second set of experiments of GRAM and integration are shown in Table 4.1. Six replicate runs of the standard and four replicate runs of the sample data were analyzed. GRAM performed significantly better than both integration methods for the analysis of all four analytes in terms of accuracy and precision. Integration methods were unable to quantify two components with unresolved signals.

Another important advantage of GRAM analysis is a reduced dependence on peak boundary detection. In practice, peak boundaries are difficult to determine in low S/N cases. Excluding some signal or including too much baseline when selecting boundaries will result in inaccurate and/or imprecise integration. GRAM excludes baseline noise by fitting the calculated shape to the analyte signal in the data while integration methods include the baseline noise by integrating it with the analyte signal.

Thermal modulation is the prevalent technique used in GC x GC to deliver or inject eluent from the primary GC column to the secondary column [Phillips et al., 1999]. Both modulation and diaphragm valve GC x GC are suitable for quantitative analysis. In terms of trace analysis, GRAM analysis of diaphragm valve GC x GC data led to a limit of detection (LOD) of 0.6 ppm for propylbenzene. This is better than the 5 ppm reported LOD for a similar analysis of aromatic hydrocarbons with a GC x GC system that used thermal modulation [Beens et al., 1998]. It would be intriguing to examine the utility of applying chemometric methods such as GRAM to data obtained by thermal modulated GC x GC.

## 4.5 Conclusions

GRAM and other multivariate methods based on the bilinear model can extend GC x GC quantification to lower concentrations by enhancing S/N. Calibration methods using integration do not filter out noise as effectively as GRAM and rely on proper selection of peak boundaries. In addition, multivariate analysis of GC x GC data can extend chromatographic resolution by mathematically resolving overlapped

signals. Applications of GRAM should lower the detection limit of other bilinear hyphenated chromatographic systems such as GC/MS.

**Table 4.1. Quantification results using GRAM and integration for the analyses of alkyl benzene isomer mixtures. Based on 24 replicate analyses of six standard runs and four sample runs. The four target analytes at 14 ppm each in the sample are labeled. PB: propylbenzene; 1,3,5 TMB: 1,3,5-trimethylbenzene; OET: *o*-ethyltoluene; 1,2,4 TMB: 1,2,4-trimethylbenzene.**

| Analyte | Resolution[a] | | Bias (%)[b] | | | RSD (%)[e] | | |
|---|---|---|---|---|---|---|---|---|
| | First Column | Second Column | GRAM | Integration No Threshold[c] | Integration Threshold[d] | GRAM | Integration No Threshold[c] | Integration Threshold[d] |
| 1. PB | 1.2 | 1.2 | 0.11 | 3.7 | 16 | 5.8 | 8.9 | 19 |
| 2. 1,3,5 TMB | 0.83 | 0.55 | -0.82 | NA | NA | 4.0 | NA | NA |
| 3. OET | 0.83 | 0.55 | -3.4 | NA | NA | 4.0 | NA | NA |
| 4. 1,2,4 TMB | 1.2 | 0.69 | -4.1 | -14 | -34 | 5.5 | 11 | 10 |

[a]Determined using GRAM chromatographic profiles of analyte and closest adjacent peak
[b]Relative difference between the gravimetrically determined analyte concentrations (i.e., true values) and those from GRAM and integration
[c]Integration achieved by summing the signal within the rectangular data matrix that encompassed the analyte GC x GC peak
[d]Integration achieved by summing the signal within the rectangular data matrix that met or exceeded three times baseline noise standard deviation
[e]Relative standard deviation
NA: Not Applicable. Integration not possible because of signal overlap

