Fundamental Considerations in One and Two-Dimensional Chromatography for Improved Chemometric Analysis

David K. Pinkerton

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2017

Reading Committee:

Robert E. Synovec, Chair

Matthew F. Bush

Bo Zhang

Program Authorized to Offer Degree:

Chemistry

University of Washington

Abstract

Fundamental Considerations in One and Two-Dimensional Chromatography for Improved Chemometric Analysis

David K. Pinkerton

Chair of the Supervisory Committee:
Professor Robert E. Synovec
Department of Chemistry

One and two-dimensional chromatography offer well-established instrumental platforms that produce separations of chemical compounds in complex samples in a variety of fields and applications and when coupled with spectral detection, results in powerful instruments capable of producing informative data that facilitates quantitative results with high chemical selectivity. However, chemometric data analysis is often required to fully extract useful information from the data. This dissertation presents several research investigations on the interconnectedness between the fundamental theoretical, experimental, and instrumental considerations for one and two-dimensional chromatography, particularly gas chromatography, and the application of advanced chemometric data analysis methods. First, a new method to predict the probability of

quantitative success for one-dimensional chromatography based on sample complexity, separation peak capacity, n_c, and the algorithm specific limit of chemometric resolution using a simulation based study of multivariate curve resolution alternating least squares for demonstration. For comprehensive two dimensional gas chromatography the relationship of phase ratio, β, between the primary and secondary separation dimensions and the implications of β on 2D peak capacity, $n_{c,2D}$, were examined. The β ratio, $\beta_R = {}^1\beta/{}^2\beta$, is defined as a quantitative metric to facilitate this study. Overall, β_R substantially affected $n_{c,2D}$ by influencing retention factors on the ^{2}D column and thereby changing the modulation period necessary for proper ^{2}D column separations. Modulation period selection is then further investigated as a part of a new method to determine the true modulation ratio, M_R, from the measurable effective modulation ratio, M_R*, in comprehensive two-dimensional gas chromatography, GC×GC, which can be generalized to other comprehensive two-dimensional chromatographic methods. The method was developed through the investigation of modulator induced band broadening, as a function of ${}^1W_b$ and the selected modulation period, P_M. Finally, a new quantitative metric, trilinearity deviation ratio (TDR), is introduced to describe the trilinearity of two-dimensional chromatography data with spectral detection for the purpose of predicting the performance of the chemometric method parallel factor analysis (PARAFAC). It is found that use of a modulation period in the 1 s to 2 s range simultaneously optimizes ${}^1n_c$, $n_{c,2D}$, and TDR to facilitate low quantitative errors with PARAFAC.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

DEDICATION

For my wife Jacquie. Thank you for your continuous love, support, patience and encouragement. I love you.

Chapter 1. Introduction to Chromatography and Chemometric Data Analysis[1]

1.1 INTRODUCTION

One-dimensional (1D) chromatography and two-dimensional (2D) chromatography offer well-established instrumental platforms that produce separations of chemical compounds in complex samples in a variety of fields and applications. The most commonly implemented chromatographic instruments are 1D gas chromatography (GC), 1D liquid chromatography (LC), comprehensive 2D gas chromatography (GC × GC), and comprehensive 2D liquid chromatography (LC × LC) [1–4]. Coupling or "hyphenating" chromatographic instruments to spectroscopic detectors like multi-wavelength absorbance diode array detection (DAD), mass spectrometry (MS), and time-of-flight mass spectrometry (TOFMS), results in powerful instruments capable of producing informative data that facilitates quantitative results with high chemical selectivity [5,6]. While there are other chromatographic instruments and methods coupled with various detectors, this research presented herein focuses on GC–MS, LC–DAD, GC × GC–TOFMS and LC × LC–DAD, with a particular focus on gas chromatography, since the majority of research that employs chemometric resolution is performed with these systems.

[1]Portions of this Chapter have been adapted from D.K. Pinkerton, K.M. Pierce, R.E.Synovec, "Chemometric Resolution of Complex Higher Order Chromatographic Data with Spectral Detection" from Resolving Spectral Mixtures: With Applications from Ultrafast Time-Resolved Spectroscopy to Super-Resolution Imaging, Edited by Cyril Ruckebusch (2016).

Concurrent with these chromatographic instruments producing very large and information-rich data (typically 50-500 MB per sample injection onto the instrument), there are significant challenges in working with multiple chromatograms of this size, especially when analyzing complex samples. Different chemicals that possess similar chemical functional groups often have similar spectroscopic and chromatographic properties that make it difficult for analysts to physically resolve, identify, and quantify the chemical analytes in a complex mixture. It is the accurate and precise identification and quantification of the chemical analytes in complex mixtures that is paramount to converting the experimental data into useful information.

When physical resolution by chromatography is not fully achieved, analysts often turn to chemometrics to achieve mathematical resolution, to unmix the signal contributions from overlapped analytes. To illustrate mathematical resolution, Figure 1.1 (left) shows the observed signals for a simulated unresolved two-component mixture, LC–DAD (top) and GC × GC–TOFMS (bottom), wherein each chemical analyte is represented in the chromatographic domain(s) by a Gaussian elution profile with total signal proportional to concentration. Many chemometric resolution techniques are available which, when operated under proper conditions and given proper information, may be able to extract the pure component elution profiles and the pure component spectral profiles, as shown in Figure 1.1 (right). The mathematically resolved, pure component profiles for each analyte in the sample can then be individually examined and quantified. Figure 1.1 (top) is an example of resolving second order data, while Figure 1.1 (bottom) is an example of resolving third order data. A 1D chromatography instrument coupled with a univariate detector (e.g. GC-FID) would produce first-order data. Even higher dimensions of data can be achieved when multiple second order or multiple third order chromatographic runs

are simultaneously analyzed, thus adding a sample dimension to the data. The order of the data determines which chemometrics can be applied to resolve overlapping analyte peaks and different orders of data provide different advantages, sometimes referred to as the second order advantage, bilinear data, or third order advantage, trilinear data [7]. In general, when aiming to achieve mathematical resolution of overlapping analyte peaks, the advantage is that higher order data requires a user to input less information than lower order data. For example, chemometric resolution methods for first-order data (e.g. classical least squares analysis (CLS)) require the user to input a pure reference spectrum for each component in the mixture. An alternative approach to using CLS for first-order data would be to simultaneously analyze multiple samples to artificially increase the order of the data so higher order chemometrics can be used. Samples with second and third order advantages involve freeing the analyst from providing pure reference spectra (a big advantage for the analysis of unknown samples) and freeing the analyst from providing more than one chromatogram to analyze any component in the sample (a big advantage in the midst of analyte peak retention time shifting and other sources of uncontrollable and chemically-irrelevant variations between chromatograms).

The ability to mathematically resolve analytes within a single sample significantly relaxes the requirement for sample-to-sample retention time precision, minimizing the need to implement retention time alignment as a data preprocessing step because replicate sample injections to the instrument can be analyzed independently. However, we should be clear that producing higher order data does not necessarily ensure that the data is amenable to analysis by higher order chemometrics or that the user can make use of the second or third order advantages. Producing data that can take full advantage of these chemometric methods strongly depends on

the balance between the instrumental design, how the user operates an instrument, and the specific capabilities of the algorithm(s) implemented within the chemometric methods.

While there are specific advantages and disadvantages to each chemometric method, mathematical resolution generally requires at least some selectivity exists between analyte components in each data dimension. The ability of a data set to be decomposed into individual analyte component vectors also generally requires the data to be bilinear for second order data or trilinear for third order data, meaning each of the data dimensions are linearly independent from one another and the pixel-level data are composed of unique, reproducible, and analyte concentration-dependent signals at consistent retention times from each independent and unique analyte source that is present. To be more precise, contributions to the signal are also provided by baseline and indeterminate noise sources. Linear dependence between dimensions or inconsistencies in retention times can be removed by reducing the order of the data, but this also means that the user may not be able to fully utilize the advantages of the true dimensionality of the data. Accordingly, this chapter will first describe the instrumental designs and instrumental operation techniques that produce good quality high order data. Then the chapter will describe several chemometric methods that provide mathematical resolution for ND data, where $N > 1$, specifically, multivariate curve resolution-alternating least squares (MCR-ALS) for two-way data, generalized rank annihilation method (GRAM) for two-way data, and parallel factor analysis methods (PARAFAC and PARAFAC2) for three-way data [8–16].

1.2 Instrument Design and Utilization to Produce High Order Chemical Separations Data

Producing ND chromatographic data that engenders second order and third order advantages strongly depends on the quality of the instrument design and utilization. It is important to note that when discussing and comparing GC–MS, LC–DAD, GC × GC–TOFMS and LC × LC–DAD, or other related chromatographic instruments coupled with spectral detection, the instrumentation and experimental challenges are generally different, but the data analysis challenges for applying chemometric resolution methods are very similar.

GC–MS, LC–DAD, GC × GC–TOFMS or LC × LC–DAD should chromatographically separate as many analytes as possible with the shortest separation run time possible. However, sometimes it is not feasible to chromatographically resolve all the analytes, especially if a complex sample contains hundreds to thousands of chemical compounds and high throughput applications demand fast chromatographic acquisition rates (high experimental duty cycle). Therefore, the key goal for the analyst is to strike a balance between too little chromatographic resolution (time saving) and total chromatographic resolution (time wasting), while collecting appropriate data that can be readily analyzed and transformed into useful information to answer a specific analytical question of interest.

1.2.1 *Goals for Two-Way Data: GC–MS and LC–DAD*

The instrumental goals for 1D chromatography with spectral detection such as GC–MS and LC–DAD are to maximize selectivity, provide physical separation of analyte peaks, and keep the run time fast to improve the experimental duty cycle, so narrow chromatographic peaks

are preferred. Narrower peaks not only improve separation peak capacity, they also focus the analyte signal, improving sensitivity and signal-to-noise ratio (*S/N*). However, chemometric algorithms require some degree of selectivity between overlapped peaks when chromatographic resolution is not achieved, so each peak should be wide enough and sampled with a data acquisition rate high enough to produce at least ~ 10 to 20 spectral scans for each peak's elution profile, and provide enough chemical selectivity for chemometric algorithms to appropriately resolve analyte peaks. This requirement for 10 to 20 scans per peak-width-at-base does not mesh well with simultaneously trying to obtain narrow peaks, especially for quadrupole mass spectrometers (qMS) which are commonly used in GC–MS. A typical qMS with unit mass resolution can collect up to 2 to 4 scans/s over a 40 to 600 *m/z* range per scan. Therefore, analyte peaks in GC-qMS need to have a width of ~ 4 to 6 s to achieve ~ 10 to 20 scans/peak. In contrast, TOFMS can collect 500 full spectra/s over the same 40 to 600 *m/z* range, although most work with TOFMS is performed at 100 spectra/s. Therefore, at the 100 spectra/s data acquisition rate, GC-TOFMS provides ~ 10 to 20 spectra/peak for peaks as narrow as ~ 100 to 300 ms wide. This fast scanning rate is why the TOFMS is ideally suited for detecting the second chromatographic dimension of peaks in GC × GC–TOFMS.

When qMS is used in GC–MS, then collection of a sufficient number of spectra across the chromatographic peak width is especially important because of a spectral skew imposed by this detector. Spectral skew occurs because the qMS detector does not scan all *m/z* simultaneously, and within a single observed spectrum, later *m/z* can be measured at a different analyte concentration than earlier *m/z* because the Gaussian-shaped peak rapidly swept through the detector between observations, exposing each observation to a slightly different analyte

concentration. This causes the collection of spectra in a GC-MS chromatogram to be not only a function of retention time between spectra (totally expected, since this is chromatography after all), but also a function of retention time *within* each spectrum (potentially troubling to a chromatographer who wanted to collect bilinear data). This skewing effect is most noticeable when the peak signal is changing rapidly relative to the detector scan rate, such as the case with narrow chromatographic peaks. The TOFMS does not have this spectral skew because all *m/z* within each spectrum are analyzed at the same time, rather than scanned one-by-one while time passes. When the data are skewed and the analyte spectrum varies across the peak elution profile, the GC–MS data is not strictly bilinear Fortunately, skewed GC–qMS data can be corrected by applying an unskewing algorithm that corrects the signal intensity for each mass based on the measured intensity and the time needed to go from the start of the scan to each mass [17].

Application of mathematical resolution to LC–DAD data has the same peak sampling and spectral acquisition rate requirements as GC–MS data. A typical diode array detector offers a scan rate up to ~ 100 scans/s, but is commonly implemented at 40 scans/s over a UV-vis wavelength range of 200 to 700 nm [6,18]. The fast scan rates for DAD means that peaks only need to be ~ 0.5 s wide to achieve 10 to 20 spectra/peak, but the performance of 1D LC typically produces peak widths that range from 5 s to 1 min. A 5 s wide peak detected with a DAD spectral scan rate of 40 scans/s results in 200 spectra/peak, well beyond the minimum requirements for mathematical resolution. Indeed, the fast spectral scan rate of the DAD makes it ideally suited for the second chromatographic dimension of peaks in LC × LC–DAD. Achieving an adequate number of spectral scans across the chromatographic peaks with LC–DAD is less of a concern than in GC–MS, but UV-vis DAD is inherently less chemically selective than mass

spectrometry. Decreased detector selectivity generally requires additional chromatographic selectivity in order to resolve overlapped analyte peaks within a single chromatographic run if the analyst does not have *a priori* knowledge of the analyte spectrum or elution profile. Alternatively, this challenge is often handled by stacking multiple 2D chromatograms and analyzing the data as a single three-way array [19,20].

Combining replicates of 1D chromatography with spectral detection for simultaneous analysis can significantly reduce analysis time and has the potential to produce trilinear data that is amenable for third order chemometric methods, but there are some common non-ideal features in the data that often need to be addressed. Among these non-ideal features are differences in baseline signals and misalignment of peaks between samples. When either of these issues are present in the data, they can cause significant challenges (and shortcomings) in applying third order chemometrics. Proper tuning of the detector and running samples close together temporally can reduce these issues. However, even with great care in the instrumental analysis, these problems can persist. Because of this, replicates of two-way chromatograms can be fused end-to-end in the time domain into a longer two-way data set (spectral dimension × concatenated retention time dimension) rather than "stacked" as a three-way array (spectral dimension × retention time dimension × sample dimension). Figure 1.2 (top) provides a graphical representation of the rearrangement schemes for replicates of 1D chromatography with spectral detection. Reducing the data dimensionality in this manner can remove the deviations from trilinearity that are caused by analyte peak retention time shifts. The analyst also has the option to use peak alignment software to make the data approximately trilinear and amenable for third order chemometrics [21–23].

1.2.2 *Goals for Three-Way Data: GC × GC–TOFMS and LC × LC–DAD*

Comprehensive 2D chromatographic instruments, GC × GC–TOFMS and LC × LC–DAD, produce a third order chromatogram for each sample that is injected into the instrument. Before discussing the specific instrumental/experimental considerations required to produce sufficiently trilinear chromatograms, it is necessary to describe the general design of the instruments. In comprehensive 2D chromatography, two columns are connected in series with a modulator between them to move eluate from one column to the next, and a spectral detector is positioned at the end of the second column [4,18]. The two columns should have stationary phases that produce separations that are "complementary" and not overly correlated, in order to more fully utilize the two-dimensional peak capacity offered by comprehensive 2D chromatography. Peak capacity, at unit chromatographic resolution, is simply the separation run time in a given separation dimension divided by the average analyte peak width at base. Hence, the ideal peak capacity of a comprehensive 2D separation is the product of the peak capacities for the first column (^{1}D) and second column (^{2}D) separation dimensions, far greater than the peak capacity offered by 1D chromatography. Peak capacity is a figure-of-merit that roughly indicates to the amount of chemical information that can be gleaned from 1D and 2D chromatographic separations. The modulator serves as an automated sample injection device that interfaces the two columns, placed in-series, in the 2D chromatographic instrument. Analytes are initially separated along the ^{1}D separation dimension, with the modulator periodically injecting eluate from this ^{1}D column onto the ^{2}D separation dimension using a constant duty cycle. The ^{2}D separations occur in real-time while preserving chromatographic resolution that was achieved on ^{1}D. GC × GC–TOFMS commonly uses a thermal modulator which is capable of periodically focusing the ^{1}D eluate and improving *S/N*. LC × LC–DAD commonly uses an 8 or 10 port

switching valve as the modulator. The time length of the ^{2}D separation, is referred to as the modulation period (P_M) and it is necessarily short (on the order of one second to several seconds) relative to the ^{1}D separation time (on the order of minutes) so that each ^{1}D peak can be sampled by the modulator and separated by ^{2}D an appropriate number of times. Since the detector is positioned at the end of the second column and the second column serially followed the first column, the number of spectra detected and collected along the ^{2}D chromatographic peak profile is determined directly from the second column peak width at base ($^{2}W_b$) relative to the detector spectral scan rate. In contrast, detector sampling of the ^{1}D chromatographic peak profile is dictated by the modulation ratio (M_R). The M_R of a comprehensive 2D chromatographic peak is defined as the first column peak width at base ($^{1}W_b$) divided by P_M ($M_R = {}^{1}W_b / P_M$) [24].

For comprehensive 2D chromatography with spectral detection, generally, a M_R of ~ 2 to 4 is capable of producing a trilinear chromatogram with sufficient chemical selectivity in all three dimensions, concurrent with sufficient data density to quantitatively define the ^{1}D and ^{2}D peaks, and all the while reasonably utilizing the high peak capacity offered by comprehensive 2D chromatography. Smaller M_R values do result in noticeable broadening of ^{1}D peaks due to "undersampling" of the analyte peaks [25–27]. Larger M_R values are better at defining the ^{1}D peak and its shape, avoiding broadening, and better at detecting chemically selective signals among overlapping peaks. However, the trade-off for larger M_R is reduced P_M, which results in reduced peak capacity in ^{2}D, which in turn may cause a reduced overall 2D peak capacity. For these reasons, M_R ~ 2 to 4 is preferred to balance peak capacity utilization, while appropriately sampling peaks in both dimensions. Analysts should strive to design a comprehensive 2D separation by optimizing the peak capacity of the ^{1}D separation first, and then designing the ^{2}D

separation to have the appropriate P_M taking into account ^1W_b to provide a proper M_R. The resulting ^{2}D peak capacity at this P_M will provide a true improvement to the overall comprehensive 2D peak capacity. These concepts are discussed in the context of specific GC × GC–TOFMS and LC × LC–DAD instrumental operation parameters in the following paragraphs.

When a ^{1}D separation in GC × GC–TOFMS is sufficiently optimized and ^1W_b is kept small (^1W_b ~ 4 to 8 s) a relatively fast modulation period (P_M ~ 1 to 2 s) is required to achieve a desirable M_R ~ 4. Peak widths of ^2W_b on the ^{2}D separation range from ~ 40 to 400 ms, so the 500 spectra/s provided by the TOFMS is ideally suited for this instrumental platform, readily providing the ~ 10 to 20 spectra/peak, and a ^{2}D peak capacity range of ~ 10 to 25. An alternative GC × GC–TOFMS method is commonly practiced, with the goal of running a longer modulation period (P_M ~ 5 to 8 s) to improve the selectivity of the ^{2}D separation (i.e., to increase the ^{2}D peak capacity). Achieving the same M_R with this longer P_M would then require that the ^{1}D separation be run sub-optimally in order to produce wide enough peaks (^1W_b ~ 20 to 30 s). Ironically, the overall comprehensive 2D separation peak capacity is compromised by this second approach. Indeed, for all comprehensive 2D separations, the analyst should strive to initially optimize the ^{1}D separation, so as not to reduce separation performance. If the ^{1}D separation is basically optimized, then the use of a long P_M translates into an "undersampling" of the ^{1}D analyte peaks, resulting in a $M_R < 2$, which negates the ability to produce chemically selective third order data for chemometric analysis. To this point, a long P_M has been shown to be problematic for chemometric analysis, causing significant ^{2}D retention time shifts, defined as $\Delta^2 t_R$, between modulations of an analyte, due to the necessary use of temperature programming in GC × GC [28]. The term trilinearity deviation ratio (TDR) was defined as a metric to study this issue. The

TDR is the ^{2}D retention time shift normalized relative to the peak width (TDR = $\Delta^2 t_R$ / ^2W_b) and provides a quantitative metric for the trilinearity of the data that describes the performance of trilinear chemometrics. When TDR = 0, it indicates that the data is strictly trilinear and no shifting has occurred along the ^{2}D separations from one peaklet to another for a given analyte modulated from the ^{1}D separation, as shown in Figure 1.3 (top). Herein, we refer to each peak detected in the ^{2}D separation dimension as a peaklet, because groups of the peaklets originate from each ^{1}D dimension analyte peak. Note that the M_R is not necessarily equal to the number of peaklets observed for a given ^{1}D chromatographic peak. While it is still possible to use trilinear chemometric methods when TDR > 0, large TDR values increase the quantitative errors associated with those modeling techniques. If third order chemometric models are to be used, then the data must be sufficiently trilinear (TDR ~ 0 to 0.05 in the initial study); often the use of a long P_M ~ 5 to 8 s coupled with fast temperature programs in GC × GC results in data that is not amenable for analysis with third order chemometrics [28]. Under these circumstances, the data must be rearranged prior to chemometric analysis. Generally this is accomplished by reducing the order of the data to apply second order chemometrics, unfolding the three-way data matrix (time ^{2}D dimension × time ^{1}D dimension × spectral dimension) by concatenating each of the n modulations, along the time ^{2}D dimension, to produce a two-way matrix (spectral dimension × n modulations of time ^{2}D dimension). Figure 1.2 (bottom) is a graphical depiction of the native three-way data structure with modulations depicted as slabs (spectra × time ^{2}D) along the time ^{1}D dimension, and the unfolded two-way data with n modulations connected end-to-end. Alternatively, localized peak retention time alignment can be applied if third order chemometrics are to be used. To illustrate localized peak retention time alignment, Figure 1.3

(bottom) shows a non-trilinear chromatogram, TDR = 0.2, that would require retention time alignment prior to analysis by third order chemometrics.

The necessary data considerations for applying mathematical resolution to LC × LC–DAD are generally very similar to those of GC × GC–TOFMS; but unlike GC × GC–TOFMS, the solutions to these challenges often come from changing how the LC × LC–DAD instrument functions, rather than varying the separation parameters. Overall, there is significantly less variability in how analysts operate LC × LC–DAD instruments than compared to GC × GC–TOFMS. Typical LC × LC experiments have a total ^{1}D run time of 30 to 60 min and produce ^{1}D peaks with a $^{1}W_b$ ~ 30 s to 1 min. Unfortunately, the LC × LC valve modulation system does not provide analyte focusing to improve the *S/N*, which is a benefit provided by the GC × GC–TOFMS thermal modulator. Also, the P_M selection in state-of-the-art LC × LC–DAD is generally restricted to a relatively narrow range (P_M ~ 15 to 30 s), with ^{2}D peak widths $^{2}W_b$ ~ 0.5 to 2.5 s, resulting in a ^{2}D peak capacity range of ~ 10 to 20. The instrumental constraints of LC × LC–DAD, particularly the relatively long P_M necessary to produce adequate selectivity on the ^{2}D separation dimension, coupled with the peak widths on ^{1}D, results in M_R ~ 1 to 3, which in turn impacts applicability of some chemometric methods, so extra care may be required to ensure that the data is amenable to third order chemometric analysis for a single sample run. Recent LC × LC instrumental developments have helped to reduce the need for ^{2}D peak alignment or the use of lower order chemometric resolution methods, namely the introduction of a synchronized dual valve modulation system, typically implemented as a single 8 or 10 port valve, and independently programmable ^{2}D separation gradients [18,21,29,30].

1.3 CHEMOMETRIC RESOLUTION METHODS

The chemometric resolution methods discussed in this chapter have at least one design property in common: they all assume the observed signal profile of a sample mixture is a linear combination of the signal profiles of the pure components in the mixture and chemically-irrelevant sources of variations in observed signals are minimized. This is often a reasonable assumption for chromatography detectors, especially for absorbance spectroscopy when the spectra of unresolved components are additive according to the Beer-Lambert Law, or when ion counts provided by mass spectrometers are additive. However, chemometric resolution methods have unique benefits and tradeoffs in terms of the amount of information a user must provide and the limitations imposed by uncontrollable, chemically-irrelevant variations in the data. For example, aside from requiring the user to input the one-way observed mixed spectrum of the original sample mixture ($\boldsymbol{x}$) as a vector, CLS also requires the user to input the pure spectrum of each component in the mixture ($\boldsymbol{S}$) as a matrix. Then CLS uses the model

$$x = \boldsymbol{cS^T} + \boldsymbol{e} \tag{1.1}$$

to compute the concentrations of the components in the mixture ($\boldsymbol{c}$) vector by applying the pseudo inverse of $\boldsymbol{S}$ and minimizing the least squares error of fit ($\boldsymbol{e}$) vector between $\boldsymbol{x}$ and the $\boldsymbol{cS}^T$ model. CLS outputs the resolved quantitative spectrum for each component in the mixture, but the user needs to know the identity of the components in the mixture before CLS can successfully resolve one-way data. To resolve the overlapping signals of a two-way sample profile using MCR-ALS, the user needs to only input the number of components in the mixture (the identity of those components can remain unknown, if necessary); this is an example of a second order advantage. The second dimension of chemical selectivity provides chemically-

relevant information to the MCR-ALS algorithm that was unavailable when CLS was given only one-way data. An analogous additional advantage is provided by third order data for PARAFAC algorithms wherein fewer user inputs are required to achieve mathematical resolution of overlapped analyte peaks due to the additional data dimension contributing additional chemically selective information. Please note that for equations that deal with mathematically manipulating ND chromatograms in this chapter, lower case letters represent scalars, superscript T means the former variable is transposed, bold lower case letters represent vectors, bold uppercase letters represent matrices, and underlined bold uppercase letters represent 3D arrays.

1.3.1 *MCR-ALS*

In terms of two-way chromatographic data, the chemometric model for MCR-ALS is

$$\boldsymbol{D} = \boldsymbol{C}\boldsymbol{S}^{\boldsymbol{T}} + \boldsymbol{E} \tag{1.2}$$

where $\boldsymbol{D}$ represents a given observed two-way chromatogram with retention times in the rows dimension and wavelengths (or ion *m/z*) in the columns dimension, $\boldsymbol{C}$ contains pure predicted chromatograms for the components, $\boldsymbol{S}^T$ contains pure predicted spectra for the components, and $\boldsymbol{E}$ is the minimized error of residuals between the predicted and observed two-way chromatogram. We refer the reader to excellent references that thoroughly describe MCR-ALS [12,20,31–33]. In brief, while using appropriate constraints, initial estimated pure chromatograms and pure spectra are multiplied to yield the $\boldsymbol{C}\boldsymbol{S}^T$ model, tested for convergence, then $\boldsymbol{C}$ and $\boldsymbol{S}$ values are alternately and repeatedly iterated, multiplied, and again tested for convergence, until the convergence criterion is finally met. MCR-ALS outputs a pure two-way chromatogram for each component present in $\boldsymbol{D}$. To apply MCR-ALS, the user does not have to know the identity of the components in the mixture. Instead, the user inputs initial values for $\boldsymbol{C}$

and ***S***, and the user chooses appropriate constraints (such as unimodality, nonnegativity, or local rank) and an appropriate convergence criterion (such as threshold number of iterations, minimal value of ***E***, or an improvement threshold for lack of fit between the predicted model and ***D***). Initial values for ***C*** and ***S*** can be provided by evolving factor analysis (EFA), simple-to-use self-modeling analysis (SIMPLISMA) orthogonal projection approach (OPA), or Key Set Factor analysis (KFA), all of which, in general, try to select the purest rows or columns present in the original data as initial estimates [12].

As mentioned previously, MCR-ALS requires that the two-way input data is bilinear. In GC–MS and LC–DAD applications, if samples or replicates compose a third dimension, users can append matrices to remove the effects of retention time variations. Users can unfold the three-way array along t_R dimensions to produce a suitable two-way matrix, as in Figures 1.2 and 1.3, and then the resolved component profiles that are produced can be refolded into the original three-way form [32]. In GC–MS and LC–DAD applications, if samples or replicates compose a fourth dimension and t_R variations are problematic, then unfolding along the t_R dimension and augmenting samples or replicates into a single dimension can remove the problematic effects of analyte peak retention time shifts between chromatograms.

1.3.2 *GRAM*

Like MCR-ALS, GRAM is also applicable to resolving overlapping peaks in two-way chromatographic data. Unlike MCR-ALS, GRAM is not iterative, meaning it *does not* require the user to input convergence criteria (or iteration stopping point). GRAM outputs a final unique set of quantitative chromatograms for the resolved components in a given sample mixture

chromatogram (***M***). However, GRAM *does* require the user to input a standard mixture chromatogram (***N***). In brief, GRAM models ***M*** and ***N*** as

$$\boldsymbol{M = XC_M Y^T} \tag{1.3}$$

and

$$\boldsymbol{N = XC_N Y^T} \tag{1.4}$$

where ***X*** is a matrix containing the pure component profiles in the first chromatographic dimension, ***Y*** is a matrix containing the pure component profiles in the second chromatographic dimension, and $\boldsymbol{C_N}$ is a diagonal matrix containing the concentration of *k* known components in the standard sample. Thus, given ***N***, ***M***, $\boldsymbol{C_N}$, and *k* (*k* can be objectively and algorithmically estimated when *k* is initially unknown), it is possible to calculate the resolved pure chromatographic profiles of the components in the unknown sample mixture (***X*** and ***Y***) as well as the component's concentrations relative to the standard ($\boldsymbol{C_M}$). We refer the reader to excellent references that thoroughly describe GRAM [13,14,34–40]. In brief, GRAM mathematically resolves the components of the sample mixture, given the following requirements: the unknown sample must contain relative analyte component concentrations that differ from the standard (for example, if the two analytes of interest that overlap in the unknown sample have concentration ratio = 2:1, the standard must contain the two analytes at concentration ratio ≠ 2:1), the overlapped peaks must have some chromatographic and spectral resolution on the respective dimensions, and the analyte retention times and peak shapes (shapes, not magnitudes) must be equal between the unknown sample and the standard (the data must be bilinear). The unknown sample can contain interference peaks that are absent in the standard and vice versa. The key is that the user augments ***M*** and ***N*** into a single matrix so that all components that are common to both ***M*** and ***N*** are modeled. Thus, GRAM is initiated by either augmenting or summing ***M*** and ***N***

into a single matrix. Then the GRAM algorithm goes on as follows: it performs singular value decomposition (SVD) on that single matrix, the SVD decomposition matrices are truncated according to *k* components, an eigenvalue problem is solved for the truncated matrices yielding eigenvectors and eigenvalues. The eigenvectors and decomposition matrices from SVD are used to determine the pure component profiles (***X*** and ***Y***), and the eigenvalues and $\boldsymbol{C_N}$ are used to determine the concentrations of the unknown sample components ($\boldsymbol{C_M}$) relative to the standard mixture.

1.3.3 *PARAFAC*

As described earlier, GC × GC–TOFMS and LC × LC–DAD instruments produce trilinear data wherein the signal from an analyte is essentially the outer product of three fixed vectors: the first separation dimension profile (^{1}D), the second separation dimension profile (^{2}D), and the spectrum. PARAFAC can mathematically resolve a given trilinear structure by modeling each component as the outer product of its three fixed vectors, while simultaneously filtering out nontrilinear signals such as noise contributions to signals. The trilinear chemometric model for PARAFAC is

$$\underline{\boldsymbol{R}} = \sum_{for\ i=1\ to\ n} \boldsymbol{x}_i \otimes \boldsymbol{y}_i \otimes \boldsymbol{z}_i + \underline{\boldsymbol{E}} \tag{1.5}$$

wherein $\underline{\boldsymbol{R}}$ represents a given observed three-way chromatogram, $\boldsymbol{x}_i$ is the pure predicted ^{2}D chromatogram for the i^{th} component, $\boldsymbol{y}_i$ is the pure predicted ^{1}D chromatogram for the i^{th} component, $\boldsymbol{z}_i$ is the pure predicted spectral profile of the i^{th} component, $\underline{\boldsymbol{E}}$ is the minimized error of residuals between the predicted and observed chromatograms, and *n* is the total number of factors in the chromatographic subregion (factors are the sum of number of analyte components and number of sources of noise or baseline contributions to the total signal). We refer the reader

to excellent references that thoroughly describe PARAFAC [6,8,28,41–51] and an enhanced version called PARAFAC2 [9,15]. In brief, PARAFAC mathematically resolves overlapping components of a 2D chromatographic subregion while filtering out nontrilinear noise contributions (as long as there is not an excessive number of factors in the model) by using alternating least squares to iteratively converge upon the $\boldsymbol{x}_i$, $\boldsymbol{y}_i$, and $\boldsymbol{z}_i$ values that minimize $\underline{\underline{\boldsymbol{E}}}$.

In theory, PARAFAC performs well given the following requirements: the unknown sample array must be sufficiently trilinear (TDR ≤ 0.05), it must contain chemically selective information for each component on two of the three dimensions, the user must choose appropriate initialization values (common choices are random values or values provided by generalized eigenvalue decomposition of the raw data), and the user must choose stopping criterion (some form of minimizing $\underline{\underline{\boldsymbol{E}}}$ such as reaching a small threshold value for relative change in fit between two successive models). It is beneficial if a small maximum number of components are known to be present in the three-way data. But in the absence of this *a priori* knowledge, convergence of the algorithm and objective determination of the number of components present can be achieved by building models of increasing number of factors and automatically detecting the number of factors at which overfitting occurs, that number minus one factor is the optimal model. Optional constraints such as nonnegative signals and unimodality in peak shape are often reasonably acceptable assumptions that can be input into a PARAFAC algorithm to facilitate efficient convergence to a final solution, as well, but these are not requirements. As mentioned previously, PARAFAC requires trilinear data. But to be clear, PARAFAC is robust against some ${}^2t_R$ shifting and it is still useful for data with some deviations from trilinearity (Pinkerton et al., 2015). With GC × GC–TOFMS and LC × LC–DAD data, if ${}^2t_R$ variations are problematic, use appropriate alignment techniques before submitting the data to

PARAFAC. If samples or replicates compose a fourth dimension in the data, then unfold the ${}^2t_R$ dimension to reduce the data from four-way to three-way and align all the chromatograms to a target before doing PARAFAC. It is rare to apply PARAFAC to two-way GC-MS data (GC–MS or LC–DAD), but if samples or replicates compose a third dimension, then align all of the chromatograms to a target chromatogram before submitting the three-way data to PARAFAC. Another excellent chemometric option to consider is PARAFAC2, which is an advanced version of PARAFAC that provides as many $\boldsymbol{x}_{i,k}$ vectors as there are modulations in the three-way data, where $x_{i,k}$ is the pure predicted ^{2}D chromatogram for the i^{th} component in the k^{th} modulation. PARAFAC2 is robust against ${}^2t_R$ variations within a comprehensive 2D chromatographic run because PARAFAC2 is built to assume the covariance matrix of each second column elution profile is constant (e.g. the cross product matrix of $\boldsymbol{X}_k$, specifically $\boldsymbol{X}_k{}^T\boldsymbol{X}_k$, is constant for each component as a function of retention time). This condition is met when $\Delta^2 t_R$ is constant between pairs of $\boldsymbol{X}_k$ matrices, and there are no changes in the shape of each component peak profile.

1.4 General Overview

Skilled researchers have developed and utilized MCR-ALS, GRAM, and PARAFAC algorithms to mathematically resolve complex samples and quantify interesting analytes from environmental samples, industrial materials, fuels, and biological samples. The chemometrics enhances the detection, identification and quantification of the analyte peaks that otherwise would not be readily analyzed by more conventional data analysis methods. Hence, using chemometric resolution methods substantially broadens the scope of these instrumental platforms to provide a more in-depth and confident analysis of ever increasing complex samples. All of these chemometric methods seem to provide a satisfactory level of accuracy and precision in

applications if the data input for analysis meets specific standards of quality. Future challenges involve the general acceptance, more common implementation, and automation of chemometric software with the relatively large data sets that are produced with GC–MS, LC–DAD, GC × GC–TOFMS and LC × LC–DAD, and in particular for the comprehensive 2D separation methods. Currently most applications of chemometric resolution requires the analyst to be an expert in the chemometrics field, so most software that is commonly applied is provided by the instrument manufacturer, using algorithms that are not in the public domain and accordingly are not traditionally labeled as "chemometric" software, although sharing the same goal to convert chemical data into chemical information.

1.5 Overview of Following Chapters

The following chapters describe a total of four novel concepts that were studied in depth for this work. Each of the following chapters relate specifically to either one-dimensional gas chromatography or comprehensive two-dimensional gas chromatography; but, as indicated by the above introduction the fundamental chromatographic principles and analytical methods studied are also generally applicable to both liquid and gas chromatography with minor additional instrumental consideration. A brief abstract of each chapter is provided below.

1.5.1 *Chapter 2: Determining the Probability of Achieving Successful Quantitative Analysis for Gas Chromatography-Mass Spectrometry*

A new method is presented to determine the probability of achieving a successful quantitative analysis for gas chromatography-mass spectrometry. The proposed theory is based upon a probabilistic description of peak overlap in GC-MS separations to determine the probability of obtaining a successful quantitative analysis which has its lower limit of

chromatographic resolution R_s at some minimum chemometric resolution, R_s*; that is to say, successful quantitative analysis can be achieved when $R_s \geq R_s$*. The value of R_s* must be experimentally determined and is dependent on the chemometric method to be applied. This method makes use of the assumption that analyte peaks are independent and randomly distributed across the separation space, or are at least locally random, namely that each analyte represents an independent Bernoulli random variable, which is then used to predict the binomial probability of successful quantitative analysis. The theoretical framework is based on chromatographic saturation factor and chemometric enhanced peak capacity. For a given separation, the probability of quantitative success can be improved via two pathways, a chromatographic efficiency pathway that reduces the saturation of the sample, and a chemometric pathway that reduces R_s* and improves the chemometric enhanced peak capacity. This theory is demonstrated through a simulation based study to approximate the resolution limit, R_s*, of multivariate curve resolution-alternating least squares (MCR-ALS). For this study, the limit R_s* was found to be between 0.2 and 0.3, depending on the analytical expectations.

1.5.2 *Chapter 3: Implications of Phase Ratio for Maximizing Peak Capacity in Comprehensive Two-Dimensional Gas Chromatography Time-of-Flight Mass Spectrometry*

The relationship between the phase ratio, β, of the primary (^{1}D) and secondary (^{2}D) separation dimensions of comprehensive two-dimensional (2D) gas chromatography (GC×GC) separations, and the implications of β on realization of maximal 2D peak capacity, $n_{c,2D}$, are examined. A GC×GC chromatographic system with time-of-flight mass spectrometry, TOFMS, was otherwise held constant for the separation of a multi-component test mixture spanning a range of chemical functionalities, while only the β of the two analytical columns were changed,

$^1\beta$ for ^{1}D and $^2\beta$ for ^{2}D. Six column sets were studied using common, commercially available β values. The β ratio, $\beta_R = {}^1\beta/{}^2\beta$, is defined as a quantitative metric to facilitate this study. It is demonstrated that β_R plays a key role in maximizing $n_{c,2D}$. Overall, β_R substantially affected $n_{c,2D}$ by influencing retention factors on the ^{2}D column, ^2k, and thereby changing the modulation period, P_M, necessary for proper ^{2}D column separations. The necessary changes to P_M modify the modulation ratio, M_R, which affects the ^{1}D column peak widths and ^1n_c due to the impact of undersampling. Through changes to $^1\beta$, the range of ^2k can be controlled, with subsequent effects to both ^2n_c and ^1n_c. These effects were opposite in direction, such that improvements to ^2n_c may result in declines in ^1n_c. It is observed that due to the pseudo-isothermal nature of the ^{2}D separation, there are diminishing returns to extending the ^2n_c at the cost of ^1n_c. In this particular study, column set 3 (^{1}D: 20 m length, 250 μm i.d., 0.25 μm film; ^{2}D: 2 m, 180 μm i.d., 0.2 μm film; $\beta_R = 1.11$) with a P_M of 3 s provided the highest theoretical $n_{c,2D}$ of ~8200, though this was at a relatively low M_R of ~1.8. Column set 2 (^{1}D: 20 m length, 250 μm i.d., 0.5 μm film; ^{2}D: 2 m, 180 μm i.d., 0.2 μm film; $\beta_R = 0.56$) with a P_M of 1.5 s provided a high theoretical $n_{c,2D}$ of ~5800, at a much higher M_R of ~3.7. Though column set 2 had a lesser total peak capacity than column set 3, its higher M_R suggests that by improving the ^{1}D column efficiency (i.e., narrowing the ^{1}D column peak widths) to improve ^1n_c, can result in an increased theoretical $n_{c,2D}$.

1.5.3 *Chapter 4: Method to Determine the True Modulation Ratio for Comprehensive Two-Dimensional Gas Chromatography*

A new method is presented to determine the true modulation ratio, M_R, from the measurable effective modulation ratio, M_R*, in comprehensive two-dimensional gas chromatography, GC × GC, without the requirement for a detector at the end of the primary

column. The method was developed through the investigation of modulator induced band broadening, as a function of ${}^1W_b$ and the selected modulation period, P_M, for simulated GC × GC data, by first defining primary column ^{1}D peak(s) and simulating the modulation process. Gaussian curve fitting is used to model each modulated secondary column separation peaklet, ^{2}D, in the unfolded GC × GC data to accurately determine the maxima of the peaklet distribution, followed by Gaussian curve fitting to the maxima to determine the effective ^{1}D peak profile and width, ${}^1W_b$*. The relationship between ${}^1W_b$ and ${}^1W_b$* is studied as a function of the effective modulation ratio, M_R*, which is ${}^1W_b$* divided by P_M, in order to determine the true modulation ratio, M_R, which is ${}^1W_b$ divided by P_M. We explore how peak sampling phase (in-phase and out-of-phase) plays a role in the relationship between M_R and M_R*. Experimental validation of the simulated results is also provided, to span a range of commonly implemented conditions with typical ${}^1W_b$ (2–4.5 s) and P_M (0.25–8 s). Use of $M_R < 2$ significantly broadens the ^{1}D peak (M_R* $\geq 1.2\ M_R$) corresponding to a loss in ^{1}D peak capacity, ${}^1n_c \geq 20\%$. The new method relies upon mapping from M_R* to M_R, which is discussed in relation to peak capacity theories for GC × GC. It is found that optimizing ${}^1n_c$ in GC × GC requires that ${}^1W_b$ is minimized and must be sampled with a sufficiently short P_M (1–2 s) to minimize modulator induced band broadening and a subsequent reduction in the effective ^{1}D peak capacity.

1.5.4 *Chapter 5: Trilinearity Deviation Ratio: A New Metric for Chemometric Analysis of Comprehensive Two-Dimensional Gas Chromatography Time-of-Flight Mass Spectrometry Data*

Comprehensive two-dimensional gas chromatography coupled with time-of-flight mass spectrometry (GC × GC – TOFMS) is a well-established instrumental platform for complex samples. However, chemometric data analysis is often required to fully extract useful

information from the data. We demonstrate that retention time shifting from one modulation to the next, $\Delta^2 t_R$, is not sufficient alone to quantitatively describe the trilinearity of a single GC × GC – TOFMS run for the purpose of predicting the performance of the chemometric method parallel factor analysis (PARAFAC). We hypothesize that analyte peak width on second dimension separations, ${}^2W_b$, also impacts trilinearity, along with $\Delta^2 t_R$. The term Trilinearity Deviation Ratio, *TDR*, which is $\Delta^2 t_R$ normalized by ${}^2W_b$, is introduced as a quantitative metric to assess accuracy for PARAFAC of a GC × GC – TOFMS data cube. We explore how modulation ratio, M_R, modulation period, P_M, temperature programming rate, T_{ramp}, sampling phase (in-phase and out-of-phase), and signal-to-noise ratio, *S/N*, all play a role in PARAFAC performance in the context of *TDR*. Use of a P_M in the 1 s to 2 s range provides an optimized peak capacity for the first dimension separation (500 to 600) for a 30 min run, with an adequate peak capacity for the second dimension separation (12 to 15), concurrent with an optimized two-dimensional peak capacity (6000 to 7500), combined with sufficiently low *TDR* values (0 to 0.05) to facilitate low quantitative errors with PARAFAC (0 to 0.5 percent). In contrast, use of a P_M in the 5 s or greater range provides a higher peak capacity on the second dimension (30 to 35), concurrent with a lower peak capacity on the first dimension (100 to 150) for a 30 min run, and a slightly reduced two-dimensional peak capacity (3000 to 4500), and furthermore, the data is not sufficiently trilinear for the more retained second dimension peaks in order to directly use PARAFAC with confidence.

1.7 REFERENCES

[1] M.M. Bushey, J.W. Jorgenson, Automated instrumentation for comprehensive two-dimensional high-performance liquid chromatography of proteins, Anal. Chem. 62 (1990) 161–167. doi:10.1021/ac00201a015.

[2] P.W. Carr, D.R. Stoll, X. Wang, Perspectives on Recent Advances in the Speed of High-Performance Liquid Chromatography, Anal. Chem. 83 (2011) 1890–1900. doi:10.1021/ac102570t.

[3] J.C. Giddings, Two-dimensional separations: concept and promise, Anal. Chem. 56 (1984) 1258A–1270A. doi:10.1021/ac00276a003.

[4] Z. Liu, J.B. Phillips, Comprehensive Two-Dimensional Gas Chromatography using an On-Column Thermal Modulator Interface, J. Chromatogr. Sci. 29 (1991) 227–231. doi:10.1093/chromsci/29.6.227.

[5] K.M. Pierce, J.C. Hoggard, J.L. Hope, P.M. Rainey, A.N. Hoofnagle, R.M. Jack, B.W. Wright, R.E. Synovec, Fisher Ratio Method Applied to Third-Order Separation Data To Identify Significant Chemical Components of Metabolite Extracts, Anal. Chem. 78 (2006) 5068–5075. doi:10.1021/ac0602625.

[6] S.E.G. Porter, D.R. Stoll, S.C. Rutan, P.W. Carr, J.D. Cohen, Analysis of Four-Way Two-Dimensional Liquid Chromatography-Diode Array Data: Application to Metabolomics, Anal. Chem. 78 (2006) 5559–5569. doi:10.1021/ac0606195.

[7] K.S. Booksh, B.R. Kowalski, Theory of Analytical Chemistry, Anal. Chem. 66 (1994) 782A–791A. doi:10.1021/ac00087a718.

[8] R. Bro, PARAFAC. Tutorial and applications, Chemom. Intell. Lab. Syst. 38 (1997) 149–171. doi:10.1016/S0169-7439(97)00032-4.

[9] R. Bro, C.A. Andersson, H.A.L. Kiers, PARAFAC2—Part II. Modeling chromatographic data with retention time shifts, J. Chemom. 13 (1999) 295–309. doi:10.1002/(SICI)1099-128X(199905/08)13:3/4<295::AID-CEM547>3.0.CO;2-Y.

[10] J.C. Hoggard, R.E. Synovec, Parallel Factor Analysis (PARAFAC) of Target Analytes in GC × GC−TOFMS Data: Automated Selection of a Model with an Appropriate Number of Factors, Anal. Chem. 79 (2007) 1611–1619. doi:10.1021/ac061710b.

[11] M. Jalali-Heravi, H. Parastar, M. Kamalzadeh, R. Tauler, J. Jaumot, MCRC software: A tool for chemometric analysis of two-way chromatographic data, Chemom. Intell. Lab. Syst. 104 (2010) 155–171. doi:10.1016/j.chemolab.2010.08.002.

[12] A. de Juan, J. Jaumot, R. Tauler, Multivariate Curve Resolution (MCR). Solving the mixture analysis problem, Anal. Methods. 6 (2014) 4964–4976. doi:10.1039/C4AY00571F.

[13] S. Li, J.C. Hamilton, P.J. Gemperline, Generalized rank annihilation method using similarity transformations, Anal. Chem. 64 (1992) 599–607. doi:10.1021/ac00030a007.

[14] E. Sanchez, B.R. Kowalski, Generalized rank annihilation factor analysis, Anal. Chem. 58 (1986) 496–499. doi:10.1021/ac00293a054.

[15] T. Skov, J.C. Hoggard, R. Bro, R.E. Synovec, Handling within run retention time shifts in two-dimensional chromatography data using shift correction and modeling, J. Chromatogr. A. 1216 (2009) 4020–4029. doi:10.1016/j.chroma.2009.02.049.

[16] R. Tauler, Multivariate curve resolution applied to second order data, Chemom. Intell. Lab. Syst. 30 (1995) 133–146. doi:10.1016/0169-7439(95)00047-X.