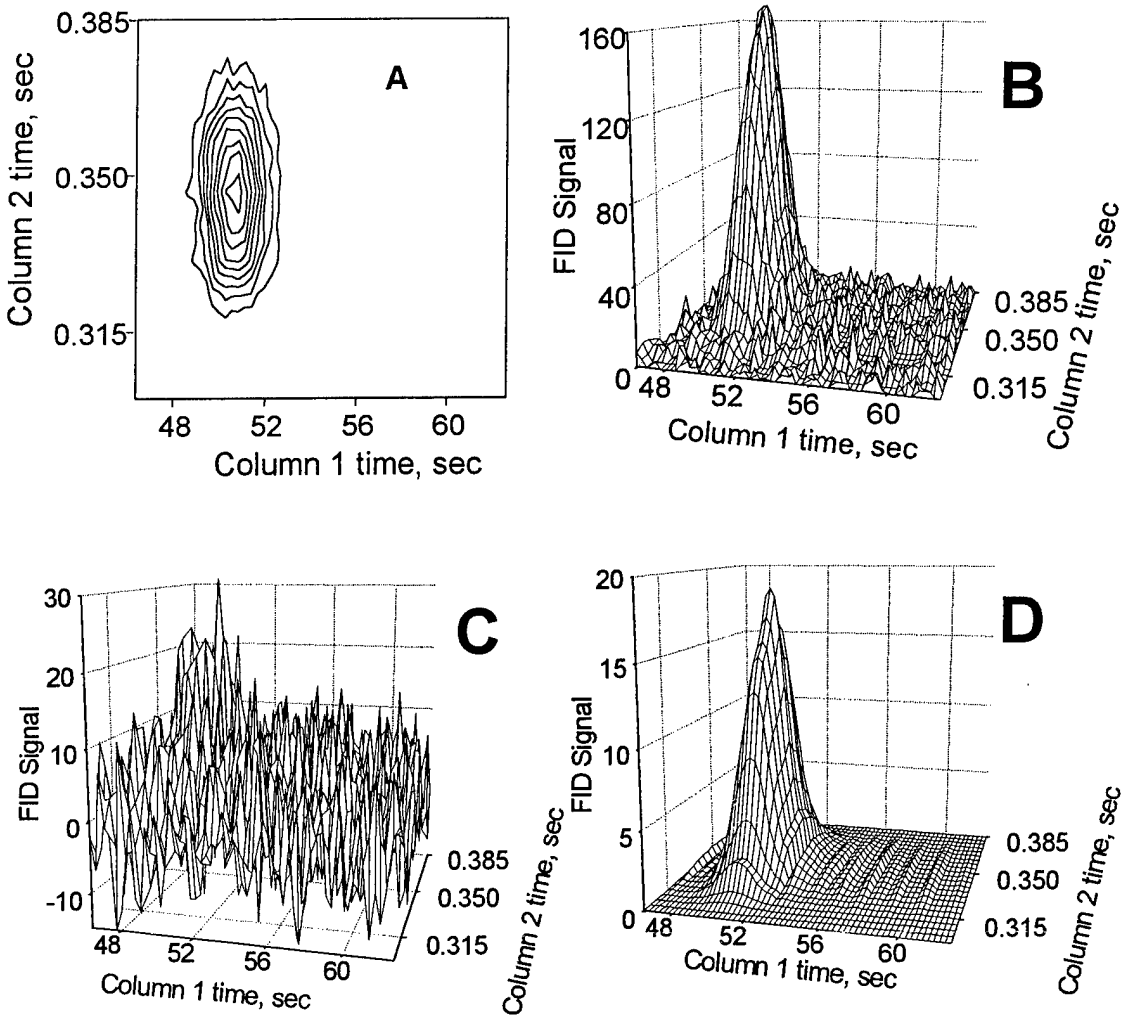



**Figure 4.1.** Representative GC x GC data for a 22-ppm propylbenzene calibration standard used in the GRAM and integration studies in the form of (A) a contour plot and (B) a three-dimensional mesh plot, (C) 2.7-ppm propylbenzene sample without being subject to bilinear signal enhancement and (D) 2.7-ppm propylbenzene with bilinear signal enhancement by GRAM. GRAM extracts the analyte signal from the presence of most of the noise by finding the bilinear signal common to both the sample and calibration standard.

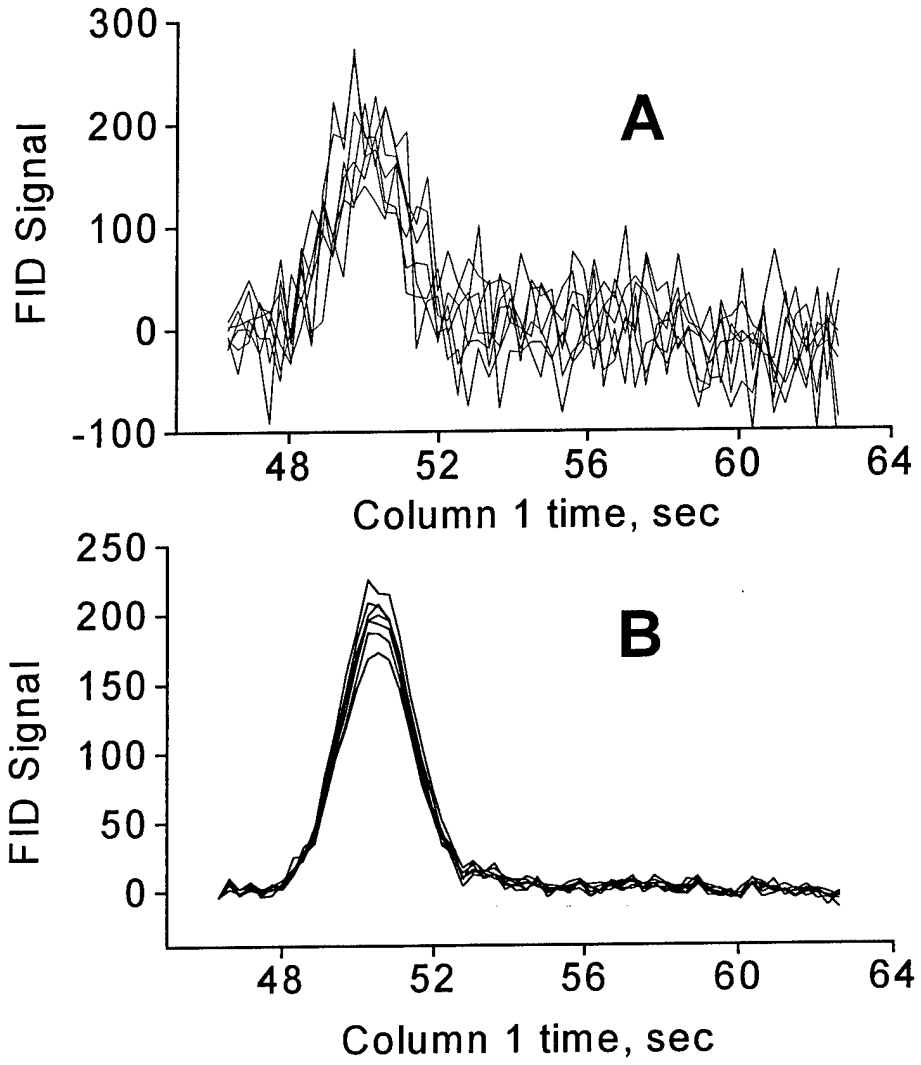



**Figure 4.2.** (A) Overlay of seven summed first column profiles for 2.7-ppm sample without GRAM analysis. (B) Overlay of seven summed first column profiles for 2.7-ppm sample with GRAM analysis. GRAM enhances S/N from 4.8 to 14.

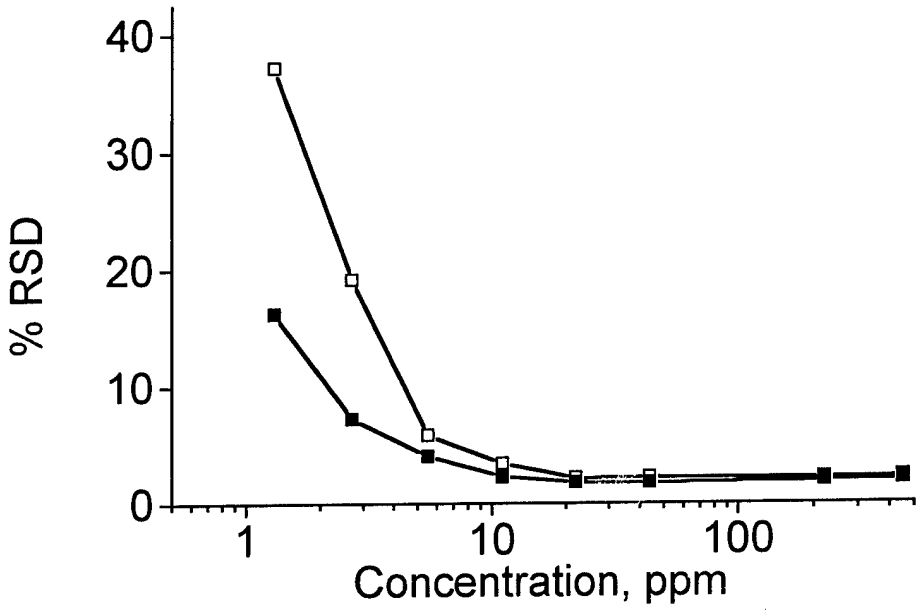



**Figure 4.3.** Quantification precision, expressed as the percent relative standard deviation (%RSD), for GRAM (■) and integration (□). Each data point is the average of at least 48 results. GRAM significantly improves precision for low concentration samples.

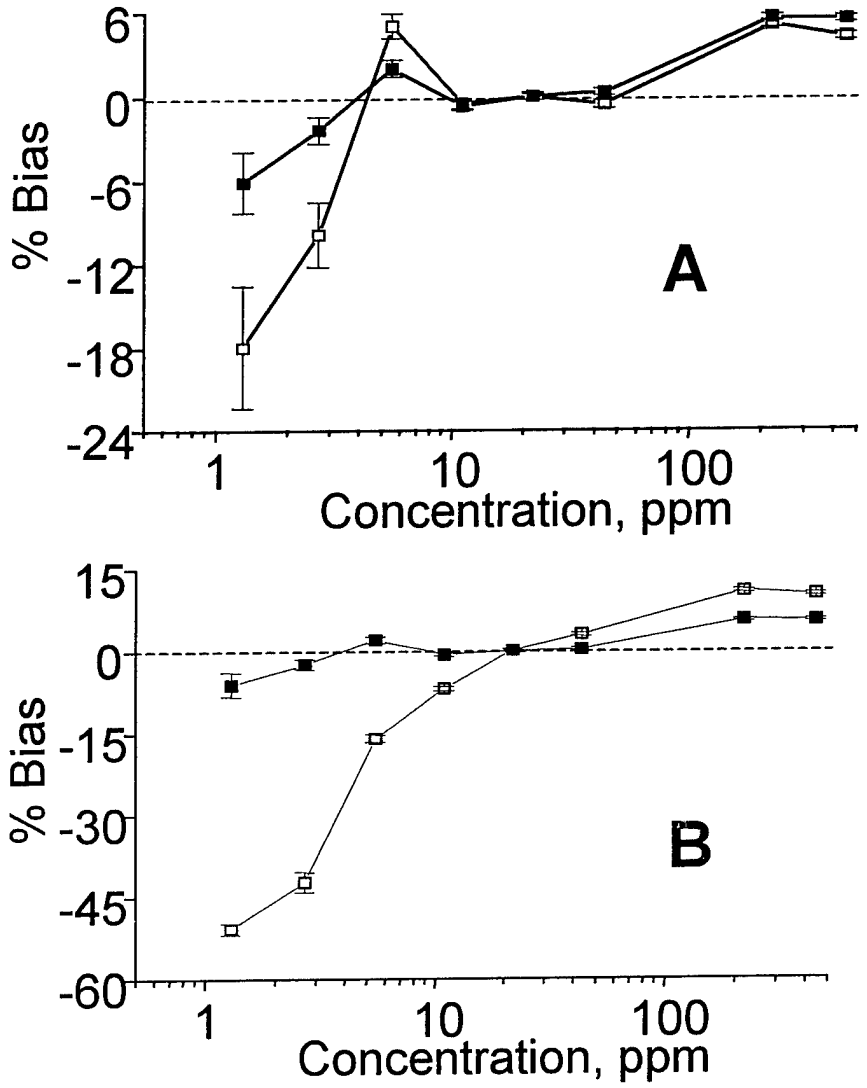



**Figure 4.4.** (A) Quantification accuracy for GRAM (■) and integration with no threshold (□) using gravimetrically measured concentrations as true values. (B) Quantification accuracy for GRAM (■) and integration with a threshold (□) using gravimetrically measured concentrations as true values. Note, change in y-axis scale due to large bias introduced when a threshold is used with integration. The % Bias is the determined concentration minus the gravimetric, then divided by the gravimetric, and expressed as a percentage.

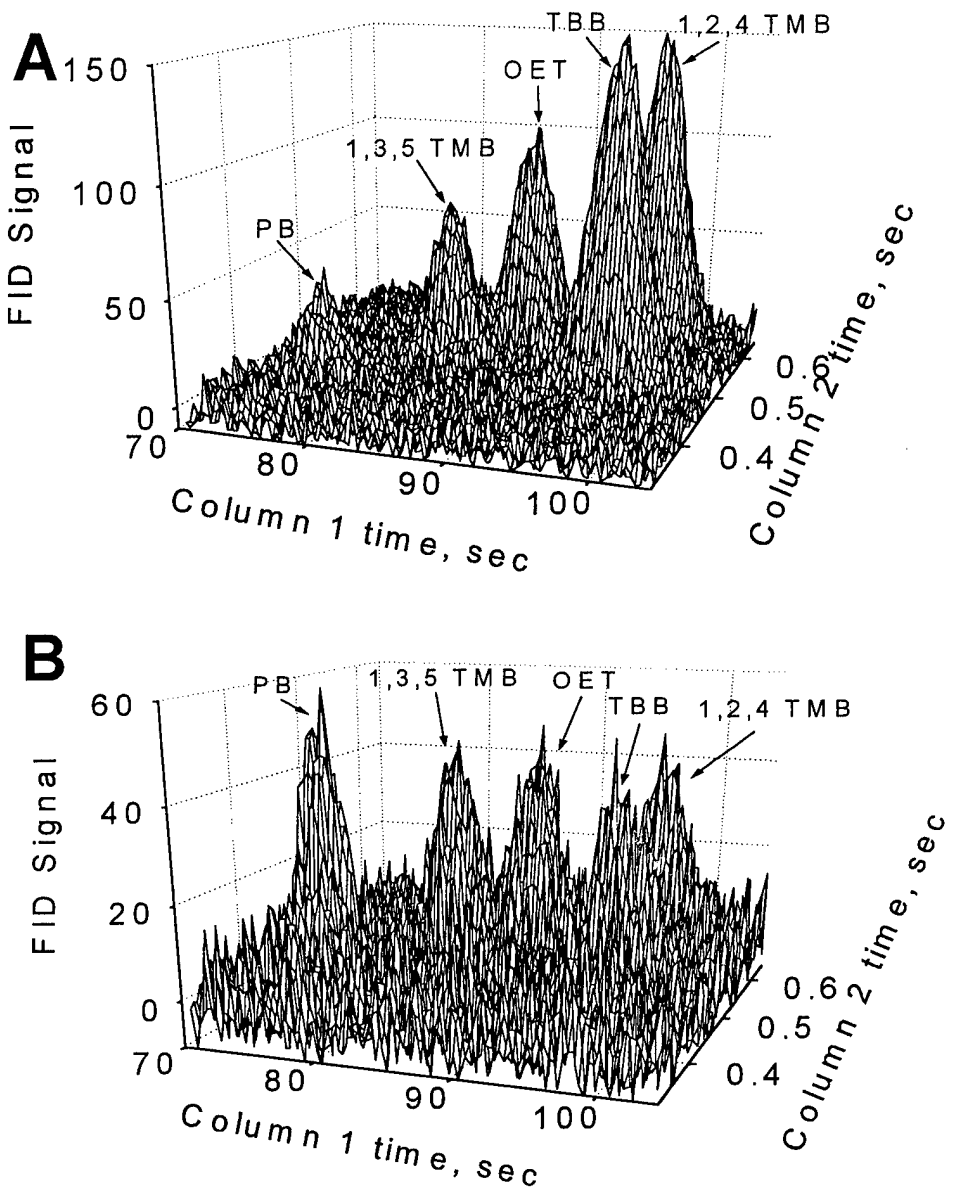



**Figure 4.5.** GC x GC data for (A) calibration standard and (B) sample used in the GRAM and integration studies. The components are labeled. PB: propylbenzene; 1,3,5 TMB: 1,3,5-trimethylbenzene; OET: *o*-ethyltoluene; 1,2,4 TMB: 1,2,4-trimethylbenzene; and TBB: *tert*-butylbenzene.