[17] W.G. Pool, J.W. de Leeuw, B. van de Graaf, A rapid routine to correct for skewing in gas chromatography/mass spectrometry, J. Mass Spectrom. 31 (1996) 213–215. doi:10.1002/(SICI)1096-9888(199602)31:2<213::AID-JMS284>3.0.CO;2-6.
[18] D.W. Cook, S.C. Rutan, D.R. Stoll, P.W. Carr, Two dimensional assisted liquid chromatography – a chemometric approach to improve accuracy and precision of quantitation in liquid chromatography using 2D separation, dual detectors, and multivariate curve resolution, Anal. Chim. Acta. 859 (2015) 87–95. doi:10.1016/j.aca.2014.12.009.
[19] A. de Juan, R. Tauler, Comparison of three-way resolution methods for non-trilinear chemical data sets, J. Chemom. 15 (2001) 749–771. doi:10.1002/cem.662.
[20] H. Parastar, R. Tauler, Multivariate Curve Resolution of Hyphenated and Multidimensional Chromatographic Measurements: A New Insight to Address Current Chromatographic Challenges, Anal. Chem. 86 (2014) 286–297. doi:10.1021/ac402377d.
[21] H.P. Bailey, S.C. Rutan, P.W. Carr, Factors that affect quantification of diode array data in comprehensive two-dimensional liquid chromatography using chemometric data analysis, J. Chromatogr. A. 1218 (2011) 8411–8422. doi:10.1016/j.chroma.2011.09.057.
[22] K.J. Johnson, B.W. Wright, K.H. Jarman, R.E. Synovec, High-speed peak matching algorithm for retention time alignment of gas chromatographic data for chemometric analysis, J. Chromatogr. A. 996 (2003) 141–155. doi:10.1016/S0021-9673(03)00616-2.
[23] J.S. Nadeau, R.B. Wilson, J.C. Hoggard, B.W. Wright, R.E. Synovec, Study of the interdependency of the data sampling ratio with retention time alignment and principal component analysis for gas chromatography, J. Chromatogr. A. 1218 (2011) 9091–9101. doi:10.1016/j.chroma.2011.10.031.
[24] W. Khummueng, J. Harynuk, P.J. Marriott, Modulation Ratio in Comprehensive Two-dimensional Gas Chromatography, Anal. Chem. 78 (2006) 4578–4587. doi:10.1021/ac052270b.
[25] J.M. Davis, D.R. Stoll, P.W. Carr, Effect of First-Dimension Undersampling on Effective Peak Capacity in Comprehensive Two-Dimensional Separations, Anal. Chem. 80 (2008) 461–473. doi:10.1021/ac071504j.
[26] R.E. Murphy, M.R. Schure, J.P. Foley, Effect of Sampling Rate on Resolution in Comprehensive Two-Dimensional Liquid Chromatography, Anal. Chem. 70 (1998) 1585–1594. doi:10.1021/ac971184b.
[27] J.V. Seeley, Theoretical study of incomplete sampling of the first dimension in comprehensive two-dimensional chromatography, J. Chromatogr. A. 962 (2002) 21–27. doi:10.1016/S0021-9673(02)00461-2.
[28] D.K. Pinkerton, B.A. Parsons, T.J. Anderson, R.E. Synovec, Trilinearity deviation ratio: A new metric for chemometric analysis of comprehensive two-dimensional gas chromatography time-of-flight mass spectrometry data, Anal. Chim. Acta. 871 (2015) 66–76. doi:10.1016/j.aca.2015.02.040.
[29] D.R. Stoll, J.D. Cohen, P.W. Carr, Fast, comprehensive online two-dimensional high performance liquid chromatography through the use of high temperature ultra-fast gradient elution reversed-phase liquid chromatography, J. Chromatogr. A. 1122 (2006) 123–137. doi:10.1016/j.chroma.2006.04.058.
[30] D.F. Thekkudan, S.C. Rutan, P.W. Carr, A study of the precision and accuracy of peak quantification in comprehensive two-dimensional liquid chromatography in time, J. Chromatogr. A. 1217 (2010) 4313–4327. doi:10.1016/j.chroma.2010.04.039.

[31] J. Kuligowski, G. Quintás, R. Tauler, B. Lendl, M. de la Guardia, Background Correction and Multivariate Curve Resolution of Online Liquid Chromatography with Infrared Spectrometric Detection, Anal. Chem. 83 (2011) 4855–4862. doi:10.1021/ac2004407.
[32] H. Parastar, J.R. Radović, M. Jalali-Heravi, S. Diez, J.M. Bayona, R. Tauler, Resolution and Quantification of Complex Mixtures of Polycyclic Aromatic Hydrocarbons in Heavy Fuel Oil Sample by Means of GC × GC-TOFMS Combined to Multivariate Curve Resolution, Anal. Chem. 83 (2011) 9289–9297. doi:10.1021/ac201799r.
[33] H.P. Bailey, S.C. Rutan, Chemometric resolution and quantification of four-way data arising from comprehensive 2D-LC-DAD analysis of human urine, Chemom. Intell. Lab. Syst. 106 (2011) 131–141. doi:10.1016/j.chemolab.2010.07.008.
[34] C.A. Bruckner, B.J. Prazen, R.E. Synovec, Comprehensive Two-Dimensional High-Speed Gas Chromatography with Chemometric Analysis, Anal. Chem. 70 (1998) 2796–2804. doi:10.1021/ac980164m.
[35] N.M. Faber, L.M.C. Buydens, G. Kateman, Generalized rank annihilation method. I: Derivation of eigenvalue problems, J. Chemom. 8 (1994) 147–154. doi:10.1002/cem.1180080206.
[36] C. Moler, G. Stewart, An Algorithm for Generalized Matrix Eigenvalue Problems, SIAM J. Numer. Anal. 10 (1973) 241–256. doi:10.1137/0710024.
[37] R.B. Poe, S.C. Rutan, Effects of resolution, peak ratio and sampling frequency in diode-array fluorescence detection in liquid chromatography, Anal. Chim. Acta. 283 (1993) 845–853. doi:10.1016/0003-2670(93)85298-X.
[38] B.J. Prazen, C.A. Bruckner, R.E. Synovec, B.R. Kowalski, Second-order chemometric standardization for high-speed hyphenated gas chromatography: Analysis of GC/MS and comprehensive GC×GC data, J. Microcolumn Sep. 11 (1999) 97–107. doi:10.1002/(SICI)1520-667X(1999)11:2<97::AID-MCS2>3.0.CO;2-Z.
[39] B.J. Prazen, R.E. Synovec, B.R. Kowalski, Standardization of Second-Order Chromatographic/Spectroscopic Data for Optimum Chemical Analysis, Anal. Chem. 70 (1998) 218–225. doi:10.1021/ac9706335.
[40] E. Sanchez, L. Scott Ramos, B.R. Kowalski, Generalized rank annihilation method, J. Chromatogr. A. 385 (1987) 151–164. doi:10.1016/S0021-9673(01)94629-1.
[41] J.D. Carroll, J.-J. Chang, Analysis of individual differences in multidimensional scaling via an n-way generalization of "Eckart-Young" decomposition, Psychometrika. 35 (1970) 283–319. doi:10.1007/BF02310791.
[42] E. Comas, R.A. Gimeno, J. Ferré, R.M. Marcé, F. Borrull, F.X. Rius, Quantification from highly drifted and overlapped chromatographic peaks using second-order calibration methods, J. Chromatogr. A. 1035 (2004) 195–202. doi:10.1016/j.chroma.2004.02.069.
[43] R.A. Harshman, University microfilms. Michigan: Ann Arbor, (1970).
[44] J.C. Hoggard, W.C. Siegler, R.E. Synovec, Toward automated peak resolution in complete GC × GC–TOFMS chromatograms by PARAFAC, J. Chemom. 23 (2009) 421–431. doi:10.1002/cem.1239.
[45] J.C. Hoggard, R.E. Synovec, Automated Resolution of Nontarget Analyte Signals in GC × GC-TOFMS Data Using Parallel Factor Analysis, Anal. Chem. 80 (2008) 6677–6688. doi:10.1021/ac800624e.
[46] J.C. Hoggard, J.H. Wahl, R.E. Synovec, G.M. Mong, C.G. Fraga, Impurity Profiling of a Chemical Weapon Precursor for Possible Forensic Signatures by Comprehensive Two-

Dimensional Gas Chromatography/Mass Spectrometry and Chemometrics, Anal. Chem. 82 (2010) 689–698. doi:10.1021/ac902247x.
[47] M.L. Oca, M.C. Ortiz, A. Herrero, L.A. Sarabia, Optimization of a GC/MS procedure that uses parallel factor analysis for the determination of bisphenols and their diglycidyl ethers after migration from polycarbonate tableware, Talanta. 106 (2013) 266–280. doi:10.1016/j.talanta.2012.10.086.
[48] D.N. Rutledge, D. Jouan-Rimbaud Bouveresse, Multi-way analysis of outer product arrays using PARAFAC, Chemom. Intell. Lab. Syst. 85 (2007) 170–178. doi:10.1016/j.chemolab.2006.06.011.
[49] M. Vosough, M. Bayat, A. Salemi, Matrix-free analysis of aflatoxins in pistachio nuts using parallel factor modeling of liquid chromatography diode-array detection data, Anal. Chim. Acta. 663 (2010) 11–18. doi:10.1016/j.aca.2010.01.039.
[50] S. Yang, J.S. Nadeau, E.M. Humston-Fulmer, J.C. Hoggard, M.E. Lidstrom, R.E. Synovec, Gas chromatography–mass spectrometry with chemometric analysis for determining 12C and 13C labeled contributions in metabolomics and 13C flux analysis, J. Chromatogr. A. 1240 (2012) 156–164. doi:10.1016/j.chroma.2012.03.072.
[51] S. Yang, M. Sadilek, R.E. Synovec, M.E. Lidstrom, Liquid chromatography–tandem quadrupole mass spectrometry and comprehensive two-dimensional gas chromatography–time-of-flight mass spectrometry measurement of targeted metabolites of Methylobacterium extorquens AM1 grown on two different carbon sources, J. Chromatogr. A. 1216 (2009) 3280–3289. doi:10.1016/j.chroma.2009.02.030.

1.8 FIGURES

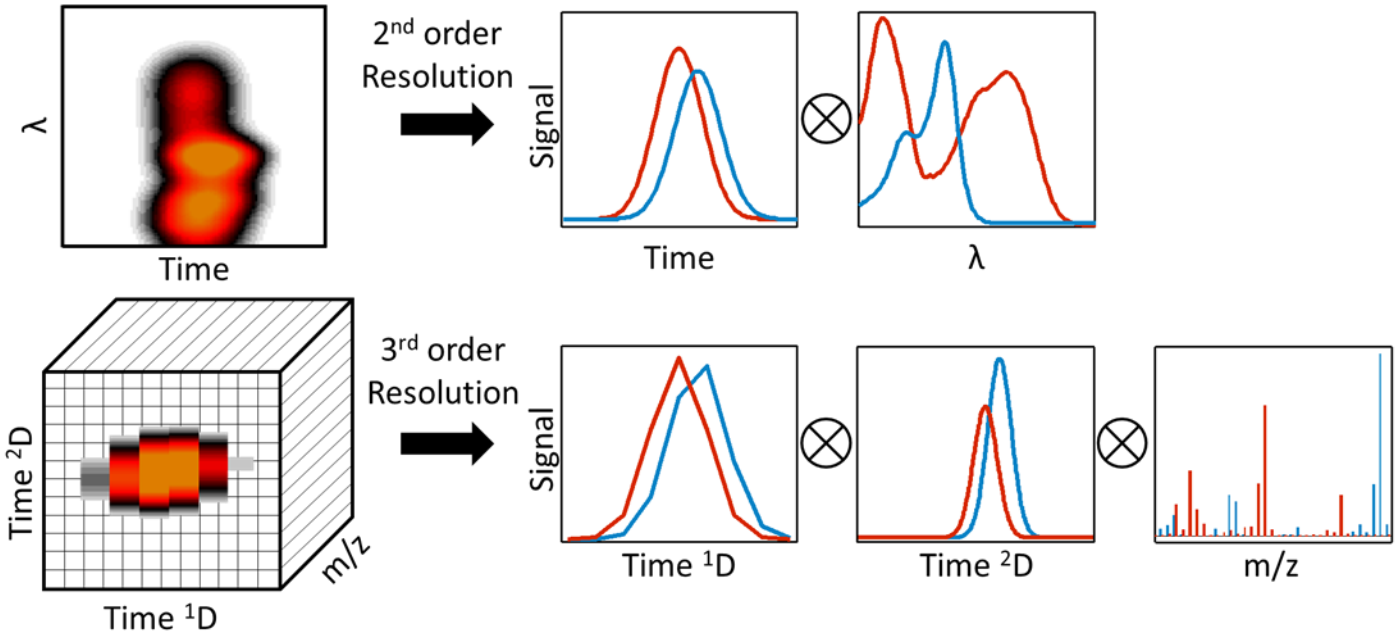

Figure 1.1. Observed sample profiles of chromatographically overlapped two component mixture (left) with LC–DAD (top) and GC × GC–TOFMS (bottom). Mathematically resolved component profiles (right) after application of 2nd order (top) and 3rd order (bottom) chemometric resolution. LC–DAD data decomposed into chromatographic elution profiles, signal versus time, and UV-vis absorbance spectra, absorbance versus wavelength (λ). GC × GC–TOFMS data decomposed into column one elution profiles, signal versus time ^{1}D, column two elution profiles, signal versus time ^{2}D, and time-of-flight mass spectra, signal versus *m/z*.

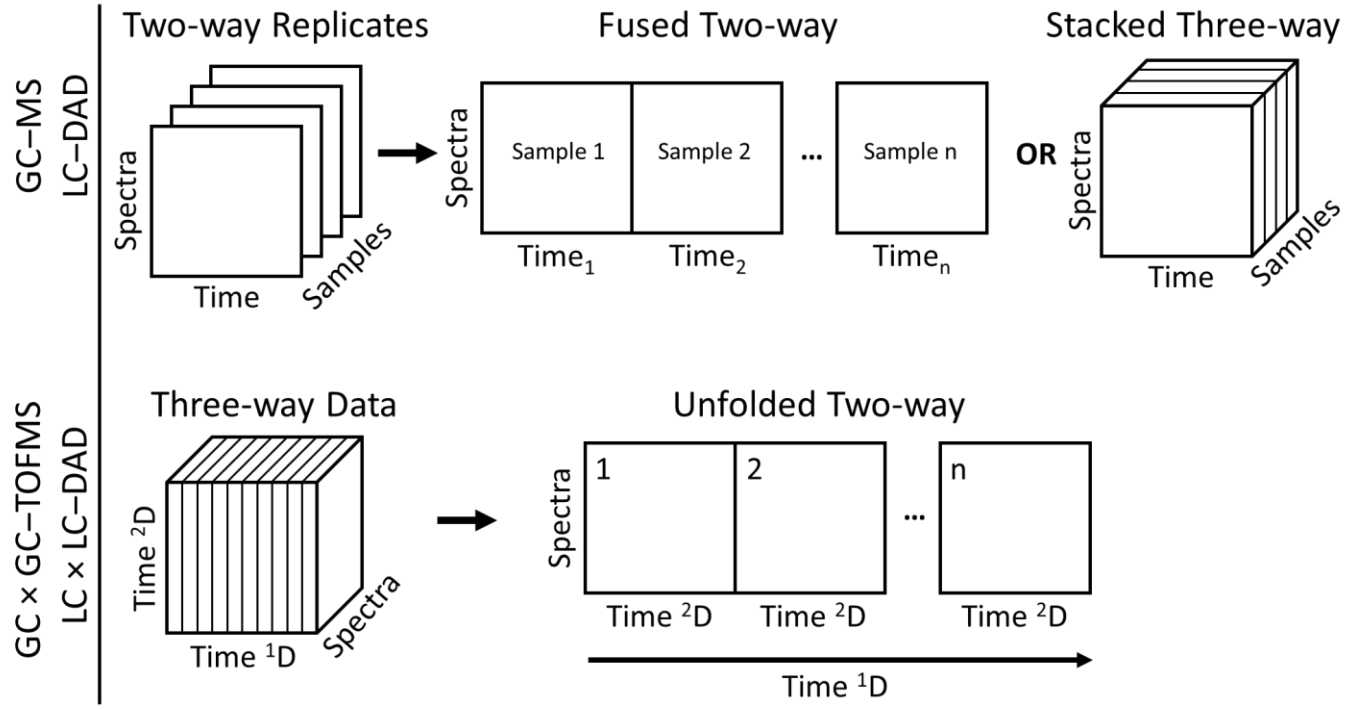

Figure 1.2 Graphical representation of data structures (left) for replicates of 1D chromatography with spectral detection (top) and single run of 2D chromatography with spectral detection (bottom), with data rearrangement schemes (right). Replicates of two-way data (top) are shown fused together end-to-end in the time dimension to create a single two-way array (spectral dimension × concatenated retention time dimension) , where $Time_n$ is the time dimension of the n^{th} sample, and stacked along the sample dimension to create a three-way array (spectral dimension × retention time dimension × sample dimension). Three-way data (bottom) can be unfolded along the time ^{1}D dimension (modulation dimension) such that each of the n modulations are concatenated, each modulation is as a two-way slab with dimensions spectra × time ^{2}D from the original three-way array.

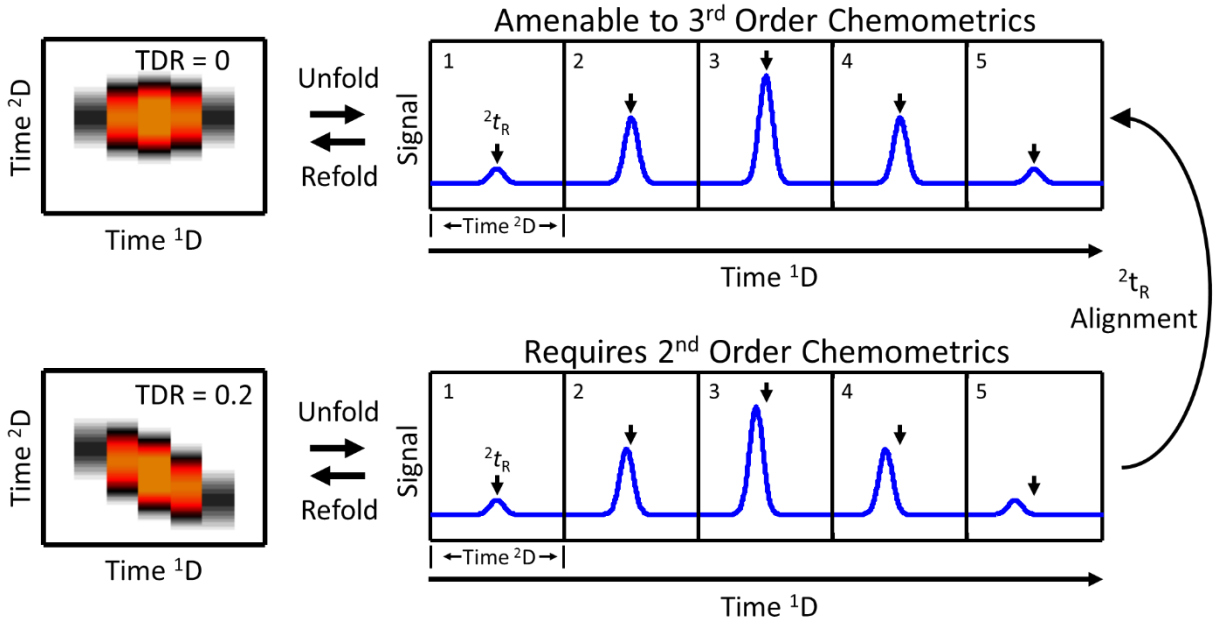

Figure 1.3 2D chromatogram of one analyte for a single m/z or wavelength taken from GC × GC–TOFMS or LC × LC–DAD sample. Trilinear data (TDR = 0) (top) and non-trilinear data (TDR = 0.2) (bottom) are shown folded (left) and unfolded (right). Numbered windows in unfolded data represent individual modulations. Arrows in windows of unfolded data indicate the ^2D retention time (^2t_R) of the first modulated ^2D peaklet.

Chapter 2. Determining the Probability of Achieving Successful Quantitative Analysis for Gas Chromatography-Mass Spectrometry

2.1 INTRODUCTION

Gas chromatography (GC) is a powerful separation technique used to separate volatile and semi-volatile compounds in complex mixtures. Often coupled with mass spectrometry (MS) as a hyphenated technique, GC-MS is a popular analytical platform for a variety of applications, such as metabolomics,[1,2] forensics,[3,4] food products,[5,6] and more.[7–9] Complex samples generate complex chromatograms, often with excessive peak overlap due to co-eluting analytes. Such chromatograms often require advance chemometrics techniques for successful deconvolution and extraction of useful information about the sample components.

Deconvolution software is often used to mathematically resolve two or more analytes that are chromatographically overlapped. Compared to simple peak height or peak area integration techniques that require a pure signal for a given analyte response, deconvolution can provide a more accurate and precise analyte peak area and a purified analyte mass spectrum, for better analyte quantification and identification. Many chemometric software algorithms have been developed to perform deconvolution on chromatographic peaks. Each algorithm functions slightly differently in its requirements of the data structure and its efficacy for certain applications. However, all deconvolution algorithms are accuracy- and precision-limited in their performance by three main factors: (1) chromatographic resolution between the target analyte and nearby/overlapping interferent(s); (2) signal-to-noise ratio (*S/N*) of the analytes of interest; and (3) mass spectral similarity between the target analyte and nearby/overlapping interferent(s).

Many metrics exist to evaluate and compare the efficacy and power of various GC separations. One popular metric is peak capacity, which is representative of the amount of information that can be contained in a given separation, defined as the maximum number of peaks that can be evenly resolved, most commonly determined at unit chromatographic resolution (R_s = 1.0), in a certain separation time.[10] However, if one can decrease the resolution at which peaks can be resolved through chemometric deconvolution, the requirement for physical separation can be loosened and the peak capacity can be "chemometrically" enhanced and increased, allowing more peaks per separation time window.

The likelihood of needing to apply chemometrics hinges upon the relative "crowdedness" of a chromatogram, that is, how many analytes are present in a separation window, referred to as the saturation factor, α°.[11] The saturation factor is mathematically defined as the number of components present in a separation divided by the peak capacity.[11] Therefore, if the peak capacity can be chemometrically enhanced, this effectively reduces the saturation factor of the separation. Consequently, the metrics chromatographers use to compare and evaluate separations, especially resolution, peak capacity and saturation factor, are interrelated by definition and can be chemometrically enhanced through the application of an effective deconvolution algorithm.

The relation between peak overlap and saturation factor has previously been reported by Davis and Giddings as part of their statistical overlap theory.[11,12] Their theory expressly states that with complex samples, the overlap of peaks can be statistically estimated for a given level of chromatographic saturation α°, specifically that the number of observed peaks can be statistically approximated. However, their theory does not include a discussion of chemometrically enhanced resolution and was suggested to fail at α° > 1. We develop and apply

a similar theory here, which is nominally consistent with that presented by Davis and Giddings, but aims to describe the effects of chemometrically enhanced resolution on peak capacity and the probability of successful chemometric deconvolution for analyte identification and quantification. Operating under the assumption that analyte peaks are randomly distributed across the separation space, the theory is based upon a probabilistic description of peak overlap in GC-MS separations to determine the probability of obtaining a successful quantitative analysis, referred to herein as obtaining "chemometric success," which has its lower limit of chromatographic resolution R_s at some minimum chemometric resolution, R_s*.

For practical implementation of the theory, chemometric success for a target analyte requires that any interfering peaks have $R_s \geq R_s$* to the target peak, in order to achieve a user specified level of quantitative accuracy and precision, and/or level of analyte identification confidence via a mass spectral library match. Hence, the theory serves as a spring board into a simulation-based investigation to determine the minimum chemometric resolution, R_s*, at which a user-selected chemometric algorithm can successfully deconvolute co-eluting GC-MS peaks, to assess how practical application of the deconvolution algorithm fits with the expected theory. Simulations include the generation of chromatograms including one target analyte and one interferent analyte (referred to as a "target-interferent pair") at several R_s and two *S/N* levels. A broad range of target-interferent pairs are evaluated, varied to span the range to include pairs with similar mass spectral character to pairs with very different mass spectral character. Any deconvolution algorithm could be applied to the simulated data to determine R_s*. For the purpose of this study, due to its ubiquitous application, multivariate curve resolution with alternating least squares (MCR-ALS) was chosen as the deconvolution algorithm.[13–17] However, other deconvolution algorithms could also be applied to determine their respective R_s*.

2.2 THEORY

The theory presented here operates under the assumption that chromatographic peaks are randomly distributed across the separation space, which is an approach with historical roots.[11,12] While this is not necessarily the case for all separations, depending on sample composition and column stationary phase, the theory presented can also be applied as local peak capacities for a particular temporal portion of the total separation time, t_{window}, by replacing the total separation time, t_{sep}, such that $t_{window} \leq t_{sep}$, assuming only local randomness and allowing for variations in peak distribution across the chromatogram. The assumption that peaks are randomly distributed across the entire separation is used for the purposes of demonstration, but all equations presented are readily applied with the assumption of local randomness.

In chromatography, the amount of information contained in a given separation is inferred from and is directly related to the peak capacity of the separation, defined as

$$n_c = \frac{t_{sep}}{W_b R_s} = \frac{t_{sep}}{W_b} \tag{2.1}$$

where t_{sep} is the total separation time and W_b is taken as the average peak width at base. The peak capacity, n_c, represents the number of evenly resolved peaks at unit resolution, $R_s = 1$, that can fit into a given separation. The resolution between two chromatographic peaks is defined by

$$R_s = \frac{t_{R,2} - t_{R,1}}{\frac{1}{2}(W_{b,1} + W_{b,2})} = \frac{\Delta t_R}{W_b} \tag{2.2}$$

where Δt_R is the difference in retention time for peaks one and two, $t_{R,1}$ and $t_{R,2}$, and W_b is the average of the individual peak widths at base for peaks one and two, $W_{b,1}$ and $W_{b,2}$. For temperature programmed separations peak widths tend to vary only slightly, such that the average width at base terms, W_b, from Equations (2.1) and (2.2) are nominally equivalent.

The application of chemometrics to chromatographic data allows for mathematical resolution of peaks that may not have been resolved physically by chromatography. The ability to mathematically resolve peaks that are physically overlapped, taken to be $R_s < 1$, significantly relaxes the requirement for chromatographic resolution across the separation and increases the number of evenly resolvable peak, resulting in a new chemometric enhanced peak capacity,

$$n_c^* = \frac{t_{sep}}{W_b R_s^*} = \frac{n_c}{R_s^*} \tag{2.3}$$

where R_s* is defined to be the minimum chemometric resolution, which is the minimum resolution between two peaks for a particular chemometric method to deconvolute them, that is, to resolve them mathematically. It should be noted that the minimum chemometric resolution, R_s*, is distinct from chromatographic resolution defined by Equation (2.2); in particular, R_s* depends on the chemometric method and must be determined experimentally, while R_s is a measured chromatographic value between two peaks. In general, R_s* suggests that the chemometric method can be successfully applied for all values $R_s \geq R_s$*. Figures 2.1(A) and (B) show simulated selective ion chromatograms of a single target analyte (solid line with $t_R = 4.5$) and the two nearest interferent peak profiles eluting with $R_s = R_s$* for R_s* = 1.0 and 0.3 respectively. The R_s* = 1.0 in Figure 2.1(A) is representative of an analysis without chemometrics while Figure 2.1(B), shows a separation analyzed by a chemometric technique with R_s* = 0.3, illustrating the substantial gain in peak capacity when chemometrics are considered.

The assumption of randomly distributed peaks generally leads to the Poisson distribution, a generalized approximation of the binomial distribution. However, the assumption

of randomly distributed peaks means that the probability of chemometric success can be determined rigorously with the binomial distribution with each chromatographic peak representing an independent Bernoulli random variable, since chemometric success requires that there are no interferent peaks at $R_s < R_s^*$ to the target analyte. For a single target analyte, the probability of chemometric failure, defined as $R_s < R_s^*$, for the target-interferent pair ($m = 1$) is defined by

$$p_{fail} = \frac{2}{n_c^*} \tag{2.4}$$

where n_c^* is as in Equation (2.3). This leads to the determination that the probability of chemometric success for a single target target-interferent pair, $R_s \geq R_s^*$, is

$$p_{success} = 1 - p_{fail} = 1 - \frac{2}{n_c^*} \tag{2.5}$$

Equation (2.5) can be generalized to samples of any complexity to define the probability of chemometric success for a target analyte with $m \geq 0$ independent interferent peaks

$$P(success) = (p_{success})^m = \left(1 - \frac{2}{n_c^*}\right)^m \qquad m \geq 0 \tag{2.6}$$

In this context, the saturation factor of a chromatographic separation is given by

$$\alpha^\circ = \frac{\#\ components}{n_c} = \frac{m+1}{n_c} \tag{2.7}$$

where m is the number of interferent peaks as in Equation (2.6) so the total number of components is m interferents + 1 target analyte. Figure 2.2(A) and (B) are examples of simulated total ion current (TIC) chromatograms with $n_c = 100$ with randomly distributed analytes, corresponding to $\alpha^o = 0.5$ and 1.0 respectively. An alternative expression for P(success), as a function of α^o, can now be provided by combining Equations (2.3) and (2.7) with Equation (2.6),

$$P(success) = \left(1 - \frac{2R_s^*}{n_c}\right)^{(\alpha^\circ n_c - 1)} \tag{2.8}$$

Clearly, *P*(success) depends on the separation peak capacity and the sample complexity but it also requires determination of the limit of chemometric resolution, R_s*, for the chosen chemometric method. The most generalized benefit of Equation (2.8) is that it allows for a quantitative comparison of chemometric method performance and chromatographic separation parameter selection for samples of varying complexity. Even though the theory appears to be based on a single target analyte, in a broader sense, the theory is generally applicable to all components in the separation (m+1), and can be iteratively applied such that each component is thought to be the target analyte in relation to the other m components.

A plot of *P*(success), via Equation (2.8), as a function of α^o for various R_s* ranging from 0.05 to 1.0 is provided in Figure 2.3(A). Figure 2.3(B) shows a zoomed in view of the plot in Figure 2.3(A) to illustrate the significant gains realized when chemometrics with lower R_s* values are applicable. Since R_s* depends on the particular chemometric method to be used and the conditions under which the method will be applied, R_s* must be determined experimentally (demonstrated herein). The plot of *P*(success) at R_s* = 1.0 represents an analysis without chemometric deconvolution, relying solely on the physical chromatographic resolution of peaks followed by traditional quantitative measurements such as peak height. If a separation is initially achieved with α^o = 1.0 (Figure 2.2(B)), *P*(success) is 0.14, thus only about 14% of the analytes could be successfully analyzed if limited by data analysis tools to follow the curved R_s* = 1.0 path. This challenge for successfully applying chromatography to complex samples was first noted by David and Giddings,[11,12] and has been an impetus for the development of more efficient separation methods (higher N), and multi-channel detectors such as MS for GC. If the chromatographic efficiency N is increased 10-fold, the α^o is reduced by $10^{1/2}$ to 0.32, resulting in a *P*(success) of 0.55, nearly a 4-fold increase in analyzable components following strictly the R_s*

= 1.0 path via improvements to chromatographic efficiency. However, in order to more closely approach a *P*(success) of 1.0, the goal being to be able to analyze 100% of the components, improvements to chromatographic efficiency should be coupled with chemometric methods so as to simultaneously decrease α^o while following an analysis path at a reduced R_s*. For example, if a chemometric method provides an R_s* = 0.3, with α^o = 1.0, the *P*(success) jumps from 0.14 to 0.55, and then if in concert the chromatographic efficiency was increased 10-fold, now α^o = 0.32, the *P*(success) is 0.83, well on the way to approaching 1.0.

2.3 EXPERIMENTAL

All data were simulated and analyzed using Matlab R2015b (Mathworks, Inc., Natwick, MA, USA), with the modeling conditions summarized in Table 1. Three-second long GC-TOFMS chromatograms were simulated consisting of one analyte and one interferent at a mass spectral collection rate of 100 Hz. Peaks were modeled as Gaussians with the same peak area and a width at base (± 2σ) of 1 s. Random Gaussian-distributed noise was generated independently for each mass channel, *m/z*, over the *m/z* range of 20-360 The parameters for the noise were determined such that the mean and standard deviation of the noise would provide the desired signal-to-noise ratio (*S/N*) in the total ion current (TIC) of the peak. Two *S/N* values were studied, the high *S/N* was defined as *S/N* of 100 (hereafter, "*S/N* 100") and the low *S/N* was defined as *S/N* of 10 (hereafter, "*S/N* 10"), where the magnitude of the noise, *N*, was defined as 3σ, or three times the standard deviation of the noise. With the *S/N* defined relative to the TIC for a given analyte, the *S/N* at a sensitive *m/z* would be proportionately higher than the *S/N* of the TIC, depending upon the total number of unique mass fragments and the relative distribution of

mass fragments in the analyte spectrum. These concepts are illustrated for 2,3,4 trimethylpentane and 3,4 diethylhexane in Figure S2.1 in Supplemental.

In order to create a large enough sample of simulations for study, a total of 45 analytes were chosen to serve as the targets and interferents with their mass spectra taken from the NIST database. Here, a "target" is taken to be an analyte of interest and "interferent" an analyte that is chromatographically overlapped with the desired target. A list of the 45 analytes can be found in Table S1 in the Supplemental. Since two of the 45 analytes were used for each simulation (one target, one interferent), a total of 990 simulations were created at each *S/N* ($_{45}C_2$, or 45 choose 2). These 45 analytes were chosen such that there was a relatively diverse distribution of target-interferent pairs, ranging from pairs with very similar mass spectra to pairs with very dissimilar mass spectra. A mass spectral match value (MV) was calculated between each target and interferent to determine whether there was a "High" MV (that is, the mass spectra were very similar) "Mid" MV, or "Low" MV (that is, the mass spectra are very dissimilar). MV was calculated based on the equation outlined by Stein.[18] "High" MV target-interferent pairs were those with $600 \leq MV < 1000$; "Mid" MV pairs were those with $300 \leq MV < 600$; and "Low" MV pairs were those with $MV < 300$. The calculated match values for all 990 target-interferent pairs are provided in a histogram in Figure S2.2 in Supplemental, with the distribution indicative of the broad range of MV target-interferent pairs studied.

In addition to varying the target-interferent pair and the *S/N*, the resolution, R_s, between the target and interferent peaks was varied in order to determine the minimum resolution for the deconvolution algorithm, R_s*, to successfully deconvolute the target from the interferent. Sixteen resolution values (R_s = 0.02, 0.04, 0.06, 0.08, 0.10, 0.12, 0.14, 0.16, 0.18, 0.20, 0.25, 0.30, 0.35, 0.40, 0.45, and 0.5) were chosen. A R_s = 0.2 is illustrated in Figure S2.3 Supplemental, pointing

out that by definition the peak height ratio of target-interferent in the TIC was held constant at 1:1. For each target-interferent pair (990 options), a chromatogram of each R_s (16 options) at each *S/N* (2 options) was generated, resulting in a total of 31,680 total chromatograms generated. All simulation parameters are provided in Table 1.

Next, an MCR-ALS model was created for each simulation. One component, two component and three component models were generated, and it was determined that the two-component models effectively modeled both target and interferent while pulling out the noise in the residuals. All chemometric models shown and discussed hereafter will refer to the two-component models. Models were generated without any constraints. All chromatograms and models were generated over the course of 36 hours on a personal computer with an Intel® Core™ i7-4770 processor (3.4 GHz), 24 GB of random access memory (RAM), a 250 GB Samsung 840 solid state hard drive, and with Windows 7 SP1 as the operating system.

Two parameters were calculated to evaluate each deconvolution model. First, the peak area of the chromatographic component from the MCR-ALS model was compared to the simulated peak area via a percent error calculation defined as

$$\%error = \frac{A_{Model} - A_{Sim}}{A_{Sim}} \times 100 \tag{2.9}$$

where the A_{Model} is the area of the peak resulting from the chemometric deconvolution model and A_{Sim} is the true area of the peak generated during the simulation. Percent error was calculated for the target in each of the 31,680 models. For clarity, the absolute value of all errors positive is determined, for use in subsequent figures and discussion. Secondly, the match value (MV, with possible values ranging from 0 to 1000) was calculated between the mass spectrum of the target

generated from the loadings of the model and the original library mass spectrum for the target from the NIST library based on the equation outlined by Stein. [18]

2.4 Results and Discussion

We now consider the simulation-based study to determine an approximate R_s* for the chemometric deconvolution algorithm. Figure 2.4 summarizes (A) the average MV and (B) the average percent error (%error, Equation (2.9)) results for the 31,680 chemometric models. The "Low" MV results are shown in green, the "Mid" MV in blue and the "High" MV in red, with *S/N* 100 results as solid lines and *S/N* 10 results in dashed lines. Deconvolution examples at R_s = 0.2 and 0.3, each at *S/N* 10 and *S/N* 100, with the loadings provided in the Figure S2.4 in Supplemental for a representative "Low" MV target-interferent pair: 1-chloro-5-methyl hexane (target), and decane (interferent). Figure 2.4(A) shows the average target loading-to-library MV results, where an increase in R_s results in an increase in the MV, as expected, due to a more facile and accurate extraction of the pure analyte mass spectra by the model. The *S/N* 100 results obviously have higher target loading-to-library MV at all R_s simulated, starting at an average MV of approximately 950 at the lowest R_s and capping out at a maximum average MV at 990 at the larger R_s values, while the *S/N* 10 simulations start at much lower average MV (as low as 648 for the "Low" MV at R_s = 0.02) and converge on a maximum average MV of about 915. This pattern is likely due to noisy *m/z* contributing spurious signals that are extracted with the true signal in the models. The extent of this inclusion of noisy *m/z* increases as R_s decreases and the summed signal of these noisy *m/z* increases due to excessive peak overlap between the target and interferent.

Contrary to initial expectation, the "Low" MV target-interferent pairs have chemometric models that generally result in a lower target loading-to-library MV to the library spectrum than the "High" MV pairs. One might assume that target-interferent pairs that are very alike ("High" MV pairs) would be more difficult to correctly deconvolute and result in lower MV between the extracted and library spectrum. However, because the target and interferent share so many *m/z* for those "High" MV pairs, any *m/z* signal attributed to the target that actually came from the interferent in the model will not appreciably affect the target loading-to-library MV because so many of the *m/z* are shared. On the other hand, those "Low" MV target-interferent pairs have many *m/z* that are not shared whatsoever, so any signal from an interferent mistakenly attributed to the target will automatically decrease the target loading-to-library MV.

It is always important when looking at the results of a chemometric model that the mass spectral loadings are not the only model loadings regarded to determine the efficacy of the model, even though it might be the easiest to "match" to a library spectrum to assess the model accuracy used to deconvolute an unknown. Again, relying solely on the mass spectral loadings may prevent the analyst from discovering signal falsely attributed to the target of interest when the interferent has very similar mass spectral features. Fortunately, the chromatographic peak profile loadings which provide the quantitative information regarding the chemometric models are provided in conjunction with the mass spectral loadings. Looking at either the mass spectral loadings or the peak profile loadings alone, examples of which are provided in Figure S2.4 in Supplemental, will not provide a full and accurate picture of the quality of model generated; but rather, both should be viewed and utilized simultaneously.

From the peak profile loadings determined in this study, the average peak area %error (Equation (2.9)) results are readily determined, and are provided in Figure 2.4(B). The *S/N* 10

results show a sharp increase in the %error between R_s of 0.3 and 0.2, especially for the "High" MV case, where mass spectral signals are easily misattributed to the target peak resulting in a large %error. While the *S/N* 100 cases have much lower average %error, there is an obvious increase in the %error for the "Low" and "Mid" MV subsets below a R_s of 0.3. The cause of the slight peak in %error for the "Low" MV, *S/N* 100 subset between $R_s = 0.3$ and $R_s = 0.1$ is unknown, although this feature did not impact the findings in this study

Based upon Figure 2.4(A) and (B), a conservative value for the "experimentally" determined (via simulation) chemometric resolution is a R_s* of 0.3. This estimate of R_s* is defined herein where there is a "significant" increase from an average %error of 20%, which was arbitrarily selected as a maximum allowable %error, and/or observation of a "significant" decrease in target loading-to-library MV in relation to the maximum MV at $R_s = 0.5$. This conservative estimate of R_s* of 0.3 is principally limited by the *S/N* 10 results for the "High" MV subset (significant %error increase), and the *S/N* 10 results for the "Low" and "Mid" MV subsets (significant target loading-to-library MV decrease). Depending upon the analytical requirements a higher %error and/or lower MV may be tolerated. Indeed, since real samples are comprised of a wide range of target-interferent *S/N* and MV conditions, the results in Figure 2.4(A) and (B) suggest that a more appropriate R_s* of 0.2 is may be reasonable since the average target loading-to-library MV curves are all above 800 (a useful threshold), and the %errors are all comfortably below 20%, except for the *S/N* 10 results for the "High" MV subset, which for real samples these chromatographic overlap situations are comprised of chemically and mass spectrally similar compounds, eg. isomers, and a large %error may be acceptable. While the target-to-interferent peak height ratio was 1:1 in this study, for real samples any ratio that favors the target will

improve (decrease) R_s* while a ratio that favors the interferent will increase R_s*. The aspect of peak height ratio has not been considered in this study for brevity, and further study is warranted.

We now return to the theoretically derived *P*(success) as a function of α^o for various R_s*, Equation (2.8) in Figure 2.3(A), to put into context the analytical benefit of the "experimentally" determined R_s* via simulation with MCR-ALS. The conservative "experimentally" determined R_s* of 0.3 provides the same analytical benefit as increasing the chromatographic efficiency, *N*, by a factor of ~ 10 (while staying on the R_s* = 1.0 path). The benefit to increasing the *P*(success) depends highly on the chromatogram saturation factor α^o. If α^o = 1.0, the *P*(success) jumps from 0.14 to 0.55 following either a chemometric enhancement path by jumping vertically from the R_s* = 1.0 to 0.3 curve, or by the strictly increasing *N* 10-fold on the curved R_s* = 1.0 path. Both outcomes are the same, and a *P*(success) of 0.55 is not generally acceptable. For this reason, often the analyst seeks to provide separations with a reduced α^o, such as 0.5. At a R_s* of 0.3, with an α^o of 0.5, the *P*(success) jumps from 0.37 (with R_s* = 1.0) to 0 .75. Using the reasonable "experimentally" determined R_s* of 0.2, equivalent to a 25-fold increase in *N*, larger gains in *P*(success) are anticipated. At a α^o of 1.0, *P*(success) now jumps from 0.14 to 0.67, for R_s* of 1.0 to 0.2. Now for α^o of 0.5, *P*(success) jumps from 0.35 to 0.82, for R_s* of 1.0 to 0.2, much closer to the desired >0.90 range.

2.5 CONCLUSIONS

The theory is presented that allows the analyst to estimate the probability of obtaining successful chemometric deconvolution of target analyte peaks in GC-MS. The theory takes into account the separation peak capacity, n_c, the sample complexity via the saturation factor, α^o, and also the limit of chemometric resolution, R_s*, for the chosen chemometric method. The results

highlight the challenges associated with obtaining an acceptable analysis, defined as achieving a satisfactory quantitative %error and confident analyte identification. While the chemometric method studied herein performed admirably, with an R_s* in the range of 0.2 to 0.3 depending upon analytical expectations, in general, more improvements to chemometric deconvolution methods to reduce R_s* are warranted. Since the MCR-ALS models were generated without constraints in this study, improvement to R_s* may be achieved if suitable constraints can be readily implemented, such as initial peak profile estimates for the more challenging overlap situations. For example, the previously reported mass cluster method[1,19,20] shows promise for extracting peak profiles with an R_s* of ~ 0.05, but further development is warranted. Indeed, if an R_s* of ~ 0.05 can be routinely achieved with chemometric deconvolution, then for a separation with an α^o of 0.5 the *P*(success) would be 0.95, and with an α^o of 1.0 the *P*(success) would be 0.91, essentially reaching the desired goal of a >0.90% *P*(success).

2.7 REFERENCES

[1] B.C. Reaser, S. Yang, B.D. Fitz, B.A. Parsons, M.E. Lidstrom, R.E. Synovec, Non-targeted determination of 13C-labeling in the Methylobacterium extorquens AM1 metabolome using the two-dimensional mass cluster method and principal component analysis, J. Chromatogr. A. 1432 (2016) 111–121. doi:10.1016/j.chroma.2015.12.088.

[2] J. Lisec, F. Hoffmann, C. Schmitt, C. Jaeger, Extending the Dynamic Range in Metabolomics Experiments by Automatic Correction of Peaks Exceeding the Detection Limit, Anal. Chem. 88 (2016) 7487–7492. doi:10.1021/acs.analchem.6b02515.

[3] C.G. Fraga, G.A. Pérez Acosta, M.D. Crenshaw, K. Wallace, G.M. Mong, H.A. Colburn, Impurity Profiling to Match a Nerve Agent to Its Precursor Source for Chemical Forensics Applications, Anal. Chem. 83 (2011) 9564–9572. doi:10.1021/ac202340u.

[4] J.M. Płotka, C. Morrison, D. Adam, M. Biziuk, Chiral Analysis of Chloro Intermediates of Methylamphetamine by One-Dimensional and Multidimentional NMR and GC/MS, Anal. Chem. 84 (2012) 5625–5632. doi:10.1021/ac300503g.

[5] T. Sasaki, E. Koshi, H. Take, T. Michihata, M. Maruya, T. Enomoto, Characterisation of odorants in roasted stem tea using gas chromatography–mass spectrometry and gas chromatography-olfactometry analysis, Food Chem. 220 (2017) 177–183. doi:10.1016/j.foodchem.2016.09.208.

[6] N. Sadoughi, L.M. Schmidtke, G. Antalick, J.W. Blackman, C.C. Steel, Gas Chromatography–Mass Spectrometry Method Optimized Using Response Surface Modeling for the Quantitation of Fungal Off-Flavors in Grapes and Wine, J. Agric. Food Chem. (2015). doi:10.1021/jf505444r.

[7] B. Krakowska, I. Stanimirova, J. Orzel, M. Daszykowski, I. Grabowski, G. Zaleszczyk, M. Sznajder, Detection of discoloration in diesel fuel based on gas chromatographic fingerprints, Anal. Bioanal. Chem. 407 (2014) 1159–1170. doi:10.1007/s00216-014-8332-4.

[8] M. del Olmo, A. González-Casado, N.A. Navas, J.L. Vilchez, Determination of bisphenol A (BPA) in water by gas chromatography-mass spectrometry, Anal. Chim. Acta. 346 (1997) 87–92. doi:10.1016/S0003-2670(97)00182-7.

[9] C. Rodier, O. Vandenabeele-Trambouze, R. Sternberg, D. Coscia, P. Coll, C. Szopa, F. Raulin, C. Vidal-Madjar, M. Cabane, G. Israel, M.F. Grenier-Loustalot, M. Dobrijevic, D. Despois, Detection of martian amino acids by chemical derivatization coupled to gas chromatography: In situ and laboratory analysis, Adv. Space Res. 27 (2001) 195–199. doi:10.1016/S0273-1177(01)00047-3.

[10] K. Robards, P.R. Haddad, P.E. Jackson, Principles and Practice of Modern Chromatographic Methods, Elsevier, Ltd., 2004.

[11] J.M. Davis, J.C. Giddings, Statistical theory of component overlap in multicomponent chromatograms, Anal. Chem. 55 (1983) 418–424. doi:10.1021/ac00254a003.

[12] J.M. Davis, J.C. Giddings, Statistical method for estimation of number of components from single complex chromatograms: theory, computer-based testing, and analysis of errors, Anal. Chem. 57 (1985) 2168–2177. doi:10.1021/ac00289a002.