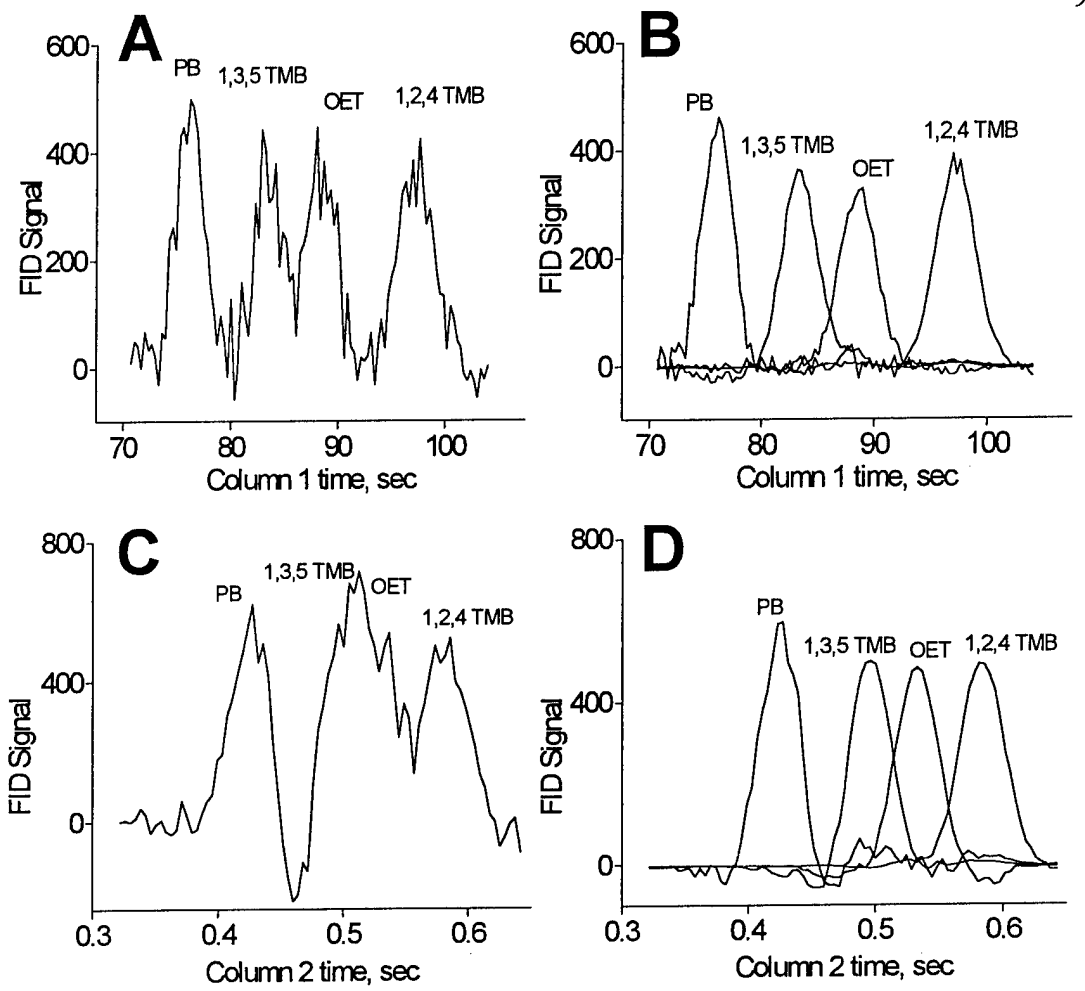



**Figure 4.6.** Profiles of the four component sample mixture. (A) Summed first column profiles without bilinear signal enhancement analysis. (B) Overlay of summed first column profiles with bilinear signal enhancement using GRAM. (C)

# Chapter 5: Final Conclusion and Future Work

The overall goal of this dissertation is to enhance the analysis capabilities of GC x GC by applying chemometric methods to extract component signals in the presence of significant interference, noise, or both. This goal was achieved through the verification of the five hypotheses introduced in section 1.3 and tested in chapters 2, 3, and 4. In chapter 2, Monte Carlo simulations of comprehensive 2-D separations are used to prove the hypothesis that GRAM, through the quantification of unresolved peaks, can increase the number of analyzable peaks in GC x GC separations. In addition, chapter 2 supports the hypothesis that interpolated retention time alignment can further increase the number of analyzable peaks in comprehensive 2-D separations. In summary, chapter 2 demonstrates, through the use of simulated data, that the bilinear chemometric methods of GRAM and an objective retention time alignment algorithm can enhance the analysis capabilities of GC x GC. In chapter 3, the analysis enhancement provided by GRAM and the objective retention time alignment algorithm is demonstrated on real world GC x GC data. In this chapter, the above chemometric methods and the standard addition method were used to successfully quantifying and deconvolute the GC x GC signals of two jet fuel components in the presence of interfering signals. Furthermore, the material in chapter 3 demonstrates the 2-D alignment of GC x GC data by a modified retention time alignment algorithm. This modified algorithm should extend the analysis enhancement capabilities of GRAM to less reproducible comprehensive 2-D separation techniques. Lastly, in chapter 4 the hypothesis that GRAM analysis improves the precision, accuracy, and LOD of GC x GC is proven. In this chapter, GRAM was used to quantify the signals of both resolved and unresolved GC x GC peaks that initially had a low S/N in the original GC x GC data. GRAM analysis resulted in significantly better quantitative precision and accuracy than traditional peak integration. It also lowered the LOD by a factor of three. Altogether, the works in chapters 2, 3, and 4 support this dissertation's goal. However, more work can be done to further demonstrate the analysis enhancement provided by chemometrics when it is applied to GC x GC and other bilinear 2-D techniques.

At least four projects stemming from the works in chapters 2, 3, and 4 are possible. The first stems from chapter 2. In this chapter, randomly distributed peaks were used to model comprehensive 2-D separations. A possible project might entail applying a more realistic model of comprehensive 2-D separations. Davis and co-workers have produced several papers where they have tried to more realistically model 2-D separations [Davis, 1991; Rowe and Davis, 1995; Shi and Davis, 1993]. The second and third possible projects stem from chapter 3. One project's objective could be to further shorten the separation time of the jet fuel sample. The current GC x GC separation in chapter 3 is five times faster than a traditional single column separation. The first column separation time is currently the limiting factor in terms of shorter separation times. The first column separation time can be reduced by possibly switching to a narrower and shorter first column while still maintaining the same chromatographic resolution on the first column. However, the peak widths on the first column would be narrower. Assuming the second column separation time is left unchanged, the narrower first column peaks will have less second column separations across their widths. Fortunately, computer simulations modeled after the jet fuel data in chapter 3, indicate that when interpolated retention time alignment is used prior to GRAM analysis, the number of second column separation across a first column peak can be reduced from 23 to five without a significant loss in quantification precision and accuracy. Another project could demonstrate the benefits of the 2-D retention time alignment algorithm by using it prior to the GRAM analysis of 2-D bilinear data that need alignment on both dimensions. The last possible project stems from the work in chapter 4. In this chapter, the LOD of GC x GC was improved through bilinear signal enhancement. This improvement should be demonstrated with other 2-D techniques that can benefit from a lower LOD such as gas chromatography with mass spectrometric detection.

# Bibliography


Adlard, E. R. Chromatography in the Petroleum Industry; Elsevier Science: Amsterdam, 1995.

Anderson, A. H.; Gibb, T. C.; Littlewood, A. B. Computer Analysis of Unresolved Non-Gaussian Gas Chromatograms by Curve Fitting. *Analytical Chemistry* **1970a**, *42*, 434-440.

Anderson, A. H.; Gibb, T. C.; Littlewood, A. B. Computer Resolution of Unresolved Convoluted Gas-Chromatographic Peaks. *Journal of Chromatographic Science* **1970b**, *8*, 640-646.

Antalek, B.; Windig, W. Generalized Rank Annihilation Method Applied to a Single Multicomponent Pulsed Gradiant Spin Echo NMR Data Set. *Journal of the American Chemical Society* **1996**, *118*, 10331-10332.

Bahowick, T. J.; Synovec, R. E. Correlation of Quantitative Analysis Precision to Retention Time Precision and Chromatographic Resolution for Rapid, Short-Column Analysis. *Analytical Chemistry* **1995**, *67*, 631-40.

Beebe, K. R.; Pell, R. J.; Seasholtz, M. B. Chemometrics: A Practical Guide; John Wiley & Sons, Inc.: New York, 1998.

Beens, J.; Blomberg, J.; Schoenmakers, P. J. Proper Tuning of Comprehensive Two-Dimensional Gas Chromatography (GC x GC) to Optimize the Separation of Complex Oil Fractions. *Journal of High Resolution Chromatography* **2000**, *23*, 182-188.

Beens, J.; Boelens, H.; Tijssen, R.; Blomberg, J. Quantitative Aspects of Comprehensive Two-Dimensional Gas Chromatography (GCxGC). *Journal of High Resolution Chromatography* **1998**, *21*, 47-54.

Bertsch, W. Multidimensional Gas Chromatography. In *Multidimensional Chromatography, Techniques and Applications*; H. J. Cortes, Ed.; Marcel Dekker: New York, 1990; pp 74-144.

Bertsch, W. Two-Dimensional Gas Chromatography. Concepts, Instrumentation, and Applications - Part 1: Fundamentals, Conventional Two-Dimensional Gas Chromatography, Selected Applications. *Journal of High Resolution Chromatography* **1999**, *22*, 647-665.

Bijlsma, S.; Louwerse, D. J.; Smilde, A. K. Estimating rate constants and pure UV-vis spectra of a two-step reaction using trilinear models. *Journal of Chemometrics* **1999**, *13*, 311-329.

Blomberg, J.; Schoenmakers, P. J.; Beens, J.; Tijssen, R. Comprehensive Two-Dimensional Gas Chromatography (GCxGC) and Its Applicability to the Characterization of Complex (Petrochemical) Mixtures. *Journal of High Resolution Chromatography* **1997**, *20*, 539-544.

Booksh, K. S.; Kowalski, B. R. Theory of Analytical Chemistry. *Analytical Chemistry* **1994**, *66*, 782A-791A.

Braithwaite, A.; Smith, F. J. Chromatographic Methods, 5 ed.; Blackie Academic & Professional: Glasgow, 1996.

Bruckner, C. A. Rapid Chromatographic Analysis Using Novel Detection Systems and Chromatographic Techniques, University of Washington, 1998.

Bruckner, C. A.; Prazen, B. J.; Synovec, R. E. Comprehensive Two-Dimensional High Speed Gas Chromatography with Chemometric Analysis. *Analytical Chemistry* **1998**, *70*, 2796-2804.

Bushey, M. M.; Jorgenson, J. W. Automated Instrumentation for Comprehensive Two-Dimensional High-Performance Liquid Chromatography of Proteins. *Analytical Chemistry* **1990a**, *62*, 161-167.

Bushey, M. M.; Jorgenson, J. W. Automated Instrumentation for Comprehensive Two-Dimensional High-Performance Liquid Chromatography/Capilary Zone Electrophoresis. *Analytical Chemistry* **1990b**, *62*, 978-984.

Chen, S.; Lee, M. L. Automated Instrumentation for Comprehensive Isotachophoresis-Capillary Zone Electrophoresis. *Analytical Chemistry* **2000**, *72*, 816-820.

Davis, J. M. Statistical Theory of Spot Overlap in Two-Dimensional Separations. *Analytical Chemistry* **1991**, *63*, 2141-2152.

Davis, J. M. Statistical Theory of Overlap for Variable Poisson Density : Relaxation of Constraints of Randomness. *Analytical Chemistry* **1994**, *66*, 735-746.

Davis, J. M. Simple Correction for Perturbations of Peak Amplitudes in Statistical Models of Overlap. *Chromatographia* **1996**, *42*, 367-377.

Davis, J. M.; Giddings, J. C. Statistical Theory of Component Overlap in Multicomponent Chromatograms. *Analytical Chemistry* **1983**, *55*, 418-424.

Delinger, S. L.; Davis, J. M. Experimental Verification of Parameters Calculated with the Statistical Model of Overlap from Chromatograms of Synthetic Multicomponent Mixture. *Analytical Chemistry* **1990**, *62*, 436-443.

Dietz, E. A. Shifting of Gas Chromatographic Retention Times Due to Solvent Effects - A Study Using Sulfur Chemiluminescence Detection. *Journal of High Resolution Chromatography* **1996**, *19*, 485-491.