[13] H.P. Bailey, S.C. Rutan, P.W. Carr, Factors that affect quantification of diode array data in comprehensive two-dimensional liquid chromatography using chemometric data analysis, J. Chromatogr. A. 1218 (2011) 8411–8422. doi:10.1016/j.chroma.2011.09.057.
[14] A. de Juan, J. Jaumot, R. Tauler, Multivariate Curve Resolution (MCR). Solving the mixture analysis problem, Anal. Methods. 6 (2014) 4964–4976. doi:10.1039/C4AY00571F.
[15] J. Kuligowski, G. Quintás, R. Tauler, B. Lendl, M. de la Guardia, Background Correction and Multivariate Curve Resolution of Online Liquid Chromatography with Infrared Spectrometric Detection, Anal. Chem. 83 (2011) 4855–4862. doi:10.1021/ac2004407.
[16] H. Parastar, R. Tauler, Multivariate Curve Resolution of Hyphenated and Multidimensional Chromatographic Measurements: A New Insight to Address Current Chromatographic Challenges, Anal. Chem. 86 (2014) 286–297. doi:10.1021/ac402377d.
[17] H. Parastar, J.R. Radović, M. Jalali-Heravi, S. Diez, J.M. Bayona, R. Tauler, Resolution and Quantification of Complex Mixtures of Polycyclic Aromatic Hydrocarbons in Heavy Fuel Oil Sample by Means of GC × GC-TOFMS Combined to Multivariate Curve Resolution, Anal. Chem. 83 (2011) 9289–9297. doi:10.1021/ac201799r.
[18] S.E. Stein, An integrated method for spectrum extraction and compound identification from gas chromatography/mass spectrometry data, J. Am. Soc. Mass Spectrom. 10 (1999) 770–781. doi:10.1016/S1044-0305(99)00047-1.
[19] B.D. Fitz, B.C. Reaser, D.K. Pinkerton, J.C. Hoggard, K.J. Skogerboe, R.E. Synovec, Enhancing Gas Chromatography–Time of Flight Mass Spectrometry Data Analysis Using Two-Dimensional Mass Channel Cluster Plots, Anal. Chem. 86 (2014) 3973–3979. doi:10.1021/ac5004344.
[20] B.D. Fitz, R.E. Synovec, Extension of the two-dimensional mass channel cluster plot method to fast separations utilizing low thermal mass gas chromatography with time-of-flight mass spectrometry, Anal. Chim. Acta. 913 (2016) 160–170. doi:10.1016/j.aca.2016.01.045..

2.8 TABLES

Table 2.1 Chromatographic conditions used in GC-MS data simulations for chemometric analysis.

	Conditions studied
R_s	0.02; 0.04; 0.06; 0.08; 0.10; 0.12; 0.14; 0.16; 0.18; 0.2; 0.25; 0.30; 0.35; 0.40; 0.45; 0.50
S/N	10; 100
Number of analytes	45
Number of analyte-interferent combinations	990
Total number of Chemometric Models	31,680

2.9 FIGURES

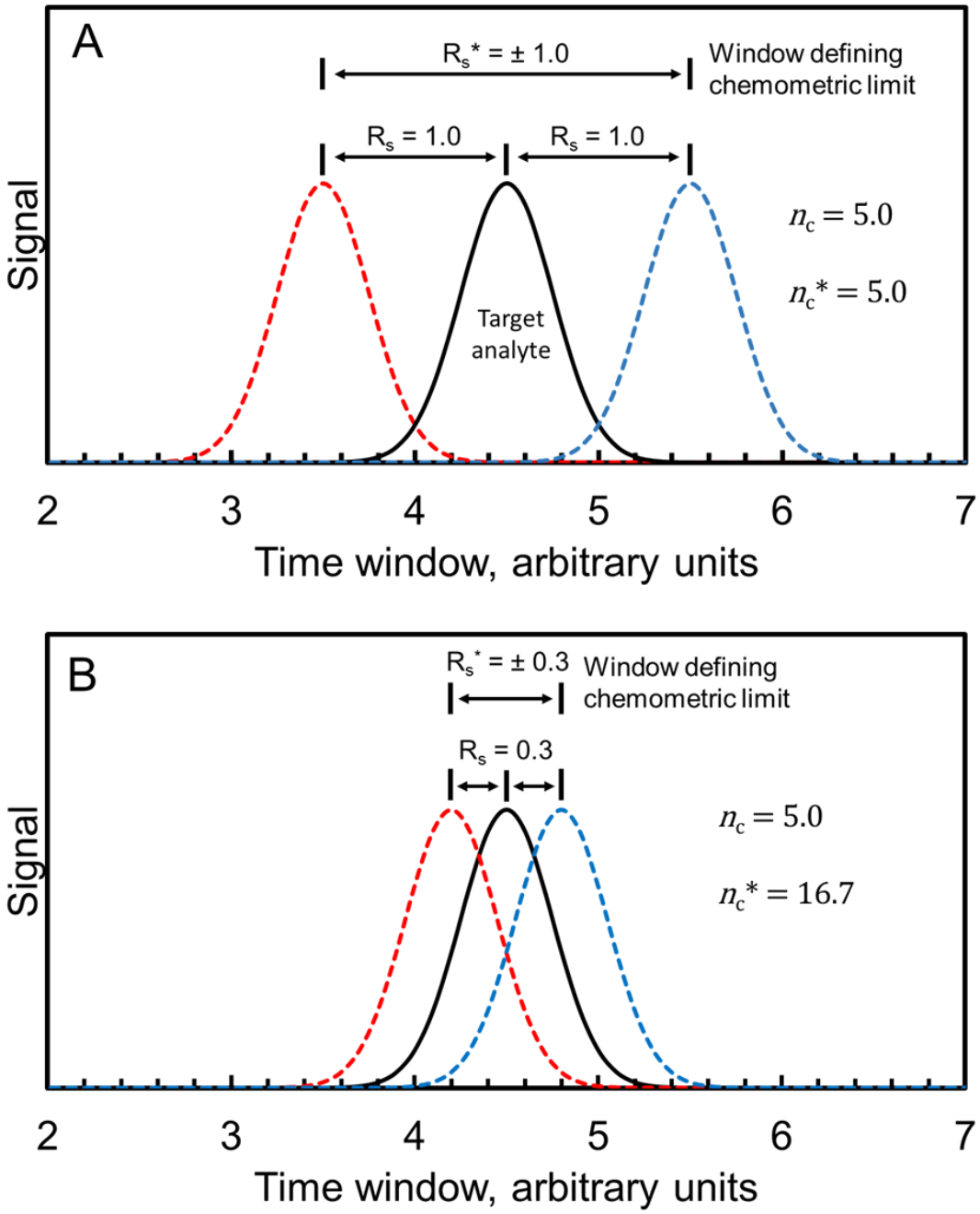

Figure 2.1 Simulated selective ion chromatograms of a single target analyte (solid line with $t_R = 4.5$) with two nearest interferent peaks (dashed lines) at $R_s = R_s^*$, all peaks ${}^1w_b = 1$ with constant concentration, $t_{window} = 5$ and $n_c = 5$. Signal and time in arbitrary units. (A) Interferent peaks at $R_s = 1$, consistent with the illustrated $R_s^* = 1$ and $n_c^* = n_c = 5$. (B) Interferent peaks at $R_s = 0.3$ consistent with the illustrated $R_s^* = 0.3$ and $n_c^* = 16.7$

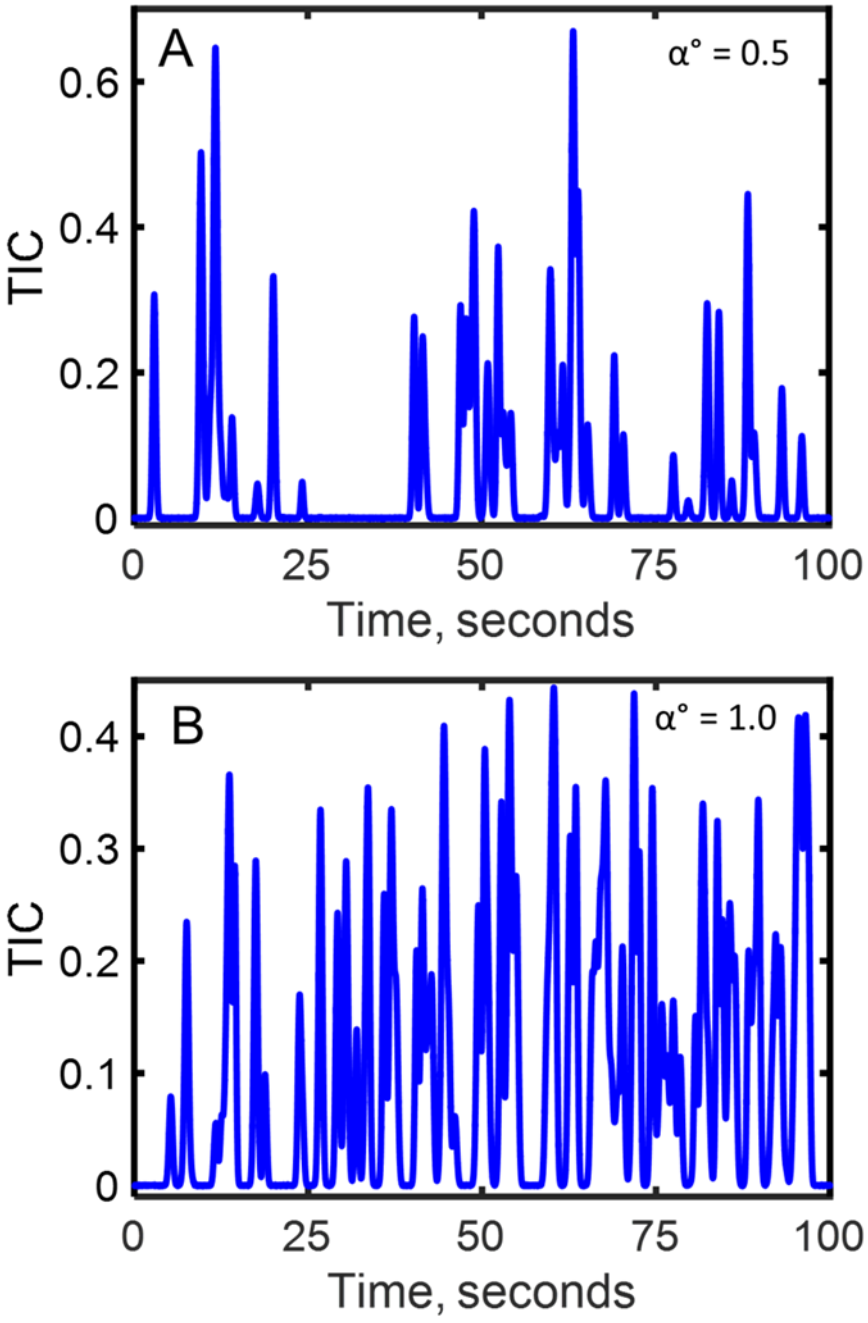

Figure 2.2 Representative simulated total ion current (TIC) chromatograms with n_c = 100 at (A) α^o = 0.5, or 50 randomly distributed analytes; (B) α^o = 1.0, or 100 randomly distributed analytes.

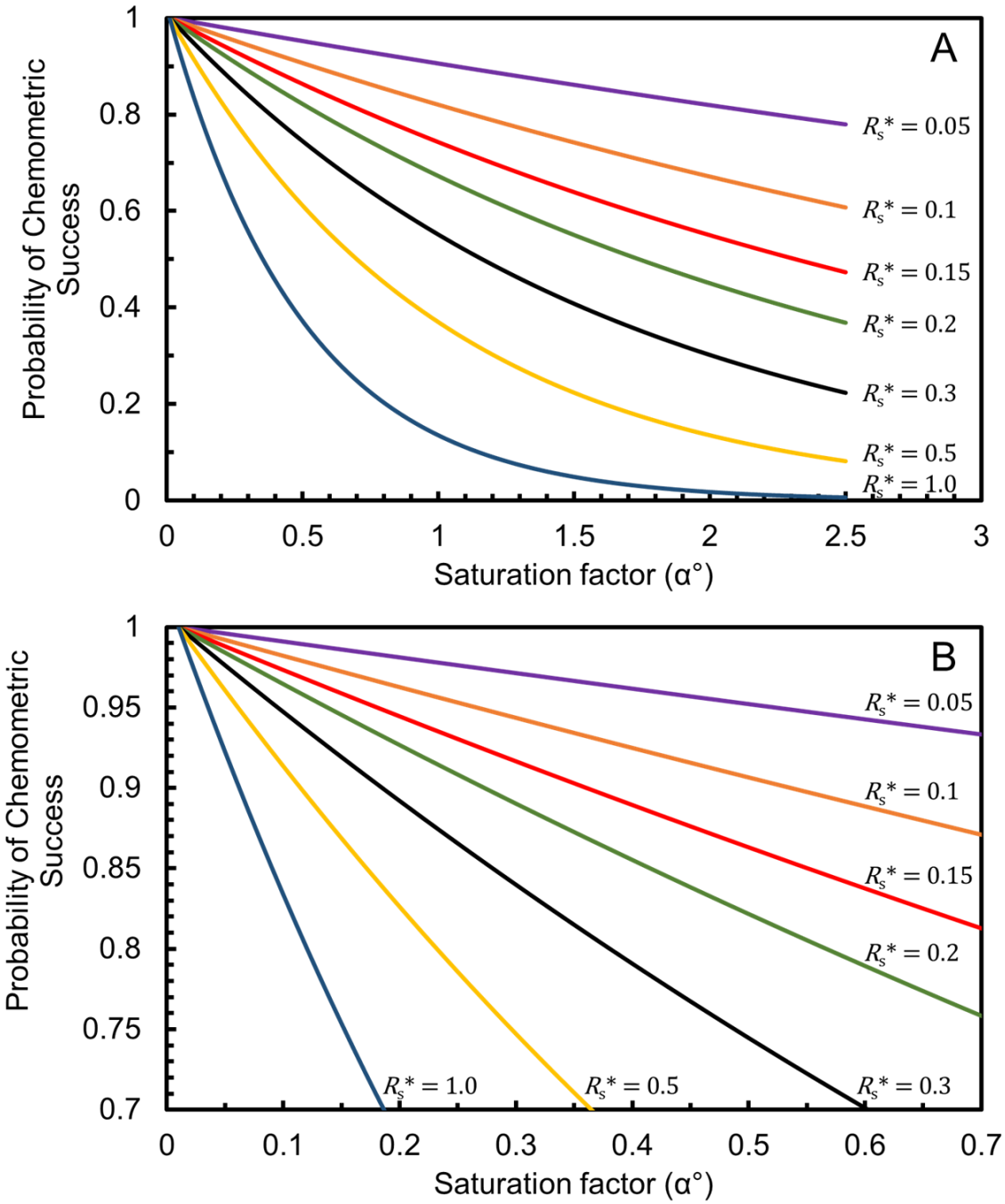

Figure 2.3 Plot of the probability of chemometric success, as a function of the saturation factor, α^o, Equation (2.8), for several chemometric resolutions, R_s^* = 0.05, 0.1, 0.15, 0.2, 0.3, 0.5, and 1.0. Saturation factor, α^o, is varied by fixing n_c and varying the number of interferent peaks, m. (B) is a zoom in of (A) for α^o = 0.0 to 0.7.

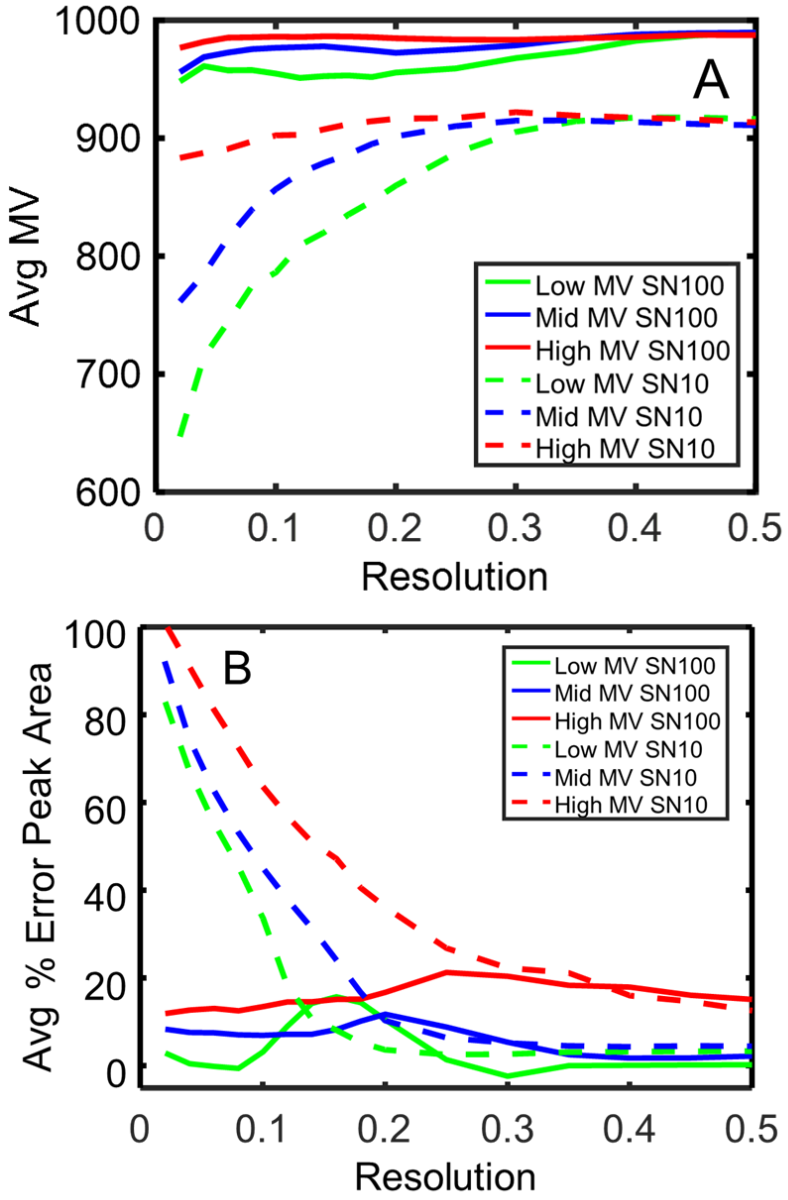

Figure 2.4 Summary of the model results versus resolution, R_s, where the solid lines correspond to the simulations at *S/N* 100 and the dashed lines to the simulations at *S/N* 10; the line colors correspond to the similarity between the target and interferent, with "Low" MV target-interferent pairs (very different mass spectra) in green, "Mid" MV pairs in blue and "High" MV (very similar mass spectra) pairs in red. (A) The average target loading-to-library MV results between the mass spectrum extracted by the model and the library spectrum used to generate the simulated data. (B) The average peak area percent error results (%error, Equation (2.9)) between the target analyte peak area extracted by the model and the peak area originally simulated.

2.10 SUPPORTING INFORMATION

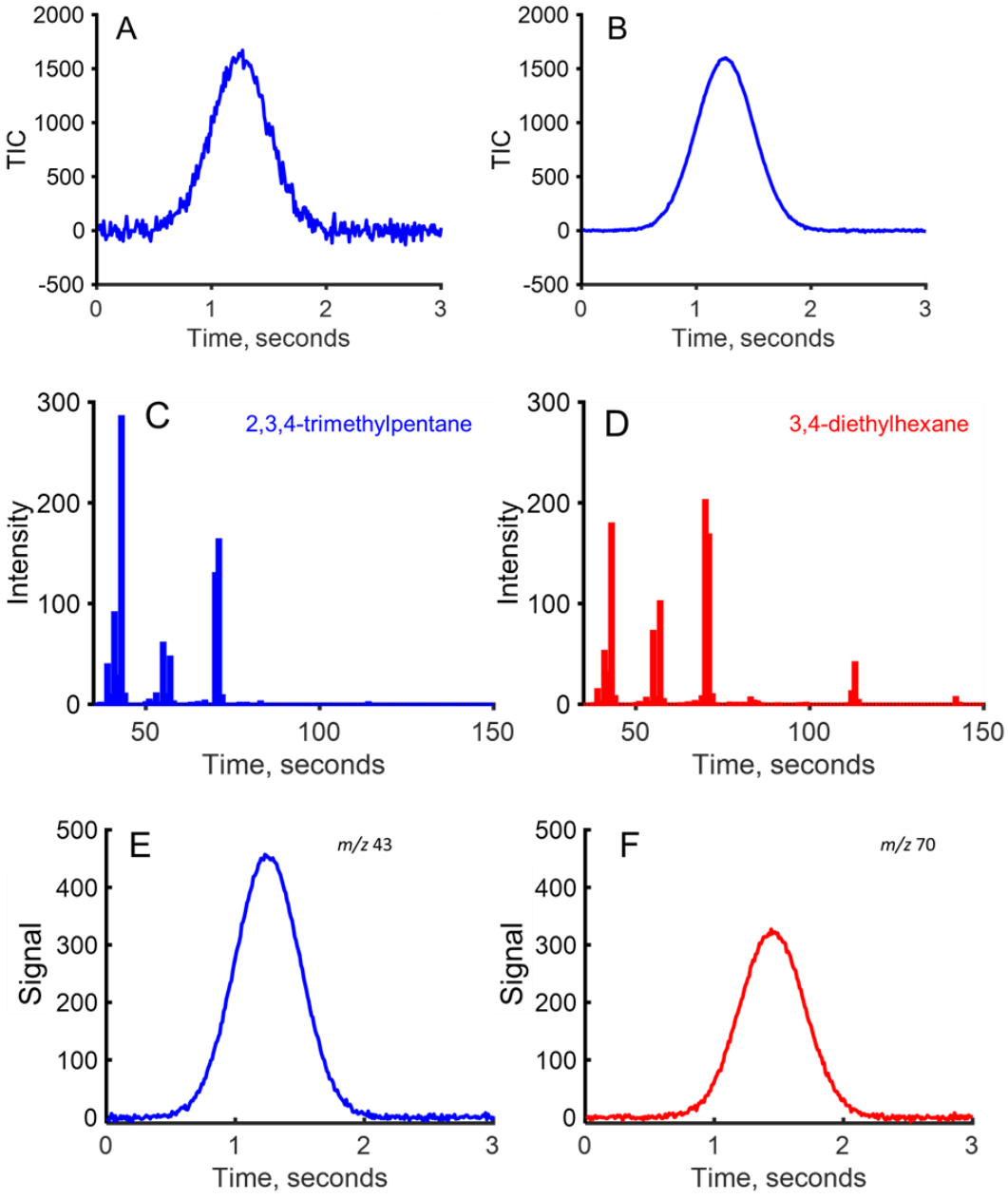

Figure S2.1 Simulated total ion current (TIC) chromatograms of a target and interferent at R_s = 0.2 with (A) *S/N* 10; and (B) *S/N* 100. The library mass spectra of (C) the target and (D) the interferent. The selective ion chromatogram (SIC) of (E) the target (*m/z* 43) and (F) the interferent (*m/z* 70), showing that there are slightly different *S/N* values at each *m/z* depending on the signal designated by the mass spectrum, but the *S/N* designation is based off the TIC.

Table S2.1 Table of 45 analytes used for the target-interferent pairs.

Number	Compound
1	Hexane
2	Heptane
3	Octane
4	Nonane
5	Decane
6	Undecane
7	Dodecane
8	Tridecane
9	Tetradecane
10	Pentadecane
11	Hexadecane
12	Pristane
13	1-Chlorohexane
14	1-chlorobutane
15	carbon tetrachloride
16	cyclooctane
17	cis-1,2-dimethylcyclohexane
18	2,3,4-trimethylpentane
19	2-methylpentane
20	1-octanol
21	1-nonanol
22	1-decanol
23	1-dodecanol
24	1-tetradecanol
25	1-hexadecanol
26	1-heptene
27	dodecene
28	1-undecene
29	propylbenzene
30	p-xylene
31	o-xylene
32	m-xylene
33	1-bromo-2-ethylhexane
34	1-chloro-5-methylhexane
35	2,3,6,7-tetramethyloctane
36	3,4-diethylhexane
37	butanoic acid
38	docosane
39	heneicosane
40	heptanoic acid
41	hexanoic acid
42	pentacosane
43	pentanoic acid
44	tetracosane
45	tricosane

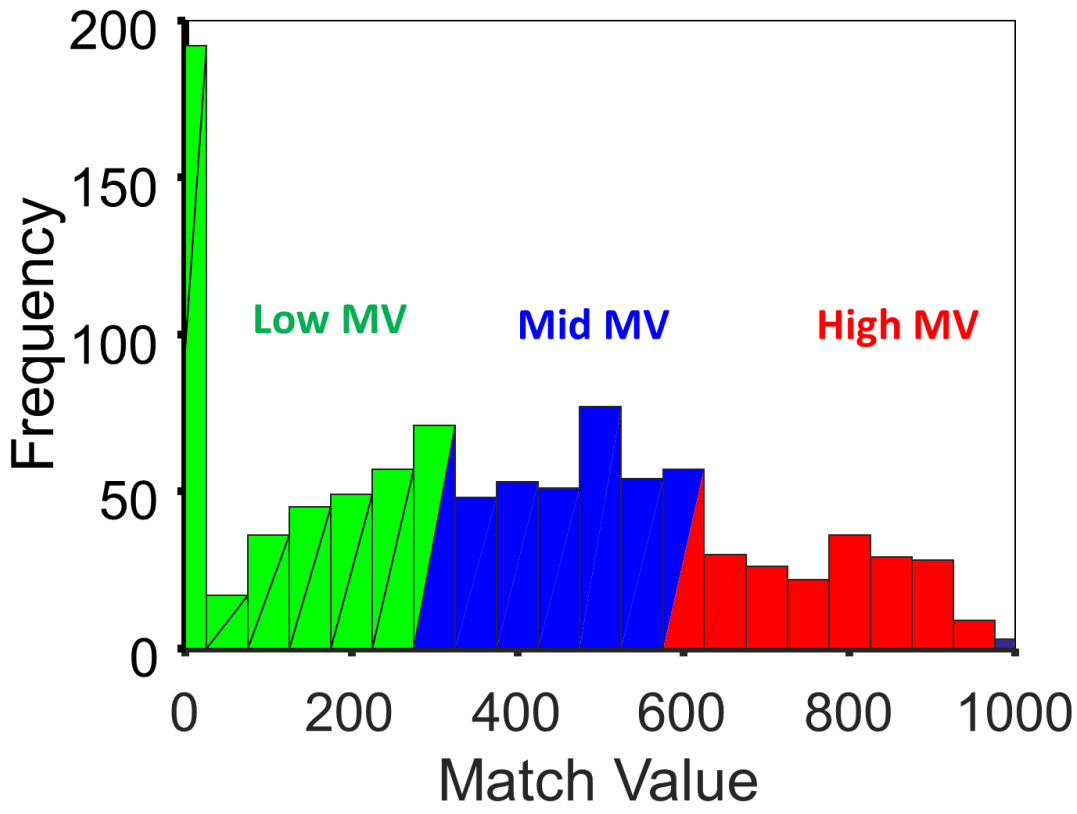

Figure S2.2 Histogram of the frequency of the match values (MV) occurring between the target-interferent pairs where Low MV (green) are defined as MV < 300 between the target and interferent, Mid MV (blue) are defined as 300 ≤ MV < 600, and High MV (red) are defined as 600 ≤ MV < 1000.

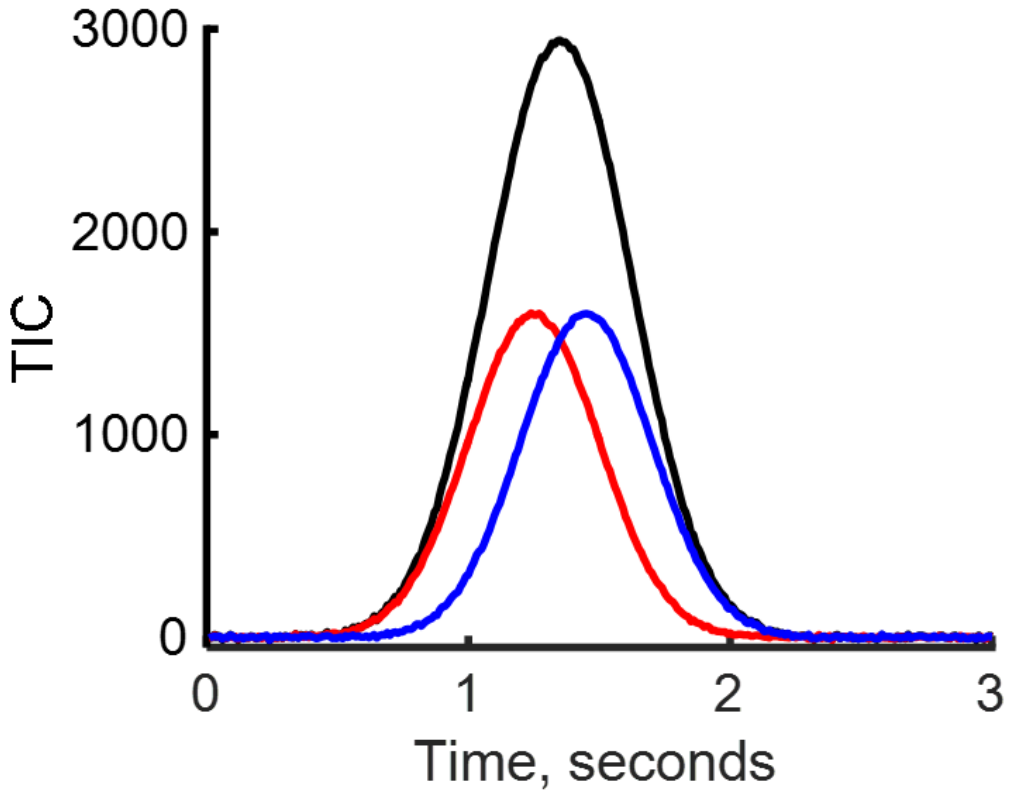

Figure S2.3 Elution profiles of a pure target (blue) and pure interferent (red) that, when coeluting at a $R_S = 0.2$, would sum to create the TIC (black).

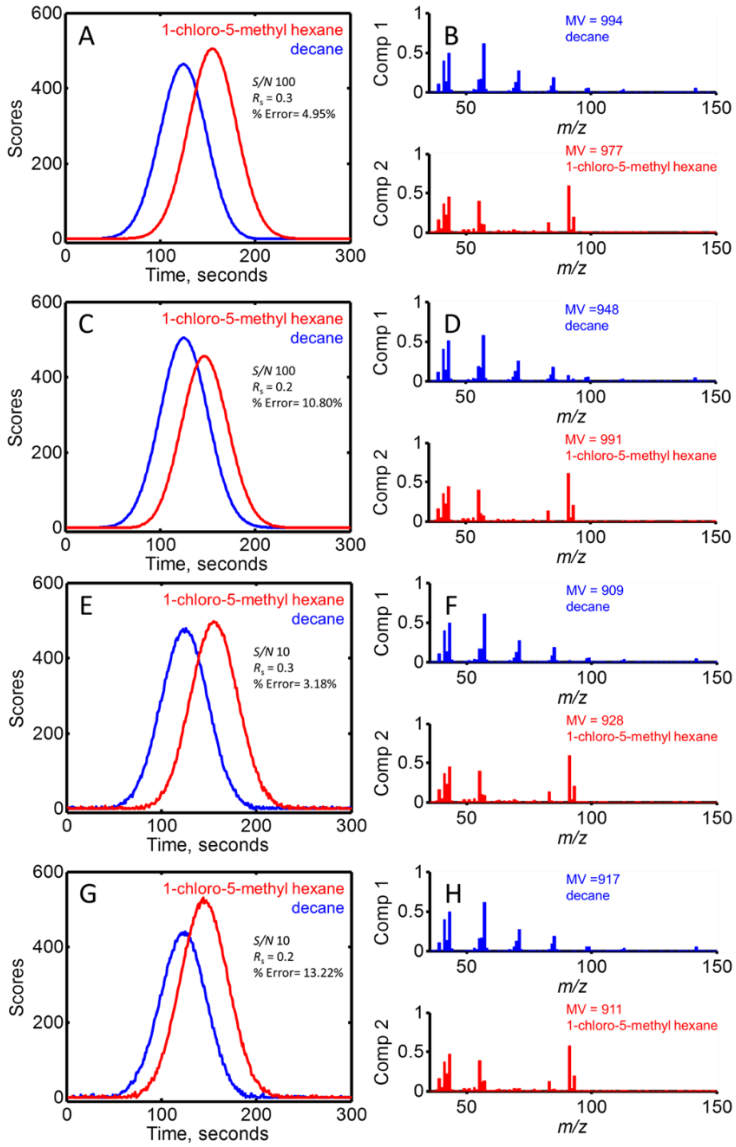

Figure S2.4 Deconvolution examples of representative "Low" MV target-interferent pair, 1-chloro-5-methyl hexane (target), and decane (interferent), at *R*S = 0.2 and 0.3 at *S/N* 100 and *S/N* 10. (A) R_s = 0.3 and *S/N* 100. The loadings corresponding to the chromatographic dimensions of target, 1-chloro-5-methyl hexane (red), and interferent, decane (blue). (B) R_s = 0.3 and *S/N* 100. The loadings corresponding to the mass spectral dimensions with the match values (MV) shown compared to the library used for the simulation. (C) R_s = 0.2 and *S/N* 100. Loadings as in (A). (D) R_s = 0.2 and *S/N* 100. Loadings as in (B). (E) R_s = 0.3 and *S/N* 10. Loadings as in (A). (F) R_s = 0.3 and *S/N* 10. Loadings as in (B). (G) R_s = 0.2 and *S/N* 10. Loadings as in (A). (H) R_s = 0.2 and *S/N* 10. Loadings as in (B).

Chapter 3. Implications of Phase Ratio for Maximizing Peak Capacity in Comprehensive Two-Dimensional Gas Chromatography Time-of-Flight Mass Spectrometry[2]

3.1 INTRODUCTION

Comprehensive two-dimensional (2D) gas chromatography coupled with time-of-flight mass spectrometry (GC×GC–TOFMS) is a powerful instrumental platform for the qualitative and quantitative analysis of complex samples [1–9]. By providing additional separation efficiency with similar run times to that of one-dimensional gas chromatography (1D-GC), GC×GC provides superior capability for resolving critical analyte peaks in complex samples [1,2,5,8]. An important figure-of-merit for chromatographic separations is peak capacity, n_c, which is generally defined as the time of the separation window divided by the average peak width-at-base (4σ width) [10]. Broadly, n_c describes the number of peaks that will fit into a given separation, generally defined at unit chromatographic resolution. Thus, n_c provides a performance-oriented metric to facilitate comparisons of chromatographic systems [11].

As a rule, analysts strive to maximize n_c, or to maintain adequate n_c while reducing separation run time. Compared to 1D-GC, the 2D peak capacity $n_{c,2D}$ with GC×GC can in principle provide approximately an order of magnitude increase in peak capacity while keeping the run time constant [5]. Recent work in GC×GC has focused on improving $n_{c,2D}$ by means of

[2]This Chapter has been reproduced from B.A. Parsons, D.K.Pinkerton, R.E. Synovec, Implications of Phase Ratio for Maximizing Peak Capacity in Comprehensive Two-Dimensional Gas Chromatography Time-of-Flight Mass Spectrometry, J. Chromatogr. A. Manuscript JCA-16-2350, *currently submitted.*

theoretical modeling [5,12,13] which provides better understanding to facilitate instrumental improvements and better experimental outcomes [14,15]. Indeed, there are many factors to be considered in the selection of parameters for GC×GC [16]. Recently, it has been recognized that high $n_{c,2D}$ separations for GC×GC may be achieved through careful selection of the dimensions of the primary, ^{1}D, and secondary, ^{2}D, capillary columns [5,14,15,17]. Column dimensions have also been considered recently for control of the pressures and flows of the ^{1}D and ^{2}D columns in the context of impacting $n_{c,2D}$ [18,19]. However, another critical parameter for column selection, phase ratio, β, has received comparatively little attention in GC×GC method development.

Phase ratio β, for a wall-coated open tubular (WCOT) capillary column, is the ratio of the volume of mobile phase, V_m, to the volume of stationary phase, V_s, and is determined by $\beta = d_c/4d_f$, where d_c is the inside diameter (i.d.) of the capillary and d_f is the thickness of the stationary phase film. Phase ratio is an important quantity for the practice of GC, as it allows for manipulation of k through selection of column film thickness relative to the capillary diameter. In practice, β allows for method considerations such as phase ratio focusing [20] and control of elution temperatures [21], while holding the chemical selectivity of the stationary phase constant. Further, by influencing k, in certain situations β may affect R_s [22]. Notably, compared to parameters such as column length or oven temperature, the variety of β values available for method development using commercial columns is much more limited. Arguably, the most common β is 250; an example of which for the ^{1}D column of GC×GC is a column of 250 μm inner i.d. with a 0.25 μm film thickness. Columns are generally available with smaller β (having a relatively thicker film) for analysis of highly volatile compounds, where phase ratio focusing helps to provide narrow peaks even at ambient or near-ambient GC oven temperatures [20]. Columns with larger β (having a relatively thin film) are also available for applications having

higher molecular mass analytes, such as in polyaromatic hydrocarbon (PAH) analyses, in which the larger β facilitates separation of very high boiling point analytes [21].

In temperature programmed separations, the rule of thumb is that for every factor of two change in β, the elution temperature, T_e, changes by ~ 10 ˚C. For example, suppose a given analyte elutes in a temperature programmed separation at 100 ˚C on a column with a β of 250, holding all other parameters constant, if a thicker film column of the same stationary phase with a β of 125 were substituted, one would expect T_e to increase to approximately 110 ˚C, so in 1D-GC, the change in T_e with β has a relatively modest impact. However, in GC×GC the effect of β on the temperature-programmed ^{1}D separation has a much greater significance, since the temperatures of the ^{1}D and ^{2}D columns are closely linked by the temperature programming, commonly with the same temperature programming rate, combined with a constant temperature offset. While some GC×GC instrumental designs allow for independent control of the temperature programming rate, this is not common. Hence, since the ^{2}D separation is pseudo-isothermal, the elution temperature, ^1T_e, of an analyte from the ^{1}D separation determines the temperature of the ^{2}D separation [5,12,13]. This relationship is a key focus of this study.

Herein, the relationship between β for the ^{1}D and ^{2}D dimensions of GC×GC separations, and the implications of β on realization of maximal $n_{c,2D}$ is experimentally examined. A complex sample mixture comprised of 115 analytes of various molecular masses and functional groups was used, while adjusting only the modulation period, P_M, as required by the degree of analyte retention on the ^{2}D separation, in order to assess and maximize the use of the 2D separation space. In this study an "appropriate" P_M was defined as when the full ^{2}D separation space was filled with peaks just up to the point of observing wraparound of the longer retained peaks into the lesser retained peaks. To the best extent possible, all other chromatographic conditions were

held constant such as the carrier gas flow rate, oven temperature program, and inlet and detector settings. Holding all other variables constant in this way allows for an analysis of the effects of β without convoluting the analysis with contributions from other chromatographic variables. Aside from application-specific columns, commercially-available columns are relatively limited in choice of β for a given column dimension and stationary phase composition. A non-polar × polar column configuration was utilized, based upon two commonly available column types: 250 μm i.d. ^{1}D columns with -5MS (diphenyl dimethyl polysiloxane) films and 180 μm i.d. ^{2}D columns with -200 (trifluoropropylmethyl polysiloxane) films. From the selected vendor for the chosen column dimensions, -5MS columns are available in 0.1, 0.25, 0.5, and 1.0 μm film thicknesses (with β of 625, 250, 125, and 62.5, respectively), while the -200 columns are limited to 0.2 and 0.4 μm films (with β of 225 and 112.5, respectively). The findings of this study are independent of the stationary phase selection and the column configuration. Taking into account the β commercially available, this study aims to examine the implications of a series of six GC×GC configurations chosen from the aforementioned columns. In this performance comparison, the focus was on the effects of β on n_c for both the ^{1}D and ^{2}D columns, and subsequently the 2D peak capacity $n_{c,2D}$ that was realized for each configuration.

Since changes in $^{1}\beta$ alter $^{1}T_e$, the choice of $^{1}\beta$ substantially impacts the temperature of the ^{2}D separation for each analyte, and thus alters ^{2}k. This effect is particularly significant if the ^{1}D and ^{2}D temperatures are equal. Accordingly, for this study the ^{2}D oven was set to a temperature offset of +5 ˚C versus the ^{1}D oven. Due to the relationship of ^{2}k with β, both directly via $^{2}\beta$ and indirectly via $^{1}\beta$, we propose that the β ratio, β_R, defined as $^{1}\beta$ divided by $^{2}\beta$ for the ^{1}D and ^{2}D separations, conveniently describes the relative ^{2}k on each column set, as increases in β_R will

increase ^2k. It will be demonstrated that β_R coupled with an appropriate, and not excessive P_M, plays a key role in maximizing the $n_{c,2D}$ obtained.

3.2 THEORY

For an isothermal separation, an analyte retention factor, k, may be determined from the retention time, t_R, and dead time, t_0 as

$$k = \frac{t_R - t_0}{t_0} \tag{3.1}$$

and β may be calculated for a WCOT column from the column inner diameter, d_c, and film thickness, d_f, as [22]

$$\beta = \frac{d_c}{4 \cdot d_f} \tag{3.2}$$

For isothermal separations, modifications to β, manifested as changes in k while holding other chromatographic parameters constant, are provided through the van't Hoff equation

$$\ln k = -\frac{\Delta H^\circ}{RT} + \frac{\Delta S^\circ}{R} - \ln \beta \tag{3.3}$$

where ΔH° is the enthalpy of vaporization of the analyte, R is the ideal gas constant, T is the temperature of the 2D separation for a specific modulation, and ΔS° is the entropy of the analyte. The analyte peak width-at-base, w_b, depends upon k according to [13,23,24],

$$w_b \approx w_{b,k} + w_{b,0} \tag{3.4}$$

where $w_{b,k}$ is the k dependent contribution to the peak width for $k > 0$ due to on-column band broadening, dominated by mass transfer in the mobile phase, and $w_{b,0}$ is the minimum peak width of the analyte, at $k = 0$, often dominated by "extra-column" band broadening due to injection.

Changes in w_b from altering β are principally of concern for their impact on peak capacity, n_c, herein defined at unit chromatographic resolution,

$$n_c = \frac{\Delta t_R}{w_b} \tag{3.5}$$

where Δt_R is the difference in retention time between the first and last eluting peak. For isothermal separations, or pseudo-isothermal such as those involving the ^{2}D column in GC×GC, w_b is not constant and increases linearly with k according to Eq. (3.4) when $w_{b,0}$ is sufficiently less than $w_{b,k}$ [13,23,24]. Thus, the peak capacity on ^{2}D, ^2n_c, must be determined iteratively to obtain an accurate result. The Giddings definition of ideal 2D peak capacity is applied, $n_{c,2D}$, with comparisons only made between one set of GC×GC conditions to another [10],

$$n_{c,2D} = {}^1n_c \times {}^2n_c \tag{3.6}$$

Several authors have discussed improvements to Eq. (3.6) to account for considerations of statistical peak overlap theory and/or the geometry inherent to multidimensional separations [15,25,26]. These considerations are highly useful for comparisons of GC×GC to 1D-GC, or for comparisons of substantially different GC×GC systems. However, for the purpose of understanding the impact of β in terms of $n_{c,2D}$, Eq. (3.6) is sufficient. Indeed, it is the effect of β on the relative retention (via k) and on n_c for both the ^{1}D and ^{2}D separations, and subsequently the impact on $n_{c,2D}$ that we seek to explore. For this purpose, the β ratio, or β_R, is defined,

$$\beta_R = {}^1\beta / {}^2\beta \tag{3.7}$$

for the ^{1}D and ^{2}D separations, respectively. It will be demonstrated that β_R plays a key role in understanding how ^1n_c and ^2n_c relate to each other for the purpose of increasing the $n_{c,2D}$.

3.3 EXPERIMENTAL

3.3.1 *GC×GC Column Sets*

Capillary columns were obtained from Restek (Bellefonte, PA). A total of five columns were used: three ^{1}D columns and two ^{2}D columns. The ^{1}D columns were 20.0 m × 250 µm i.d. and were 0.25 µm (Rtx®-5MS), 0.5 µm (Rtx®-5MS), and 1.0 µm (Rtx®-5) films; both Rtx®-5 and Rtx®-5MS films were Crossbond® diphenyl dimethyl polysiloxane. The ^{2}D columns were 2.0 m × 180 µm i.d. and were 0.2 µm and 0.4 µm Rtx®-200 films, which are Crossbond® trifluoropropylmethyl polysiloxane. The five columns were combined to make six column combinations, summarized in Table 3.1 to cover the β_R values typically encountered, Eq. (3.7). Three additional columns were applied for GC–FID isothermal data collection to provide van't Hoff plots of selected analytes in order to validate the ^{2}k values for a subset of analytes in the test mixture: a 10.0 m × 250 µm × 0.25 µm Rtx®-5MS, a 10.0 m × 180 µm × 0.2 µm Rtx®-200, and a 10.0 m × 180 µm × 0.5 µm Rtx®-200 (details and results provided in Supplemental).

3.3.2 *GC×GC–TOFMS Data Collection and Analysis*

The six column sets were installed into a Pegasus 4D system (LECO, St. Joseph, MI), consisting of an Agilent 6890N gas chromatograph with 7683 autosampler (Agilent Technologies, Palo Alto, CA) coupled to a Pegasus III TOFMS with a quad-jet thermal modulator and ^{2}D oven. For the purposes of avoiding variation in the ^{2}D column length and installation between column sets, all column sets using the 0.2 µm film ^{2}D column were run together, changing only the ^{1}D columns between runs. After the data collection for the first three column sets was complete, the TOFMS was vented and the 0.2 µm film ^{2}D column was exchanged for the 0.4 µm film. The acquisitions for the column sets using the 0.4 µm film ^{2}D

column were then completed, changing only the ^{1}D column between column sets. The ^{2}D column installation was carefully performed to ensure that the column lengths were consistent between the 0.2 µm and 0.4 µm film columns.

Instrumental parameters were set based on prior reports, to place the findings of this study in context of previous results [3–8,12,14,15]. The GC inlet was set to 275 °C and the transfer line was set to 285 °C. The ^{1}D oven was held at 40 °C for 1 min and then increased at 5 °C/min to the end of the temperature program (see Table 3.1), where it was held for 1 min, for a total run time ranging from 40 to 44 min. The ^{2}D oven and modulator block followed the same temperature program as the ^{1}D oven, with +5 °C and +20-40 °C offsets, respectively. The modulator temperature offsets, modulation period, and hot and cold pulse times are provided in Table 3.1. The GC instrument was set to maintain a constant (ambient temperature and pressure corrected) flow rate of 2 ml/min at the outlet of the ^{2}D column, with helium (Praxair, Grade 5.0) as the carrier gas. The ion source was set to 225 °C, the electron impact energy was 70 eV, and the detector voltage was set to 1690 V. Mass channels, *m/z* 33-300, were collected at 500 spectra/s after a 10 s acquisition delay. A 0.1 µl injection of a 115 component mixture was made with a 0.5 µl autosampler syringe (Hamilton, Reno, NV) in split mode with a split ratio of 200:1. The sample was injected in quadruplicate on each column set.

The GC×GC–TOFMS data collected with the six column sets were principally analyzed using ChromaTOF v3.32 (LECO, St. Joseph, MI). Peak processing methods were built to find peaks in the chromatograms, and subject them to automated deconvolution, peaklet assignment, and mass spectral library matching, along with measurement of ^{1}D retention time, $^{1}t_{R}$, ^{2}D retention time, $^{2}t_{R}$, and the ^{2}D width-at-half-height, $^{2}w_{1/2}$. Peak identification assignments were manually inspected and corrected, as needed. The peak tables were then further processed in

Microsoft Excel 2013 to correct ^2t_R for wrap-around, as well as to calculate the ^{2}D width-at-base, ^2w_b (with ^2w_b equal to $1.7 \times {}^2w_{1/2}$) and retention factor, ^2k. The ^{2}D dead time, $^2t_0$ was calculated for each column set using the ChromaTOF flow calculator, and was estimated to be 1.00 s at a ^{1}D oven temperature of 150 °C. A total of 70 analyte peaks from the 115 component mixture were selected for in-depth analysis and were measured for each of the six column combinations (see Table 3.2). GC×GC–TOFMS data were also imported into MATLAB 2015b (The MathWorks, Natick, MA) using an in-house data converter, peg2mat3p8 [27], for the purposes of measuring the ^{1}D widths-at-base, ^1w_b, and plotting.

3.3.3 *Determination of 1D Peak Capacity*

The ^1w_b was measured for each analyte based on a method previously described [28–30], utilizing curve fitting to a Gaussian profile to determine the parameters of the ^{1}D peak profile represented by the profile of the peaklets in the unfolded GC×GC data. While the previously reported method was developed for accurate measurement of ^1t_R [28], our purpose here was to accurately measure ^1w_b. It is important to note that the peak widths measured herein have been broadened by modulation [14,15,29–32], and do not require an additional undersampling correction for peak width, albeit without the correction for statistical peak overlap [33]. Briefly, the Curve Fitting Toolbox (The MathWorks, Natick, MA) was used to fit Gaussian profiles to each of the ^{2}D peaklets, belonging to a given analyte. The amplitude (*a*) and position (*b*) terms of the equation for each peaklet were then used to fit a ^{1}D peak profile to the modulated peaklets. The ^1w_b was then calculated from the standard deviation (*c*) term. This method was applied to tetradecane using *m/z* 57, taken as a representative analyte, which was free of interferences from neighboring peaks, and which eluted near the middle of the ^{1}D separation. To determine ^1n_c via

Eq. (3.5), the observed separation window was calculated by subtracting the retention time of the first eluting analyte (2-propanol) from that of the last eluting analyte (1-eicosanol). For each column set, the separation window was divided by ${}^{1}w_b$ for tetradecane. This method allows for reasonable comparison of ${}^{1}n_c$ between column sets.