Dimandja, J.-M. D.; Stanfill, S. B.; Grainger, J.; Donald G. Patterson, J. Application of Comprehensive Two-Dimensional Gas Chromatography (GC x GC) to the Qualitative Analysis of Essential Oils. *Journal of High Resolution Chromatography* **2000**, *23*, 208-214.

Dongarra, J. J.; Moler, C. B.; Bunch, J. R.; Stewart, G. W. LINPACK: User's Guide.; Society for Industrial and Applied Mathematics: Philadelphia, 1979.

Dunbar, B. S. Two-Dimensional Electrophoresis and Immunological Techniques; Plenum Press: New York, 1987.

Faber, K.; Kowalski, B. R. Critical-Evaluation of 2 F-Tests for Selecting the Number of Factors in Abstract Factor-Analysis. *Analytica Chimica Acta* **1997**, *337*, 57-71.

Faber, K.; Lorber, A.; Kowalski, B. R. Generalized Rank Annihilation Method: Standard Errors in the Estimated Eigenvalues if the Instrumental Errors are Heteroscedastic and Correlated. *Journal of Chemometrics* **1997**, *11*, 95-109.

Faber, N. M.; Buydens, L. M. C.; Kateman, G. Aspects of Pseudorank Estimation Methods Based on an Estimate of the Size of the Measurement Error. *Analytica Chimica Acta* **1994a**, *296*, 1-20.

Faber, N. M.; Buydens, L. M. C.; Kateman, G. Aspects of Pseudorank Estimation Methods Based on the Eigenvalues of Principal Component Analysis of Random Matrices. *Chemometrics and Intelligent Laboratory Systems* **1994b**, *25*, 203-226.

Faber, N. M.; Buydens, L. M. C.; Kateman, G. Generalized Rank Annihilation Method. I: Determination of Eigenvalue Problems. *Journal of Chemometrics* **1994c**, *8*, 147-154.

Felinger, A. Data Analysis and Signal Processing in Chromatography; Elsevier: Amsterdam, 1998.

Foley, J. P.; Crow, J. A.; Thomas, B. A.; Zamora, M. Unavoidable Flow-Rate Errors in High-Performance Liquid Chromatography. *Journal of Chromatography* **1989**, *478*, 287-309.

Fraga, C. G.; Prazen, B. J.; Synovec, R. E. Comprehensive Two-Dimensional Gas Chromatography (GC x GC) 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. *Analtyical Chemistry* **2000a**, *72*, 4154-4162.

Fraga, C. G.; Prazen, B. J.; Synovec, R. E. Enhancing the Limit of Detection for Comprehensive Two-Dimensional Gas Chromatography (GC x GC) Data using Bilinear Chemometric Analysis. *Journal of High Resolution Chromatography* **2000b**, *23*, 215-224.

Frysinger, G. S.; Gaines, R. B. Comprehensive Two-Dimensional Gas Chromatography with Mass Spectrometric Detection (GC x GC/MS) Applied to the Analysis of Petroleum. *Journal of High Resolution Chromatography* **1999**, *22*, 251-255.

Frysinger, G. S.; Gaines, R. B. Determination of Oxygenates in Gasoline by GC x GC. *Journal of High Resolution Chromatography* **2000**, *23*, 197-201.

Frysinger, G. S.; Gaines, R. B.; Ledford Jr., E. B. Quantitative Determiniation of BTEX and Total Aromatic Compounds in Gasoline by comprehensive Two-Dimensional Gas Chromatography (GC x GC). *Journal of High Resolution Chromatography* **1999**, *22*, 195-200.

Gaines, R. B.; Frysinger, G. S.; Hendrick-Smith, M. S.; Stuart, J. D. Oil Spill Source Identification by Comprehensive Two-Dimensional Gas Chromatography. *Environmental Science and Technology* **1999**, *33*, 2106-2112.

Gaines, R. B.; Ledford Jr., E. B.; Stuart, J. D. Analysis of Water Samples for Trace Levels of Oxygenate and Aromatic Compounds Using Headspace Solid-Phase Microextraction and Comprehensive Two-Dimensional Gas Chromatography. *Journal of Microcolumn Separations* **1998**, *10*, 597-604.

Giddings, J. C. Two-Dimensional Separations: Concept and Promise. *Analytical Chemistry* **1984**, *56*, 1258A-1270A.

Giddings, J. C. Concepts and Comparisions in Multidimensinoal Separation. *Journal of High Resolution Chromatography & Chromatography Communications* **1987**, *10*, 319-323.

Giddings, J. C. Unified Separation Science; John Wiley & Sons, Inc.: New York, 1991.

Giddings, J. C. Sample Dimensionality: a Predictor of Order-Disorder in Component Peak Distribution in Multidimensional Separation. *Journal of Chromatography A* **1995**, *703*, 3-15.

Goedert, M.; Guiochon, G. Sources of Error in Measurement of Retention Times. *Analytical Chemistry* **1970**, *42*, 962-968.

Goldberg, B. A Numerical Method of Resolving Peak Areas in Gas Chromatography. *Journal of Chromatographic Science* **1971**, *9*, 287-292.

Goodman, K. J.; Brenna, J. T. Curve Fitting for Restoration of Accurcay for Overlapping Peaks in Gas Chromatography/Combustion Isotope Ratio Mass Spectrometry. *Analytical Chemistry* **1994**, *66*, 1294-1301.

Grant, D. W. Capillary Gas Chromatography; John Wiley & Sons Ltd: West Sussex, UK, 1996.

Grob, K. On-Column Injection in Capillary Gas Chromatography : Basic Technique, Retention Gaps, Solvent Effects, 2nd ed.; Hüthig: Heidelberg, FRG, 1991.

Harris, D. C. Quantitative Chemical Analysis, 4th ed.; W. H. Freeman and Company: New York, 1995.

Hieftje, G. M. Signal-to-Noise Enhancement througth Instrumental Techniques. Part 1. *Analytical Chemistry* **1972**, *44*, 81A-88A.

Ho, C.-N.; Christian, G. D.; Davidson, E. R. Application of the Method of Rank Annihilation to Quantitative Analyses of Multicomponent Fluorescence Data from the Video Fluorometer. *Analytical Chemistry* **1978**, *50*, 1108-1113.

Ho, C.-N.; Christian, G. D.; Davidson, E. R. Application of Method of Rank Annihilation to Fluorescent Multicomponent Mixtures of Ploynuclear Aromatic Hydrocarbons. *Analytical Chemistry* **1980**, *52*, 1071-1079.

Ho, C.-N.; Christian, G. D.; Davidson, E. R. Simultaneous Multicomponent Rank Annihilation and Applications to Multicomponent Fluorescent Data Acquired by the Video Fluorometer. *Analytical Chemistry* **1981**, *53*, 92-98.

Hooker, T. F.; Jorgenson, J. W. A Transparent Flow Gating Interface for the Coupling of Microcolumn LC with CZE in a Comprehensive Two-Dimensional System. *Analytical Chemistry* **1997**, *69*, 4134 -4142.

James, C. S. Analytical Chemistry of Foods; Blackie Acadmeic and Professional: New York, 1995.

Juan, A. d.; Rutan, S. C.; Tauler, R.; Massart, D. L. Comparison Between the Direct Trilinear Decomposition and the Multivariate Curve Resolution - Alternating Least Sqaures Methods for the Resolution of Three-Way Data Sets. *Chemometrics and Intelligent Laboratory Systems* **1998**, *40*, 19-32.

Khaledi, M. G. High-Performance Capillary Electrophoresis: Theory, Techniques, and Applications; John Wiley & Sons, Inc.: New York, 1998.

Kinghorn, R. M.; Marriott, P. J. High Speed Cryogenic Modulation-A Technlogy Enabling Comprehensive Multidemensional Gas Chromatography. *Journal of High Resolution Chromatography* **1999**, *22*, 235-238.

Larmann, J. P. J.; Lemmo, A. V.; Moore, A. W., Jr.; Jorgenson, J. W. Two-Dimensional Separations of Peptides and Proteins by Comprehensive Liquid Chromatography-Capillary Electrophoresis. *Electrophoresis* **1993**, *14*, 439-447.

Ledford Jr., E. B.; Phillips, J. B.; Xu, J.; Gaines, R. B.; Blomberg, J. Ordered Chromatograms: A Powerful Methodology in Gas Chromatography. *American Laboratory* **1996**, *28*, 22-25.

Lee, A. L.; Lewis, A. C.; Bartle, K. D.; Quaid, J. B. M.; Marriott, P. J. A Comparison of Modulating Interface Technologies in Comprehensive Two-Dimensional Gas Chromatography (GC x GC). *Journal of Microcolumn Separations* **2000**, *12*, 187-193.

Lee, T. A.; Headley, L. M.; Hardy, J. K. Noise Reduction of Gas Chromatography /Mass Spectrometry Data Using Principal Component Analysis. *Analytical Chemistry* **1991**, *63*, 357-360.

Lemmo, A. V.; Jorgenson, J. W. Transverse Flow Gating Interface for the Coupling of Microcolumn LC with CZE in a Comprehensive Two-Dimensional System. *Analytical Chemistry* **1993**, *65*, 1576-1581.

Li, J. A Comutational Study on the Effect of Random Noise on the Precision and Accuracy of the Integration of Chromatographic Peaks. *Analytica Chimica Acta* **1999**, *388*, 187-199.

Li, S.; Hamilton, J. C.; Gemperline, P. J. Generalized Rank Annihilation Method Using Similarity Transformations. *Analytical Chemistry* **1992**, *64*, 599-607.

Liu, Y.-M.; Sweedler, J. V. Two-Dimensional Separations: Capillary Electrophoresis Coupled to Channel Gel Electrophoresis. *Analytical Chemistry* **1996**, *68*, 3928-3933.

Liu, Z.; Ostrovsky, I.; Farnsworth, P. B.; Lee, M. L. Instrumentation for Comprehensive Two-Dimensional Capillary Supercritical Fluid-Gas Chromatography. *Chromatographia* **1993**, *35*, 567-573.

Liu, Z.; Patterson Jr., D. G.; Lee, M. L. Geometric Approach to Factor Analysis for the Estimation of Orthogonality and Practical Peak Capacity in Comprehensive Two-Dimensional Separations. *Analytical Chemistry* **1995**, *67*, 3840-3845.

Liu, Z.; Sirimanne, S. R.; Patterson Jr., D. G.; Needham, L. L.; Phillips, J. B. Comprehensive Two-Dimensional Gas Chromatography for the Fast Separation and Determination of Pesticides Extracted from Human Serum. *Analytical Chemistry* **1994**, *66*, 3086-3092.

Lorber, A. Quantifying Chemical Composition from Two Dimensional Data Arrays. *Analytica Chimica Acta* **1984**, *164*, 293-297.

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

Mathworks Inc., T. M. W. Using MATLAB: Natick ,MA, 1998.

Miller, J. M. Chromatography: Concepts and Contrasts; John Wiley & Sons: New York, 1988.

Mittermayr, C. R.; Frischenschlager, H.; Rosenberg, E.; Grasserbauer, M. Filtering and Integration of Chromatographic Data: A Tool to Improve Calibration. *Fresenius Journal of Analytical Chemistry* **1997**, *358*, 456-464.

Moore, A. W. J.; Jorgenson, J. W. Rapid Comprehensive Two-Dimensional Separations of Peptides via RPLC-Optically Gated Capillary Zone Electrophoresis. *Analytical Chemistry* **1995**, *67*.

Murphy, R. E.; Schure, M. R.; Foley, J. P. Effect of Sampling Rate on Resolution in Comprehensive Two-Dimensional Liquid Chromatography. *Analytical Chemistry* **1998**, *70*, 1585-1594.

Nagels, L. J.; Creten, W. L.; Vanpeperstraete, P. M. Determination Limits and Distribution Function of Ultraviolet Absorbing Substances in Liquid Chromatographic Analysis of Plant Extract. *Analytical Chemistry* **1983**, *55*, 216-220.

Opiteck, G. J.; Jorgenson, J. W.; Moseley, M. A.; Anderegg, R. J. Two-Dimensional Microcolumn HPLC Coupled to a Single-Quadrupole Mass Spectrometer for the Elucidation of Sequence Tags and Peptide Mapping. *Journal of Microcolumn Separations* **1998a**, *10*, 365-375.

Opiteck, G. J.; Lewis, K. C.; Jorgenson, J. W.; Anderegg, R. J. Comprehensive On-Line LC/LC/MS of Proteins. *Analytical Chemistry* **1997**, *69*, 1518 -1524.

Opiteck, G. J.; Ramirez, S. M.; Jorgenson, J. W.; Moseley, M. A. Comprehensive Two-Dimensional High-Performance Liquid Chromatography for the Isolation of Overexpressed Proteins and Proteome Mapping. *Analytical Biochemistry* **1998b**, *258*, 349-361.

Ouchi, G. I. Peak Detection and Integration. *LC-GC* **1991**, *8*, 628-633.

Phillips, J. B.; Gaines, R. B.; Blomberg, J.; van der Wielen, F. W. M.; Dimandja, J. M.; Green, V.; Granger, J.; Patterson, D.; Racovalis, L.; de Geus, H. J.; de Boer, J.; Haglund, P.; Lipsky, J.; Sinha, V.; Ledford Jr., E. B. A Robust Thermal Modulator for Comprehensive Two-Dimensional Gas Chromatography. *Journal of High Resolution Chromatography* **1999**, *22*, 3-10.