3.3.4 *Determination of ${}^{2}D$ Peak Capacity*

Determining ${}^{2}n_c$ is more complicated than that of ${}^{1}n_c$, as the pseudo-isothermal nature of the ${}^{2}D$ separation precludes the assumption that the peak width is constant, such that simple application of Eq. (3.5) would not provide an accurate measure of ${}^{2}n_c$. Instead, it is necessary to iteratively model the ${}^{2}D$ separation [13,24], using a function describing the dependence of ${}^{2}w_b$ on ${}^{2}k$, as determined by the linear best fit equation obtained from the measurements of ${}^{2}k$ and ${}^{2}w_b$ for the 70 analytes that were included in the study (Table 3.2). Briefly, the iterative approach determines the retention times and widths of a series of Gaussian peaks with adjacent peaks all at ${}^{2}R_s$ of 1. To achieve this unit resolution, the peak spacing progressing increases as the ${}^{2}w_b$ increases as a function of ${}^{2}k$. By setting the retention time of the first peak and last peak in the series commensurate with that observed in the experimental data, the number of peaks that fit in the separation window is determined, thus providing ${}^{2}n_c$.

3.4 Results and Discussion

The primary focus is on column sets 1-3, since these column sets facilitate a study of the interrelationship between P_M and β_R, using the thinner (more commonly applied) film thickness on ${}^{2}D$. We begin by investigating the selection of an appropriate P_M to fully utilize the 2D separation space for a given β_R. Next, the pseudo-isothermal nature of each ${}^{2}D$ separation in a

GC×GC separation is explored, as this has important consequences on the accurate determination of the ^{2}D peak capacity, ^2n_c. We then consider the impact of the appropriate P_M for a given β_R, as this substantially impacts the sampling of the ^{1}D peaks, and hence, the ^{1}D peak capacity, ^1n_c. All of these aspects are then combined in a section dealing two-dimensional peak capacity, $n_{c,2D}$. Finally, the findings for the column sets 1-3 are compared to the findings for column sets 4-6 to assess the impact of the ^{2}D stationary phase film thickness in the context of β_R.

3.4.1 *Retention Time Range for ^{2}D, and Implications for P_M for various β_R*

A critical parameter in method development for GC×GC separations, especially for this study of the relationship between β_R and $n_{c,2D}$, is the modulation period P_M. Figure 3.1(A) is a total ion current (TIC) chromatogram from column set 3 collected with a P_M of 6 s. Column set 3 is commonly used by many practitioners (^{1}D: 20 m length, 250 μm i.d., 0.25 μm film; ^{2}D: 2 m, 180 μm i.d., 0.2 μm film; β_R = 1.11). Our study is based on the portion of the chromatogram after the ^{2}D retention times of the n-alkanes reached a plateau after ~ 4.6 min in which the band of ^{2}D retention times have stabilized, with the unaligned region prior to ~ 4.6 min a consequence of using the single-ramp temperature program [33]; this region spans cyclohexanol to 1-eicosanol, and includes 70 analytes (see Table 3.2). The band of peaks in Fig. 3.1(A) with ^1t_R ~ 4.6 to 36.5 min on ^{1}D, and a ^2t_R of ~1.9 s on ^{2}D, includes most of the n-alkanes in the 115 component mixture, from nonane to eicosane. As expected for the non-polar × polar configuration studied here, this n-alkane band represents the least retained analytes on ^{2}D. The peak at ^1t_R of ~ 23 min and ^2t_R of ~ 4.6 s is the most retained analyte on ^{2}D, diethyl phthalate.

For column set 3, using a P_M of 6 s, there is unutilized 2D separation space. A more appropriate P_M would be shorter, such that the ^{2}D separation time does not exceed that of the

most retained analyte, diethyl phthalate; this would suggest a P_M of approximately 5 s. However, while the P_M of 5 s would eliminate excessive "empty space" in the upper portion of the 2D chromatogram, there would still be substantial unused space in the lower portion. A more appropriate P_M would be based on the difference in retention time between the most and least retained analytes; subtracting the ${}^2t_R$ for the n-alkane band (~ 1.9 s) from that of diethyl phthalate (~ 4.6 s), and allowing for the ${}^2w_b$ of those peaks (~ 0.12 s and ~ 0.24 s, respectively), suggests a suitable P_M of 3 s. Under these conditions, partial ^{2}D wraparound is allowed, and the most retained analytes on ^{2}D have retention times exceeding that of the P_M. Figure 3.1(B) is a representative TIC chromatogram from column set 3 using a P_M of 3 s. Under these conditions, all of the 2D separation space is essentially utilized. However, due to wraparound on ^{2}D, the observed elution order in Fig. 3.1(B) is misleading; the n-alkane band is still at ${}^2t_R$ of ~1.9 s, but diethyl phthalate now appears below the n-alkane band, at ${}^2t_R$ of ~1.5 s. Reregistration of the separation, to correct for partial wrap around, in Fig. 3.1(C) is readily produced and restores the n-alkane band to the bottom and diethyl phthalate to the top of the 2D chromatogram concurrent with full use of the separation space. For ease of visual interpretation, all further 2D chromatograms are presented in the reregistered form of Fig. 3.1(C); however, all data was collected in the form of Fig. 3.1(B), with the appropriate P_M chosen based on the relative retention times of the n-alkane band and diethyl phthalate. Finally, Fig. 3.2 shows the reregistered TIC chromatograms for column sets 1, 2, 3, and 6. Using knowledge of ${}^2t_0$ and P_M, along with each analyte peak's observed ${}^2t_R$, one can readily calculate the true ${}^2t_R$ for all 70 analyte peaks included in the subsequent analysis. In Fig. 3.2 full use of the 2D separation space is achieved, albeit with the need to apply various P_M as dictated by the use of the various column sets (1, 2, 3, and 6) which provide a range of β_R (0.28, 0.56, 1.11 and 2.22, respectively).

3.4.2 *Pseudo-isothermal Conditions of the ^{2}D Separation*

In GC×GC, the short ^{2}D separation relative to the ^{1}D separation, in the context of common implementation of moderate oven temperature programming rates, results in pseudo-isothermal ^{2}D separations within each P_M, and ^2t_R shift to shorter times from one modulation to the next [13]. The temperature change between modulations (ΔT) may be calculated from the programmed temperature ramp for the ^{1}D and ^{2}D ovens (T_{ramp}) and modulation period as $\Delta T = T_{\mathrm{ramp}} \cdot P_M$. For column set 3, there is only a differential in temperature of 0.25 ˚C within the span of each ^{2}D separation (i.e., each modulation). The most impactful consequence of the pseudo-isothermal nature of the ^{2}D separations is that the relatively small change in temperature within each P_M causes ^2w_b to increase linearly with ^2k [13,23,24] per Eq. (3.4).

The linear relationship between ^2w_b and the ^2k for each analyte is readily observed in the data. Figure 3.3(A) is a section of the reregistered 2D TIC chromatogram from the column set 3 separation in Fig. 3.1(C), from which six representative analytes having different ^2t_R were selected within a ^{1}D retention window of ~ 4 min. This difference in ^{1}D retention does not impact the study, as will be demonstrated herein. The six selected analytes, numbered in order of increasing ^2t_R were hexadecane, 1-tetradecanol, methyl dodecanoate, 2-pentadecanone, benzophenone, and diethyl phthalate. The combined ^{2}D chromatogram in Fig. 3.3(B) was generated by selecting the most abundant modulation for the six 2D peaks and extracting the ^{2}D peak for each analyte using a selective *m/z*. The six extracted peak profiles were summed together to generate a "single" ^{2}D extracted ion chromatogram (XIC) that would be expected had all six analytes co-eluted on the ^{1}D column and then been separated on the ^{2}D column. The ^2k and ^2w_b for each analyte peak in Fig. 3.3(B) were measured, and summarized in the figure inset, with the plot of ^2w_b as a function of ^2k showing a linear trend. Because peak width is also a

function of chemical constants for a given analyte, such as the analyte diffusion coefficient in the gas phase at the column outlet, $D_{g,0}$ [34], some scatter around the line of best fit is expected; however, the overall observation of linear increase of ${}^2w_b$ with ${}^2k$ per Eq. (3.4) due to the pseudo-isothermal nature of the ^{2}D separations.

To more comprehensively evaluate the linear increase of ${}^2w_b$ with ${}^2k$ per Eq. (3.4), a total of 70 analytes were measured spanning the range of cyclohexanol to 1-eicosanol, corresponding to 4.6 min to 36.5 (~ 32 min window) on ^{1}D for column set 3. Figure 3.4(A) is a plot of ${}^2w_b$ as a function of ${}^2k$ for the 70 measured analytes. The markers for each ${}^2w_b$ and ${}^2k$ pair have been assigned according to the ${}^1t_R$ ranges for the analytes in three groups; in order of ${}^1t_R$, analytes 1-23 (4.6 min – 8.7 min), 24-46 (8.7 min – 15.8 min), and 47-70 (15.8 min – 36.5 min) are represented by circles, triangles, and squares, respectively (see Table 3.2). This figure demonstrates the consistent linear dependence of ${}^2w_b$ on ${}^2k$, as well as the independence of this phenomenon from ${}^1t_R$. One of the assumptions made in the determination of ${}^2k$ from the GC×GC separations was that ${}^2t_0$ could be accurately calculated using the column flow calculator included with the ChromaTOF instrument software. Traditionally, t_0 determinations are performed by injection of an unretained analyte, such as methane [35]; however, this is generally not possible in thermally modulated GC×GC, though it has been demonstrated in flow-modulated GC×GC [36]. To overcome this issue, validation of the ${}^2k$ values obtained from the GC×GC analyses was performed using isothermal GC–FID analysis (via van't Hoff plots) of all six column sets (see Table S3.1 and Fig. S3.1 in the Supplemental). 11 analytes were selected comprising a total of four groups of closely eluting compounds from the ^{1}D column. Overall, ${}^2k$ is observed to be essentially identical between the two experimental platforms. For example, in Fig. 3.4(B) the plot of ${}^2k$ for each of the 11 analytes, as determined from the GC×GC separation (for column set

3) versus the *k* determined from the isothermal GC–FID separations is provided. The plot shows strong agreement between the *k* determined by each experimental platform.

We now return to the hypothesis that the phase ratio β on ^{1}D and ^{2}D, and in particular the ratio of the β, given by β_R defined in Eq. (3.7), should play an important role in GC×GC separation method development. Based upon Fig. 3.4(B) and the van't Hoff plots used for its preparation the ^1T_e were calculated for the series of 11 analytes that span the 2D chromatograms for each $^1\beta$ studied (see Supplemental Table S3.1). As is evident in Table S3.1, in GC×GC the effect of β on the temperature-programmed ^{1}D separation has a very large impact, since the temperatures of the ^{1}D and ^{2}D columns are closely linked by the temperature programming. Since the ^{2}D separation is pseudo-isothermal, the ^1T_e of an analyte from the ^{1}D separation determines the temperature of that analyte for the ^{2}D separation. For example, decreasing $^1\beta$ by increasing the ^{1}D column film thickness significantly increases the ^1T_e. In turn, this increase in ^1T_e will result in a significant decrease in the observed ^2k range at fixed $^2\beta$. This effect can be studied using β_R via Eq. (3.7) which serves as a simple metric to design and compare GC×GC separations. Figure 3.5(A-C) provide the ^2w_b versus ^2k plots for column sets 3, 2, and 1, which cover the three β_R of 1.11, 0.56, and 0.28, respectively. Overall, a similar linear relationship between ^2w_b and ^2k for each column set, with nearly the same slopes and y-intercepts. Essentially, the three sets of data overlay on top of each other in strong support of the dependence on β_R. The major difference between column sets is the range of ^2k present in each column set, which increases with β_R. Further, it is observed that the starting ^2k increases with β_R as well, shifting the best fit lines up and to the right. These observations have a significant impact on peak capacity in GC×GC separations.

3.4.3 *Peak Capacity on ^2D*

For the purpose of determining ^2n_c, the retention time window should not be defined strictly by the P_M, rather, it should be defined by the difference of retention time between the most and least retained analytes at a given P_M. Furthermore, one must take into account the ^2k-dependent band broadening on ^2w_b, which is further dependent on β_R as demonstrated in Fig. 3.5. Utilizing Eq. (3.4) to relate ^2w_b as a function of ^2k, and the range of ^2k encountered which defines a retention time window for column sets 3, 2 and 1 (data from Fig. 3.5, summarized in Table 3.3), three chromatograms have been simulated at unit R_S and are provided in Fig. 3.6, in which the peak count provides an accurate ^2n_c [13]. An interesting finding is that the least retained analytes (the n-alkanes) have non-negligible retention on ^{2}D, and their retention increases with β_R.

Under the specified typical conditions applied, there are diminishing returns from experimentally extending the ^2k range through manipulation of β_R. In the case of column set 1 (β_R = 0.28), which has the smallest ^2k range, the ^2n_c is 12 with a P_M of 1 s; in column set 2 (β_R = 0.56), the ^2n_c is 14 with a P_M of 1.5 s; finally, in column set 3 (β_R = 1.11), P_M is tripled to 3 s with the widest ^2k range, but ^2n_c only modestly increases to 21. The band broadening with ^2k results in progressively smaller gains in ^2n_c as the ^2k range is increased. The gains to ^2n_c are further decreased since the ^2k range is not just extending, it is also shifting such that the minimum ^2k is increasing, as observed in the ^2w_b versus ^2k plots in Fig. 3.5, in which the lines of best fit both extend and shift to higher ^2k as β_R increases. This effect further deprecates the ^2n_c gained at higher β_R as the minimum ^2k increases for the least-retained n-alkanes.

3.4.4 *Peak Capacity on ^1D*

While the change in ^2n_c is the most obvious impact of the effect of β on the ^2k range, there is also a significant effect on ^1n_c by the influence of ^2k range on the modulation ratio, M_R, defined as the ratio of the ^{1}D peak width at base to the modulation period. The chosen P_M affects ^1n_c by way of broadening ^{1}D peaks as a function of M_R [5,12,13]. Briefly, at high M_R the width of a modulated peak approaches that of the peak prior to modulation; that is, at sufficiently high M_R, there is essentially no band broadening due to modulation. However, under typical GC×GC operational parameters, M_R tends to be smaller (eg., in the range of 1.5 to 3 for most practitioners), which impacts ^1n_c. One may predict the impact on ^1n_c, if knowledge of P_M and ^1w_b prior to modulation is provided, using concepts introduced by Davis and Giddings referred to as statistical peak overlap theory [32]. In common practice of GC×GC, however, ^1w_b prior to modulation is not known, since the signal is measured after the modulation and subsequent ^{2}D separations. Therefore, the ^1w_b measured following modulation provides the effective width for all subsequent analysis [30]. The measured ^1w_b following modulation takes into account the band broadening that results from modulation, but does not consider other matters which are addressed by statistical peak overlap approaches. However, the methods applied herein are adequate to facilitate a relative comparison of ^1w_b and ^1n_c between column sets.

Using the method described in Experimental, ^1w_b was calculated for tetradecane in each column set, using the data collected with an appropriate P_M for the particular column set. Additionally, data was collected with longer P_M to eliminate analyte wraparound, which simplifies the interpretation of GC×GC chromatograms, but has notable impacts on ^1w_b and ^1n_c. In the case of column set 3, substantial modulator-induced broadening of the tetradecane peak was observed. At a P_M of 3 s, the ^1w_b was 5.5 s, while at a P_M of 6 s, the ^1w_b was 7.5 s. It is worth

emphasizing that, between these conditions, only P_M was changed. Similar degrees of modulator-induced band broadening were observed for column sets 1 and 2 when the longer P_M were used; these results are summarized in Table 3.3.

Comparing the degree of modulator-induced band broadening and its impacts to ${}^1n_c$ is more difficult between column sets, as the ^{1}D column film thickness is expected to influence ${}^1w_b$ prior to modulation. For example, column set 1 utilizes a 1.0 µm film on the ^{1}D column, while that of column set 3 is 0.25 µm. Prior to modulation, column set 1 is expected to produce peaks with wider ${}^1w_b$, all other things equal, than column set 3. Further, under the same experimental conditions, the ^{1}D column retention range between the most and least retained analytes is larger with thicker films. Typically, one would expect the peak capacity to be similar between the ^{1}D columns used, as the shorter retention time range the column with the thinner film would be made up for by its narrower peaks. After considering modulator-induced band broadening, however, the longer P_M required for column set 3 eliminates any ${}^1w_b$ advantages that the thinner film may have provided. Between column sets 1 and 3, the measured M_R is substantially reduced from 5.5 to 1.8, respectively. Ultimately, the combination of the band broadening effects of the ^{1}D column film thickness and the modulator result in measured ${}^1w_b$ which are approximately the same between column sets 1, 2, and 3. When M_R drops more substantially, such as in analyses using an overly long P_M, the increase in ${}^1w_b$ and decrease in ${}^1n_c$ are more significant.

3.4.5 *2D Peak Capacity*

Two-dimensional peak capacity $n_{c,2D}$, Eq. (3.6), is useful for comparing the ability of GC×GC systems to separate complex mixtures. Table 3.3 summarizes the performance of all of the column sets at their appropriate P_M, i.e., defined as using the full 2D separation space. In this

study, column set 3 (β_R = 1.11), provided the best $n_{c,2D}$ of 8200. This column set provided a ^1n_c of 390 and the largest ^2n_c at 21, and has a ^2k range amenable to a short P_M of 3 s concurrent with a M_R for tetradecane of 1.8. This column set utilized a thin film of 0.25 μm on the ^{1}D column, which provides narrower peaks that are somewhat broadened by the modulator. In contrast, the very commonly applied conditions in the field are like those provided in entry 3-a (β_R = 1.11) in Table 3.3, with the P_M of 6 s, no wraparound (see Fig. 3.1(A)), in which an extremely low M_R of 1.2 is observed. The $n_{c,2D}$ is also significantly reduced to 6100, as there is a tremendous loss in ^1n_c due to the undersampling of ^{1}D separation. The ^1w_b are significantly broadened by modulator-induced undersampling in this case, and not by intentionally broadening the ^{1}D peaks.

Column set 2 (β_R = 0.56) had a smaller ^2k range and a smaller ^2n_c of 14 compared to column set 3, while providing a ^1n_c of 420 and a $n_{c,2D}$ of 5900. However, the appropriate P_M for column set 2 is also shorter at 1.5 s, allowing an improved M_R of 3.7 for tetradecane. At this M_R, one would expect relatively low modulator-based band broadening. Also, the ^1w_b can be more precisely measured due to the improved data density across the ^{1}D separation. While the limited ^2k range of column sets 1 and 2 provides for a short P_M and enhanced M_R relative to column set 3, the M_R is perhaps larger than needed. Indeed, a $n_{c,2D}$ of 7200 has been reported with a ^1n_c of 600 and a ^2n_c of 12 using Column set 2 conditions as defined in Table 3.1; other separation conditions were tuned to increase the $n_{c,2D}$ while maintaining a M_R of ~ 2.5 [37].

Column set 1 (β_R = 0.28) had the smallest ^2k range and the smallest ^2n_c of 12, and provided a ^1n_c of 460 and a $n_{c,2D}$ of 5500. The appropriate P_M for column set 1 is also shorter at 1.0 s, with an excessive M_R of 5.5 for tetradecane. While in this study a β_R = 0.28 appears to not provide promising 2D separations, a $n_{c,2D}$ of 6000 has been achieved with a ^1n_c of 400 coupled with a ^2n_c of 15 in only a 7 min separation, using a β_R = 0.25 based upon 100 μm i.d. columns

[15]. The separation conditions were tuned to provide lower ^2k values, coupled with a P_M of 500 ms in order to increase the $n_{c,2D}$ while maintaining an average M_R of ~ 2.

In summary, for the test mixture studied here, column set 3 provided the maximum $n_{c,2D}$ while using an appropriate P_M. Operating at extensive P_M so as to avoid wraparound has dramatic negative implications, since ^2n_c stays constant, with substantial decreases in ^1n_c. The effect is particularly apparent if attempts are made to improve M_R by means of increasing ^1w_b prior to modulation. The differences in M_R and $n_{c,2D}$ between column sets 1, 2, and 3 demonstrate that there are inherent tradeoffs between these metrics. Depending on what is planned for the data following collection, analysts may prefer the M_R provided by column sets 1 and 2, or the $n_{c,2D}$ provided by column set 3. For example, separations having a higher M_R may be preferred to avoid undersampling narrower peaks that elute in earlier portions of the chromatogram, or to enable more accurate and precise measurement of ^1t_R. Another case for higher M_R could be where any lost resolution on the ^{1}D column is unlikely to be made up for on the ^{2}D column, such as in the example of chiral GC×GC [38], where a high M_R may be needed to maintain quantitative accuracy for critical enantiomeric peak pairs.

3.4.6 *Impacts of Increasing Film Thickness on the ^{2}D Column*

The aforementioned data were collected with column sets 1-3, which utilized a thin film of 0.2 µm on the ^{2}D column (β of 225), which is representative of most GC×GC configurations. In contrast, column sets 4-6 used a 0.4 µm film on the ^{2}D column (β of 112.5), which is less commonly utilized in GC×GC. Data were collected on column sets 4-6 to allow for the study of more extensive ^2k ranges, and their implications for P_M, M_R, and n_c. As summarized in Table 3.3, the observed ^2k ranges in column sets 4-6 were substantially increased by the thicker ^{2}D column

stationary phase, compared to the thinner film counterpart in column sets 1-3. Commensurate with the increase in the range of ${}^2k$, the appropriate P_M also increased. In contrast to findings that ${}^2n_c$ increased with ${}^2k$ range within column sets 1-3, albeit with diminishing returns, when comparing between column sets using the same ^{1}D column (e.g. comparing sets 1 and 4), there are subdued benefits to ${}^2n_c$. Though using the thicker film on the ^{2}D column significantly increases ${}^2k$, the ${}^2n_c$ is only slightly improved.

The restrained gains in ${}^2n_c$ for column sets 4-6 relative to column sets 1-3 may be attributed to two causes. First, with the thicker film, the starting ${}^2k$ is higher, resulting in a larger ${}^2w_b$ of the first eluting peak in the series, and increasing from there. Second, retained peaks on thicker films have non-negligible broadening from inhibited mass transfer in the stationary phase. The increased on-column band-broadening reduces the potential for increased ${}^2k$ range to improve ${}^2n_c$, especially in light of reductions to ${}^1n_c$ due to the modulator-induced band broadening that occurs at the longer appropriate P_M required by column sets 4-6. As reported in Table 3.3, column sets 4-6 achieve similar ${}^2n_c$ to the sets 1-3. However, the longer P_M required results in lower M_R, wider ${}^1w_b$, and ultimately, reduced $n_{c,2D}$. If the data were collected with a longer P_M to avoid ^{2}D column wraparound, reductions in $n_{c,2D}$ would be even more substantial.

3.5 CONCLUSIONS

For the practice of high peak capacity GC×GC, a variety of instrumental factors, including both fixed hardware design and variable instrument parameters, must be carefully chosen to maximize $n_{c,2D}$. Many of these factors have received previous attention in the context of $n_{c,2D}$ with the exception of the phase ratio β selected for both the ^{1}D and ^{2}D columns. In this study, the effects of changes in β were examined while holding instrument parameters other than

the modulation period P_M constant. Overall, β substantially affected $n_{c,2D}$ by influencing the ^{2}D retention factors ^2k, and thereby changing the P_M necessary for proper ^{2}D column separations. The necessary changes to P_M modify the modulation ratio M_R, which affects the ^{1}D column peak widths and ^1n_c. Through changes to $^1\beta$, the range of ^2k may be controlled, with subsequent effects to both ^2n_c and ^1n_c. These effects were opposite in direction, such that gains in ^2n_c may result in losses in ^1n_c. Due to the pseudo-isothermal nature of the ^{2}D column separation, there are diminishing returns to extending the ^2n_c at the cost of ^1n_c. The β ratio, β_R, defined as $^1\beta$ divided by $^2\beta$ for the ^{1}D and ^{2}D separations, was introduced as an important parameter for tuning GC×GC separation conditions. β_R conveniently describes the relative ^2k on each column set, as increases in β_R will increase ^2k. It was found that β_R must be considered in conjunction with an appropriate, and not excessive P_M, in order to maximize the $n_{c,2D}$ obtained.

3.7 REFERENCES

[1] Z. Liu, J.B. Phillips, Comprehensive two-dimensional gas chromatography using an on-column thermal modulator interface, J. Chromatogr. Sci. 29 (1991) 227–231. doi:10.1093/chromsci/29.6.227.

[2] R.M. Kinghorn, P.J. Marriott, Comprehensive two-dimensional gas chromatography using a modulating cryogenic trap, J. High Resolut. Chromatogr. 21 (1998) 620–622. doi:10.1002/(SICI)1521-4168(19981101)21:11<620::AID-JHRC620>3.0.CO;2-#.

[3] B.A. Parsons, D.K. Pinkerton, B.W. Wright, R.E. Synovec, Chemical characterization of the acid alteration of diesel fuel: Non-targeted analysis by two-dimensional gas chromatography coupled with time-of-flight mass spectrometry with tile-based Fisher ratio and combinatorial threshold determination, J. Chromatogr. A. 1440 (2016) 179–190. doi:10.1016/j.chroma.2016.02.067.

[4] S. Prebihalo, A. Brockman, J. Cochran, F.L. Dorman, Determination of emerging contaminants in wastewater utilizing comprehensive two-dimensional gas-chromatography coupled with time-of-flight mass spectrometry, J. Chromatogr. A. 1419 (2015) 109–115. doi:10.1016/j.chroma.2015.09.080.

[5] M.S. Klee, J. Cochran, M. Merrick, L.M. Blumberg, Evaluation of conditions of comprehensive two-dimensional gas chromatography that yield a near-theoretical maximum in peak capacity gain, J. Chromatogr. A. (n.d.). doi:10.1016/j.chroma.2015.01.031.

[6] N.P. Vasquez, M. Crosnier de bellaistre-Bonose, N. Lévêque, E. Thioulouse, D. Doummar, T. Billette de Villemeur, D. Rodriguez, R. Couderc, S. Robin, C. Courderot-Masuyer, F. Moussa, Advances in the metabolic profiling of acidic compounds in children's urines achieved by comprehensive two-dimensional gas chromatography, J. Chromatogr. B. 1002 (2015) 130–138. doi:10.1016/j.jchromb.2015.08.006.

[7] P.Q. Tranchida, P. Donato, F. Cacciola, M. Beccaria, P. Dugo, L. Mondello, Potential of comprehensive chromatography in food analysis, TrAC Trends Anal. Chem. 52 (2013) 186–205. doi:10.1016/j.trac.2013.07.008.

[8] J.V. Seeley, S.K. Seeley, Multidimensional gas chromatography: fundamental advances and new applications, Anal. Chem. 85 (2013) 557–578. doi:10.1021/ac303195u.

[9] A. Sampat, M. Lopatka, M. Sjerps, G. Vivo-Truyols, P. Schoenmakers, A. van Asten, Forensic potential of comprehensive two-dimensional gas chromatography, TrAC Trends Anal. Chem. 80 (2016) 345–363. doi:10.1016/j.trac.2015.10.011.

[10] J.C. Giddings, Unified separation science, Wiley, New York, 1991. https//catalog.hathitrust.org/Record/002450066 (accessed June 8, 2016).

[11] J.M. Davis, J.C. Giddings, Statistical theory of component overlap in multicomponent chromatograms, Anal. Chem. 55 (1983) 418–424. doi:10.1021/ac00254a003.

[12] L.M. Blumberg, F. David, M.S. Klee, P. Sandra, Comparison of one-dimensional and comprehensive two-dimensional separations by gas chromatography, J. Chromatogr. A. 1188 (2008) 2–16. doi:10.1016/j.chroma.2008.02.044.

[13] D.K. Pinkerton, B.A. Parsons, T.J. Anderson, R.E. Synovec, Trilinearity deviation ratio: A new metric for chemometric analysis of comprehensive two-dimensional gas

chromatography time-of-flight mass spectrometry data, Anal. Chim. Acta. 871 (2015) 66–76. doi:10.1016/j.aca.2015.02.040.
[14] R.B. Wilson, W.C. Siegler, J.C. Hoggard, B.D. Fitz, J.S. Nadeau, R.E. Synovec, Achieving high peak capacity production for gas chromatography and comprehensive two-dimensional gas chromatography by minimizing off-column peak broadening, J. Chromatogr. A. 1218 (2011) 3130–3139. doi:10.1016/j.chroma.2010.12.108.
[15] B.D. Fitz, R.B. Wilson, B.A. Parsons, J.C. Hoggard, R.E. Synovec, Fast, high peak capacity separations in comprehensive two-dimensional gas chromatography with time-of-flight mass spectrometry, J. Chromatogr. A. 1266 (2012) 116–123. doi:10.1016/j.chroma.2012.09.096.
[16] A. Mostafa, M. Edwards, T. Górecki, Optimization aspects of comprehensive two-dimensional gas chromatography, J. Chromatogr. A. 1255 (2012) 38–55. doi:10.1016/j.chroma.2012.02.064.
[17] M.M. Koek, B. Muilwijk, L.L.P. van Stee, T. Hankemeier, Higher mass loadability in comprehensive two-dimensional gas chromatography–mass spectrometry for improved analytical performance in metabolomics analysis, J. Chromatogr. A. 1186 (2008) 420–429. doi:10.1016/j.chroma.2007.11.107.
[18] D. Peroni, H.-G. Janssen, Comprehensive two-dimensional gas chromatography under high outlet pressure conditions: A new approach to correct the flow-mismatch issue in the two dimensions, J. Chromatogr. A. 1332 (2014) 57–63. doi:10.1016/j.chroma.2014.01.051.
[19] P.Q. Tranchida, M. Maimone, F.A. Franchina, T.R. Bjerk, C.A. Zini, G. Purcaro, L. Mondello, Four-stage (low-)flow modulation comprehensive gas chromatography–quadrupole mass spectrometry for the determination of recently-highlighted cosmetic allergens, J. Chromatogr. A. 1439 (2016) 144–151. doi:10.1016/j.chroma.2015.12.002.
[20] M.S. Klee, GC Inlets an introduction, Agilent Technologies, Inc., Wilmington, DE USA, 2005. http://www.agilent.com/cs/library/usermanuals/public/5958-9468_041007.pdf.
[21] L. Mahé, M. Courtiade, C. Dartiguelongue, J. Ponthus, V. Souchon, D. Thiébaut, Overcoming the high-temperature two-dimensional gas chromatography limits to elute heavy compounds, J. Chromatogr. A. 1229 (2012) 298–301. doi:10.1016/j.chroma.2012.01.030.
[22] L.M. Blumberg, Theory of Fast Capillary Gas Chromatography Part 4: Column performance vs. liquid film thickness, J. High Resolut. Chromatogr. 22 (1999) 501–508. doi:10.1002/(SICI)1521-4168(19990901)22:9<501::AID-JHRC501>3.0.CO;2-N.
[23] G.M. Gross, B.J. Prazen, J.W. Grate, R.E. Synovec, High-speed gas chromatography using synchronized dual-valve injection, Anal. Chem. 76 (2004) 3517–3524. doi:10.1021/ac049909g.
[24] R.B. Wilson, J.C. Hoggard, R.E. Synovec, High throughput analysis of atmospheric volatile organic compounds by thermal injection – isothermal gas chromatography – time-of-flight mass spectrometry, Talanta. 103 (2013) 95–102. doi:10.1016/j.talanta.2012.10.013.
[25] L.M. Blumberg, Accumulating resampling (modulation) in comprehensive two-dimensional capillary GC (GC×GC), J. Sep. Sci. 31 (2008) 3358–3365. doi:10.1002/jssc.200800424.
[26] J.M. Davis, Statistical theory of spot overlap in two-dimensional separations, Anal. Chem. 63 (1991) 2141–2152. doi:10.1021/ac00019a014.
[27] J.C. Hoggard, peg2mat3p8, 2011. http://depts.washington.edu/synlab/software/.

[28] J.L. Adcock, M. Adams, B.S. Mitrevski, P.J. Marriott, Peak modeling approach to accurate assignment of first-dimension retention times in comprehensive two-dimensional chromatography, Anal. Chem. 81 (2009) 6797–6804. doi:10.1021/ac900960n.
[29] W.C. Siegler, B.D. Fitz, J.C. Hoggard, R.E. Synovec, Experimental study of the quantitative precision for valve-based comprehensive two-dimensional gas chromatography, Anal. Chem. 83 (2011) 5190–5196. doi:10.1021/ac200302b.
[30] D.K. Pinkerton, B.A. Parsons, R.E. Synovec, Method to determine the true modulation ratio for comprehensive two-dimensional gas chromatography, J. Chromatogr. A. 1476 (2016) 114–123. doi:10.1016/j.chroma.2016.11.015.
[31] W. Khummueng, J. Harynuk, P.J. Marriott, Modulation ratio in comprehensive two-dimensional gas chromatography, Anal. Chem. 78 (2006) 4578–4587. doi:10.1021/ac052270b.
[32] J.M. Davis, D.R. Stoll, P.W. Carr, Effect of first-dimension undersampling on effective peak capacity in comprehensive two-dimensional separations, Anal. Chem. 80 (2008) 461–473. doi:10.1021/ac071504j.
[33] L.M. Blumberg, M.S. Klee, Elution parameters in constant-pressure, single-ramp temperature-programmed gas chromatography, J. Chromatogr. A. 918 (2001) 113–120. doi:10.1016/S0021-9673(01)00659-8.
[34] B.D. Fitz, B.C. Reaser, D.K. Pinkerton, J.C. Hoggard, K.J. Skogerboe, R.E. Synovec, Enhancing gas chromatography–time of flight mass spectrometry data analysis using two-dimensional mass channel cluster plots, Anal. Chem. 86 (2014) 3973–3979. doi:10.1021/ac5004344.
[35] S. Vezzani, G. Castello, D. Pierani, Measurement and prediction of dead times and column diameter in capillary gas chromatography by using air, methane and some solvents, J. Chromatogr. A. 811 (1998) 85–96. doi:10.1016/S0021-9673(98)00215-5.
[36] M.S. Klee, L.M. Blumberg, Measurement of retention in comprehensive two-dimensional gas chromatography using flow modulation with methane dopant, J. Chromatogr. A. 1217 (2010) 1830–1837. doi:10.1016/j.chroma.2010.01.027.
[37] R.E. Mohler, K.M. Dombek, J.C. Hoggard, E.T. Young, R.E. Synovec, Comprehensive two-dimensional gas chromatography time-of-flight mass spectrometry analysis of metabolites in fermenting and respiring yeast cells, Anal. Chem. 78 (2006) 2700–2709. doi:10.1021/ac052106o.
[38] Y.F. Wong, R.N. West, S.-T. Chin, P.J. Marriott, Evaluation of fast enantioselective multidimensional gas chromatography methods for monoterpenic compounds: Authenticity control of Australian tea tree oil, J. Chromatogr. A. 1406 (2015) 307–315. doi:10.1016/j.chroma.2015.06.036.

3.8 TABLES

Table 3.1 The column sets utilized in this report, with film thicknesses and corresponding β ratio, β_R, defined in Eq. (3.7). The ^{1}D columns were all 20.0 m × 250 µm i.d., and the ^{2}D columns were all 2.0 m × 180 µm i.d. The modulator and oven parameters used for each column set. The hot and cold pulse times are listed for each stage in the dual-stage thermal modulator. The maximum temperature of the ^{1}D oven program, $^1T_{max}$, and the thermal offset applied to the modulator block, $T_{mod,offset}$, are also provided.

.

Column Set	^1d_f (µm)	^2d_f (µm)	β_R	P_M (s)	Hot Pulse (s)	Cold Pulse (s)	$^1T_{max}$ (°C)	$T_{mod,offset}$ (°C)
1	1.00	0.2	0.28	1	0.4	0.10	250	20
2	0.50	0.2	0.56	1.5	0.4	0.35	240	20
3	0.25	0.2	1.11	3	0.6	0.90	230	40
4	1.00	0.4	0.56	3	0.6	0.90	250	40
5	0.50	0.4	1.11	5	1.25	1.25	240	40
6	0.25	0.4	2.22	8	2.0	2.00	230	40

Table 3.2 Numbered list of 70 measured analytes (in the complex mixture of 115 analytes) as separated on column set 3 (^{1}D: 20 m length, 250 μm i.d., 0.25 μm film; ^{2}D: 2 m, 180 μm i.d., 0.2 μm film; β_R = 1.11).

Column Set 3					
Peak Number	Name	^1t_R (s)	^2t_R (s)	^2k	^2w_b (s)
1	cyclohexanol	277	2.528	1.528	0.132
2	3-heptanone	286	3.362	2.362	0.172
3	o-xylene	289	2.29	1.29	0.125
4	2-heptanone	292	3.7	2.7	0.199
5	nonane	298	1.736	0.736	0.091
6	2-heptanol	301	2.44	1.44	0.130
7	cyclooctane	325	1.886	0.886	0.095
8	anisole	325	2.65	1.65	0.122
9	1-nonyne	334	1.972	0.972	0.091
10	methyl hexanoate	337	2.92	1.92	0.152
11	bromobenzene	340	2.462	1.462	0.124
12	2,6-dimethyloctane	346	1.846	0.846	0.101
13	1-bromohexane	349	2.38	1.38	0.111
14	propylbenzene	376	2.298	1.298	0.124
15	1,3,5-trimethylbenzene	400	2.34	1.34	0.129
16	3-octanone	436	3.65	2.65	0.151
17	t-butylbenzene	439	2.386	1.386	0.134
18	1,2,4-trimethylbenzene	439	2.44	1.44	0.124
19	decane	451	1.862	0.862	0.092
20	iso-butylbenzene	463	2.366	1.366	0.113
21	sec-butylbenzene	469	2.3	1.3	0.110
22	butylcyclohexane	502	1.982	0.982	0.104
23	benzyl alcohol	514	3.098	2.098	0.146
24	1-bromoheptane	517	2.512	1.512	0.115
25	5-decyne	529	1.892	0.892	0.092
26	n-butylbenzene	544	2.396	1.396	0.114
27	1-octanol	574	2.656	1.656	0.131
28	adamantane	583	2.192	1.192	0.114
29	1-undecene	607	1.972	0.972	0.086
30	2-nonanone	613	3.98	2.98	0.178
31	undecane	619	1.912	0.912	0.093
32	methyl octanoate	667	3.046	2.046	0.142
33	1,3,5-trichlorobenzene	670	2.614	1.614	0.120

34	1,6-dichlorohexane	676	3.562	2.562	0.157
35	1-bromooctane	691	2.544	1.544	0.114
36	1-nonanol	745	2.64	1.64	0.131
37	naphthalene	757	3.178	2.178	0.162
38	dodecene	775	1.988	0.988	0.099
39	methyl salicylate	781	3.66	2.66	0.170
40	2-decanone	781	3.878	2.878	0.195
41	dodecane	790	1.922	0.922	0.093
42	1,2,3-trichlorobenzene	805	2.908	1.908	0.132
43	1-geranoil	880	2.77	1.77	0.136
44	ethyl salicylate	907	3.616	2.616	0.170
45	1-decanol	910	2.606	1.606	0.132
46	2-undecanone	946	3.706	2.706	0.181
47	1,1'-bicyclohexyl	949	2.182	1.182	0.100
48	tridecane	952	1.924	0.924	0.095
49	cyclohexylbenzene	976	2.526	1.526	0.117
50	1,2,4,5-tetrachlorobenzene	991	2.728	1.728	0.138
51	methyl decanoate	991	2.894	1.894	0.151
52	2-dodecanone	1102	3.556	2.556	0.187
53	tetradecane	1105	1.924	0.924	0.093
54	1-dodecanol	1216	2.528	1.528	0.128
55	pentadecane	1252	1.916	0.916	0.093
56	methyl dodecanoate	1288	2.736	1.736	0.147
57	diethyl phthalate	1390	4.546	3.546	0.235
58	hexadecane	1390	1.906	0.906	0.100
59	benzophenone	1432	3.798	2.798	0.177
60	1-tetradecanol	1495	2.436	1.436	0.125
61	heptadecane	1522	1.902	0.902	0.102
62	2-pentadecanone	1525	3.19	2.19	0.177
63	pristane	1531	1.968	0.968	0.112
64	phenanthrene	1618	3.49	2.49	0.163
65	octadecane	1648	1.892	0.892	0.102
66	1-hexadecanol	1747	2.366	1.366	0.123
67	nonadecane	1768	1.884	0.884	0.102
68	eicosane	1882	1.876	0.876	0.103
69	1-octadecanol	1975	2.312	1.312	0.123
70	1-eicosanol	2185	2.256	1.256	0.121

Table 3.3 Summary of results. Column sets 1-6 (defined in Table 3.1) are experimental results using an appropriate P_M to avoid ^{2}D wrap around. Values for 3-a indicate experimental results from Colum set 3 using a commonly applied longer P_M (6 s instead of 3 s). The ^1t_R range for the ^{1}D separation was essentially constant for each column set at ~ 40 min.

Column Set	β_R	^1w_b (s)	P_M (s)	M_R	^2k range	^2w_b range (ms)	^1n_c	^2n_c	$n_{c,2D}$
1	0.28	5.5	1	5.5	0.27 - 1.09	67-100	460	12	5500
2	0.56	5.5	1.5	3.7	0.51 - 2.12	88-163	420	14	5900
3	1.11	5.5	3	1.8	0.74 - 3.54	89-217	390	21	8200
4	0.56	6.5	3	2.2	0.97 - 3.20	111-272	390	13	5100
5	1.11	7.3	5	1.5	1.47 - 5.85	157-399	320	17	5400
6	2.22	9.6	8	1.2	2.06 - 9.87	219-566	220	21	4600
3-a	1.11	7.5	6	1.2	0.74 - 3.54	89-217	290	21	6100

3.9 FIGURES

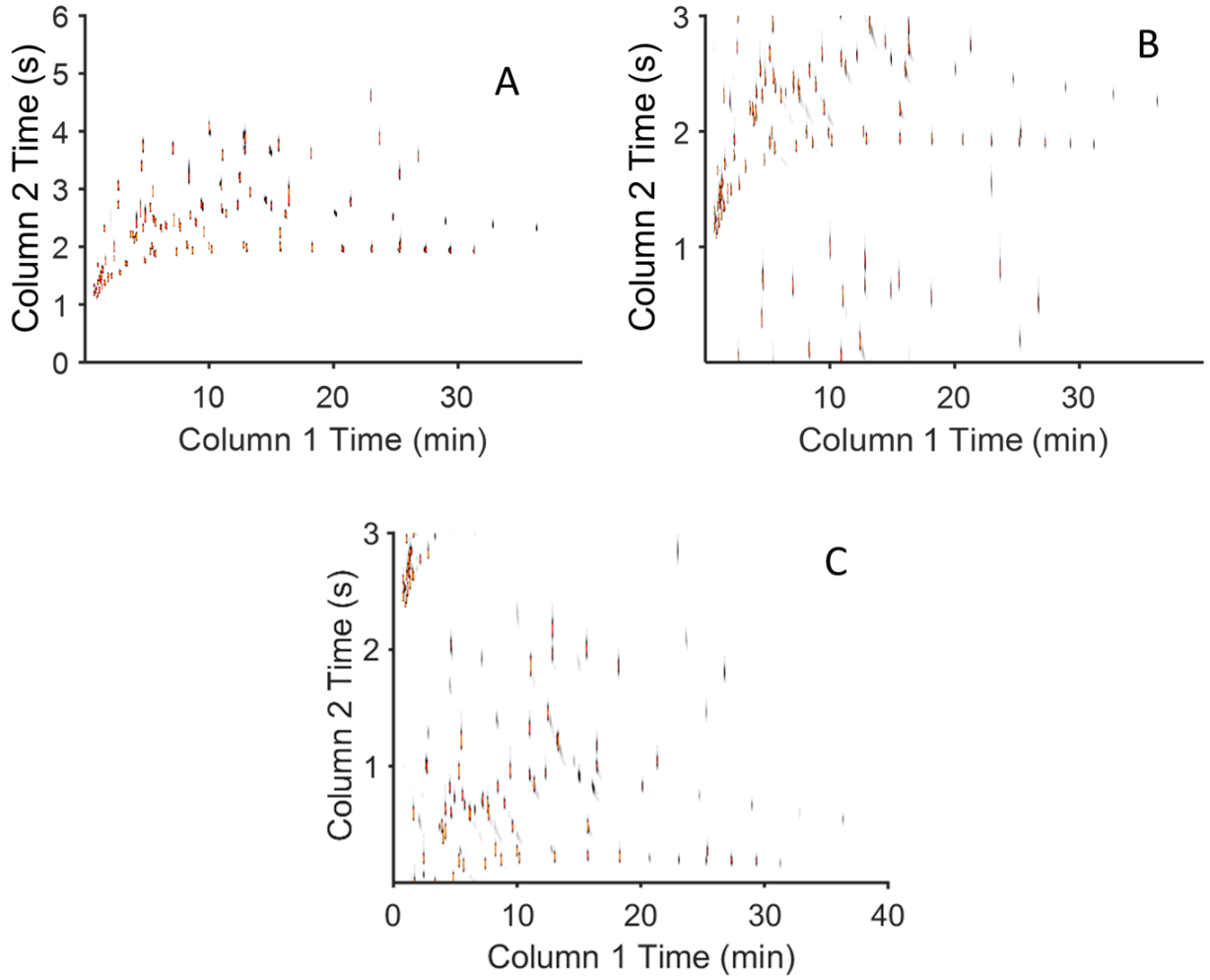

Figure 3.1 Representative total ion current (TIC) chromatograms from the separation of the 115 component mixture on column set 3 (^{1}D: 20 m length, 250 μm i.d., 0.25 μm film; ^{2}D: 2 m, 180 μm i.d., 0.2 μm film; β_R = 1.11). (A) using a P_M of 6 s, which allows for no wraparound of analytes. (B) using an optimal P_M of 3 s, which allows for partial wraparound to better utilize the 2D separation space. (C) reregistered chromatogram in which (B) has been replotted such that the least retained analytes appear at the bottom of the chromatogram.

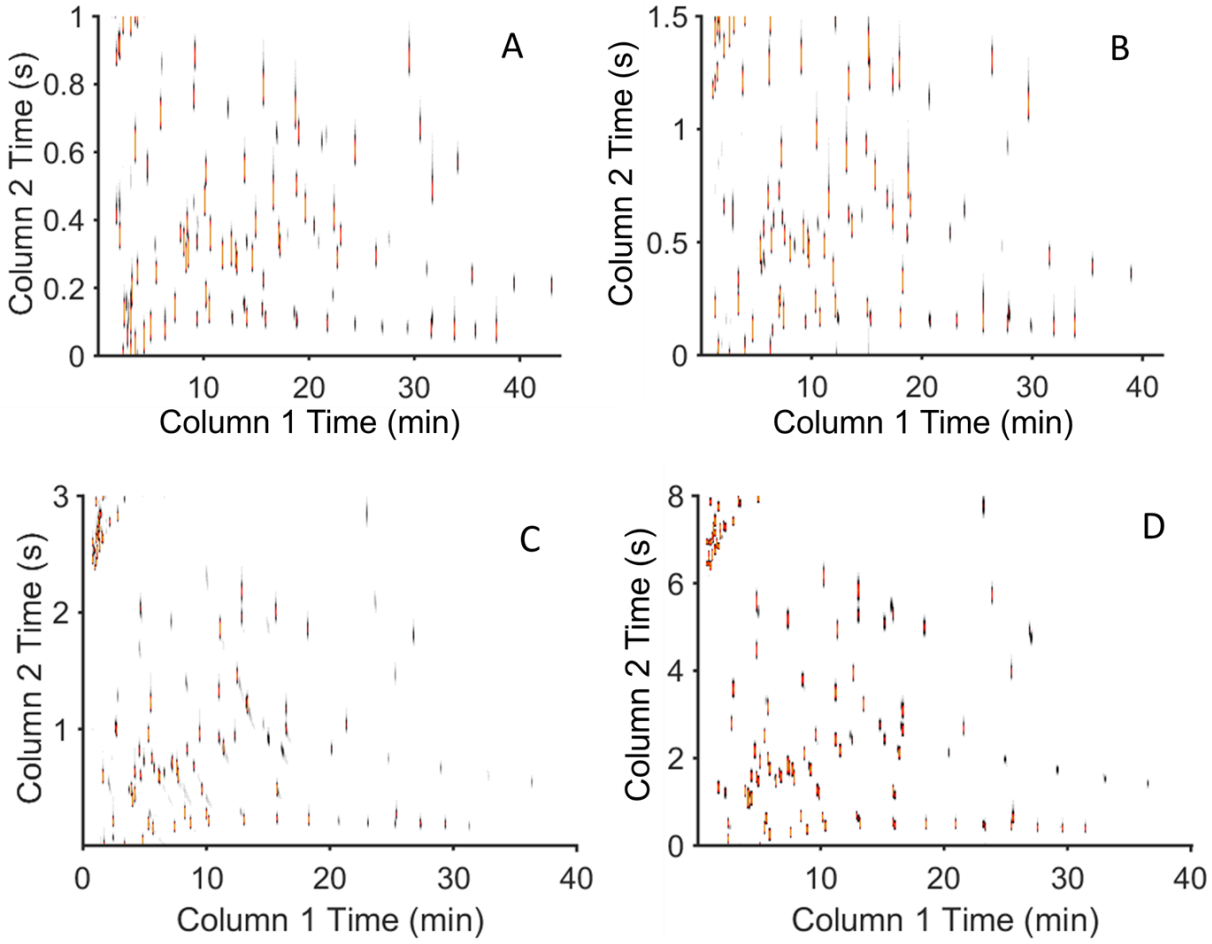

Figure 3.2 Representative reregistered TIC chromatograms for four column sets; the ^{1}D columns were all 20.0 m × 250 µm i.d., and the ^{2}D columns were all 2.0 m × 180 µm i.d.: (A) column set 1 (1.0 µm ^{1}D film, 0.2 µm ^{2}D film, β_R = 0.28). (B) column set 2 (0.5 µm ^{1}D film, 0.2 µm ^{2}D film, β_R = 0.56). (C) column set 3 (0.25 µm ^{1}D film, 0.2 µm ^{2}D film, β_R = 1.11). (D) column set 6 (0.25 µm ^{1}D film, 0.4 µm ^{2}D film, β_R = 2.22).

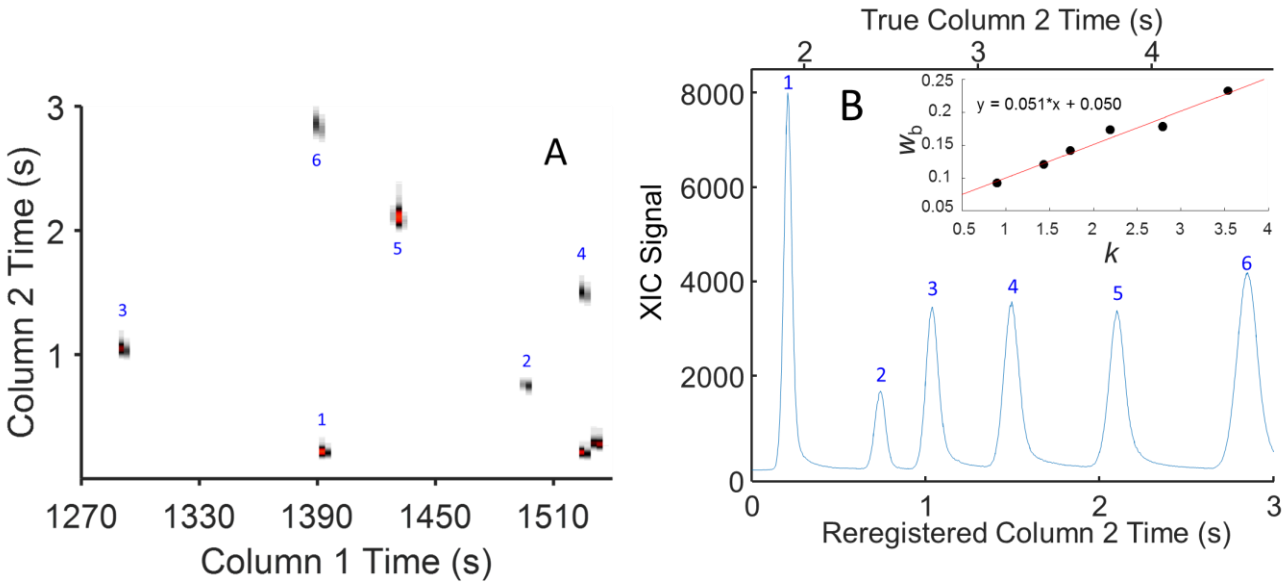

Figure 3.3 (A) A section of the reregistered separation using column set 3 (^{1}D: 20 m length, 250 μm i.d., 0.25 μm film; ^{2}D: 2 m, 180 μm i.d., 0.2 μm film; β_R = 1.11), from which six peaks were selected to demonstrate the relationship between ^2t_R and ^2w_b. The six analytes in order of increasing ^2t_R are: hexadecane, 1-tetradecanol, methyl dodecanoate, 2-pentadecanone, benzophenone, and diethyl phthalate. (B) The extracted ion chromatogram (XIC Signal) obtained by summing the ^{2}D column peak profiles from the most intense modulations of the six 2D peaks using a selective *m/z* for each analyte. The inset figure plots the ^2w_b versus ^2k for the six peaks with a linear best fit line (red).

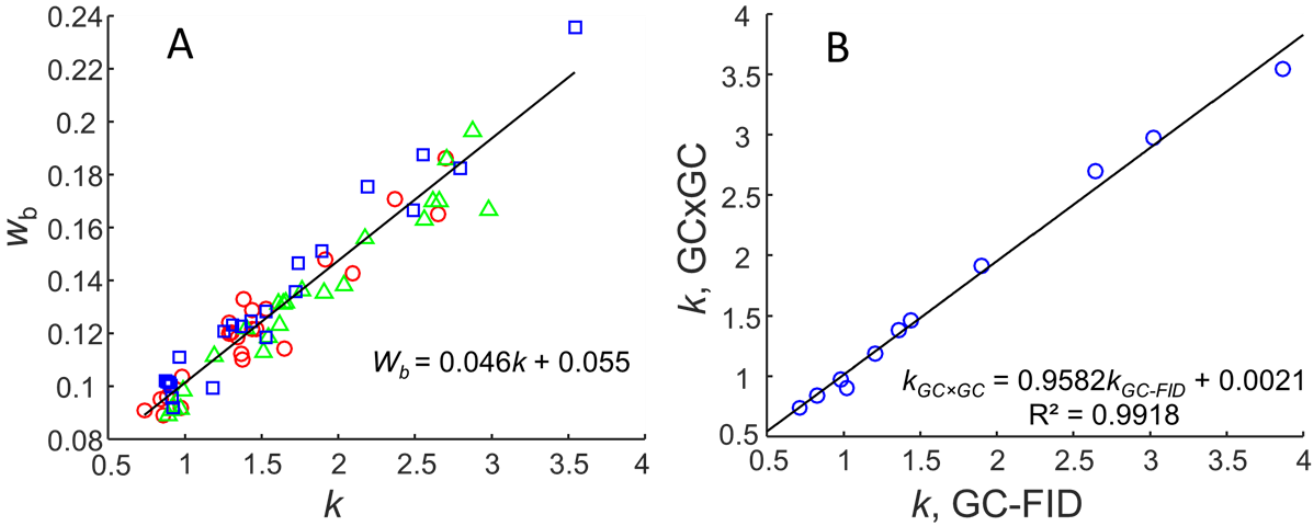

Figure 3.4 (A) Plot of ${}^2w_b$ versus ${}^2k$ for the 70 measured analytes, as separated on column set 3 (^{1}D: 20 m length, 250 μm i.d., 0.25 μm film; ^{2}D: 2 m, 180 μm i.d., 0.2 μm film; β_R = 1.11). The markers for each ${}^2w_b$ and ${}^2k$ pair have been assigned according to the represented analyte ${}^1t_R$; in order of ${}^1t_R$, analytes 1-23 (${}^1t_R$ = 4.6-8.7 min), 24-46 (${}^1t_R$ = 8.7-15.8 min), and 47-70 (${}^1t_R$ = 15.8-36.5 min) are represented by circles, triangles, and squares, respectively. A numbered list of the analytes with ${}^1t_R$, ${}^2t_R$, ${}^2k$, and ${}^2w_b$ is available in Table 3.3. (B) Plot of ${}^2k$ for each of the 11 analytes (Table S3.1 in Supplemental), as determined from the column set 3 GC×GC separation versus those determined from the isothermal GC–FID separations (Fig. S3.1 in Supplemental).

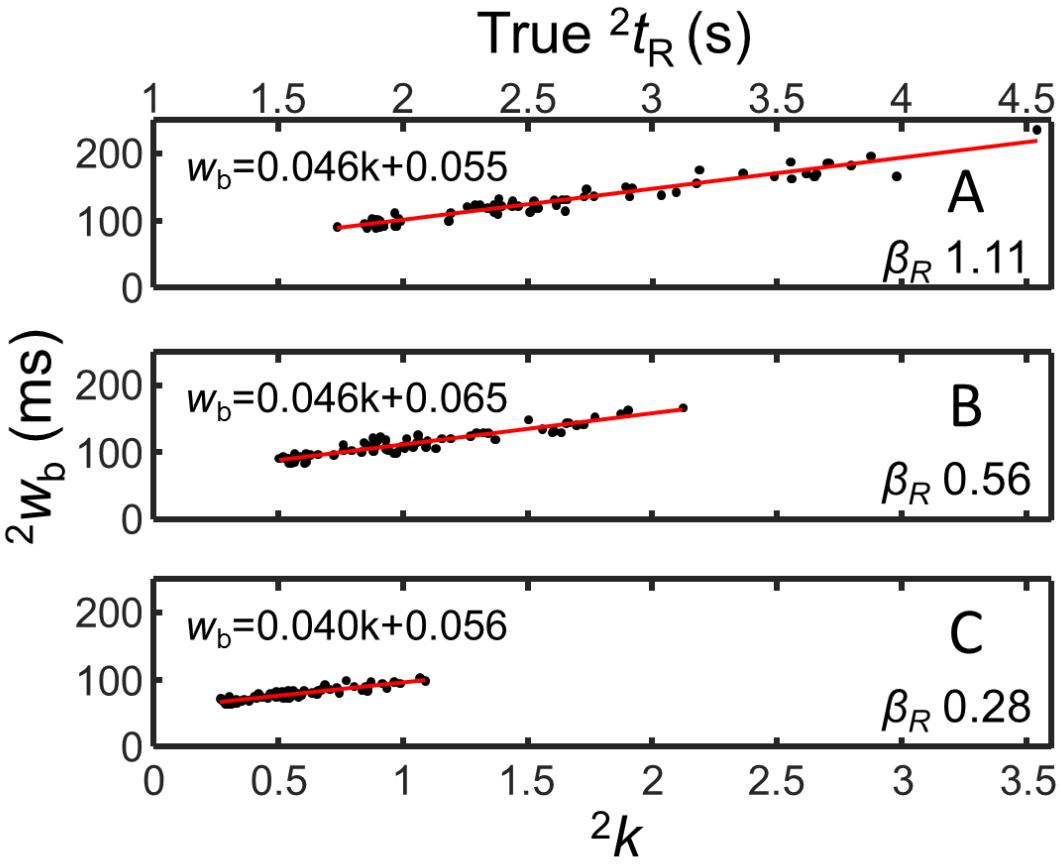

Figure 3.5 The ^2w_b versus ^2k plots for column sets 1, 2 and 3. The ^{1}D columns were all 20.0 m × 250 μm i.d., and the ^{2}D columns were all 2.0 m × 180 μm i.d.. (A) column set 3 (0.25 μm ^{1}D film, 0.2 μm ^{2}D film, β_R = 1.11). (B) column set 2 (0.5 μm ^{1}D film, 0.2 μm ^{2}D film, β_R = 0.56). (C) column set 1 (1.0 μm ^{1}D film, 0.2 μm ^{2}D film, β_R = 0.28).

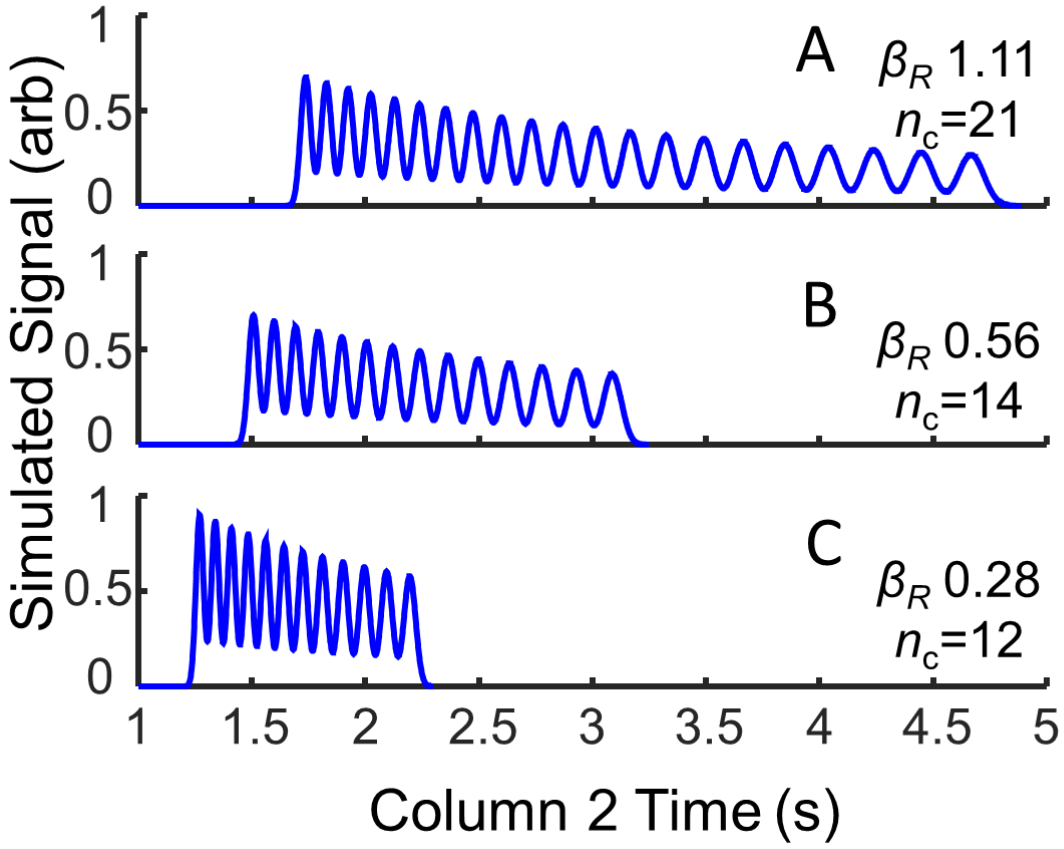

Figure 3.6 Simulated isothermal ^{2}D column chromatograms used to determine ^2n_c for column sets 1, 2 and 3. The ^{1}D columns were all 20.0 m × 250 μm i.d., and the ^{2}D columns were all 2.0 m × 180 μm i.d.. (A) column set 3 (0.25 μm ^{1}D film, 0.2 μm ^{2}D film, β_R = 1.11). (B) column set 2 (0.5 μm ^{1}D film, 0.2 μm ^{2}D film, β_R = 0.56). (C) column set 1 (1.0 μm ^{1}D film, 0.2 μm ^{2}D film, β_R = 0.28).

3.10 SUPPORTING INFORMATION

The measurements of ^{2}k from the GC×GC chromatograms for all six column sets were validated via van't Hoff plots by analyzing a subset of the 70 measured analytes by isothermal GC–FID. The 20.0 m × 250 µm × 0.5 µm Rtx®-5MS and 20.0 m × 250 µm × 1.0 µm Rtx®-5 ^{1}D columns, and the three additional 10.0 m columns described in Section 3.3.1, were individually installed into an Agilent 6890N gas chromatograph with a 7683 autosampler (Agilent Technologies, Palo Alto, CA), equipped with a flame ionization detector (FID). Eleven selected analytes (Table S3.1) from the 115 component mixture were each diluted to 1 ppth (w/w) in hexane and injected. The GC inlet was set to 250 °C and the split ratio was set to 200:1 with a column flow of 2.5 ml/min of hydrogen. The oven was held isothermal at the desired temperature until the given analyte eluted. For each analyte and column, the oven temperature was varied in 10 °C increments to collect temperature-dependent retention data. A 5 µl autosampler syringe (Hamilton, Reno, NV) was used to inject 1 µl of sample for the 250 µm i.d. columns and 0.1 µl for the 180 µm i.d. columns. Column dead times, t_0, were determined at each oven temperature by an injection of 1 µl of methane. Syringe rinses using acetone and hexane were performed after each analyte injection to prevent carry-over between analytes. The FID was operated at a temperature of 250 °C, with flows of 40.0 ml/min hydrogen, 450.0 ml/min air, and 45.0 ml/min nitrogen makeup gas. The detector reported data to the instrument at 200 Hz.

Data from the isothermal injections for the eleven selected analytes and methane, at the various oven temperatures, were exported as a .csv from MSD Chemstation D.03.00.611. (Agilent Technologies, Palo Alto, CA). Because the Chemstation software records time only to the nearest 0.001 min (60 ms), it was necessary to write in-house software in MATLAB 2015b to

find analyte peak apices in the raw data in order to calculate accurate retention times to the appropriate to the sampling period of the detector, 5 ms. The *k* was calculated for each analyte using the retention time of methane at the given column/temperature combination to determine the column dead time, t_0. Run temperatures were converted to Kelvin and used to generate van't Hoff plots, which plot the natural log of *k* (or alternatively, K_D) versus inverse Kelvin to determine thermodynamic parameters for the analyte. Linear best fits were applied to the van't Hoff plot for each analyte to allow for interpolation of *k* at the ^{2}D elution temperature, ^2T_e, for that analyte from each GC×GC column set, as summarized in Table S1. Within each group, all analytes eluted within ~1-2 ˚C of each other, and included co-elutions on the ^{1}D column that are resolved by the ^{2}D column. Each of these analytes was analyzed by isothermal GC–FID using a longer segment (10.0 m) of the ^{2}D columns, at a series of temperatures with 10 ˚C increments to determine *k* as a function of temperature, and to allow interpolation of the *k* for each analyte at the temperature corresponding to ^2T_e.

Figure S3.1 includes the comparisons of ^2k from the GC×GC experiment versus that of the isothermal GC–FID analysis for all six column sets. Overall, ^2k is predicted with reasonable accuracy between the two experimental platforms, with deviations from unity likely due to systematic bias in the determination of $^2t_0$ in the GC×GC experiment and/or flow deviations during its temperature programmed separation. Briefly, the bias in slope is consistent within column sets 1-3, and within column sets 4-6, which utilize ^{2}D columns with 0.2 μm and 0.4 μm films, respectively. This provides further evidence that the bias is likely due to a slight systematic error in the calculated $^2t_0$.

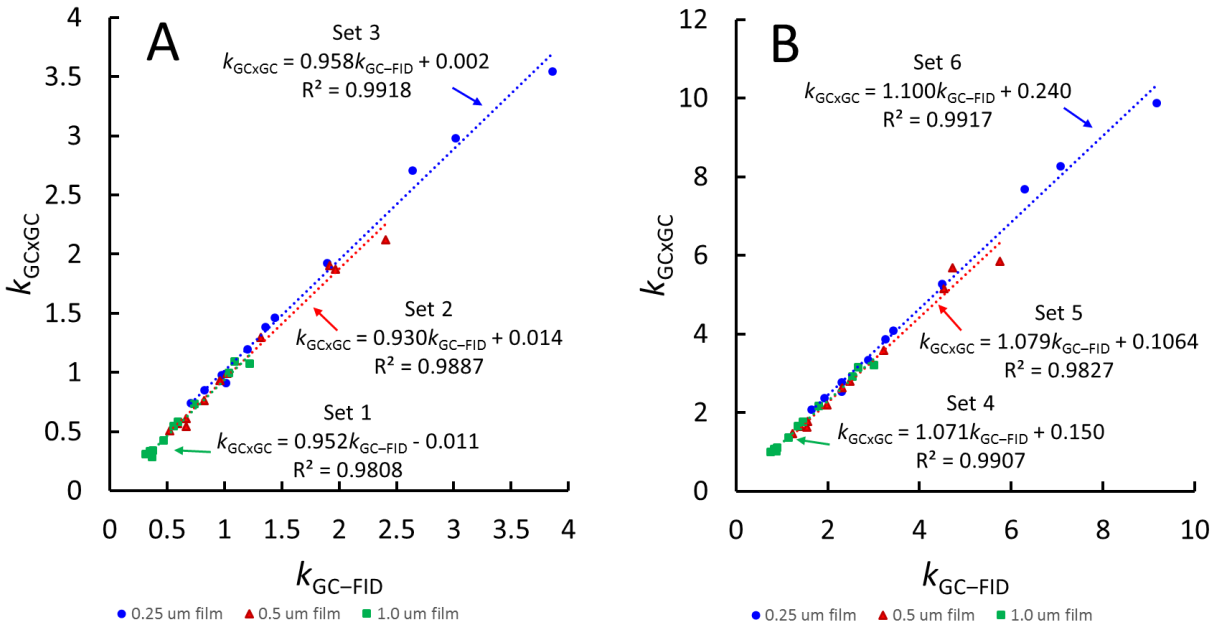

Figure S3.1 Plots of ^{2}k for each of the 11 analytes, as determined from the GC×GC separations for each column set versus those determined from the isothermal GC–FID separations.

Table S3.1 The calculated ^{1}D column elution temperatures (°C) for the 11 selected analytes, provided for each of the three ^{1}D column film thicknesses applied.

	0.25 µm film	0.5 µm film	1.0 µm film
2-heptanone	59.3	67.1	81.7
nonane	59.8	68.3	83.0
methyl hexanoate	63.1	72.0	87.1
2,6-dimethyloctane	63.8	73.1	88.5
bromobenzene	63.3	72.8	88.9
1-bromohexane	64.1	73.5	89.2
1-undecene	85.6	96.6	113.8
2-nonanone	86.1	96.8	114.0
adamantane	83.6	95.6	114.3
hexadecane	150.8	163.6	182.5
diethyl phthalate	150.8	163.7	183.4

Chapter 4. Method to Determine the True Modulation Ratio for Comprehensive Two-Dimensional Gas Chromatography[3]

4.1 INTRODUCTION

Comprehensive two-dimensional (2D) gas chromatography, GC × GC, is a powerful instrumental platform, well suited for the analysis of complex mixtures [1–7]. Successful separation and analysis requires that the instrumental platform translate the complex chemical samples into useful chemical information. In chromatography the amount of information contained in a given separation is inferred from, or correlated to, the peak capacity [8–14]. Herein, the peak capacity, n_c, represents the number of evenly resolved peaks at unit resolution, $R_s = 1$, that can fit into a given separation, where the n_c can be calculated as the separation time, t_{sep}, divided by the average chromatographic peak width-at-base, W_b. GC × GC is particularly well suited for the analysis of complex samples because of the increased peak capacity available in the 2D space, relative to a one-dimensional (1D) GC separation, where the ideal peak capacity of a GC × GC separation, $n_{c,2D}$, with independent separation mechanisms is commonly defined as the product of the peak capacities of the two separation dimensions, ${}^1n_c$ and ${}^2n_c$ [8,14–16]. This definition of $n_{c,2D}$ is commonly used to quantitatively compare the peak capacities of comprehensive two-dimensional separations. It should be noted, however that the ideal peak capacity $n_{c,2D}$ is not generally the experimentally observed peak capacity for a GC × GC separation since the two separation mechanisms are not strictly independent. To address the issue

[3]This Chapter has been reproduced from D.K. Pinkerton, B.A. Parsons, R.E. Synovec, Method to determine the true modulation ratio for comprehensive two-dimensional gas chromatography, J. Chromatogr. A. 1476 (2016) 114–123. doi:10.1016/j.chroma.2016.11.015.

of increasing sample complexity, a great deal of research has been performed to improve and optimize the peak capacity in GC × GC [8,9,11,13,14,17].

Indeed, peak capacity optimization is a constant challenge and area of focus in the chromatography community. Unless care is taken in the GC × GC instrument experimental design, the peak capacity ${}^{1}n_c$ of the first dimension, ^{1}D, is dependent on the peak capacity ${}^{2}n_c$ of the second dimension, ^{2}D, and a compromised $n_{c,2D}$ may result [11,15]. For example, use of a relatively long modulation period, P_M (5–8 s) to facilitate an increase in ${}^{2}n_c$, will necessarily reduce ${}^{1}n_c$. This reduction in ${}^{1}n_c$ is commonly believed to be due to a cognitive effort by the analyst to reduce the separation efficiency, N, on ^{1}D, by purposely broadening the ${}^{1}W_b$ peaks in order to maintain a suitable modulation ratio, M_R, which is ${}^{1}W_b$ divided by P_M, to achieve a comprehensive GC × GC separation [18,19]. However, use of a relatively long P_M can have the same influence on the effective N on ^{1}D simply due to undersampling, which results in an effective broadening of the ${}^{1}W_b$ [14,16,20–22]. When the ^{1}D separation has N optimized, and is then coupled with a suitably short P_M (1–3 s) it has been shown that an optimized $n_{c,2D}$ can be achieved [8,15,17,23,24]. In practice, many analysts tend to select a relatively long P_M for the purpose of providing a larger ${}^{2}n_c$; while a relatively short P_M will necessarily provide a smaller ${}^{2}n_c$, it is concurrent with optimizing $n_{c,2D}$ [8,15,17,23,24]. The trade-off here is selectivity (and peak capacity), balancing the chemical selectivity needs for the two separation dimensions to address different types of samples and sample complexities. Since the modulation process represents a sampling of the ^{1}D peak profile, the analyst must consider ${}^{1}W_b$ when selecting an appropriate P_M for the GC × GC experimental design to achieve the desired separation.

The importance of having an adequate number of samples across the ^{1}D peak width, ^1W_b, and the consequences of undersampling due to the modulation frequency cannot be overstated [14, 19, 16, 20, 22]. The primary consequence being that undersampling the ^{1}D peak profile results in significant ^{1}D peak broadening and translating to a proportional reduction in ^1n_c. Use of a long P_M (5–8 s) which is commonly practiced, necessarily means the ^{1}D separation is either performed so as to produce peaks wide enough to have enough modulations across the ^1W_b (20–24 s) to achieve a truly comprehensive GC × GC separation with a M_R of 2–4 [19], or the ^{1}D peaks are initially narrow (2–5 s) and are subsequently undersampled by the modulator, and a relatively uncomprehensive GC × GC separation with a M_R less than to much less than 2 is produced. While conventional thought is that the former is practiced, it is much more likely that the latter is what is indeed practiced. This issue is one of the key aspects of our current study.

To account for undersampling by the modulation process, Davis et al. introduced β, which depends on M_R, as a 2D peak capacity correction factor to determine the effective ^{1}D separation peak capacity, incorporating both undersampling by the modulator and statistical peak overlap theory [14,16]. However, calculation of β requires that the true M_R is known (not the "apparent" or effective M_R following modulation, referred to herein as M_R*), with the true M_R depending upon the premodulation ^1W_b, which is not measured in a typical GC × GC experiment, without the addition of an inline detector at the end of ^{1}D prior to modulation. What is readily measured however, is the post modulation ^{1}D peak width, which is the effective peak width, ^1W_b* (needed to calculate M_R*), that has been impacted by modulation induced band broadening [25].

Herein, we present a new method to determine the true modulation ratio, M_R, of a peak GC × GC, without the requirement for a detector at the end of the primary column to measure the premodulation peak width, ${}^1W_b$. The proposed method is described through an investigation of the consequences of the experimental design for the practice of GC × GC, related to modulator induced band broadening as a function of ${}^1W_b$ and the selected P_M. To do so, the effective ^{1}D peak width, ${}^1W_b$*, measured at the detector, is quantitatively related to the premodulation peak width, ${}^1W_b$. The effective modulation ratio, M_R*, is determined using the ${}^1W_b$*, and is then used to determine the modulation ratio, M_R; then, the true modulation ratio, M_R, which, along with P_M, can be used to calculate ${}^1W_b$. The method utilizes Gaussian curve fitting to initially model each modulated ^{2}D peaklet (each ^{2}D peak for a given modulated analyte), followed by fitting a Gaussian curve to the tops of the modeled ^{2}D peaklets for a given analyte in the unfolded GC × GC data to determine the effective ^{1}D peak profile [25]. The method of determining M_R from M_R* is defined and studied with simulated data, by defining the ^{1}D peak, simulating the modulation process, and comparing the resulting modulated ^{1}D peak width, ${}^1W_b$*, from the unfolded GC × GC data to the unmodulated ^{1}D peak width, ${}^1W_b$. Only in-phase and out-of-phase modulation of the GC × GC peaks were studied since these two extreme cases have been shown to bracket the full range sampling phase observed in GC × GC data [7, 22, 24, 25]. The results are then compiled in a look-up table that can be used to determine M_R from M_R*. The same Gaussian curve fitting approach to measure ${}^1W_b$* is then applied to experimentally collected GC × GC data spanning a range of commonly implemented conditions to provide a validation of the results from the simulation study. Percent loss in ${}^1n_c$ values are then calculated from the results of the simulation study, comparing the premodulation peak capacity, ${}^1n_c$, that would in principle be measured prior to modulation, to the effective ^{1}D peak capacity, ${}^1n_c$*, measured based upon

the modulated separation data at the detector. Finally, a discussion of the effective peak capacity calculated with β, ${}^1n_{c,Eff}$, in the context of M_R* and ${}^1n_c$* is provided in order to demonstrate how to correctly apply the β theory [14,16] to experimentally obtained GC × GC data with post modulation detection.

4.2 THEORY

The concentration profile of an ideal, symmetrical, ^{1}D chromatographic peak can be modeled mathematically by a Gaussian distribution

$$ {}^1C(t) = \frac{{}^1A}{{}^1\sigma\sqrt{2\pi}} e^{\left(\frac{-(t-{}^1t_R)^2}{2({}^1\sigma)^2}\right)} \quad (4.1) $$

where ${}^1C(t)$ is the ^{1}D analyte concentration profile as a function of time, t, ${}^1A$ is the total peak area, representative of the total analyte concentration, ${}^1\sigma$ is the standard deviation of the ^{1}D peak width-at-base, ${}^1W_b$, where ${}^1W_b$ is $\pm 2{}^1\sigma$, and ${}^1t_R$ is the ^{1}D retention time. Fig. 4.1(A) is a typical ^{1}D peak profile modeled by Eq. (4.1) where ${}^1A = \sqrt{2\pi}$, such that the signal at the peak apex is $1/{}^1\sigma$, ${}^1t_R$ = 49.5 s, and ${}^1\sigma$ = 1 s.

Modulation of the ^{1}D peak onto the ^{2}D column can be modeled from ${}^1C(t)$ by first calculating a series of modulation injection times, t_n. The modulation injection times can be calculated from the modulation period, P_M, and the start time of the modulator, t_0, defined by

$$ t_n = (nP_M) + t_0 \quad (4.2) $$

where n is the modulation number. The vertical dashed lines in Fig. 4.1(A) show the locations of each modulation within the time window with P_M = 1 s. In the time between modulation events, the eluent from the ^{1}D separation is trapped at the modulator such that the area of the ^{1}D peak injected onto the ^{2}D column during the modulation event at time t_n can be calculated from

$$^{2}A_n = \int_{t_{n-1}}^{t_n} {}^{1}C(t)\, dt \tag{4.3}$$

where ^{1}C(t) is the time-dependent ^{1}D concentration profile as in Eq. (4.1).

The series of modulated peak areas for a given analyte, $^{2}A_{\mathrm{n}}$, can then be used to produce a series of ^{2}D Gaussian peaks, referred to herein as peaklets, with

$$^{2}C_n(t) = \frac{^{2}A_n}{^{2}\sigma\sqrt{2\pi}} e^{\left(\frac{-(t-{}^{2}t_R)^2}{2(^{2}\sigma)^2}\right)} \tag{4.4}$$

where $^{2}C_{\mathrm{n}}(t)$ is the time-dependent concentration profile of each ^{2}D peaklet following modulation at time t_{n}, $^{2}\sigma$ is the standard deviation of the ^{2}D peak width $^{2}W_{\mathrm{b}}$, with a retention time $^{2}t_{\mathrm{R}}$.

The number of ^{2}D peaklets observed for a given ^{1}D peak depends on $^{1}W_{\mathrm{b}}$, P_{M}, $^{2}W_{\mathrm{b}}$, analyte concentration, detector data acquisition frequency, noise, and sampling phase, and is not a quantitatively useful metric because of the variability in sampling phase within a GC × GC experiment. The term modulation ratio, M_{R}, of a 2D chromatographic peak is defined as

$$M_{\mathrm{R}} = \frac{^{1}W_{\mathrm{b}}}{P_{\mathrm{M}}} \tag{4.5}$$

where $^{1}W_{\mathrm{b}}$ is the analyte ^{1}D peak width-at-the-base, and P_{M} is the modulation period [18]. In practice, determination of M_{R} is not straight forward for GC × GC since it requires knowledge of the analyte ^{1}D peak width, $^{1}W_{\mathrm{b}}$, prior to modulation. Generally, $^{1}W_{\mathrm{b}}$ is not measured directly in GC × GC, since the ^{1}D peak signal is generally not measured prior to modulation.

As can be seen in Figs. 4.1(B-D), the series of peak maxima and retention times, $t_n + {}^{2}t_{\mathrm{R}}$, for modulated GC × GC peaklets can be used to model the effective ^{1}D peak. The effective ^{1}D peak width $^{1}W_{\mathrm{b}}{}^{*}$ is used to define the experimentally observed, effective modulation ratio

$$M_{\mathrm{R}}{}^{*} = \frac{^{1}W_{\mathrm{b}}{}^{*}}{P_{\mathrm{M}}} \tag{4.6}$$

In general, M_R and M_R^* are not equal because ${}^1W_b^* \geq {}^1W_b$. However, when M_R is sufficiently large, i.e. when ${}^1W_b$ is large relative to P_M, the effective modulation ratio, M_R^*, converges to M_R. In the common practice of GC × GC, M_R will be less than, to much less than, M_R^*. This discrepancy between the M_R and M_R^*, if not accounted for, will lead to an inaccurate determination of the effective ^{1}D peak capacity [14,16].

The peak capacity in the absence of undersampling of the ^{1}D separation in GC × GC is defined as

$$ {}^1n_c = \frac{t_{sep}}{{}^1W_b} \tag{4.7} $$

where t_{sep} is the total separation time and ${}^1W_b$ is taken as the average ^{1}D peak width at base. Likewise, the peak capacity of the ^{2}D separation in GC × GC is defined as

$$ {}^2n_c = \frac{P_M}{{}^2W_b} \tag{4.8} $$

Peak capacity for the 2D GC × GC separation in the absence of undersampling, $n_{c,2D}$, is the product of ^{1}D and ^{2}D peak capacities, ${}^1n_c$ and ${}^2n_c$ respectively, from Eqs. (4.7) and (4.8),

$$ n_{c,2D} = {}^1n_c * {}^2n_c \tag{4.9} $$

Since modulator induced band broadening only impacts the ^{1}D peak width in a GC × GC separation, we limit the scope of this peak capacity discussion to the impact of the modulation process on ${}^1n_c$.

Likewise, the effective ^{1}D width of the modulated GC × GC peak with any finite degree of undersampling, ${}^1W_b$*, can be used to determine the effective ^{1}D peak capacity, given by

$$ {}^1n_c{}^* = \frac{t_{sep}}{{}^1W_b{}^*} \tag{4.10} $$

where ${}^1n_c$* is not the ideal peak capacity, but is the effective peak capacity taking into account that ${}^1W_b$* has been artificially broadened by the modulation process. ${}^1n_c$* can also be used to calculate the effective 2D peak capacity, $n_{c,2D}$* given by

$$n_{c,2D}{}^* = {}^1n_c{}^* * {}^2n_c \tag{4.11}$$

Making use of the relationship between M_R and M_R* via Eqs. (4.5) and (4.6), the percent loss in ${}^1n_c$ strictly due to modulation is defined as the difference between the ideal ^{1}D peak capacity, ${}^1n_c$, and the effective ^{1}D peak capacity, ${}^1n_c$*, normalized by dividing by ${}^1n_c$,

$$\%\ Loss\ in\ {}^1n_c = 100\ \% * \frac{{}^1n_c - {}^1n_c{}^*}{{}^1n_c} = 100\ \% * \left(1 - \frac{{}^1W_b}{{}^1W_b{}^*}\right) \tag{4.12}$$

An alternative calculation for effective peak capacity has been reported that takes into account the combined effects of M_R and statistical peak overlap theory [14,16], given by

$${}^1n_{c,Eff} = \frac{{}^1n_c}{\beta} \tag{4.13}$$

where ${}^1n_c$ is the ^{1}D peak capacity as in Eq. (4.7), and β is a correction term accounting for the combined effects of undersampling (i.e. due to modulation) and statistical peak overlap theory,

$$\beta = \sqrt{1 + 3.4\left(\frac{1}{M_R}\right)^2} \tag{4.14}$$

While β can only be applied in the context of 2D separations and is, in general, applied to $n_{c,2D}$, β is a function of, and a correction for, the ^{1}D peak width ${}^1W_b$, so for the purposes of this study the consequences of β are discussed as in Eq. (4.13). It should also be noted that because β accounts for statistical peak overlap theory it is independent of peak sampling phase. This definition for effective peak capacity is problematic for implementation to GC × GC data, because both ${}^1n_{c,Eff}$ and β are calculated with the unmodulated ^{1}D peak width ${}^1W_b$, which is generally not known and cannot be measured directly without an in-line detector before the

modulator in a GC × GC experiment. In the absence of an in-line detector, ${}^1W_b$* and M_R* should not be used in calculating β and ${}^1n_{c,Eff}$ since ${}^1W_b$* and β both account for modulator induced band broadening in the ^{1}D peak. If ${}^1W_b$* and M_R* are erroneously used to calculate β and ${}^1n_{c,Eff}$, the percentage error lost in ${}^1n_{c,Eff}$ can be determined by

$$\% \, Error \; in \; {}^1n_{c,Eff} = 100 \, \% * \left(1 - \frac{\frac{{}^1n_c{}^*}{\beta^*}}{\frac{{}^1n_c}{\beta}} \right) \tag{4.15}$$

where β* is defined as in Eq. (4.14), however is calculated with M_R* instead of M_R. Substituting in Eqs. (4.7), (4.10), and (4.14), Eq. (4.15) can be rearranged to give

$$\% \, Error \; in \; {}^1n_{c,Eff} = 100 \, \% * \left(1 - \left(\frac{{M_R}^2+3.4}{{M_R}^{*2}+3.4} \right)^{\frac{1}{2}} \right) \tag{4.16}$$

Clearly, the percent error in ${}^1n_{c,Eff}$ depends only on M_R and M_R*. However, the dependence of M_R* on sampling phase, as a function of M_R, means that sampling phase must also be considered for percent error in ${}^1n_{c,Eff}$. We propose a method for determining M_R from M_R* for both in-phase and out-of-phase modulation sampling, with intermediate phases being bracketed by the two extremes. M_R can be determined from M_R* by defining a ratio of M_R to M_R* as a function of M_R* and measuring M_R* in the modulated data. Furthermore, by mapping M_R* to M_R, the analyst can accurately and confidently determine ${}^1n_{c,Eff}$ with Eq. (4.13) and (4.14).

4.3 EXPERIMENTAL

4.3.1 *Simulated GC × GC Data*

All data simulations and analyses were performed in Matlab R2012b (The Mathworks, Inc., Natick, MA, U.S.A.). GC × GC chromatograms were simulated at a 500 Hz collection rate.

All simulations were generated in the absence of noise. The first step in the GC × GC–TOFMS data simulations was to define a ^{1}D Gaussian analyte peak profile with specified width, ^1W_b, defined as the peak width at $\pm 2^1\sigma$, in units of time, retention time, ^1t_R, and analyte concentration, ^1A, as in Eq. (4.1). For all simulations presented, ^1W_b = 4 s ($^1\sigma$ = 1 s), ^1t_R = 49.5 s, and $^1A = \sqrt{2\pi}$, other values for ^1W_b, ^1t_R, and ^1A were also simulated, not shown for brevity. A series of modulation injection times t_n for the selected modulation period, P_M, and modulator start time, t_0, as in Eq. (4.2), were then selected to achieve the desired sampling phase, in-phase or out-of-phase. In-phase sampling is equivalent to when ^1t_R of a pre-modulation ^{1}D peak occurs halfway between two modulation events, $^1t_R = t_n + ½*P_M$ for some n as in Eq. (4.2), out-of-phase sampling occurs when the ^1t_R coincides with the modulation event, $^1t_R = t_n$ for some n as in Eq. (4.2), such that there were two identical modulated ^{2}D peaks of maximal intensity. With ^1t_R held constant between simulations, sampling phase was controlled by changing the start time of the modulator, more specifically the time of the first modulation event, t_0. Simulations were generated with P_M = 0.5–15 s in 0.1 s increments (M_R = 8–0.26) sampled both in-phase and out-of-phase. Modulated peaklets were generated by integrating the simulated ^{1}D peak profile between adjacent modulation times, t_{n-1} and t_n, and ^{2}D Gaussian peaklets were simulated in the subsequent modulation according to Eq. (4.4), with ^{2}D retention time ^2t_R = 0.25 s, ^{2}D peak standard deviation $^2\sigma$ = 50 ms, and ^2A_n as defined by Eq. (4.3).

4.3.2 *Instrumental Parameters*

The GC × GC–TOFMS instrumental platform consisted of an Agilent 6890N gas chromatograph equipped with an Agilent 7683 autoinjector (Agilent Technologies, Palo Alto, CA) coupled with a LECO Pegasus III TOFMS, and equipped with a 4D thermal modulator

upgrade (LECO, St. Joseph, MI). A single 41.3 m x 180 μm i.d. x 0.4 μm Rxi-5MS film column (Restek, Bellefonte, PA) was installed on the instrument, threaded through the modulator block and the secondary oven directly to the detector. The instrumental platform was used both with the modulator turned off, in the one-dimensional (1D) GC – TOFMS mode, to collect and measure true ^{1}D peak widths, $^{1}W_{b}$, and with the modulator turned on, in the 2D GC × GC–TOFMS mode. In the GC × GC–TOFMS mode, the ^{1}D column was 41.0 m, the ^{2}D column in the secondary oven was 0.1 m, with 0.2 m as the transfer line to the detector, all a part of the same 41.3 m x 180 μm i.d. x 0.4 μm Rxi-5MS film column (Restek, Bellefonte, PA). A short ^{2}D column with the same stationary phase as the ^{1}D column was applied to minimize retention on ^{2}D and allow for a wide range of P_{M} and M_{R} to be readily studied, while facilitating the measurement of the $^{1}W_{b}$ when the modulator was not in operation.

Two sets of ^{1}D separation conditions were used, along with a range of P_{M} in order to produce peaks with a range of $^{1}W_{b}$ (and $^{1}W_{b}$*), and M_{R} (and M_{R}*). Helium was used as the carrier gas for both sets of temperature/flow conditions. For the first set of conditions the oven was held at 125 °C for 0.5 min and then increased at 5.5 °C/min to 165 °C, at a constant flow rate of 1.44 ml/min. For the second set of conditions the oven was held at 130 °C for 0.5 min and then increased at 2.0 °C/min to 170 °C, with a constant flow rate of 0.5 ml/min. A 1 μl injection of a three component test mixture of decane, undecane, and dodecane in dichloromethane was made in split mode with a split ratio of 200:1. Dichloromethane was used as the solvent because it has a minimal number of common fragment ions with the compounds in the n-alkane mixture. The transfer line temperature was set to 235 °C and the inlet and ion source were set to 225 °C. Mass channels, *m/z* 41–180, were collected at 500 spectra/s. Both sets of conditions were run in the GC – TOFMS and GC × GC–TOFMS modes. When the instrument was run with the

modulator on, GC × GC–TOFMS, the modulator and secondary oven were kept 20 °C higher than the primary oven to minimize additional retention on column after modulation. Table 4.1 summarizes the modulation periods and approximate modulation ratios studied herein. Each set of sample conditions were run in quintuplicate.

4.3.3 *Peak Modeling*

All peak modeling for simulated and experimental data was performed using the Matlab R2012b Curve Fitting Toolbox (The Mathworks, Inc., Natick, MA, U.S.A.). For experimentally collected data with mass spectral detection, a single selective *m/z*, 43, was used for analysis in order to reduce the level of noise in the data and to minimize background interference from the solvent. Prior to analysis all GC–TOFMS and GC × GC–TOFMS chromatograms were baseline corrected. For all experimental 1D separations, GC–TOFMS mode, peaks were measured by directly fitting individual 1D peaks with Gaussian curves based on Eq (4.1). For 2D separations, GC × GC–TOFMS mode, Gaussian curve fitting was initially applied to modulated peaklets, ^{2}D peaks, to extract peaklet maxima and retention times. Although the exact maxima and retention times were known *a priori* for the simulated data, in general with experimental data this is not the case, so for consistency the same curve fitting procedure was applied for both simulated and experimental data. Using the Matlab R2012b Curve Fitting Toolbox a modulated peak with *n* detectable modulations can be modeled simultaneously using an *n*-term Gaussian function, a series of *n* independent Gaussian functions summed together

$$C_{mod}(t) = a_1 * e^{\left(-\left(\frac{(t-b_1)}{c_1}\right)^2\right)} + \cdots + a_n * e^{\left(-\left(\frac{(t-b_n)}{c_n}\right)^2\right)} \qquad n \leq 8 \qquad (4.17)$$

where $C_{mod}(t)$ is the time dependent concentration profile of the series of modulated peaklets, a_n is the peak maximum for peaklet *n*, b_n is the retention time of peaklet *n*, and c_n is $^{2}\sigma_n$ times $2^{1/2}$.

The series of a_n and b_n terms are then vectorized and forwarded back to the curve fitting toolbox as peak maxima and retention time pairs, used to measure the effective ^{1}D peak profile, ^1W_b*, as shown in Fig. 4.1, overlaid with the simulated ^{2}D peaklets and unmodulated ^{1}D peak.

4.4 Results and Discussion

To investigate the effect of modulating a ^{1}D peak onto the ^{2}D separation dimension on the resulting ^{1}D peak width, ^1W_b*, we begin by modeling the modulation process with simulated GC × GC data. The unmodulated ^{1}D peak profile (^1W_b = 4 s) used in the modulation simulations is provided in Fig. 4.1(A). Figs. 4.1(B-D) are simulated 2D GC × GC chromatograms, with P_M = 1, 2, and 4 s and M_R = 4, 2, and 1 respectively, for the unmodulated ^{1}D peak shown in Fig. 4.1(A) sampled in-phase; out-of-phase sampling was also simulated, but not shown for brevity. In Figs. 4.1(B-D), the leftmost Gaussian curve represents the unmodulated ^{1}D peak profile prior to modulation, Eq. (4.1), with the modulated ^{2}D peaklets offset to the right, Eqs. (4.2-4.4), and the corresponding Gaussian curve also offset to the right representing the measured ^{1}D peak profile based on a fit to the peak maxima of the modulated ^{2}D peaklets, by applying Eq. (4.1), at each retention time, indicated by the dots. Because the modulator samples the ^{1}D peak profile with regular frequency defined by P_M, the modulated peaklets represent a distribution that is directly related to the original ^{1}D peak. In this sense, the modulation process is analogous to the mathematical principle of Riemann sums, a discrete sampling of a continuous distribution that is not necessarily equal to, but is representative of the original distribution. And, as with Riemann sums, as the time interval between samples converges to zero, the representative distribution becomes equal to the original distribution, i.e. the Riemann sum converges to the integral of the function. The retention time and peak height can be used to model the distribution of the ^{1}D peak

profile because successive ^{2}D peaklets for a given analyte will generally have the same width ^2W_b on the ^{2}D separation, meaning that the area of the of the peaklet is directly proportional to its height. However, successfully fitting a Gaussian curve to the ^{1}D profile in this way requires precise determination of the ^{2}D peak height and retention time. Furthermore, Gaussian curve fitting requires at least three points to fit the curve to; namely, there must be at least one modulated peaklet for each degree of freedom (^1A, $^1\sigma$, and ^1t_R) in the Gaussian function, Eq. (4.1).

The results of the Gaussian peak fitting for the simulated GC × GC data are shown in Fig. 4.2. Figure 4.2(A) is the modulator induced broadening ratio (^1W_b* / ^1W_b) as a function of true modulation ratio, M_R, for simulated GC × GC data of a single peak with ^1W_b = 4 s, sampled both in-phase (solid line) and out-of-phase (dashed line), with P_M ranging from 0.5 s to 15 s. All other chromatographic parameters were held constant between simulations, including total peak area, ^1t_R, ^2t_R, ^1W_b, and ^2W_b. Indeed, Fig. 2 shows that ^1W_b* / ^1W_b increases as M_R decreases. For $M_R \geq 2$ there is no discernable difference in ^1W_b* / ^1W_b for in-phase and out-of-phase sampling. However, when $M_R < 2$ sampling phase becomes a progressively more significant contributor to the magnitude of ^1W_b* / ^1W_b. Notably, $M_R < 1$ results in significant band broadening, ^1W_b* / ^1W_b ≥ 1.3, for both in-phase and out-of-phase sampling and for $M_R < 0.75$ the broadening for out of phase sampling increases almost vertically. This steep broadening for out-of-phase sampling occurs because the number of observable peaklets is effectively less than the three required to model a Gaussian peak, that is to say for each of the two major peaklets $^2A_1 = {}^2A_2 = ½\ {}^1A$.