Poe, R. B.; Rutan, S. C. Effects of Resolution, Peak Ratio, and Sampling Frequency in Diode-Array Fluorescence Detection in Liquid Chromatography. *Analytica Chimica Acta* **1993**, *283*, 845-853.

Prazen, B. J. Development of High Speed Hyphenated Chromatographic Analyzers and Second Order Data Analysis Technqiues, University of Washington, 1998.

Prazen, B. J.; Bruckner, C. A.; Synovec, R. E.; Kowalski, B. R. Enhanced Chemical Analysis using Parallel Column Gas Chromatography with Single-Detector Time-of-Flight Mass Spectrometry and Chemometric Analysis. *Analytical Chemistry* **1999a**, *71*, 1093-1099.

Prazen, B. J.; Bruckner, C. A.; Synovec, R. E.; Kowalski, B. R. Second Order Chemometric Standardization for High Speed Hyphenated Gas Chromatography: Analysis of GC/MS and Comprehensive GCxGC Data. *Journal of Microcolumn Separations* **1999b**, *11*, 97-107.

Prazen, B. J.; Synovec, R. E.; Kowalski, B. R. Standardization of Second-Order Chromatographic/Spectroscopic Data for Optimum Chemical Analysis. *Analytical Chemistry* **1998**, *70*, 218-225.

Quigley, W. W. C.; Fraga, C. G.; Synovec, R. E. Comprehensive LC x GC for Enhanced Headspace Analysis. *Journal of Microcolumn Separations* **2000**, *12*, 160-166.

Ramos, L. S.; Sanchez, E.; Kowalski, B. R. Generalized Rank Annihilation Method: II. Analysis of Bimodal Chromatographic Data. *Journal of Chromatography* **1987**, *385*, 165-180.

Rosenthal, D. Theoretical Limitations of Gas Chromatographic / Mass Spectrometric Identification of Multicomponent Mixtures. *Analytical Chemistry* **1982**, *54*, 63-66.

Rowe, K.; Bowlin, D.; Zou, M.; Davis, J. M. Application of 2-D Statistical Theory of Overlap to Three Separation Types: 2-D Thin Layer Chromatography, 2-D Gas

Chromatography, and Liquid Chromatography/Capillary Electrophoresis. *Analytical Chemistry* **1995**, *67*, 2994-3003.

Rowe, K.; Davis, J. M. Relaxation of Randomness in Two-Dimensional Statistical Model of Overlap: Theory and Verification. *Analytical Chemistry* **1995**, *67*, 2981-2993.

Samuel, C.; Davis, J. M. Application of Statistical Overlap Theory to Gas Chromatograms Simulated on Nonpolar Stationary Phases with Commercial Software. *Journal of Chromatography A* **1999**, *842*, 65-77.

Sánchez, E.; Kowalski, B. R. Generalized Rank Annihilation Factor Analysis. *Analytical Chemistry* **1986**, *58*, 496-499.

Sánchez, E.; Kowalski, B. R. Tensorial Resolution: A Direct Trilinear Decomposition. *Journal of Chemometrics* **1990**, *4*, 29-45.

Sánchez, E.; Ramos, L. S.; Kowalski, B. R. Generalized Rank Annihilation Method: I. Application to Liquid Chromatography-Diode Array Ultraviolet Detection Data. *Journal of Chromatography* **1987**, *385*, 151-64.

Schure, M. R. Resolution Enhancement of Chromatographic Data. *Journal of Chromatography* **1991**, *550*, 51-69.

Shi, W.; Davis, J. M. Test of Theory of Overlap for Two-Dimensional Separations by Computer Simulations of Three-Dimensional Concentration Profiles. *Analytical Chemistry* **1993**, *65*, 482-492.

Slonecker, P. J.; Li, X.; Ridgway, T. H.; Dorsey, J. G. Information Orthogonality of Two-Dimensional Chromatographic Separations. *Analytical Chemistry* **1996**, *68*, 682.

Statheropoulos, M.; Pappa, A.; Karamertzanis, P.; Meuzelaar, H. L. C. Noise Reduction of Fast, Repetitive GC/MS Measurements Using Principal Component Analysis (PCA). *Analytica Chimica Acta* **1999**, *401*, 35-43.

Sun, X. Y.; Singh, H.; Millier, B.; Warren, C. H.; Aue, W. A. Noise, Filters, and Detection Limits. *Journal of Chromatography A* **1994**, *687*, 259-281.

Taraszewski, W. J.; Haworth, D. T.; Pollard, B. D. Application of Static Signal-to-Noise Theory to the Detection and Integration of Dynmaic Signals. *Analytica Chimica Acta* **1984**, *157*, 73-82.

Venkatramani, C. J.; Xu, J.; Phillips, J. B. Separation Orthogonality in Temperature-Programmed Comprehensive Two-Dimensional Gas Chromatography. *Analytical Chemistry* **1996**, *68*, 1486-92.

Wilson, B. E.; Sánchez, E.; Kowalski, B. R. An Improved Algorithm for the Generalized Rank Annihilation Method. *Journal of Chemometrics* **1989**, *3*, 493-498.

Windig, W.; Antalek, B. Resolving Nuclear Magnetic Resonance Data of Complex Mixtures by Three-Way Methods: Examples of Chemical Solutions and the Human Brain. *Chemometrics and Intelligent Laboratory Systems* **1999**, *46*, 207-219.

Wise, B. M.; Gallagher, N. B. PLS-Toolbox Version 2.0; Eigenvector Research, Inc.: Manson, WA, 1998.

Young, D. S.; Tracy, R. P. Clinical Applications of Two-Dimensional Electrophoresis. *Journal of Chromatography A* **1995**, *698*, 163-179.

# Vita

Carlos Gerardo Moyano Fraga was born on August 6, 1971 in Houston, Texas. In 1989 he graduated from Sterling High School, Houston, Texas as the valedictorian. He obtained his Bachelor of Science in chemistry from the United States Air Force Academy in 1993. Upon graduation he attended in the University of Illinois, Urbana-Champaign where he worked in the research group of the late Prof. Timothy Nieman. After completing his requirements for a Master of Science in chemistry in 1994, he was assigned to the Technical Operations Division, McClellan AFB, California. While at TOD, he was chief of the nuclear chemistry section. He managed and performed research in support of international nuclear treaty monitoring. He also had a 60-day tour of duty as a United Nations inspector in Baghdad, Iraq in 1995. He came to the University of Washington to purse a Ph.D. in Analytical Chemistry in the autumn of 1992 as part of an Air Force assignment.