Figure 4.2(B) depicts the same modulator induced broadening results as in Fig. 4.2(A), plotted versus the effective modulation ratio, M_R*, since M_R* is what an analyst would measure in a typical experiment, and, in general, M_R is not directly measurable in a GC × GC. This view

of the data allows the analyst to approximate the degree to which the measured ${}^1W_b$* has been broadened by the modulation process based on the experimentally measured M_R* and could be extended to approximate the unmodulated width of the ^{1}D peak ${}^1W_b$. Figure 4.2(B) indicates that, for out-of-phase sampling, $M_R^* \geq 0.8$ for all values of M_R, and that even when the ^{1}D is severely undermodulated, ${}^1W_b << P_M$, the analyst will measure ${}^1W_b^* = P_M$, resulting in a significant loss of information, quantifiable as a loss in the ^{1}D peak capacity, ${}^1n_c$.

Experimental data are now presented for evaluation of the simulations modeled above. The three component test mixture of decane, undecane, and dodecane was run under two sets of separation conditions in both the 1D GC–TOFMS and 2D GC × GC–TOFMS modes in order to produce a range of ^{1}D peak widths. In the 2D mode, separations were run with a range of modulation periods, P_M, to study a range of modulation ratios. Table 4.1 provides a complete summary of the measured unmodulated 1D peak widths, ${}^1W_b$, with standard deviations for the five replicate injections, as well as the P_M studied and the resulting modulation ratios. The RSDs for the unmodulated 1D peaks widths were all ≤ 1.0 %. Figure 4.3 shows a series of representative 2D GC × GC–TOFMS peaks, selected to cover a range of typical experimentally observed ${}^1W_b$, P_M, M_R, and M_R*, overlaid with the corresponding unmodulated 1D peak. The unmodulated 1D peak profiles in Fig. 4.3 have been rescaled by a factor of five for visualization purposes only, it should be noted that rescaling in this way does not affect the width of the observed peak. Variations for peak sampling phase for a given analyte under identical separation conditions may occur due to minor variations in ^{1}D retention time, ${}^1t_R$, or in the modulator start time, t_0. As a result, the exact locations of the modulation events, indicated by vertical dashed lines, in Fig. 4.3 are approximated to produce the observed sampling phase. There is some lack of precision in trying to measure the modulated ^{2}D retention times and peak maxima directly

from raw peak data because the narrow ^{2}D peak widths, 50 ms to 60 ms, mean that ^{2}D peaks, sampled by the detector, are also subject to a variety of sampling phases. However, the lack of precision in the retention time and peak maxima is easily overcome by Gaussian curve fitting the peaklets rather than measuring directly.

Figure 4.4 is an overlay of the curve fitting peak measurement results for the experimentally collected GC × GC–TOFMS data with the simulated GC × GC data, shown previously in Fig. 4.2. Overall, Figs. 4.4(A) and (B) indicate that the experimentally collected data are consistent with the results from fitting the theoretically simulated data. All simulations used to prepare Figs. 4.2 and 4.4 were generated in the absence of noise, which allowed for rigorous investigation of the theoretical effects of the modulator on the ^{1}D peak profile. Simulations were also defined to be sampled either perfectly in-phase, or perfectly out-of-phase, as these are the two extreme cases and have been shown to bracket the range of sampling phases [7, 22, 24, 25]. However, experimental data is necessarily collected with some noise, and the sampling phase is determined by the combination of ^1t_R, t_0, and P_M, with in-phase or out-of-phase being only a subset of all sampling phase possibilities. The presence of noise and sampling phase uncertainty results in a narrow distribution of broadening ratios for replicates of each set of conditions. These distributions are observed in Fig. 4.4(A) as a series of five points arranged vertically, often overlapping, since the conditions were all run in quintuplicate and the ^1W_b RSD $\leq 1.0\%$. Despite the close agreement between the simulated and experimental results, there are some deviations observed in the experimental data set. In the presence of noise, Gaussian fitting severely undermodulated, $M_R < 1$, peaks can be problematic if the Gaussian peak modeling requirement of three detectable peaks to be used in fitting is not met.

It is clear from Fig. 4.4 that both simulated and experimental GC × GC chromatograms show significant modulator induced band broadening when the ^{1}D peak is not sufficiently sampled by the modulator. Table 4.1 shows that ^1W_b does not vary significantly between samples for a given instrumental configuration run with the same separation conditions, so the analyst must select an appropriate P_M to adequately sample the eluting ^{1}D peaks. Maximizing the amount of chemical information in a separation means the analyst should seek to keep ^1W_b small, such as ^1W_b = 4–6 s (or even smaller); however, if a large modulation period is selected, P_M = 5–6 s, in an attempt to increase ^2n_c, the result is a low modulation ratio, M_R = 0.67–1.2, with broadening ratio, ^1W_b* / ^1W_b, ranging from 1.25–2.0. Effectively, the use of a long P_M undoes any increase in ^1n_c that would be gained from optimizing the ^{1}D separation, and the nonlinear relationship between ^2n_c and P_M, due to the pseudo-isothermal separation conditions on ^{2}D, results in only a modest increase in ^2n_c [15].

Making use of the relationship between M_R and M_R* defined in Eqs. (4.5) and (4.6) and determined in Figs. 4.2 and 4.4, Fig. 4.5 shows the percent loss in ^1n_c due strictly to undersampling via implementation of an inappropriately long P_M as defined in Eq. (4.12). Figure 4.5 quantifies the consequences of an inappropriate modulation period selection for a given ^{1}D peak, expressed as the loss of peak capacity. It is worth noting that the percent losses in ^1n_c shown in Fig. 4.5 are nominally consistent with the loss predicted by Davis et al. [14]. Figure 4.5 along with Table 4.2 also allows the analyst to use the effective M_R* (which is readily determined directly from the typical GC × GC experimental data), to determine the true modulation ratio, M_R, that would otherwise require the width of the original ^{1}D peak prior to modulation, ^1W_b (which is not typically measured in the GC × GC experiment), through the plot of M_R /M_R* as a function of M_R*. For example, a separation run with a long modulation period,

P_M = 6 s, might produce modulated peaks such that ${}^1W_b^*$ = 6 s, corresponding to M_R^* = 1. The effective modulation ratio M_R^* can then be used, along with Fig. 4.5 and Table 4.2, to determine that this corresponds to a true $M_R \leq 0.66\ M_R^*$, i.e. $M_R \leq 0.66$ and ${}^1W_b$ = 3.96 s, representing a 34% loss in ${}^1n_c$. The same peak, however, sampled with P_M = 1 s has M_R^* = 3.8 and a corresponding M_R = 3.6, meaning only a 5% loss in ${}^1n_c$ due to modulator induced band broadening. Clearly, appropriate P_M selection must be considered as a part of ^{1}D separation optimization, determined by the ${}^1W_b$ of the eluting ^{1}D peaks, since the effective ^{1}D peak capacity, ${}^1n_c^*$, is calculated based on the modeled ^{1}D peak width, which has already been broadened by the modulation process.

It may be of interest to analysts to apply the β correction term in order to obtain a correction for undersampling of the 1D separation, in which the correction incorporates undersampling and statistical peak overlap theory, via Eq. (4.13). The definition for effective peak capacity ${}^1n_{c,Eff}$ in Eq. (4.13) is problematic for implementation to GC × GC data since both ${}^1n_{c,Eff}$ and β should be calculated with the unmodulated ^{1}D peak width ${}^1W_b$, which is generally not known and cannot be measured directly without an in-line detector before the modulator. In the absence of an in-line detector, ${}^1W_b^*$ and M_R^* should not be used in calculating β and ${}^1n_{c,Eff}$ since ${}^1W_b^*$ and β both account for modulator induced band broadening in the ^{1}D peak. If ${}^1W_b^*$ and M_R^* are erroneously used to calculate β and ${}^1n_{c,Eff}$, the percentage error lost in ${}^1n_{c,Eff}$ can be determined as defined in Eq. (4.16). Figure 4.6 is a plot of Eq. (4.16) as a function of M_R for both in-phase and out-of-phase sampling, simulated with the results from Fig. 4.2. Unlike the modulator induced broadening ratio, the percent error in ${}^1n_{c,Eff}$ does not decrease monotonically with M_R, but the largest observed errors, 11 % and 8 % for in-phase and out-of-phase respectively, occur when the ^{1}D peaks are severely undermodulated, M_R = 1.25 and 1.66. Indeed,

if β is to be used to calculate ${}^1n_{c,Eff,}$ the premodulation peak width ${}^1W_b$ must be known prior to calculation. Although ${}^1W_b$ is not generally measured, the effective peak width ${}^1W_b$*, which is readily measured, coupled with the modulation period, P_M, and Fig. 4.5 or Table 4.2, can be used to determine ${}^1W_{b,}$ the width of the unmodulated ^{1}D peaks through the relationship between M_R and M_R*. For demonstration, we now consider a previously published separation of an organic extraction of microalgae based bio-oils run with P_M = 6 s [28], such that the approximate M_R* = 1.25 (${}^1W_b$* = 7.5 s) with an approximate average ${}^2W_b$ = 500 ms, resulting in ${}^1n_c$* = 640, ${}^2n_c$ = 12, and $n_{c,2D}$* of approximately 7,680 for an 80 min separation. Making use of Table 4.2 and Fig. 4.5, for an M_R* = 1.25, we determine that M_R / M_R* = 0.7. Therefore M_R = 0.88 and ${}^1W_b$ = 5.28 s which suggests an approximate ${}^1n_c$ = 900 in 80 min prior to the ^{1}D broadening due to undersampling with β = 2.32 via Eq. (4.14), so the ${}^1n_{c,Eff}$ = 388 based upon Eq. (4.13). Clearly, under these separation conditions, the analyst is accepting a significant level of broadening on the ^{1}D separation with the concurrent loss of ^{1}D peak capacity (30% loss according to Table 4.2), for the purpose of applying the P_M = 6 s which provided a relatively modest ^{2}D peak capacity.

4.5 CONCLUSIONS

We present a new method to determine the true modulation ratio, M_R, of a peak in GC × GC, without the requirement for a detector at the end of the primary column as a function of the effective modulation ratio, M_R*, calculated from the effective peak width, ${}^1W_b$*, and the modulation period, P_M. We demonstrate the consequences of modulation period, P_M, selection for a GC × GC separation as it relates to modulator induced band broadening, by mapping from the effective modulation ratio, M_R*, of a modulated ^{1}D peak to the true modulation ratio, M_R. We demonstrated a method that uses Gaussian curve fitting to model the effective ^{1}D peak width,

^1W_b*, by initially modeling each modulated ^{2}D peaklet in the unfolded GC × GC data to accurately determine the maxima of the ^{2}D peaklets, followed by Gaussian curve fitting to the maxima to determine the effective ^{1}D peak profile. The relationship between M_R and M_R* was first determined with simulated data and was validated with experimental data, making use of both the GC × GC–TOFMS and GC–TOFMS modes of the instrument. Sampling phase was objectively studied by considering both in-phase and out-of-phase sampling, finding that there is effectively no difference in the modulator induced band broadening between sampling phases for $M_R \geq 2$. The proposed method of mapping between M_R and M_R* was then used to calculate a percent loss in ^1n_c, comparing the premodulation peak capacity, ^1n_c, to the effective peak capacity, ^1n_c*, from the results of the simulation study. The method presented herein makes use of the average broadening ratio to determine the average M_R for a GC × GC separation that is independent of sampling phase, though it may be possible to precisely determine the sampling-phase of the modulated peak from the ^{2}D peak heights in order to more accurately calculate M_R for a single peak; however, this is beyond the scope of the current work and would require further study. We also demonstrate that the mapping from M_R* to M_R allows the analyst to properly apply the β correction factor to determine the effective peak capacity, accounting for both modulator induced band broadening and statistical peak overlap theory. The concepts presented herein are not limited to GC × GC with cryogenic modulation, and should be readily extendable to other comprehensive two-dimensional separation techniques with total transfer modulation for the entire measurable range of M_R, and also applicable to techniques using partial transfer modulation as long as the M_R is greater than about 2 to ensure quantitative modulation sampling [22,27].

4.6 REFERENCES

[1] Z. Liu, J.B. Phillips, Comprehensive Two-Dimensional Gas Chromatography using an On-Column Thermal Modulator Interface, J. Chromatogr. Sci. 29 (1991) 227–231. doi:10.1093/chromsci/29.6.227.

[2] M. Adahchour, L.L.P. van Stee, J. Beens, R.J.J. Vreuls, M.A. Batenburg, U.A.T. Brinkman, Comprehensive two-dimensional gas chromatography with time-of-flight mass spectrometric detection for the trace analysis of flavour compounds in food, J. Chromatogr. A. 1019 (2003) 157–172. doi:10.1016/S0021-9673(03)01131-2.

[3] C.A. Bruckner, B.J. Prazen, R.E. Synovec, Comprehensive Two-Dimensional High-Speed Gas Chromatography with Chemometric Analysis, Anal. Chem. 70 (1998) 2796–2804. doi:10.1021/ac980164m.

[4] J. Beens, M. Adahchour, R.J.J. Vreuls, K. van Altena, U.A. Th. Brinkman, Simple, non-moving modulation interface for comprehensive two-dimensional gas chromatography, J. Chromatogr. A. 919 (2001) 127–132. doi:10.1016/S0021-9673(01)00785-3.

[5] R.M. Kinghorn, P.J. Marriott, Comprehensive Two-Dimensional Gas Chromatography Using a Modulating Cryogenic Trap, J. High Resolut. Chromatogr. 21 (1998) 620–622. doi:10.1002/(SICI)1521-4168(19981101)21:11<620::AID-JHRC620>3.0.CO;2-#.

[6] M. Adahchour, J. Beens, R.J.J. Vreuls, U.A.T. Brinkman, Recent developments in comprehensive two-dimensional gas chromatography (GC × GC): IV. Further applications, conclusions and perspectives, TrAC Trends Anal. Chem. 25 (2006) 821–840. doi:10.1016/j.trac.2006.03.003.

[7] J.V. Seeley, F. Kramp, C.J. Hicks, Comprehensive Two-Dimensional Gas Chromatography via Differential Flow Modulation, Anal. Chem. 72 (2000) 4346–4352. doi:10.1021/ac000249z.

[8] R.B. Wilson, W.C. Siegler, J.C. Hoggard, B.D. Fitz, J.S. Nadeau, R.E. Synovec, Achieving high peak capacity production for gas chromatography and comprehensive two-dimensional gas chromatography by minimizing off-column peak broadening, J. Chromatogr. A. 1218 (2011) 3130–3139. doi:10.1016/j.chroma.2010.12.108.

[9] L.M. Blumberg, Comprehensive two-dimensional gas chromatography: metrics, potentials, limits, J. Chromatogr. A. 985 (2003) 29–38. doi:10.1016/S0021-9673(02)01416-4.

[10] L.M. Blumberg, Multidimensional Gas Chromatography: Theoretical Considerations, in: L. Mondello (Ed.), Compr. Chromatogr. Comb. Mass Spectrom., John Wiley & Sons, Inc., 2011: pp. 13–63. http://onlinelibrary.wiley.com/doi/10.1002/9781118003466.ch2/summary (accessed June 6, 2016).

[11] L.M. Blumberg, F. David, M.S. Klee, P. Sandra, Comparison of one-dimensional and comprehensive two-dimensional separations by gas chromatography, J. Chromatogr. A. 1188 (2008) 2–16. doi:10.1016/j.chroma.2008.02.044.

[12] J.B. Phillips, J. Beens, Comprehensive two-dimensional gas chromatography: a hyphenated method with strong coupling between the two dimensions, J. Chromatogr. A. 856 (1999) 331–347. doi:10.1016/S0021-9673(99)00815-8.

[13] X. Li, D.R. Stoll, P.W. Carr, Equation for Peak Capacity Estimation in Two-Dimensional Liquid Chromatography, Anal. Chem. 81 (2009) 845–850. doi:10.1021/ac801772u.

[14] J.M. Davis, D.R. Stoll, P.W. Carr, Effect of First-Dimension Undersampling on Effective Peak Capacity in Comprehensive Two-Dimensional Separations, Anal. Chem. 80 (2008) 461–473. doi:10.1021/ac071504j.
[15] D.K. Pinkerton, B.A. Parsons, T.J. Anderson, R.E. Synovec, Trilinearity Deviation Ratio: A New Metric for Chemometric Analysis of Comprehensive Two-Dimensional Gas Chromatography Time-of-Flight Mass Spectrometry Data, Anal. Chim. Acta. 871 (2015) 66–76. doi:10.1016/j.aca.2015.02.040.
[16] J.M. Davis, D.R. Stoll, P.W. Carr, Dependence of Effective Peak Capacity in Comprehensive Two-Dimensional Separations on the Distribution of Peak Capacity between the Two Dimensions, Anal. Chem. 80 (2008) 8122–8134. doi:10.1021/ac800933z.
[17] B.D. Fitz, R.B. Wilson, B.A. Parsons, J.C. Hoggard, R.E. Synovec, Fast, high peak capacity separations in comprehensive two-dimensional gas chromatography with time-of-flight mass spectrometry, J. Chromatogr. A. 1266 (2012) 116–123. doi:10.1016/j.chroma.2012.09.096.
[18] W. Khummueng, J. Harynuk, P.J. Marriott, Modulation Ratio in Comprehensive Two-dimensional Gas Chromatography, Anal. Chem. 78 (2006) 4578–4587. doi:10.1021/ac052270b.
[19] P.J. Marriott, Z. Wu,, P. Schoenmakers, Nomenclature and Conventions in Comprehensive Multidimensional Chromatography, LC-GC Eur. 16 (2003) 335–339.
[20] R.E. Murphy, M.R. Schure, J.P. Foley, Effect of Sampling Rate on Resolution in Comprehensive Two-Dimensional Liquid Chromatography, Anal. Chem. 70 (1998) 1585–1594. doi:10.1021/ac971184b.
[21] J.A. Murray, Qualitative and quantitative approaches in comprehensive two-dimensional gas chromatography, J. Chromatogr. A. 1261 (2012) 58–68. doi:10.1016/j.chroma.2012.05.012.
[22] J.V. Seeley, Theoretical study of incomplete sampling of the first dimension in comprehensive two-dimensional chromatography, J. Chromatogr. A. 962 (2002) 21–27. doi:10.1016/S0021-9673(02)00461-2.
[23] M.S. Klee, J. Cochran, M. Merrick, L.M. Blumberg, Evaluation of conditions of comprehensive two-dimensional gas chromatography that yield a near-theoretical maximum in peak capacity gain, J. Chromatogr. A. 1383 (2015) 151–159. doi:10.1016/j.chroma.2015.01.031.
[24] B.A. Parsons, D.K. Pinkerton, B.W. Wright, R.E. Synovec, Chemical characterization of the acid alteration of diesel fuel: Non-targeted analysis by two-dimensional gas chromatography coupled with time-of-flight mass spectrometry with tile-based Fisher ratio and combinatorial threshold determination, J. Chromatogr. A. 1440 (2016) 179–190. doi:10.1016/j.chroma.2016.02.067.
[25] J.L. Adcock, M. Adams, B.S. Mitrevski, P.J. Marriott, Peak Modeling Approach to Accurate Assignment of First-Dimension Retention Times in Comprehensive Two-Dimensional Chromatography, Anal. Chem. 81 (2009) 6797–6804. doi:10.1021/ac900960n.
[26] J.V. Seeley, N.J. Micyus, S.V. Bandurski, S.K. Seeley, J.D. McCurry, Microfluidic Deans Switch for Comprehensive Two-Dimensional Gas Chromatography, Anal. Chem. 79 (2007) 1840–1847. doi:10.1021/ac061881g.

[27] W.C. Siegler, B.D. Fitz, J.C. Hoggard, R.E. Synovec, Experimental Study of the Quantitative Precision for Valve-Based Comprehensive Two-Dimensional Gas Chromatography, Anal. Chem. 83 (2011) 5190–5196. doi:10.1021/ac200302b.
[28] H.E. Toraman, K. Franz, F. Ronsse, K.M. Van Geem, G.B. Marin, Quantitative analysis of nitrogen containing compounds in microalgae based bio-oils using comprehensive two-dimensional gas-chromatography coupled to nitrogen chemiluminescence detector and time of flight mass spectrometer, J. Chromatogr. A. 1460 (2016) 135–146. doi:10.1016/j.chroma.2016.07.009.

4.7 TABLES

Table 4.1 Measured ^1W_b of unmodulated experimental data and calculated M_R values for modulated GC × GC–TOFMS experiments run with varying modulation periods, P_M.

	^1W_b (s)			P_M (s)	M_R		
	C10	C11	C12		C10	C11	C12
5.5 °C/min Separation	2.08 ± 0.02	2.19 ± 0.01	2.41 ± 0.005	0.25	8.32	8.76	9.64
				0.5	4.16	4.38	4.82
				0.75	2.77	2.92	3.21
				1	2.08	2.19	2.41
				1.5	1.39	1.46	1.61
				2	1.04	1.10	1.21
				2.5	0.83	0.88	0.96
				3	0.69	0.73	0.80
				4	0.52	0.55	0.60
				5	0.42	0.44	0.48
2.0 °C/min Separation	3.25 ± 0.02	3.7 ± 0.02	4.35 ± 0.01	0.5	6.50	7.40	8.70
				1	3.25	3.70	4.35
				1.5	2.17	2.47	2.90
				2	1.63	1.85	2.18
				2.5	1.30	1.48	1.74
				3	1.08	1.23	1.45
				3.5	0.93	1.06	1.24
				4	0.81	0.93	1.09
				5	0.65	0.74	0.87
				8	0.41	0.46	0.54

Table 4.2 Summary of results from Fig. 4.5, used to determine M_R from M_R*.

M_R*	M_R/M_R* Average	M_R Average	Average % Loss in ^1n_c
1	0.66	0.66	-34
1.25	0.7	0.88	-30
1.5	0.75	1.13	-25
1.75	0.79	1.38	-21
2	0.82	1.64	-18
2.25	0.85	1.91	-15
2.5	0.88	2.20	-12
2.75	0.9	2.48	-10
3	0.92	2.76	-8
3.25	0.93	3.02	-7
3.5	0.94	3.29	-6
3.75	0.95	3.56	-5
4	0.96	3.84	-4
4.5	0.97	4.37	-3
5	0.975	4.88	-2.5
5.5	0.98	5.39	-2
6	0.99	5.94	-1

4.8 FIGURES

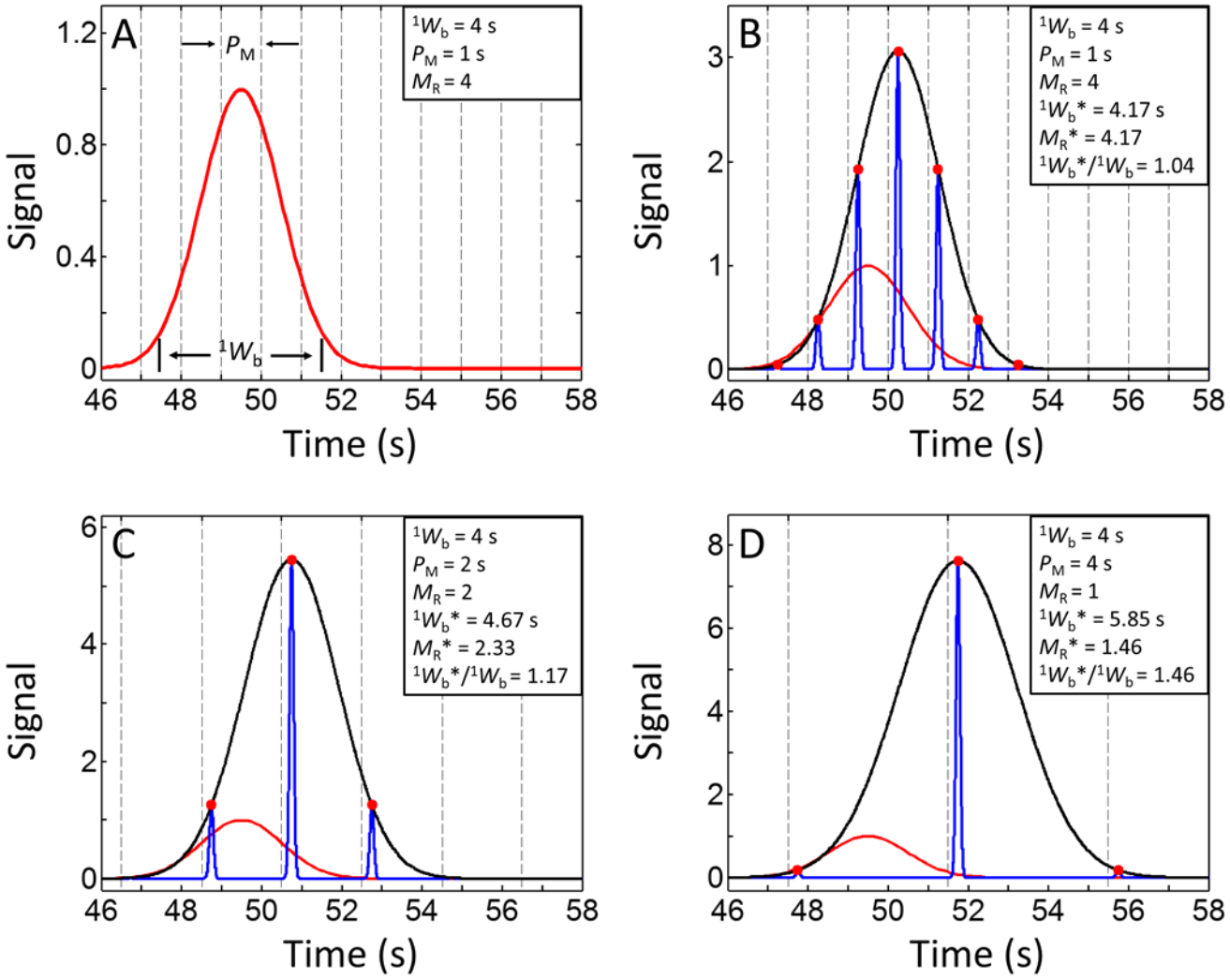

Figure 4.1 Simulation of unmodulated ^{1}D Gaussian peak, plotted according to Eq. (4.1), modulated ^{2}D peaklets, plotted according to Eqs. (4.2-4.4), and fitted Gaussian curve, Eq. (4.1), of ^{1}D profile from modulated ^{2}D peaklets, with relevant variables: P_M, ^1W_b, M_R, ^1W_b*, M_R*, and ^1W_b*/^1W_b. All peaks are sampled in-phase. Circles (●) represent the points used for the Gaussian curve fitting to the modulated peaklets based on retention times and peak heights of modulated peaklets. (A) Unmodulated ^{1}D elution profile of peak with retention time ^1t_R = 49.5 s and ^1W_b = 4 s. Peak amplified by factor of 2.5 for visualization purposes. The location of each modulation, P_M = 1 s, is shown by vertical dashed lines. The ^{1}D peak shown in (A) is used throughout Fig. 4.1. (B) Modulated GC × GC peaklets and Gaussian fit of data from (A) with (B) P_M = 1 s, (C) P_M = 2 s, (D) P_M = 4 s.

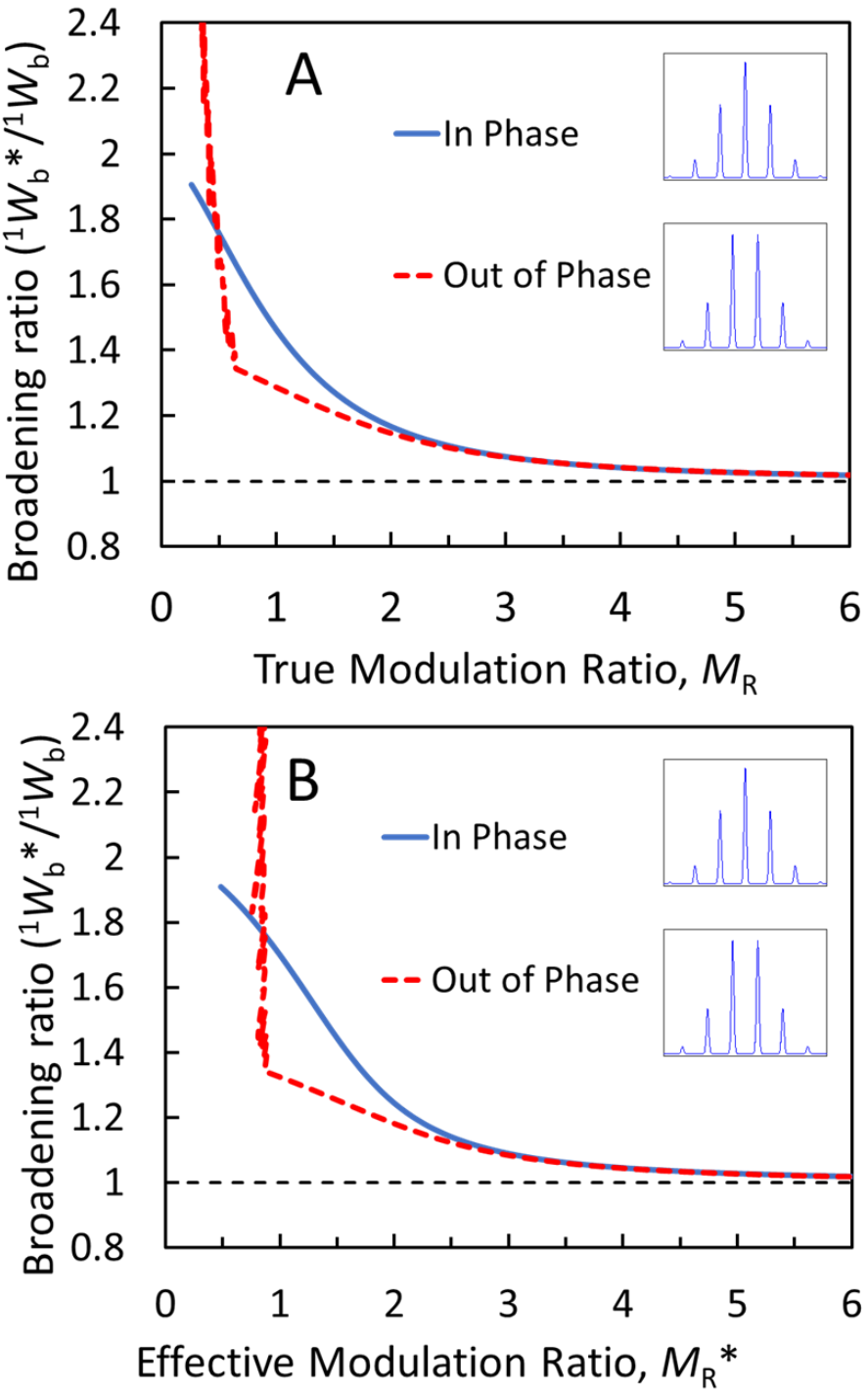

Figure 4.2 Broadening ratio (^1W_b*/^1W_b) from simulated ^{1}D and ^{2}D data with Gaussian curve fitting plotted as a function of (A) true modulation ratio, M_R, and (B) effective modulation ratio, M_R*, for in-phase and out-of-phase sampling. Insets are chromatographic representations of in-phase and out-of-phase sampling. Horizontal dashed line at broadening ratio of 1 means ^1W_b* = ^1W_b.

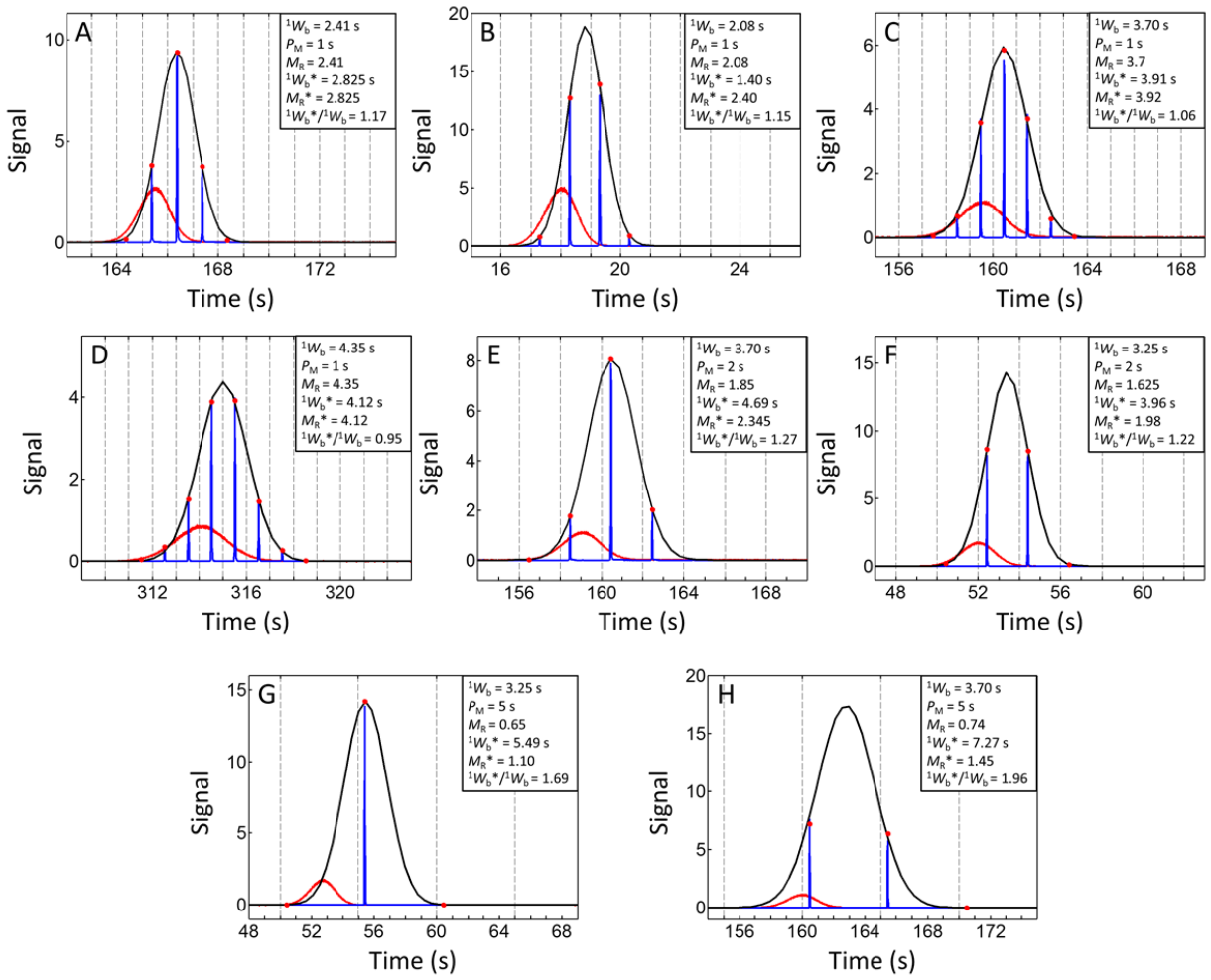

Figure 4.3 Overlays of collected data: unmodulated 1D peak, amplified by a factor of 5 for visualization, modulated 2D peaklets with peak maxima and retention times shown, (●), and fitted ^{1}D Gaussian peak profile, Eq. (4.1). Approximate modulation locations shown by vertical dashed lines, as in Fig. 4.1 with relevant variables: P_M, ^1W_b, M_R, ^1W_b*, M_R*, and ^1W_b*/^1W_b. (A) Narrow ^{1}D peak, short modulation, P_M = 1 s, in-phase sampling. (B) The same separation conditions as (A) with out-of-phase sampling. (C) Wide ^{1}D peak, short modulation, P_M = 1 s, in-phase sampling. (D) The same separation conditions as (C) with out-of-phase sampling. (E) Wide ^{1}D peak, medium modulation, P_M = 2 s, in-phase sampling. (F) The same separation conditions as (E) with out-of-phase sampling. (G) Wide ^{1}D peak, long modulation, P_M = 5 s, in-phase sampling. (H) The same separation conditions as (G) with out-of-phase sampling.

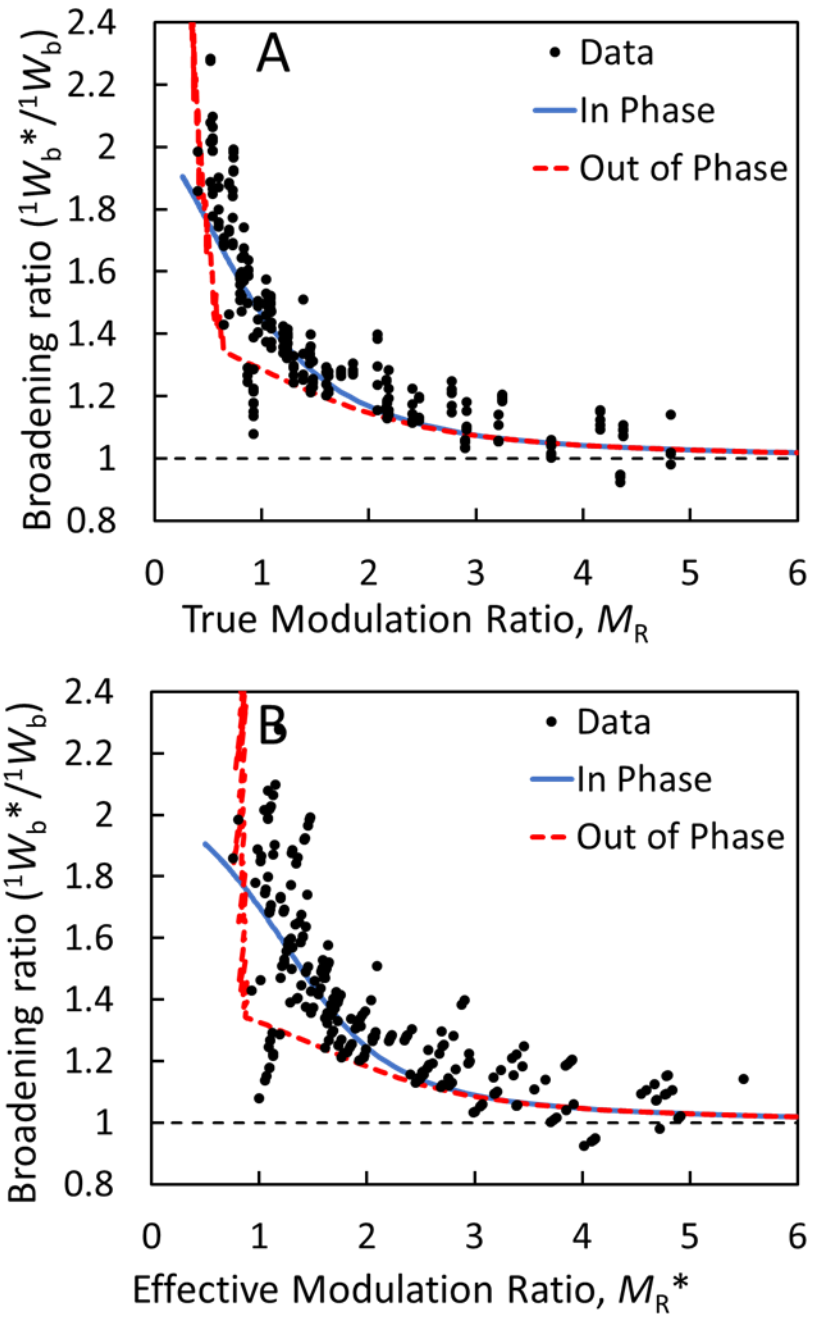

Figure 4.4 Broadening ratio (^1W_b*/^1W_b) as a function of (A) true modulation ratio, M_R, and (B) effective modulation ratio, M_R*, as in Fig. 4.2, but with peak width measurements from experimental data (●) plotted over the simulated results from Fig. 4.2.

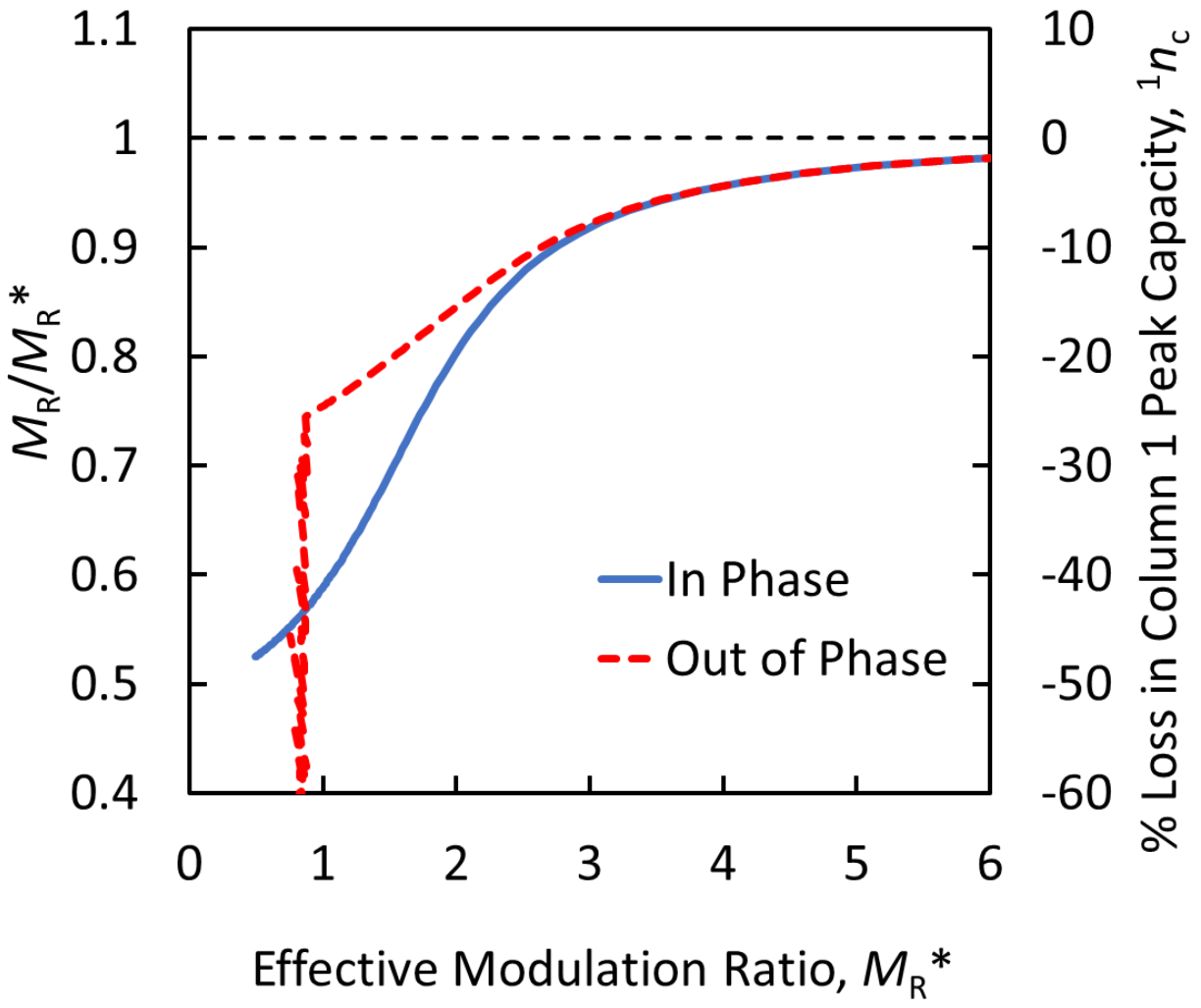

Figure 4.5 Plot of M_R/M_R* and percentage loss in ^{1}D peak capacity due to modulation dependent band broadening as a function of effective modulation ratio M_R*, Eq. (4.12). Curves for both in-phase, solid line, and out-of-phase, dashed line, sampling are shown. Results summarized in Table 4.2.

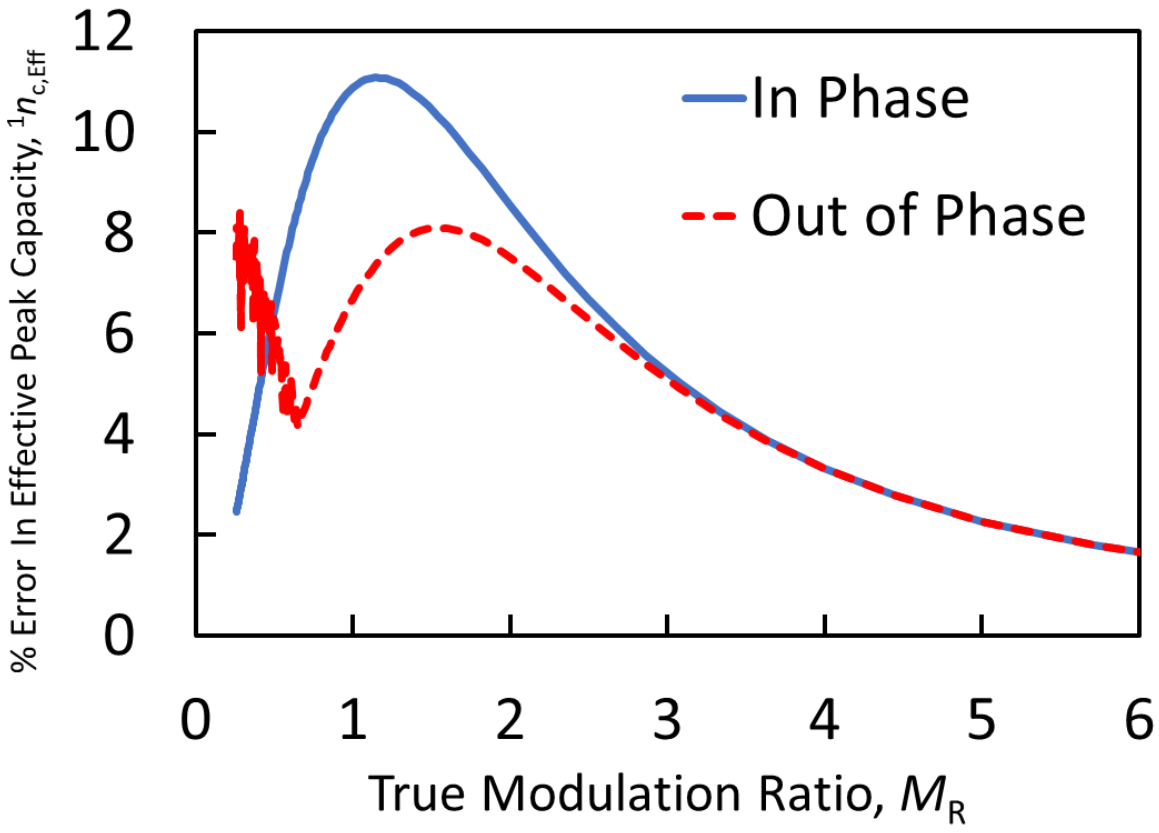

Figure 4.6 Plot of Eq. (4.16), the percent error in effective peak capacity, $^1n_{c,Eff,}$ calculated with effective modulation ratio, M_R*, and measured peak capacity, ^1n_c*, as opposed to true modulation ratio, M_R, and ideal peak capacity, ^1n_c, as a function of the true modulation ratio, M_R. Curves for both in-phase, solid line, and out-of-phase sampling, dashed line, are shown.

Chapter 5. Trilinearity Deviation Ratio: A New Metric for Chemometric Analysis of Comprehensive Two-Dimensional Gas Chromatography Time-of-Flight Mass Spectrometry Data[4]

5.1 INTRODUCTION

Comprehensive two-dimensional gas chromatography, GC × GC, is a well-established and useful instrumental platform for analyzing complex samples in a wide variety of fields and applications [1–10]. Coupling GC × GC with time-of-flight mass spectrometry, GC × GC – TOFMS, results in a particularly powerful instrument capable of generating highly selective quantitative data [9–23]. For each sample run, the GC × GC – TOFMS instrument produces a very large, dense, information-rich cube of data, typically 100 to 500 MB per sample run depending on the run time and data collection rate. There are, however, significant challenges in working with and analyzing data sets of this size and complexity. To address these challenges, a great deal of analytical research has been performed in the development of chemometric software to analyze GC × GC – TOFMS data [4,9,11,24–35].

Indeed, there are many chemometric methods amenable to the three-dimensional data structure provided by GC × GC – TOFMS: parallel factor analysis (PARAFAC), multivariate curve resolution – alternating least squares (MCR-ALS), PARAFAC2, trilinear decomposition (TLD), etc. [11,12,17,22,23,33–51]. While each of these chemometric methods used for

[4]This Chapter has been reproduced from D.K. Pinkerton, B.A. Parsons, T.J. Anderson, R.E. Synovec, Trilinearity Deviation Ratio: A New Metric for Chemometric Analysis of Comprehensive Two-Dimensional Gas Chromatography Time-of-Flight Mass Spectrometry Data, Anal. Chim. Acta. 871 (2015) 66–76. doi:10.1016/j.aca.2015.02.040.

analyzing GC × GC – TOFMS data have their own strengths and weaknesses, their analytical goals are generally the same; primarily, some combination of analyte deconvolution, identification, and quantification. In particular, this study will focus on Parallel Factor Analysis (PARAFAC), because it is one of the most highly used chemometric methods for GC × GC – TOFMS data [11,12,17,22,23,23,33–46]. PARAFAC has been demonstrated to be extremely effective for analyte deconvolution, identification and quantification, making use of the third order advantage when the data produced by the GC × GC – TOFMS instrument possesses sufficient trilinearity [12,23–25,33–46,52]. However, PARAFAC has been shown to have difficulties appropriately modeling data when the analyte signal is not sufficiently trilinear [48,50,52–55]. PARAFAC is studied herein to investigate the chromatographic conditions that produce GC × GC – TOFMS data within a given single sample run that is sufficiently trilinear for PARAFAC, as well as the limitations of using PARAFAC when the data is not sufficiently trilinear. While other chemometric methods do not require trilinearity in the data, in particular MCR-ALS and PARAFAC2, PARAFAC does require trilinearity. When the data is sufficiently trilinear and an appropriate number for model factors have been chosen, PARAFAC produces unique solutions and is not subject to the rotational ambiguities of second order methods [48,50,52–55]. Mathematically this means that the estimated PARAFAC model cannot be rotated without a loss of fit [52]. Currently, ambiguity exists across the interface between the chromatography and chemometrics fields regarding the interrelationship between the common practice of implementing GC × GC – TOFMS, the resulting data structure provided by the instrument, and the subsequent chemometric data analysis performance. This report is aimed at providing insight on the important issues related to these interrelationships.

Temperature programming is used with GC × GC – TOFMS as a way to deal with the general elution problem. However, the primary impact of temperature programming is only observed along the first column separation dimension (^{1}D). The time length of the second column separation dimension (^{2}D), which is referred to as the modulation period [56], P_M, (commonly 1 s to 8 s) is small in the context of temperature programming rates typically used for GC × GC separations, T_{ramp} in °C/min. Hence, while the ^{1}D separation is being temperature programmed, the temperature change ΔT in °C during each individual ^{2}D separation is very small, such that each ^{2}D separation is essentially a pseudo-isothermal separation at a temperature ΔT greater than the modulation preceding it. For example, with a commonly used T_{ramp} of 5 °C/min, and a P_M of 6 s, each successive ^{2}D separation is only 0.5 °C warmer than the previous separation; essentially no temperature programming occurring on the second dimension of the GC × GC separation. Nonetheless the analyte ^{2}D retention time, ^2t_R, depends on ΔT, which suggests the ^2t_R for a given analyte may decrease from one modulation to the next, defined as Δ^2t_R, which is the change in ^{2}D retention time between successive modulations. For analytes that are essentially unretained on ^{2}D, the shifting in retention time will not be observed ($\Delta^2t_R = 0$), and the Δ^2t_R is progressively more pronounced for analytes that are more retained on ^{2}D. The differences in ^2t_R between successive modulations means the data is no longer strictly trilinear, but the impact of the deviation from trilinearity on chemometric data analysis methods has not been well defined and rigorously studied. In previous studies, deviations from trilinearity brought on by Δ^2t_R have been handled by applying a ^{2}D peak alignment algorithm to the data prior to chemometric analysis [48,49,57]. To reduce the amount of preprocessing time and data distortion introduced by peak alignment, it is important to identify the chromatographic conditions where peak alignment is not needed for analysis by PARAFAC, as addressed herein.

In this report, we demonstrate that while shifting in retention time between modulations, $\Delta^2 t_{R,}$ is the principal source of non-trilinearity within a single GC × GC – TOFMS data cube, $\Delta^2 t_R$ alone is not sufficient to quantitatively describe the trilinearity of the data structure for the purposes of predicting the performance of a chemometric method such as PARAFAC. We hypothesize that the analyte base peak width on ^{2}D separations, defined as ^2W_b, also impacts the trilinearity of the data structure along with $\Delta^2 t_R$ [58]. We introduce the term Trilinearity Deviation Ratio, *TDR*, a peak width normalized retention time shift, which will be shown to quantitatively describe accuracy for PARAFAC analyses of GC × GC – TOFMS data. Simulated GC × GC – TOFMS data is used, and is subjected to iterative shifts in modulation-to-modulation retention time, $\Delta^2 t_R$; the trilinear PARAFAC model is used to quantitatively evaluate the data. These models are compared with the model of perfectly trilinear data for the given instrumental conditions. Each GC × GC – TOFMS chromatographic run, which we refer to herein as a GC × GC – TOFMS data cube, is analyzed by PARAFAC independently, and is made up of two chromatographic separation dimensions ^{1}D and ^{2}D and a mass spectral dimension. PARAFAC is performed on small sections of the GC × GC – TOFMS data cube (100 ms to 400 ms on ^{2}D, 4 to 6 modulations on ^{1}D, and all collected *m/z* for non-targeted analysis) to reduce the rank of the data in the analysis window. Because each GC × GC – TOFMS data cube is analyzed independently, sample run-to-run deviations from trilinearity are avoided such that the principal source of non-trilinearity within the data is $\Delta^2 t_R$.

We also explore how modulation ratio, M_R [56], plays a role in the performance of PARAFAC in the context of the *TDR*. Analytes are simulated to have varied ^2W_b, M_R, sampling phase, and background noise, to identify the impact of each variable on data trilinearity and PARAFAC performance. Only in-phase and out-of-phase modulation of the GC × GC peaks

were studied since these two extreme cases have been shown to bracket the full range of quantitative errors introduced during quantitative analysis of GC × GC data [8,59–61]. *TDR* values are then predicted for common sets of instrumental conditions and the results are related to the theoretical ^{1}D, ^{2}D and two-dimensional peak capacities for the GC × GC separations under the same conditions. To summarize, we evaluate how several interrelated instrumental variables in the experimental design contribute to the trilinearity in the GC × GC – TOFMS data structure and the subsequent effect when using PARAFAC. The interrelated variables include P_M, T_{ramp}, ΔT, $^{2}t_R$, $\Delta^{2}t_R$, $^{2}W_b$, M_R, sampling phase (in-phase and out-of-phase), and signal-to-noise, *S/N*. The *TDR* metric is introduced and evaluated herein for GC × GC – TOFMS data. In the context of this wide range of experimental design conditions, we critically examine the use of PARAFAC for a single analyte peak within a single GC × GC – TOFMS data cube. Focusing this study to purely resolved peaks facilitates a thorough testing of the wide range of experimental design conditions, by changing one experimental design variable at a time. However, co-elution of multiple peaks within a single GC × GC – TOFMS data cube with varying degrees of peak overlap is common in GC × GC separations. While peak overlap and deconvolution are not treated herein, this study establishes an objective framework for further studies to investigate the effects of *TDR* on PARAFAC of GC × GC – TOFMS data cubes with peak overlap.

5.2 THEORY

Before discussing trilinearity for GC × GC – TOFMS data, which can be simplified to bilinearity for GC × GC, it is necessary to define relevant chromatographic parameters and the relationships between them. The modulation ratio, M_R, of a GC × GC peak is defined as

$$M_R = \frac{^{1}W_b}{P_M} \tag{5.1}$$

where ${}^1W_b$ is the analyte ^{1}D peak width measured at the base (±2σ, in units of time) and P_M is the modulation period [56]. The M_R is used to describe how many times an analyte peak is modulated from ^{1}D onto ^{2}D, but is not necessarily equal to the number of ^{2}D peaklets observed for a given ^{1}D peak [56]. The relationship between analyte retention and temperature for the ^{2}D separations can be expressed using the van't Hoff equation, given by

$$\ln {}^2k' = \ln\left(\frac{{}^2t_R - {}^2t_0}{{}^2t_0}\right) = \frac{\Delta H}{R \cdot T} + C \tag{5.2}$$

where ${}^2k$' is the analyte retention factor, ${}^2t_R$ is the retention time of the analyte, and ${}^2t_0$ is the dead time (all on ^{2}D); ΔH is the enthalpy of vaporization of the analyte, R is the ideal gas constant, T is the temperature of the ^{2}D separation for a specific modulation, which increases by ΔT for each successive modulation, and C is essentially a constant for each analyte, which includes an entropy term for the analyte and the phase volume ratio for the ^{2}D separation.

Here we consider changes in ^{2}D retention time between modulations, $\Delta^2 t_R$, brought upon by the use of temperature programming on ^{1}D. The theoretical change in retention time between modulations is found by comparing successive modulations of a given analyte with respect to Eq. (5.2). Deriving an expression for $\Delta^2 t_R$ using Eq. (5.2) for successive analyte modulations N and $N+1$ along the temperature program (steps not shown for brevity) gives

$$\Delta {}^2t_R = \frac{-\Delta H}{R}\left(\frac{\Delta T}{T_{N+1}T_N}\right)({}^2k' \cdot {}^2t_0) \tag{5.3}$$

where ΔT is the change in temperature between modulation N and modulation $N+1$, T_N is the temperature of the pseudo-isothermal modulation N, and T_{N+1} is the temperature of the pseudo-isothermal modulation $N+1$. The temperature change between modulations, ΔT, is defined as

$$\Delta T = T_{ramp} \cdot P_M \tag{5.4}$$

T_{ramp} is the temperature programming rate for the ^{1}D separation and P_M is provided in Eq. (5.1).

We hypothesize that the quantitative error for PARAFAC is dependent not only on the modulated retention time shift, $\Delta^2 t_R$, but also on the peak width, ^2W_b ($\pm 2\sigma$, in units of time), on ^{2}D, such that wider peaks can accommodate retention time shifts more so than narrow peaks [58]. The Trilinear Deviation Ratio, *TDR*, a dimensionless quantity, is proposed as a quantitative metric that will provide the analyst with a means to directly relate chromatographic information to the PARAFAC performance, defined by

$$TDR = \frac{\Delta^2 t_R}{^2W_b} \tag{5.5}$$

^2W_b, which is a function of several parameters but principally for this discussion the analyte retention factor $^2k'$, can be derived starting with the Golay equation for isothermal separations since each ^{2}D separation is essentially isothermal during the given P_M [62–64]

$$H = \frac{2D_{g,o}jf}{\bar{u}} + \frac{1+6\cdot{}^2k\prime+11\cdot{}^2k'^2}{96(1+{}^2k')^2}\frac{d_c^2\bar{u}f}{D_{g,o}j} + \frac{2\cdot{}^2k'd_f^2\bar{u}}{3(1+{}^2k')^2D_s} \tag{5.6}$$

where $D_{g,o}$ is the analyte diffusion coefficient in the gas phase at the column outlet, j is the James-Martin gas compressibility correction factor, f is a correction term developed by Giddings to correct for the effects of gas compressibility along the length of the column, $\bar{u}$ is the average linear velocity of the carrier gas, $^2k'$ is as described above, d_c is the column internal diameter, d_f is the stationary phase thickness, and D_s is the analyte diffusion coefficient in the stationary phase. Since H is quantitatively the length variance of the analyte on-column peak width per column length, L, the translation to the detected peak width in units of time, under isothermal conditions, ^2W_b, is given by [63,64]

$$ {}^2W_b = 4\left(\frac{2D_{g,0}jf\left(1+{}^2k'\right)^{2}\cdot{}^2t_0^2}{L^2}+\frac{\left(1+6\cdot{}^2k'+11\cdot{}^2k'^2\right)d_c^2f\cdot{}^2t_0}{96D_{g,0}j}+\frac{2\cdot{}^2k'd_f^2\cdot{}^2t_0}{3D_s}\right)^{1/2} \tag{5.7} $$

For isothermal separations the mass transfer term in the mobile phase dominates (middle term), and specifically it has been shown that it is the $11\cdot{}^2k'^2$ portion of the mobile phase mass transfer term. Therefore, the ${}^2W_\mathrm{b}$ has been shown to simplify to a linear equation given by [63,64]

$$ {}^2W_\mathrm{b} \approx {}^2W_{\mathrm{b,k'}} + {}^2W_{\mathrm{b},0} = 4\cdot{}^2k'\left(\frac{11d_c^2f\cdot{}^2t_0}{96D_{g,0}j}\right)^{1/2} + {}^2W_{\mathrm{b},0} \tag{5.8} $$

where ${}^2W_{\mathrm{b,k'}}$ is the ${}^2k'$ dependent width for ${}^2k' > 0$ due to on-column band broadening, and ${}^2W_{\mathrm{b},0}$ is the minimum peak width of the analyte, at ${}^2k' = 0$, primarily due to modulator-dominated "extra-column" band broadening. A more detailed expression for estimating *TDR* as a function of ${}^2k'$, can now be provided by combining Eqs. (5.3) and (5.8) with Eq. (5.5)

$$ TDR \approx \frac{\frac{-\Delta H}{R}\left(\frac{\Delta T}{T_{\mathrm{N}+1}T_\mathrm{N}}\right)\left({}^2k'\cdot{}^2t_0\right)}{4\cdot{}^2k'\left(\frac{11d_c^2f\cdot{}^2t_0}{96D_{g,0}j}\right)^{1/2} + {}^2W_{\mathrm{b},0}} \tag{5.9} $$

The inverse relationship between ${}^2W_\mathrm{b}$ and *TDR* means that a theoretical maximum *TDR*, referred to as TDR_Max, can be found in the limit that the modulator-dominated band broadening, ${}^2W_{\mathrm{b},0}$, becomes sufficiently negligible relative to ${}^2W_{\mathrm{b,k'}}$ in Eq. (5.9) at suitably high ${}^2k'$. The expression for TDR_Max under these conditions becomes

$$ TDR_\mathrm{Max} \leq \frac{-\Delta H}{4R}\left(\frac{\Delta T}{T_{\mathrm{N}+1}T_\mathrm{N}}\right)\left(\frac{96D_{g,0}j\cdot{}^2t_0}{11d_c^2f}\right)^{1/2} \tag{5.10} $$

It should be noted that Eq. (5.10) does not depend on ${}^2k'$; when ${}^2W_{\mathrm{b},0}$, in Eq. (5.9), is sufficiently negligible, ${}^2k'$ cancels in Eq. (5.9). Eqs. (5.8), (5.9) and (5.10) are readily amenable to evaluate the interrelationships between the GC × GC – TOFMS data structure, the performance of PARAFAC with data collected under various GC × GC separation conditions,

and ultimately how the separations conditions tie directly into GC × GC peak capacity considerations.

5.3 EXPERIMENTAL

5.3.1 *Simulated GC × GC – TOFMS Data*

All data simulations and analyses were performed in Matlab R2012b (The Mathworks, Inc., Natick, MA, U.S.A.), with the modeling conditions summarized in Table 1. GC × GC – TOFMS chromatograms (i.e., data cubes) were simulated at a 500 Hz mass spectral collection rate, with each 2 ms interval referred to herein as a data point. The 500 Hz scan rate was selected since it is the maximum allowable scan rate for the LECO TOFMS on their GC × GC – TOFMS instrument (LECO, St. Joseph, MI). The first step in the GC × GC – TOFMS data simulations was to define a ^{1}D analyte peak (Gaussian profile) with specified width, $^{1}W_{b}$ defined as the peak width at ±2σ, in units of time. The modulated ^{2}D peaks were then generated as a series of Gaussian peaks with specified $^{2}W_{b}$, the spacing between peak maxima defined by the modulation period, P_{M}. Peak heights for the modulated ^{2}D peaks were defined by the value of the apparent ^{1}D Gaussian analyte peak at each modulated peak maximum location. In-phase sampling is equivalent to when a modulated ^{2}D peak location coincides with the peak maximum of the ^{1}D peak, and out-of-phase sampling was achieved by shifting ^{2}D peak locations by ½ P_{M}, such that there were two identical modulated ^{2}D peaks of maximal intensity each +/- ½ P_{M} from the maximum of the ^{1}D peak. Figs. 5.1(A) and (B) show raw "unfolded" GC × GC data for in-phase and out-of-phase modulation sampling, respectively, both at a M_{R} = 4. Additionally, the relevant GC × GC parameters are defined in Figs. 5.1(C) and (D), in which P_{M}, T_{ramp}, ΔT, $\Delta^{2}t_{R}$, $^{2}W_{b}$ and $^{1}W_{b}$ are illustrated. Note that the P_{M} illustrated in Fig. 5.1 was purposely small relative to the

peak width for clarity. In practice, the P_M is generally much larger than the ^{2}D peak widths to provide suitable peak capacity. Once the peak profiles were generated, the mass spectrum of the target analyte, tert-butylbenzene, (NIST MS main library) was applied to the separation as the outer product of the unfolded GC × GC chromatographic data (vector) and the mass spectrum vector. The vector form of the simulated data was then cut at intervals of P_M and stacked side-by-side in a matrix to construct a two-dimensional separation space for each *m/z*.

Random Gaussian-distributed noise was generated independently for each data point in the chromatographic separation space for each simulated *m/z*, 30-175. The parameters for the Gaussian-distributed noise, mean and standard deviation, were determined such that the standard deviation of the noise would provide the desired signal-to-noise ratio (*S/N*) for the simulated analyte GC × GC – TOFMS signal, and the mean was set to be at least five times the value of the standard deviation, in order to provide an experimentally reasonable baseline signal which also served to preserve the non-negativity of the data.

Deviations from trilinearity were studied by incrementally decreasing the ^2t_R for each ^{2}D separation for each successive modulation across the target analyte ^{1}D elution peak, i.e., by generating Δ^2t_R for each successive modulation. The ^2t_R for each modulation was reduced by $p(n-1)$ data points, where p is the deviation from trilinearity, in data points, and n is the modulation number for a given ^{2}D peak such that the first modulation of the target analyte has $n = 1$.

PARAFAC was performed using the PLS Toolbox, version 7.3.1 (Eigenvector Research, Inc., Wenatchee, WA, U.S.A) on individual GC × GC – TOFMS data cubes, each with two chromatographic separation dimensions ^{1}D and ^{2}D and mass spectral dimension. For the purposes of this study, targeted PARAFAC was used such that only non-zero *m/z* from the target analyte mass spectrum were included. Implementing targeted PARAFAC reduces the total

number variables of *m/z* for the PARAFAC model by excluding *m/z* with zero intensity in the known library spectrum. Non-negativity constraints were applied to all dimensions of the PARAFAC model with the stricter unimodality constraint applied to the mode corresponding to the ^{1}D peak. All simulations at infinite *S/N* required a one factor PARAFAC model, and data with noise was modeled by two factor models. Trilinear GC × GC – TOFMS data cube chromatograms were reconstructed from the outer product of the PARAFAC loadings and summed across all three dimensions of the reconstructed GC × GC – TOFMS data cube GC × GC – TOFMS data for peak quantification. PARAFAC peak quantification accuracy was assessed by subtracting the known simulated peak signal from the PARAFAC sum signal, normalizing the difference to the simulated peak signal and multiplying by one hundred to express it as a percentage.

5.3.2 *Instrumental Parameters*

GC × GC – TOFMS data was collected in order to provide experimental *TDR* values for comparison to the *TDR* modeling and its relevance to PARAFAC analyses. The GC × GC – TOFMS instrumental platform consisted of an Agilent 6890N gas chromatograph equipped with an Agilent 7683 autoinjector (Agilent Technologies, Palo Alto, CA) coupled with a LECO Pegasus III TOFMS, and equipped with a 4D thermal modulator upgrade (LECO, St. Joseph, MI). The ^{1}D column was a 20 m x 250 μm i.d. x 0.5 μm RTX-5MS film (Restek, Bellefonte, PA) and the ^{2}D column was a 2 m x 180 μm i.d. x 0.2 μm RTX-200 film (Restek, Bellefonte, PA). The P_M for the separation was 1 s. Mass channels, m/z 41-340, were collected at 100spectra/s. A 1 μL injection of a diesel sample was made in split mode with a split ratio of 200:1 and the sample was injected in quadruplicate. The ^{1}D column was held at 50 °C for 0.25

min and then increased with T_{ramp} of 5 °C/min to 300 °C, where it was held for 5 min. The ^{2}D column was initially set at 55 °C and followed the same temperature program as the ^{1}D column. Hence, the ΔT between modulations was 0.083 °C.

5.4 Results and Discussion

We begin by studying the effect of ^{2}D retention time shifting, $\Delta^{2}t_{R}$, alone on the accuracy of PARAFAC quantification for the target analyte. Figure 5.2(A) is the calculated percent error for the PARAFAC peak sum signal as a function of $\Delta^{2}t_{R}$ in ms, for simulated GC × GC – TOFMS data of tert-butylbenzene at infinite *S/N*, with $^{2}W_{b}$ ranging from 50 ms to 300 ms, treated with the iterative retention shift described above. All other chromatographic parameters were held constant between simulations. Simulations were initially generated without noise to reduce the number of variables changing with each simulation because of the dependence of multiway methods like PARAFAC on the *S/N* of the peak of interest. The percent error values were calculated for each PARAFAC peak sum signal relative to the known simulated peak sum signal from the input GC × GC – TOFMS data cube. Since the analyte concentration was held constant for the simulated input GC × GC – TOFMS data (across all of the various experimental design conditions), the quantitative error introduced by PARAFAC modeling was determined relative to the known concentration of the simulated input GC × GC – TOFMS data. There are many parameters that must be considered to evaluate the fit of a PARAFAC model to the input GC × GC – TOFMS data like model loadings, core consistency, model residuals, etc. These parameters all provide useful information, which must be considered when using PARAFAC, and together describe the overall performance of the PARAFAC model. However, in order to provide

quantitative results from the analysis performed, percent error for the PARAFAC peak sum signal is used as the metric for PARAFAC model performance in this study.

Figure 5.2(A) shows that PARAFAC peak sum signal decreases with the negative percent errors increasing as $\Delta^2 t_R$ increases. Additionally, peaks with narrower ${}^2W_b$ produce larger errors for a given $\Delta^2 t_R$. The contour plots shown in Fig. 5.2 represent analyte peaks with ${}^2W_b$ of 300 ms (labeled I) and 100 ms (labeled II), each with a $\Delta^2 t_R$ = 20 ms; the observed PARAFAC percent errors are approximately 0.9 percent and 6 percent respectively for the two peaks. Based upon Fig. 5.2(A), for PARAFAC analyses of GC × GC – TOFMS data, $\Delta^2 t_R$ alone is not sufficient to describe the extent to which the data has deviated from trilinearity or the effect it will have on PARAFAC performance, and the dependence on ${}^2W_b$ must also be considered.

Indeed, as hypothesized in Eq. (5.5), the various percent error versus $\Delta^2 t_R$ curves collapse to a single unifying curve in Fig. 5.2(B) when percent error is plotted as a function of *TDR*, which we refer to as a "*TDR* curve." The observed relationship in Fig. 5.2(B) is similar to, but distinct from chromatographic resolution, R_s, between two peaks and is not to be confused with R_s. Resolution describes a width dependent separation between two adjacent peaks within the same plane, while *TDR* is a width dependent separation within a single analyte peak (along ^{1}D) between consecutive modulations onto ^{2}D. Several other target analytes were studied (along with tert-butylbenzene), and the *TDR* curves such as the curve in Fig. 5.2(B) were found to be independent of the analyte mass spectrum (other analytes not shown for brevity). It is very important to note that the results in Fig. 5.2(B) suggest that the GC × GC – TOFMS data does not have to be strictly trilinear (*TDR* exactly equal to zero) for PARAFAC to provide acceptable results with relatively low percent error. Only when the *TDR* is sufficiently large does the percent error become problematic.

The observed negative percent error for PARAFAC quantification along the *TDR* curve is due to the strict trilinearity requirement of the PARAFAC algorithm. PARAFAC decomposes the input data into three individual component vectors, comprised of two chromatographic dimensions and a mass spectral dimension. A single three-dimensional model is produced by taking the outer product of all component vectors, and this model is, by definition, strictly trilinear. When the original data input to PARAFAC has a $TDR > 0$, the PARAFAC output is a strictly trilinear model, $TDR = 0$, that captures as much of the input data as possible by minimizing the model residuals. However, small amounts of data may be lost from each modulation in the formation of the model. Because the trilinear model centers on the modulations with the greatest amount of signal, the amount of signal lost from a given modulation increases the further it is from the maximum of the ^{1}D elution peak profile.

The number of modulations across a given ^{1}D peak impacts the ability of PARAFAC to handle deviations from trilinearity. Accordingly, the impact of M_R on PARAFAC quantification along the *TDR* curve is illustrated in Fig. 5.3. *TDR* curves for analytes sampled in-phase with various modulation ratios, $M_R = 2$, 3, 4, 6 and 10, are shown in Fig. 5.3(A). Notably, PARAFAC performance decreases nonlinearly with increasing M_R along the *TDR* curves, and more so at higher *TDR*. This result is explained by the way that analyte signal is distributed between the modulations of a given ^{1}D peak. For low modulation ratios (eg., $M_R = 2$), nearly all of the analyte signal is distributed between one or two modulations, one major peak for in-phase sampling or two equally intense peaks for out-of-phase sampling, with few outer modulations containing only a small amount of the analyte signal. Whereas, peaks with large modulation ratios (g., $M_R = 10$), have analyte signal more evenly distributed over a large number of modulations. The cumulative $\Delta^2 t_R$ between the first and last modulations of a peak with $M_R = 10$ is approximately 5 times

greater than the cumulative $\Delta^2 t_R$ for a peak with $M_R = 2$ at the same *TDR*. This is illustrated in Fig. 5.3(B) for GC × GC peaks with a *TDR* = 0.05, in which a variety of M_R and the two sampling phase extremes are shown; in particular, compare peak 1 with $M_R = 10$ and peak 2 with $M_R = 2$ (both peaks sampled in-phase). Visually, at a *TDR* of 0.05 the peak shifting along ^{2}D is very modest for peak 2 at a $M_R = 2$, and as indicated in Fig. 5.3(A) the error is approximately 0 percent. In contrast, for peak 1 at a $M_R = 10$, the peak shifting along ^{2}D is more obvious, and the error is nearly 3 percent. If the GC × GC separation conditions result in a higher *TDR* = 0.10, the ^{2}D peak shifting is more obvious, as illustrated in Fig. 5.3(C), and accordingly the errors are larger (more so at the higher M_R as indicated in Fig. 5.3(A)).

As shown above, in-phase peaks with $M_R = 2$ are modeled well by PARAFAC, even for a relatively large *TDR*. However, it is evident from Fig. 5.3(C) that peak sampling phase must also be considered; this can be observed by comparing peaks 2 and 3. Both are modeled with $M_R = 2$ but with different sampling phase, peak 2 is sampled in-phase and peak 3 is sampled out-of-phase. It is likely that the percent error for PARAFAC is impacted by sampling phase, as in other quantitative methods [8,59–61]. Figure 5.4 provides the results of *TDR* curves with in-phase and out-of-phase sampling considered for peaks with $M_R = 2, 3, 4$. For $M_R = 2$ and 3 out-of-phase sampling significantly reduces the percent error for PARAFAC along the *TDR* curve, with greater differences for a smaller M_R. However, when $M_R = 4$ the *TDR* curves for in-phase and out-of-phase are almost entirely overlapped. The overlap of the two *TDR* curves at $M_R = 4$ in Fig. 5.4 shows that PARAFAC performance, while still dependent on M_R, is essentially independent of sampling phase for $M_R \geq 4$ (higher M_R not shown for brevity)

All data simulations analyzed to prepare Figs. 5.2 through 5.4 were generated with an infinite *S/N*, which facilitated the rigorous examination of the theoretical effects of each

relevant chromatographic property of interest on PARAFAC performance. Now that PARAFAC performance in terms of percent error has been established in the context of *TDR*, M_R, and the two sampling phase extremes, we now turn our attention to also include the impact of detected noise. Figures. 5.5(A) and (B) show the PARAFAC results for simulated GC × GC – TOFMS data of tert-butylbenzene at *S/N* = 2, 5, 100 and infinite, for in-phase sampling, at M_R = 4 and M_R = 2, respectively. For visualization purposes (similar to Fig. 5.1(A) and (B)), the raw "unfolded" GC × GC data at these *S/N* are provided. The *S/N* is defined as the signal height for the maximum peak within the cluster of ^{2}D peaks divided by three times the standard deviation of the Gaussian-distributed baseline noise. Inspection of Figs. 5.5(A) and (B) shows that, while the general shapes of the *TDR* curves are the same, when noise is present in the data, at any S/N ratio, the slope of the curve is greater than without noise. As with infinite *S/N*, the S/N = 100 converges to an error of 0 percent at *TDR* = 0, but has errors slightly larger than the infinite S/N case for *TDR* > 0. As expected, there is a slight positive bias in the PARAFAC results that is greater for lower S/N data [35]. This positive percent error is due to the use of non-negativity constraints in the PARAFAC model and the inability of these constraints to deal with negative noise. It should also be noted that the magnitude of the positive bias was found to be independent of the M_R. Figures. 5.5(A) and (B) also suggest that the positive bias is independent of *TDR*, as it is present even for perfectly trilinear data and the difference between the *TDR* curves of each S/N ratio is approximately constant for all *TDR*s. Analysis was also performed on out-of-phase sampled peaks and the observed bias was shown to be independent of analyte phase sampling (not shown for brevity).

We now consider utilizing GC × GC – TOFMS using typically implemented separation conditions (specifically M_R, the ^2k' range on ^{2}D, and the P_M applied), and how these

conditions manifest themselves in terms of GC × GC peak capacity, $n_{c,2D}$, and the *TDR* range (with accompanying higher percent error for PARAFAC at the higher *TDR*). The Theory section is implemented for this portion of the study, using typical separation and thermodynamic values for a generic test analyte provided in Table 5.2. For this purpose, we presume the analyst has selected an appropriate P_M such that the entire GC × GC separation space is utilized in conjunction with the ^2k' range of interest to maximize the use of the available peak capacity on ^{2}D. In practice, an optimal P_M is roughly equivalent to the ^2k' range, given by the difference in retention time of the least retained peak that can hypothetically elute at $^2t_0$, and the most retained peak that elutes with retention time at a maximum ^2k' without wraparound longer than the P_M (i.e., so as to avoid overlap with lesser retained peaks on ^{2}D). Under these commonly implemented experimental constraints relating P_M and ^2k', using the data in Table 5.2, Fig. 5.6(A) shows that increasing P_M, as a function of ^2k', results in greater changes in the ^{2}D retention time between modulations, given by Δ^2t_R (defined in Eq. (5.3)). This result is explained by simultaneous consideration of Eqs. (5.3) and (5.4), since the temperature change from one modulation to the next, ΔT, depends linearly on P_M (for a given temperature programming rate, T_{ramp}), and the ^2k' range is also roughly equal to P_M. The bottom line is that a short P_M concurrent with a smaller ^2k' range results in relatively small amounts of retention time shifting on ^{2}D (5 ms or less per modulation), whereas the opposite is true for use of a longer P_M with a larger ^2k' range, where 50 to 60 ms shifting per modulation is possible for the more retained analytes.

The *TDR* also depends on ^2W_b and not just Δ^2t_R, so next we consider the ^2W_b dependence on ^2k' using the data for typical analyte values and conditions in Table 5.2. Fig. 5.6(B) depicts the modeled ^2W_b, versus k' as described by Eq. (5.8), and illustrates the linearity of the

relationship between ${}^2k$' and ${}^2W_b$ for an isothermal separation. ${}^2W_b$ only depends on P_M to the extent that a certain ${}^2k$' range is chromatographically realized for the appropriately selected P_M. By combining the plots in Figs. 5.6(A) and (B), the theoretical *TDR* dependence on ${}^2k$' for each P_M is presented in Fig. 5.6(C), as defined in Eq. (5.9). As is evident, each curve in Fig. 5.6(C) is approaching a maximum, independent of ${}^2k$', as suggested by Eq. (5.10). Figure 5.6(C) provides considerable insight into the discussion of data structure trilinearity (or lack thereof) and the practice of using the GC × GC – TOFMS instrument. Principally, use of a longer P_M of 5 s to 6 s will result in *TDR* ranging from 0 to 0.2 with ${}^2k$' ranging from 0 to 5. In contrast, use of P_M of 1 s to 2 s will result in *TDR* ranging from 0 to only 0.05 (with very small PARAFAC percent error) coupled with a smaller ${}^2k$' range of 0 to 1.5. In order to facilitate generation of sufficiently trilinear GC × GC – TOFMS data followed by PARAFAC, separations optimized for low P_M are required, as has been implemented in previous reports [22,23,33,35,41,65,66].

Next, we relate the results presented in Fig. 5.6(C) in terms of the chosen P_M, the ${}^2k$' range, and subsequent *TDR* range, to the peak capacities produced for the ^{1}D, ^{2}D, and the overall GC × GC separation, given by ${}^1n_c$, ${}^2n_c$ and $n_{c,2D}$, respectively. This discussion will be in terms of the theoretical GC × GC peak capacity $n_{c,2D}$ that has been corrected for undersampling of the ^{1}D peaks from the modulation process, given by the product of ${}^1n_c$ with ${}^2n_c$, divided by the undersampling correction factor β [67]. Here, a short P_M refers to 1 s to 2 s, while a long P_M refers to 5 s to 6 s. It was shown in Fig. 5.6 that a short P_M results in GC × GC – TOFMS data that is more trilinear than a long P_M; however, P_M selection must also incorporate the theoretical peak capacities of both separation dimensions as well as the overall GC × GC separation. Figure 5.7(A) shows the dependence of these peak capacities on P_M, for M_R = 2 and M_R = 4 since this encompasses the range of the most commonly applied M_R. The ^{2}D peak capacity, ${}^2n_c$, is

independent of M_R and increases nonlinearly with P_M. The nonlinear increase in ${}^2n_c$ is a direct consequence of each ^{2}D separation being essentially isothermal, and since ${}^2W_b$ increases linearly with ${}^2k$' (see Fig. 5.6(B)), the condition of unit chromatographic resolution R_s between adjacent peaks dictates that the cumulative ${}^2n_c$ increases at a slower rate as ${}^2k$' increases (see Fig. 5.7(B)). The nonlinearity of the increase in ${}^2n_c$ with ${}^2k$' means that increasing the P_M from 1 s to 6 s only results in an increase in ${}^2n_c$ from 12 to 15 to between 30 and 35. While increasing P_M does result in greater chemical selectivity along ^{2}D, this comes at the cost of chemical selectivity along ^{1}D. Indeed, keeping the run time on ^{1}D constant for an objective comparison (with the representative conditions in Table 5.2), the ^{1}D peak capacity, ${}^1n_c$, is 500 to 600 for a short P_M of 1 s to 2 s, and ${}^1n_c$ is substantially reduced to 100 to 150 for a long P_M of 5 s to 6 s. Reduction in ${}^1n_c$ as P_M is increased is a direct consequence of ${}^1W_b$ needing to be experimentally broadened to provide a suitable M_R. Again, in this comparison we are comparing these two regimes in P_M objectively on common ground of keeping M_R constant. For example, if a P_M = 1 s is applied, then a ${}^1W_b$ = 4 s provides a M_R = 4. To achieve the same M_R = 4 with a P_M = 6 s would require the analyst to purposely broaden the ^{1}D peaks to ${}^1W_b$ = 24 s. However, the practice of using a relatively long P_M of 5 s to 6 s is common, and the reality is that most practitioners achieve a M_R approximately equal to 2 (with ${}^1W_b$ of 10 s to 12 s) and not M_R of 4. Ironically, implementation of the longer P_M also results in a slightly lower overall $n_{c,2D}$ compared to using a shorter P_M, since the increase in peak capacity on ^{2}D, is outweighed by the decrease in peak capacity on ^{1}D. This is evident in Fig. 5.7(A) in which $n_{c,2D}$ is 6000 to 7500 for short P_M, and $n_{c,2D}$ is 3000 to 4500 for long P_M. In summary for Figs. 5.6 and 7, a P_M in the 1 s to 2 s range provides an optimized peak capacity for the ^{1}D separation (with no purposely invoked broadening), with an adequate peak capacity for the ^{2}D separation, concurrent with an optimized overall $n_{c,2D}$, and with sufficiently low *TDR*

values that allows PARAFAC to perform successfully even for chromatographically overlapped analyte peaks [22,23,33,35,41,65,66,68]. If the analyst wishes to use a long P_M of 6 s, presumably to provide the larger peak capacity on ^{2}D while paying the price with a reduced overall peak capacity $n_{c,2D}$, then the data will not be sufficiently trilinear for peaks eluting with a high ^2k', and PARAFAC is no longer appropriate for these peaks since their *TDR* will suggest a high percent error for quantification. For these cases with a high *TDR*, either localized retention time alignment prior to applying PARAFAC is advised, or other chemometric methods such as MCR-ALS or PARAFAC2 may be preferred [48,50,53–55].

Experimental data is now presented in order to provide examples of representative GC × GC peaks for various ^2k' and their subsequent *TDR*. A GC × GC – TOFMS TIC chromatogram of a diesel separation with P_M = 1 s is shown in Figs. 5.8(A) and (B). The original chromatogram in Fig. 5.8(A) has been reregistered in Fig. 5.8(B) to remove the appearance of the wrap-around from the separation and to show the true relative retention times of the second dimension separation. For the experimental separation in Fig. 5.8(A), the average ^1W_b = 3.5 s and the average ^2W_b = 80 ms, resulting in a ^1n_c = 510, a ^2n_c = 13, and a $n_{c,2D}$ is approximately 6600, which is consistent with the modeling presented in Fig. 5.7(A). Figure 5.8(C) provides zoomed-in views of four representative analyte peaks. The measured *TDR* values are 0.02 for tetradecane, 0.03 for anthracene, and 0.04 for naphthalene and tetrahydro-naphthalene. The peaks in this data set have M_R of 3 to 4. There is some lack of precision in the ^{2}D retention time measurement because of the narrow peak widths, 70 ms to 100 ms, and the applied scan rate of 100 spectra/s. However, even with the lack of precision in the retention time measurement, which can exaggerate the observed *TDR*, the *TDR*s in this data are still relatively small. Based on these observed *TDR* values and the results from Figs. 5.2 through 5.5, PARAFAC would result in

quantitative errors of less than 1 percent while maintaining a high peak capacity and chromatographic selectivity in both separation dimensions. Indeed, severely overlapped peaks with *TDR* values ranging from 0 to 0.06 from numerous previous studies have produced linear calibration plots from quantification with PARAFAC [23,33,35,65].

5.5 CONCLUSIONS

We demonstrate that retention time shifting from one modulation to the next, $\Delta^2 t_R$, due to temperature programming with GC × GC – TOFMS, which is the principal mechanism of non-trilinearity within a single GC × GC – TOFMS data cube, is not sufficient alone to quantitatively describe the trilinearity of the data cube for the purposes of predicting the performance of PARAFAC. We demonstrate that the analyte peak width on second dimension separations, ${}^2W_b$, also impacts trilinearity, along with $\Delta^2 t_R$. The term Trilinearity Deviation Ratio, *TDR*, which is $\Delta^2 t_R$ normalized by ${}^2W_b$, was introduced as a quantitative metric to assess PARAFAC accuracy with respect to data trilinearity. PARAFAC models of GC × GC – TOFMS data cubes with *TDR* values ranging from 0 to 0.2 were evaluated quantitatively relative to the known simulated peak sum signal. We objectively studied how modulation ratio M_R, modulation period P_M, temperature programming rate, sampling phase (in-phase and out-of-phase), and *S/N*, all play a role in PARAFAC performance in the context of *TDR* for purely resolved peaks and set an objective framework for further studies to investigate the effects of *TDR* on PARAFAC of GC × GC – TOFMS data cubes with peak overlap. Use of a P_M in the 1 s to 2 s range optimizes peak capacity for the first dimension separation (500 to 600) for a 30 min run, with an adequate peak capacity for the second dimension separation (12 to 15), concurrent with an optimized two-dimensional peak capacity (6000 to 7500), combined with low *TDR* values (0 to 0.05) to

facilitate low quantitative errors with PARAFAC (0 to 0.5 percent). In contrast, use of a P_M in the 5 s or greater range provides a higher peak capacity on the second dimension (30 to 35), concurrent with a lower peak capacity on the first dimension (100 to 150) for a 30 min run, and a slightly reduced two-dimensional peak capacity (3000 to 4500); furthermore, the data is not sufficiently trilinear for the more retained second dimension peaks in order to use PARAFAC without a localized retention time alignment step, or alternatively one may elect to apply other chemometric methods such as MCR-ALS or PARAFAC2 in such cases. The *TDR* concept in principle is not limited to GC × GC – TOFMS, and should be readily extendable to other comprehensive two-dimensional separation techniques, and can be reduced to bilinearity deviation ratio, *BDR*, for univariate detection methods.

5.7 REFERENCES

[1] T. Górecki, J. Harynuk, O. Panić, The evolution of comprehensive two-dimensional gas chromatography (GC×GC), J. Sep. Sci. 27 (2004) 359–379. doi:10.1002/jssc.200301650.

[2] J.B. Phillips, J. Beens, Comprehensive two-dimensional gas chromatography: a hyphenated method with strong coupling between the two dimensions, J. Chromatogr. A. 856 (1999) 331–347. doi:10.1016/S0021-9673(99)00815-8.

[3] R.M. Kinghorn, P.J. Marriott, P.A. Dawes, Design and Implementation of Comprehensive Gas Chromatography with Cryogenic Modulation, J. High Resolut. Chromatogr. 23 (2000) 245–252. doi:10.1002/(SICI)1521-4168(20000301)23:3<245::AID-JHRC245>3.0.CO;2-E.

[4] C.A. Bruckner, B.J. Prazen, R.E. Synovec, Comprehensive Two-Dimensional High-Speed Gas Chromatography with Chemometric Analysis, Anal. Chem. 70 (1998) 2796–2804. doi:10.1021/ac980164m.

[5] Z. Liu, J.B. Phillips, Comprehensive Two-Dimensional Gas Chromatography using an On-Column Thermal Modulator Interface, J. Chromatogr. Sci. 29 (1991) 227–231. doi:10.1093/chromsci/29.6.227.

[6] R.M. Kinghorn, P.J. Marriott, Comprehensive Two-Dimensional Gas Chromatography Using a Modulating Cryogenic Trap, J. High Resolut. Chromatogr. 21 (1998) 620–622. doi:10.1002/(SICI)1521-4168(19981101)21:11<620::AID-JHRC620>3.0.CO;2-#.

[7] J. Beens, M. Adahchour, R.J.J. Vreuls, K. van Altena, U.A. Th. Brinkman, Simple, non-moving modulation interface for comprehensive two-dimensional gas chromatography, J. Chromatogr. A. 919 (2001) 127–132. doi:10.1016/S0021-9673(01)00785-3.

[8] J.V. Seeley, F. Kramp, C.J. Hicks, Comprehensive Two-Dimensional Gas Chromatography via Differential Flow Modulation, Anal. Chem. 72 (2000) 4346–4352. doi:10.1021/ac000249z.

[9] A.E. Sinha, K.J. Johnson, B.J. Prazen, S.V. Lucas, C.G. Fraga, R.E. Synovec, Comprehensive two-dimensional gas chromatography of volatile and semi-volatile components using a diaphragm valve-based instrument, J. Chromatogr. A. 983 (2003) 195–204. doi:10.1016/S0021-9673(02)01651-5.

[10] A.E. Sinha, B.J. Prazen, C.G. Fraga, R.E. Synovec, Valve-based comprehensive two-dimensional gas chromatography with time-of-flight mass spectrometric detection: instrumentation and figures-of-merit, J. Chromatogr. A. 1019 (2003) 79–87. doi:10.1016/j.chroma.2003.08.047.

[11] A.E. Sinha, C.G. Fraga, B.J. Prazen, R.E. Synovec, Trilinear chemometric analysis of two-dimensional comprehensive gas chromatography–time-of-flight mass spectrometry data, J. Chromatogr. A. 1027 (2004) 269–277. doi:10.1016/j.chroma.2003.08.081.

[12] A.E. Sinha, J.L. Hope, B.J. Prazen, E.J. Nilsson, R.M. Jack, R.E. Synovec, Algorithm for locating analytes of interest based on mass spectral similarity in GC × GC–TOF-MS data: analysis of metabolites in human infant urine, J. Chromatogr. A. 1058 (2004) 209–215. doi:10.1016/j.chroma.2004.08.064.

[13] J. Dallüge, M. van Rijn, J. Beens, R.J.J. Vreuls, U.A.T. Brinkman, Comprehensive two-dimensional gas chromatography with time-of-flight mass spectrometric detection applied to the determination of pesticides in food extracts, J. Chromatogr. A. 965 (2002) 207–217. doi:10.1016/S0021-9673(01)01324-3.

[14] X. Lu, J. Cai, H. Kong, M. Wu, R. Hua, M. Zhao, J. Liu, G. Xu, Analysis of Cigarette Smoke Condensates by Comprehensive Two-Dimensional Gas Chromatography/Time-of-Flight Mass Spectrometry I Acidic Fraction, Anal. Chem. 75 (2003) 4441–4451. doi:10.1021/ac0264224.
[15] J.-F. Focant, A. Sjödin, W.E. Turner, D.G. Patterson, Measurement of Selected Polybrominated Diphenyl Ethers, Polybrominated and Polychlorinated Biphenyls, and Organochlorine Pesticides in Human Serum and Milk Using Comprehensive Two-Dimensional Gas Chromatography Isotope Dilution Time-of-Flight Mass Spectrometry, Anal. Chem. 76 (2004) 6313–6320. doi:10.1021/ac048959i.
[16] S.M. Song, P. Marriott, A. Kotsos, O.H. Drummer, P. Wynne, Comprehensive two-dimensional gas chromatography with time-of-flight mass spectrometry (GC × GC-TOFMS) for drug screening and confirmation, Forensic Sci. Int. 143 (2004) 87–101. doi:10.1016/j.forsciint.2004.02.042.
[17] J.L. Hope, A.E. Sinha, B.J. Prazen, R.E. Synovec, Evaluation of the DotMap algorithm for locating analytes of interest based on mass spectral similarity in data collected using comprehensive two-dimensional gas chromatography coupled with time-of-flight mass spectrometry, J. Chromatogr. A. 1086 (2005) 185–192. doi:10.1016/j.chroma.2005.06.026.
[18] E. Jover, M. Adahchour, J.M. Bayona, R.J.J. Vreuls, U.A.T. Brinkman, Characterization of lipids in complex samples using comprehensive two-dimensional gas chromatography with time-of-flight mass spectrometry, J. Chromatogr. A. 1086 (2005) 2–11. doi:10.1016/j.chroma.2005.05.093.
[19] J. Dallüge, R.J.J. Vreuls, J. Beens, U.A.T. Brinkman, Optimization and characterization of comprehensive two-dimensional gas chromatography with time-of-flight mass spectrometric detection (GC×GC–TOF MS), J. Sep. Sci. 25 (2002) 201–214. doi:10.1002/1615-9314(20020301)25:4<201::AID-JSSC201>3.0.CO;2-B.
[20] L. Mondello, P.Q. Tranchida, P. Dugo, G. Dugo, Comprehensive two-dimensional gas chromatography-mass spectrometry: A review, Mass Spectrom. Rev. 27 (2008) 101–124. doi:10.1002/mas.20158.
[21] J.L. Hope, B.J. Prazen, E.J. Nilsson, M.E. Lidstrom, R.E. Synovec, Comprehensive two-dimensional gas chromatography with time-of-flight mass spectrometry detection: analysis of amino acid and organic acid trimethylsilyl derivatives, with application to the analysis of metabolites in rye grass samples, Talanta. 65 (2005) 380–388. doi:10.1016/j.talanta.2004.06.025.
[22] R.E. Mohler, K.M. Dombek, J.C. Hoggard, E.T. Young, R.E. Synovec, Comprehensive Two-Dimensional Gas Chromatography Time-of-Flight Mass Spectrometry Analysis of Metabolites in Fermenting and Respiring Yeast Cells, Anal. Chem. 78 (2006) 2700–2709. doi:10.1021/ac052106o.
[23] L.C. Marney, S.C. Kolwicz Jr., R. Tian, R.E. Synovec, Sample preparation methodology for mouse heart metabolomics using comprehensive two-dimensional gas chromatography coupled with time-of-flight mass spectrometry, Talanta. 108 (2013) 123–130. doi:10.1016/j.talanta.2013.03.005.
[24] A.E. Sinha, B.J. Prazen, R.E. Synovec, Trends in chemometric analysis of comprehensive two-dimensional separations, Anal. Bioanal. Chem. 378 (2004) 1948–1951. doi:10.1007/s00216-004-2503-7.

[25] K.M. Pierce, J.C. Hoggard, R.E. Mohler, R.E. Synovec, Recent advancements in comprehensive two-dimensional separations with chemometrics, J. Chromatogr. A. 1184 (2008) 341–352. doi:10.1016/j.chroma.2007.07.059.
[26] L. Xie, P.J. Marriott, M. Adams, Chemometric analysis of comprehensive two-dimensional gas chromatography data using cryogenic modulation, Anal. Chim. Acta. 500 (2003) 211–222. doi:10.1016/j.aca.2003.07.002.
[27] K.M. Pierce, B. Kehimkar, L.C. Marney, J.C. Hoggard, R.E. Synovec, Review of chemometric analysis techniques for comprehensive two dimensional separations data, J. Chromatogr. A. 1255 (2012) 3–11. doi:10.1016/j.chroma.2012.05.050.
[28] Z. Zeng, J. Li, H.M. Hugel, G. Xu, P.J. Marriott, Interpretation of comprehensive two-dimensional gas chromatography data using advanced chemometrics, TrAC Trends Anal. Chem. 53 (2014) 150–166. doi:10.1016/j.trac.2013.08.009.
[29] J. Dallüge, J. Beens, U.A.T. Brinkman, Comprehensive two-dimensional gas chromatography: a powerful and versatile analytical tool, J. Chromatogr. A. 1000 (2003) 69–108. doi:10.1016/S0021-9673(03)00242-5.
[30] K.M. Pierce, J.L. Hope, J.C. Hoggard, R.E. Synovec, A principal component analysis based method to discover chemical differences in comprehensive two-dimensional gas chromatography with time-of-flight mass spectrometry (GC × GC-TOFMS) separations of metabolites in plant samples, Talanta. 70 (2006) 797–804. doi:10.1016/j.talanta.2006.01.038.
[31] C.G. Fraga, C.A. Bruckner, R.E. Synovec, Increasing the Number of Analyzable Peaks in Comprehensive Two-Dimensional Separations through Chemometrics, Anal. Chem. 73 (2001) 675–683. doi:10.1021/ac0010025.
[32] C. von Mühlen, C.A. Zini, E.B. Caramão, P.J. Marriott, Applications of comprehensive two-dimensional gas chromatography to the characterization of petrochemical and related samples, J. Chromatogr. A. 1105 (2006) 39–50. doi:10.1016/j.chroma.2005.09.036.
[33] J.C. Hoggard, R.E. Synovec, Parallel Factor Analysis (PARAFAC) of Target Analytes in GC × GC−TOFMS Data: Automated Selection of a Model with an Appropriate Number of Factors, Anal. Chem. 79 (2007) 1611–1619. doi:10.1021/ac061710b.
[34] J.C. Hoggard, W.C. Siegler, R.E. Synovec, Toward automated peak resolution in complete GC × GC–TOFMS chromatograms by PARAFAC, J. Chemom. 23 (2009) 421–431. doi:10.1002/cem.1239.
[35] J.C. Hoggard, R.E. Synovec, Automated Resolution of Nontarget Analyte Signals in GC × GC-TOFMS Data Using Parallel Factor Analysis, Anal. Chem. 80 (2008) 6677–6688. doi:10.1021/ac800624e.
[36] R.E. Mohler, K.M. Dombek, J.C. Hoggard, K.M. Pierce, E.T. Young, R.E. Synovec, Comprehensive analysis of yeast metabolite GC×GC–TOFMS data: combining discovery-mode and deconvolution chemometric software, Analyst. 132 (2007) 756–767. doi:10.1039/B700061H.
[37] E.M. Humston, K.M. Dombek, J.C. Hoggard, E.T. Young, R.E. Synovec, Time-Dependent Profiling of Metabolites from Snf1 Mutant and Wild Type Yeast Cells, Anal. Chem. 80 (2008) 8002–8011. doi:10.1021/ac800998j.
[38] S. Yang, M. Sadilek, R.E. Synovec, M.E. Lidstrom, Liquid chromatography–tandem quadrupole mass spectrometry and comprehensive two-dimensional gas chromatography–time-of-flight mass spectrometry measurement of targeted metabolites of Methylobacterium

extorquens AM1 grown on two different carbon sources, J. Chromatogr. A. 1216 (2009) 3280–3289. doi:10.1016/j.chroma.2009.02.030.
[39] E.M. Humston, J.D. Knowles, A. McShea, R.E. Synovec, Quantitative assessment of moisture damage for cacao bean quality using two-dimensional gas chromatography combined with time-of-flight mass spectrometry and chemometrics, J. Chromatogr. A. 1217 (2010) 1963–1970. doi:10.1016/j.chroma.2010.01.069.
[40] E.M. Humston, J.C. Hoggard, R.E. Synovec, Utilizing the Third Order Advantage with Isotope Dilution Mass Spectrometry, Anal. Chem. 82 (2010) 41–43. doi:10.1021/ac902184b.
[41] L.R. Snyder, J.C. Hoggard, T.J. Montine, R.E. Synovec, Development and application of a comprehensive two-dimensional gas chromatography with time-of-flight mass spectrometry method for the analysis of l-β-methylamino-alanine in human tissue, J. Chromatogr. A. 1217 (2010) 4639–4647. doi:10.1016/j.chroma.2010.04.065.
[42] A.C. Beckstrom, E.M. Humston, L.R. Snyder, R.E. Synovec, S.E. Juul, Application of comprehensive two-dimensional gas chromatography with time-of-flight mass spectrometry method to identify potential biomarkers of perinatal asphyxia in a non-human primate model, J. Chromatogr. A. 1218 (2011) 1899–1906. doi:10.1016/j.chroma.2011.01.086.
[43] E.M. Humston, K.M. Dombek, B.P. Tu, E.T. Young, R.E. Synovec, Toward a global analysis of metabolites in regulatory mutants of yeast, Anal. Bioanal. Chem. 401 (2011) 2387–2402. doi:10.1007/s00216-011-4800-2.
[44] A.C. Beckstrom, P. Tanya, E.M. Humston, L.R. Snyder, R.E. Synovec, S.E. Juul, The perinatal transition of the circulating metabolome in a nonhuman primate, Pediatr. Res. 71 (2012) 338–344. doi:10.1038/pr.2011.74.
[45] S.C. Kolwicz, D.P. Olson, L.C. Marney, L. Garcia-Menendez, R.E. Synovec, R. Tian, Cardiac-Specific Deletion of Acetyl CoA Carboxylase 2 Prevents Metabolic Remodeling During Pressure-Overload Hypertrophy, Circ. Res. 111 (2012) 728–738. doi:10.1161/CIRCRESAHA.112.268128.
[46] C.G. Fraga, L.H. Sego, J.C. Hoggard, G.A.P. Acosta, E.A. Viglino, J.H. Wahl, R.E. Synovec, Preliminary effects of real-world factors on the recovery and exploitation of forensic impurity profiles of a nerve-agent simulant from office media, J. Chromatogr. A. 1270 (2012) 269–282. doi:10.1016/j.chroma.2012.10.053.
[47] J. Omar, M. Olivares, J.M. Amigo, N. Etxebarria, Resolution of co-eluting compounds of Cannabis Sativa in comprehensive two-dimensional gas chromatography/mass spectrometry detection with Multivariate Curve Resolution-Alternating Least Squares, Talanta. 121 (2014) 273–280. doi:10.1016/j.talanta.2013.12.044.
[48] T. Skov, J.C. Hoggard, R. Bro, R.E. Synovec, Handling within run retention time shifts in two-dimensional chromatography data using shift correction and modeling, J. Chromatogr. A. 1216 (2009) 4020–4029. doi:10.1016/j.chroma.2009.02.049.
[49] H. Parastar, N. Akvan, Multivariate curve resolution based chromatographic peak alignment combined with parallel factor analysis to exploit second-order advantage in complex chromatographic measurements, Anal. Chim. Acta. 816 (2014) 18–27. doi:10.1016/j.aca.2014.01.051.
[50] H. Parastar, J.R. Radović, M. Jalali-Heravi, S. Diez, J.M. Bayona, R. Tauler, Resolution and Quantification of Complex Mixtures of Polycyclic Aromatic Hydrocarbons in Heavy

Fuel Oil Sample by Means of GC × GC-TOFMS Combined to Multivariate Curve Resolution, Anal. Chem. 83 (2011) 9289–9297. doi:10.1021/ac201799r.
[51] J.R. Radović, K.V. Thomas, H. Parastar, S. Díez, R. Tauler, J.M. Bayona, Chemometrics-Assisted Effect-Directed Analysis of Crude and Refined Oil Using Comprehensive Two-Dimensional Gas Chromatography–Time-of-Flight Mass Spectrometry, Environ. Sci. Technol. 48 (2014) 3074–3083. doi:10.1021/es404859m.
[52] R. Bro, PARAFAC. Tutorial and applications, Chemom. Intell. Lab. Syst. 38 (1997) 149–171. doi:10.1016/S0169-7439(97)00032-4.
[53] H. Parastar, R. Tauler, Multivariate Curve Resolution of Hyphenated and Multidimensional Chromatographic Measurements: A New Insight to Address Current Chromatographic Challenges, Anal. Chem. 86 (2014) 286–297. doi:10.1021/ac402377d.
[54] A. de Juan, R. Tauler, Comparison of three-way resolution methods for non-trilinear chemical data sets, J. Chemom. 15 (2001) 749–771. doi:10.1002/cem.662.
[55] R. Tauler, Multivariate curve resolution applied to second order data, Chemom. Intell. Lab. Syst. 30 (1995) 133–146. doi:10.1016/0169-7439(95)00047-X.
[56] W. Khummueng, J. Harynuk, P.J. Marriott, Modulation Ratio in Comprehensive Two-dimensional Gas Chromatography, Anal. Chem. 78 (2006) 4578–4587. doi:10.1021/ac052270b.
[57] H. Parastar, M. Jalali-Heravi, R. Tauler, Comprehensive two-dimensional gas chromatography (GC × GC) retention time shift correction and modeling using bilinear peak alignment, correlation optimized shifting and multivariate curve resolution, Chemom. Intell. Lab. Syst. 117 (2012) 80–91. doi:10.1016/j.chemolab.2012.02.003.
[58] T.J. Bahowick, R.E. Synovec, Correlation of Quantitative Analysis Precision to Retention Time Precision and Chromatographic Resolution for Rapid, Short-Column Analysis, Anal. Chem. 67 (1995) 631–640. doi:10.1021/ac00099a022.
[59] J.V. Seeley, Theoretical study of incomplete sampling of the first dimension in comprehensive two-dimensional chromatography, J. Chromatogr. A. 962 (2002) 21–27. doi:10.1016/S0021-9673(02)00461-2.
[60] J.V. Seeley, N.J. Micyus, S.V. Bandurski, S.K. Seeley, J.D. McCurry, Microfluidic Deans Switch for Comprehensive Two-Dimensional Gas Chromatography, Anal. Chem. 79 (2007) 1840–1847. doi:10.1021/ac061881g.
[61] W.C. Siegler, B.D. Fitz, J.C. Hoggard, R.E. Synovec, Experimental Study of the Quantitative Precision for Valve-Based Comprehensive Two-Dimensional Gas Chromatography, Anal. Chem. 83 (2011) 5190–5196. doi:10.1021/ac200302b.
[62] M.L. Lee, F.J. Yang, K.D. Bartle, Open Tubular Column Gas Chromatography, John Wiley & Sons, Inc., New York, 1984.
[63] G.M. Gross, B.J. Prazen, J.W. Grate, R.E. Synovec, High-Speed Gas Chromatography Using Synchronized Dual-Valve Injection, Anal. Chem. 76 (2004) 3517–3524. doi:10.1021/ac049909g.
[64] R.B. Wilson, J.C. Hoggard, R.E. Synovec, High throughput analysis of atmospheric volatile organic compounds by thermal injection – isothermal gas chromatography – time-of-flight mass spectrometry, Talanta. 103 (2013) 95–102. doi:10.1016/j.talanta.2012.10.013.
[65] L.C. Marney, W. Christopher Siegler, B.A. Parsons, J.C. Hoggard, B.W. Wright, R.E. Synovec, Tile-based Fisher-ratio software for improved feature selection analysis of

comprehensive two-dimensional gas chromatography–time-of-flight mass spectrometry data, Talanta. 115 (2013) 887–895. doi:10.1016/j.talanta.2013.06.038.
[66] R.E. Mohler, B.P. Tu, K.M. Dombek, J.C. Hoggard, E.T. Young, R.E. Synovec, Identification and evaluation of cycling yeast metabolites in two-dimensional comprehensive gas chromatography–time-of-flight-mass spectrometry data, J. Chromatogr. A. 1186 (2008) 401–411. doi:10.1016/j.chroma.2007.10.063.
[67] J.M. Davis, D.R. Stoll, P.W. Carr, Effect of First-Dimension Undersampling on Effective Peak Capacity in Comprehensive Two-Dimensional Separations, Anal. Chem. 80 (2008) 461–473. doi:10.1021/ac071504j.
[68] J.C. Hoggard, J.H. Wahl, R.E. Synovec, G.M. Mong, C.G. Fraga, Impurity Profiling of a Chemical Weapon Precursor for Possible Forensic Signatures by Comprehensive Two-Dimensional Gas Chromatography/Mass Spectrometry and Chemometrics, Anal. Chem. 82 (2010) 689–698. doi:10.1021/ac902247x.

5.8 TABLES

Table 5.1 Modeling conditions for simulated GC × GC – TOFMS data cubes for analysis by PARAFAC, with results presented in Figs. 5.2 through 5.5.

Parameter	**Conditions**
Analyte	Tert-butylbenzene
Data collection rate	500 Hz (spectra/s)
$\boldsymbol{P_M}$	1 s
$\boldsymbol{^2W_b}$	50, 100, 150, 200, 250, 300 ms
$\boldsymbol{^1W_b}$	2, 3, 4, 5, 6, 10 s
$\boldsymbol{M_R}$	2, 3, 4, 6, 10
Sampling phase	In-phase, Out-of-phase
Signal-to-noise, *S/N*	Infinite, 100, 5, 2
***TDR* range**	0 to 0.2

Table 5.2 Chromatographic conditions used for modeling in Figs. 5.6 and 7 the ΔT values are defined by the T_{ramp} given and the P_M using Eq. (5.4).

Parameter	**Condition**
$\boldsymbol{\Delta H}$	45 kJ/mol
$\boldsymbol{R}$	8.314 J/(mol·K)
$\boldsymbol{T_{ranp}}$	5 °C/min
$\boldsymbol{T_N}$	200 °C
$\boldsymbol{P_M}$	1, 2, 3, 4, 5, 6 s
$\boldsymbol{\Delta T}$	0.083, 0.167, 0.25, 0.333, 0.417, 0.5 °C
$\boldsymbol{D_{g,o}}$	0.4 cm^2/s
j	0.5
f	1
$\boldsymbol{d_c}$	180 µm
$\boldsymbol{^2t_0}$	1 s
$\boldsymbol{^2W_{b,0}}$	60 ms
$\boldsymbol{M_R}$	2, 4

5.9 FIGURES

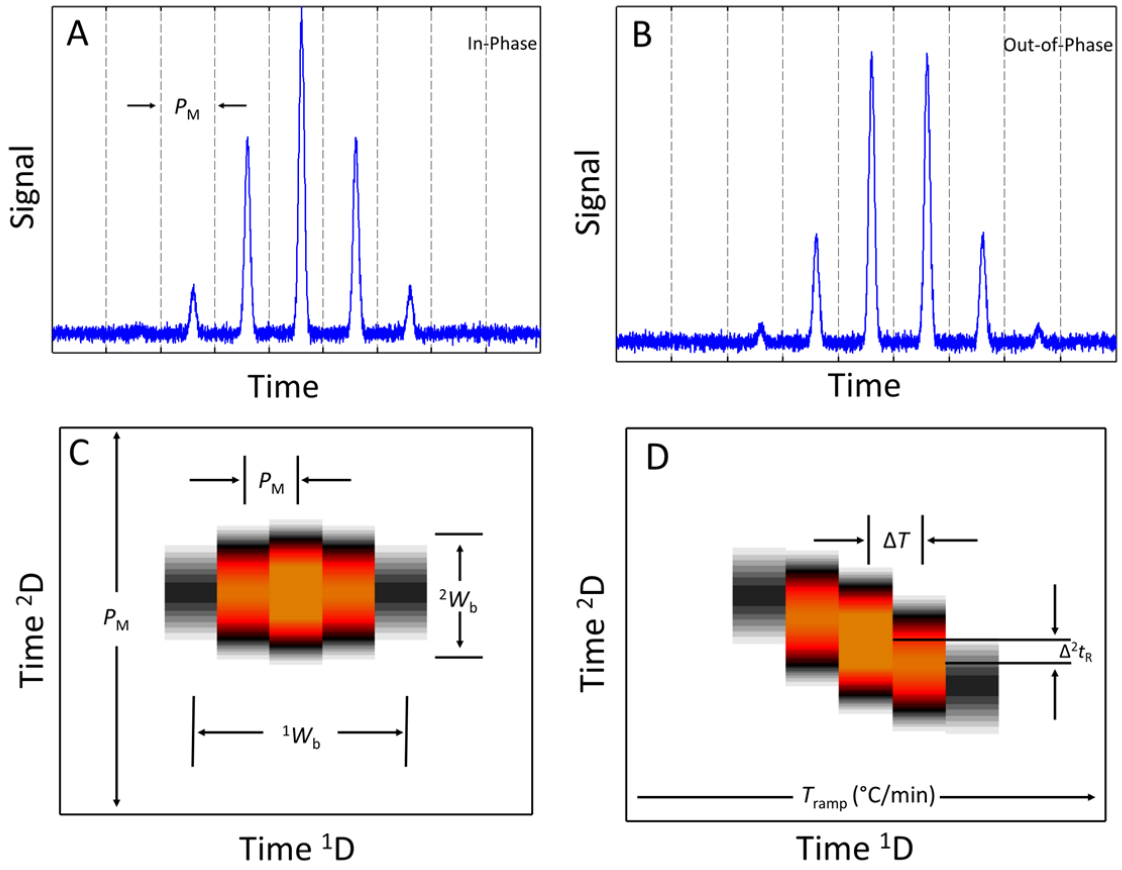

Figure 5.1 Illustrations of GC × GC data with relevant variables: P_M, T_{ramp}, ΔT, $\Delta^2 t_R$, ^2W_b and ^1W_b (A) Raw "unfolded" GC × GC data as collected from the instrument for in-phase sampling at a modulation ratio M_R = 4 per Eq. (5.1), with the location of each modulation shown by vertical dashed lines. (B) GC × GC data as in (A) for out-of-phase sampling. (C) Two-dimensional contour plot of the GC × GC data in (A) with *TDR* = 0, per Eq. (5.5). (D) Two-dimensional contour plot of GC × GC data in (A), now with ^{2}D shifting illustrated with *TDR* = 0.2, $\Delta^2 t_R$ = 40 ms and ^2W_b = 200 ms.

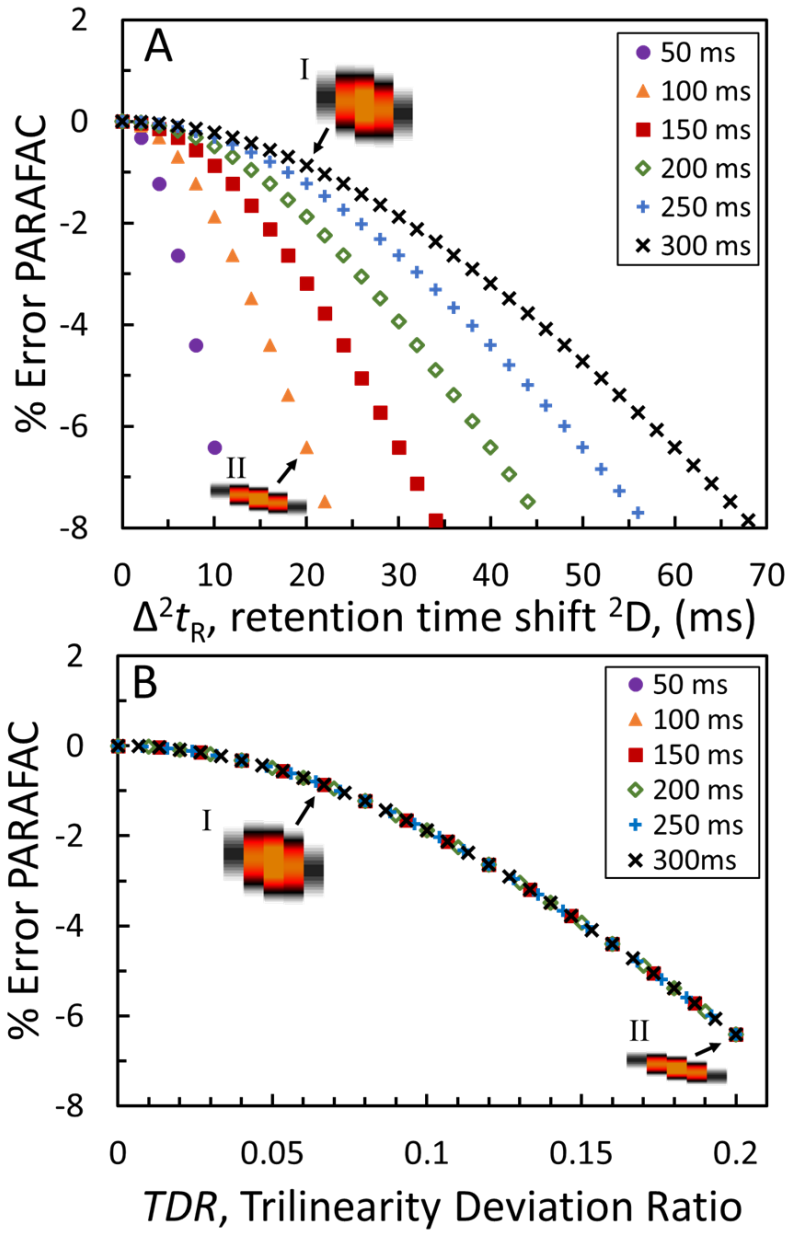

Figure 5.2 Dependence of PARAFAC performance (quantified as the percent error) on the retention time shift $\Delta^2 t_R$ and peak width ${}^2W_b$ on ^{2}D, and *TDR*, as defined in Eq. (5.5). (A) Percent error of the PARAFAC peak sum signal as a function of $\Delta^2 t_R$, in ms, for simulated tert-butylbenzene data, in-phase, $M_R = 4$, at infinite *S/N*, with various ${}^2W_b$ (50, 100, 150, 200, 250, 300 ms). The target analyte total signal was held constant between simulations. The two contour plots represent analyte peaks with ${}^2W_b$ of 300 ms (labeled I) and 100 ms (labeled II), each with a $\Delta^2 t_R = 20$ ms. (B) The same data as (A), but now the percent error is plotted as a function of *TDR*, so the $\Delta^2 t_R$ values have been normalized by ${}^2W_b$ to produce *TDR*.

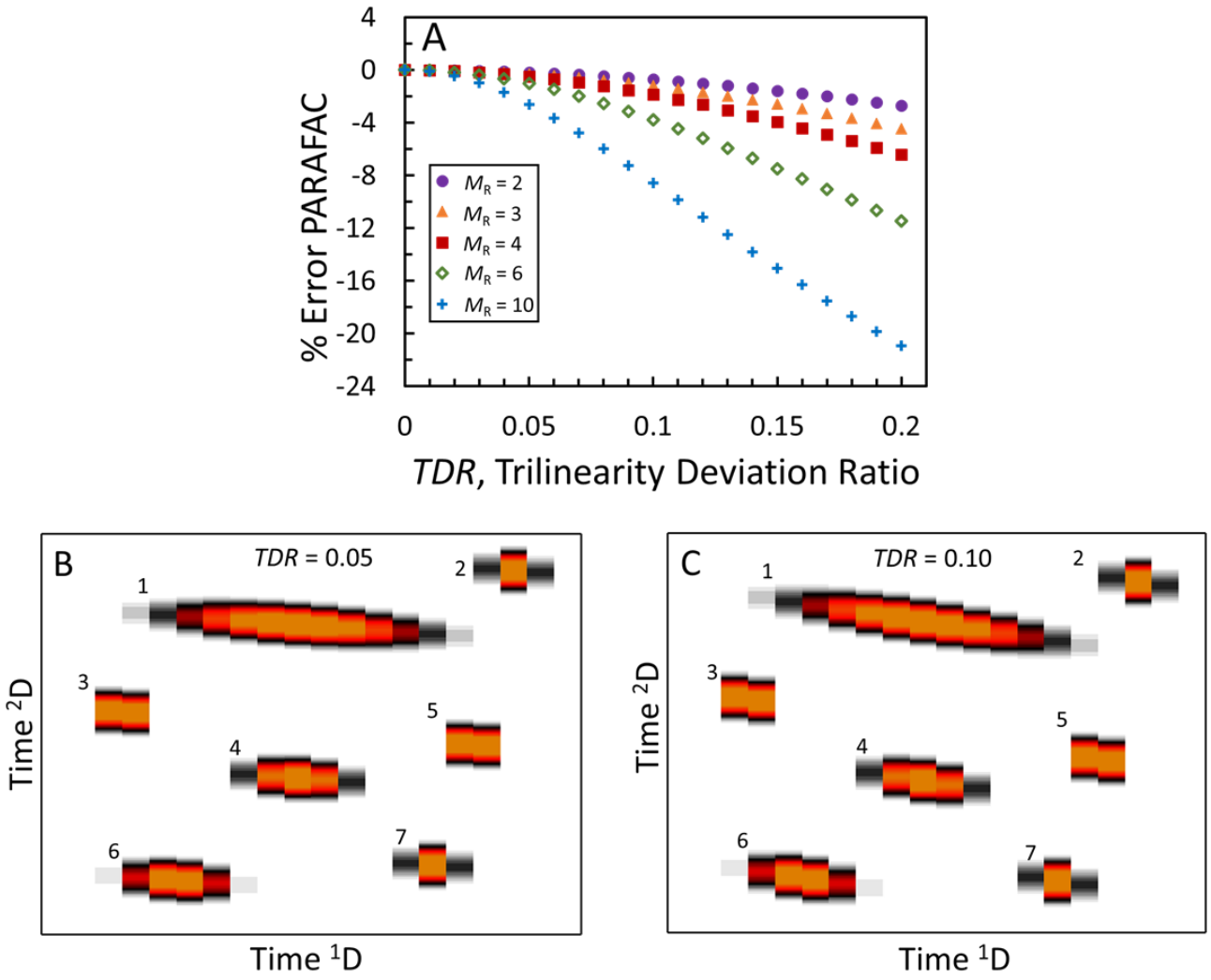

Figure 5.3 Effect of modulation ratio, M_R, on PARAFAC in the context of *TDR*. (A) Percent error as a function of *TDR* (as in Fig. 5.2) for simulated GC × GC – TOFMS data cubes of tert-butylbenzene, at infinite *S/N*, ^2W_b = 200 ms, with various modulation ratios, M_R = 2, 3, 4, 6, 10. The target analyte total signal was held constant between simulations. (B) Simulated GC × GC peaks with a variety of M_R and either in-phase or out-of-phase sampling. $^{\text{Peak \#}}$(M_R, phase): ^{1}(10, in), ^{2}(2, in), ^{3}(2, out), ^{4}(4, in), ^{5}(2, out), ^{6}(4, out), ^{7}(2, in). All peaks at a *TDR* = 0.05. (C) Simulated GC × GC peaks as in (B), but now all at a *TDR* = 0.1.

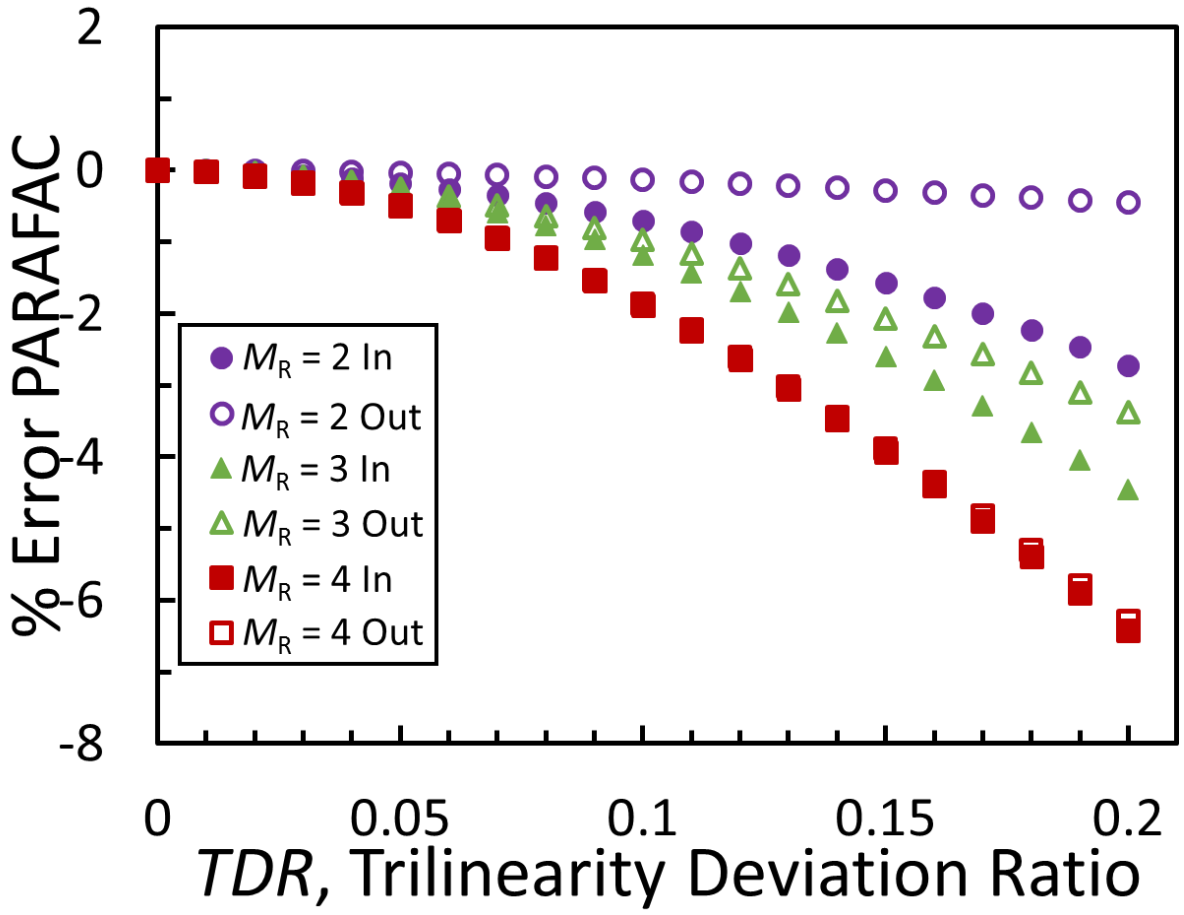

Figure 5.4 Effect of sampling phase on PARAFAC. The percent error of the PARAFAC peak sum signal is plotted as a function of *TDR* for simulated GC × GC – TOFMS data cubes of tert-butylbenzene, at infinite *S/N*, with M_R = 2, 3, 4 sampled either in-phase or out-of-phase as indicated.

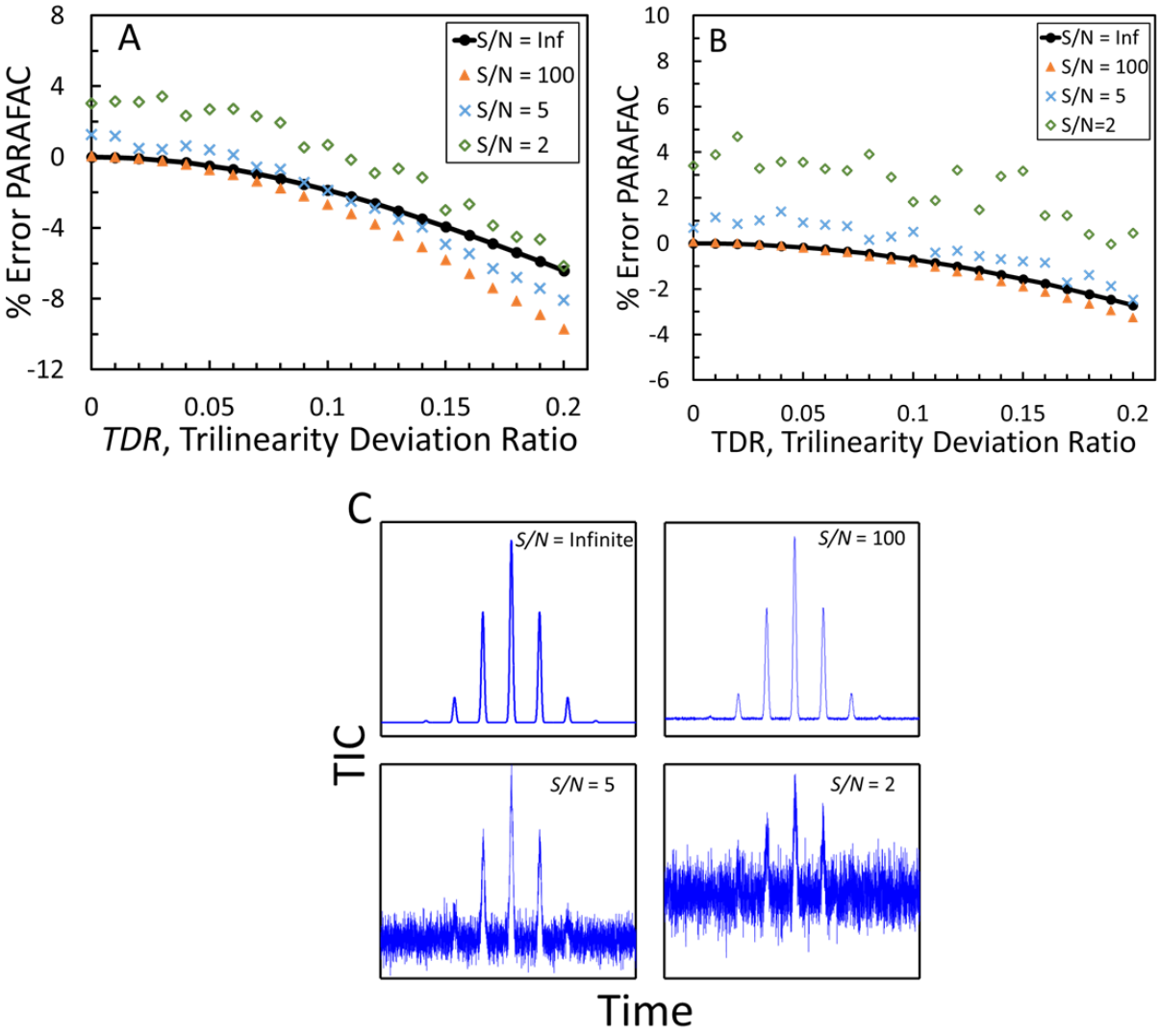

Figure 5.5 *S/N* effect on PARAFAC in context of *TDR*. The percent error of the PARAFAC peak sum signal as a function of *TDR* for simulated GC × GC – TOFMS data cubes of tert-butylbenzene, with *S/N* = 2, 5, 100 and infinite, all sampled in-phase, (A) M_R = 4 and (B) M_R = 2. Each point on each graph is the average of 10 independent PARAFAC models. (C) Simulated unfolded GC × GC – TIC data of tert-butylbenzene at *S/N* = 2, 5, 100 and infinite.

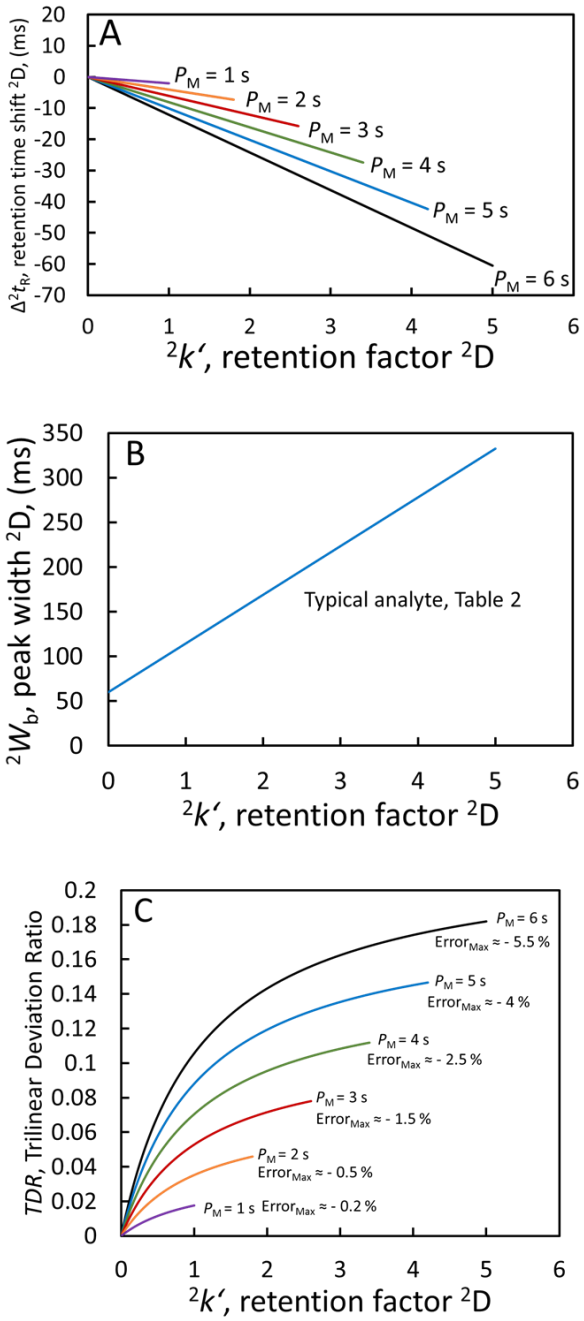

Figure 5.6 Modeled GC × GC data, using Eqs. (5.3), (5.8) and (5.9), for typical analyte and GC × GC – TOFMS parameters (see Table 5.2) for various P_M. (A) Plot of Eq. (5.3), the change in ^{2}D retention time, $\Delta^2 t_R$, as a function of retention factor, $^2k'$, for P_M = 1, 2, 3, 4, 5, 6. (B) Plot of Eq. (5.8), the 2D peak width, ^2W_b, as a function of $^2k'$. (C) Plot of Eq. (5.9), *TDR* as a function of $^2k'$, for P_M = 1, 2, 3, 4, 5, 6.

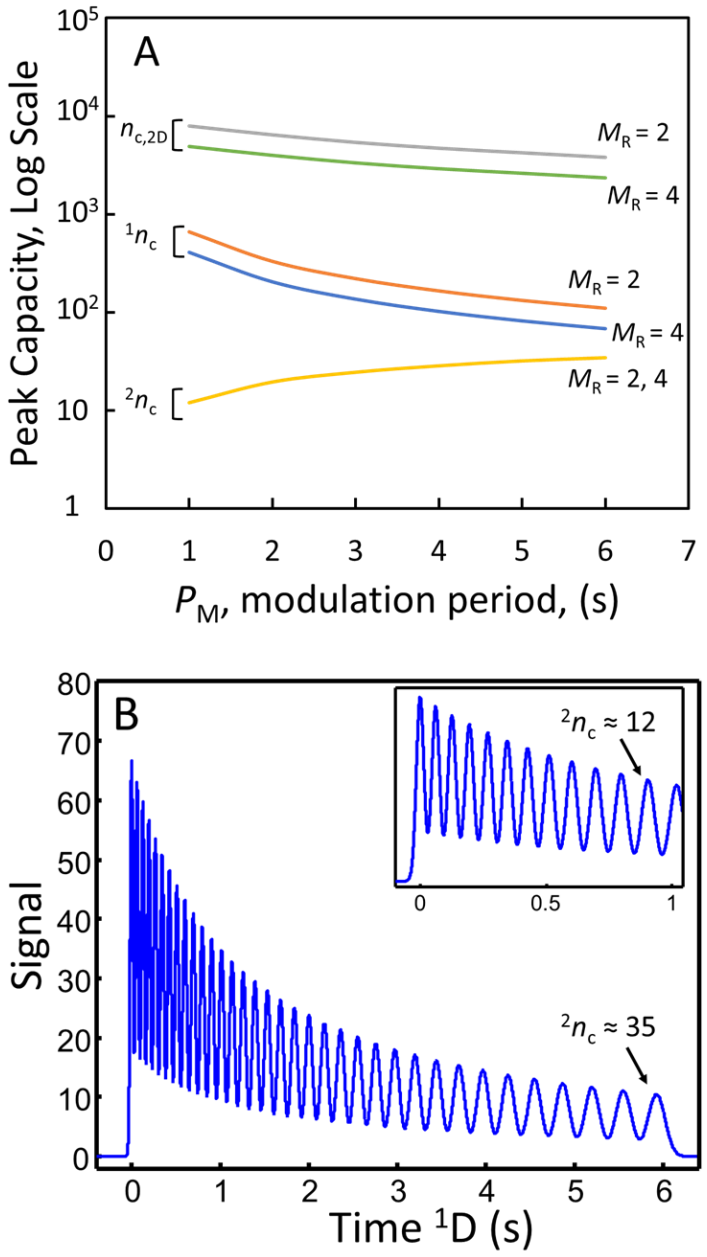

Figure 5.7 (A) Peak capacity as a function of modulation period, P_M, whereby all GC × GC peaks in a given separation have M_R = 2 or 4 as specified, with total run times of 30 min. The ^{2}D separation was simulated based on Fig. 5.6(B). All peaks were simulated at unit resolution, R_s = 1. The curves show the ^{1}D and ^{2}D peak capacities, ^1n_c and ^2n_c respectively, for each M_R, as well as the overall two-dimensional peak capacity, $n_{c,2D}$. (B) Simulated isothermal chromatogram of the ^{2}D separation based on Fig. 5.6(B), with all peaks at R_s = 1. Inset is a smaller time window from the beginning of the separation and illustrates the peak capacity for a short P_M, 1 s.

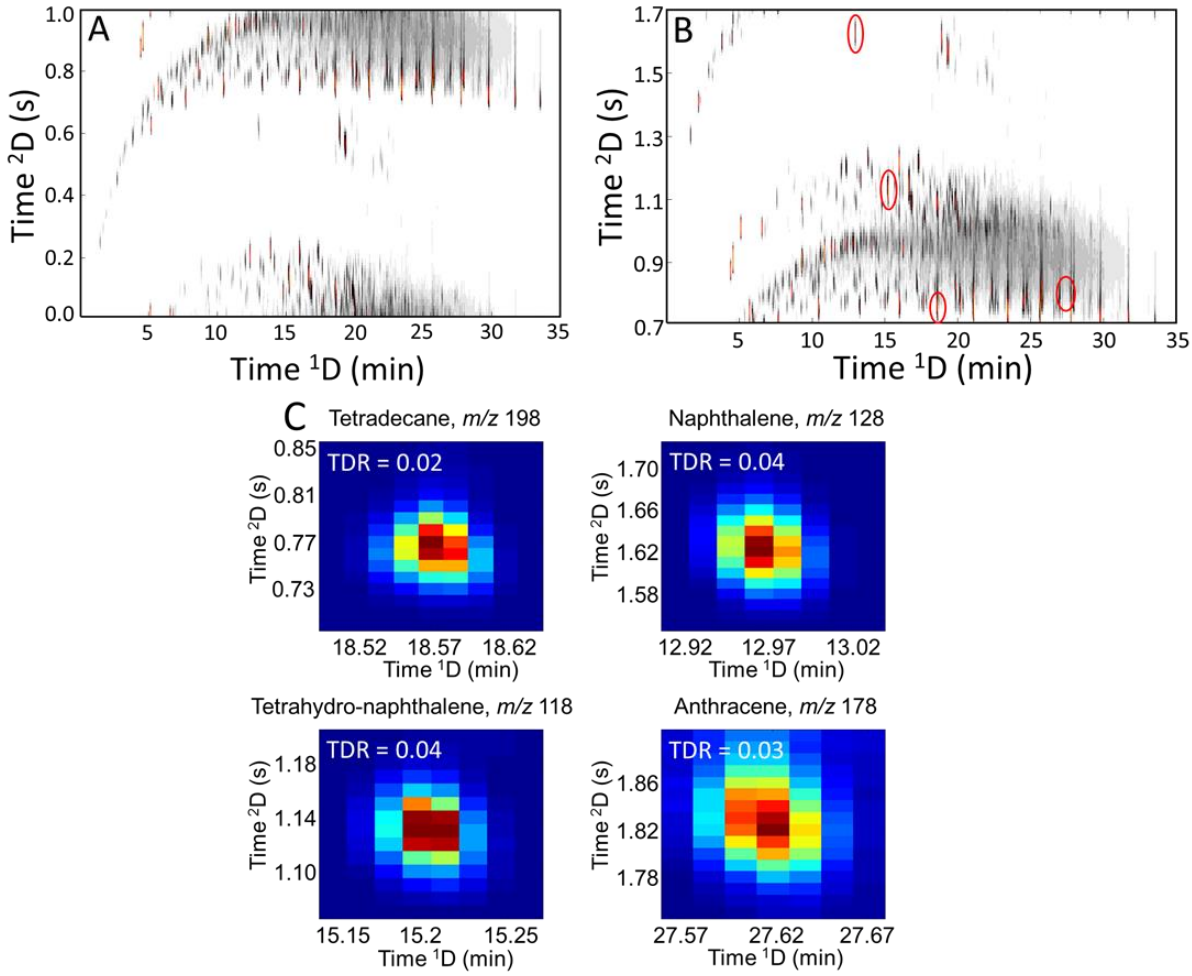

Figure 5.8 (A) Total ion current contour plot of a GC × GC – TOFMS diesel separation. (B) Same separation as in (A), reregistered on ^{2}D and the locations of four analytes used for *TDR* analysis. (C) Zoom in of each of the four analytes from (B). The ^{2}D retention times and *TDR* values are: (1) tetradecane 0.77 s, *TDR* = 0.02; (2) tetrahydro-naphthalene 1.13 s, *TDR* = 0.04; (3) naphthalene 1.62 s, *TDR* = 0.04; (4) anthracene 1.82 s, *TDR* = 0.03.

Chapter 6. Conclusions

One and two-dimensional chromatography coupled with spectral detection is a central techniques in the modern analytical chemistry laboratory, and is well suited for analyzing a wide variety of samples types across many fields and applications. Since these instruments are commonly applied to study complex samples, the resulting data is also very complex, and analyte peaks commonly overlap. Chemometric methods are often used to mathematically resolve the overlapped analyte peaks, which leads to successful analyte identification and quantification. The design and implementation of a particular instrumental platform can significantly impact various underlying aspects of the data structure, whether second order or third order, causing benefits and challenges for chemometric methods. The goal of the presented work was aimed at exploring the benefits and challenges that appear at the interface between hyphenated chromatography and chemometrics, to improve instrument design, operation and the resulting data analysis.

Chapter 2 presented a new method for determining the probability of chemometric success for GC-MS based on sample complexity and the chemometric method dependent chemometric resolution limit, R_s*. The probabilistic theory of component overlap in separations was presented defining chemometric success and calculating the probability of peak overlap in a GC-MS separation. MCR-ALS was used as a case study to demonstrate how the probabilistic theory can be integrated as a part of the chromatographic experiment through chemometrics. A simulation-based investigation was used to determine the minimum chromatographic resolution required for MCR-ALS to successfully deconvolute coeluting GC-MS peaks. Chromatograms consisting of a target analyte and interferent analyte coeluting at various *S/N* thresholds and

chromatographic resolutions were generated. MCR-ALS was applied to each simulation the results of the chemometric model, both the peak area and mass spectrum of the target, were compared to the originally simulated data to determine the efficacy of the method. The results of a simulation based study indicated that the R_s* of MCR-ALS in the range of 0.2 to 0.3 depending upon analytical expectations, in general, more improvements to chemometric deconvolution methods to reduce R_s* are warranted. Since the MCR-ALS models were generated without constraints in this study, improvement to R_s* may be achieved if suitable constraints can be readily implemented, such as initial peak profile estimates for the more challenging overlap situations..

Chapter 3 presented work to study the effects of phase ratio, β, on peak capacity, n_c, for GC×GC–TOFMS, and to determine the conditions under which $n_{c,2D}$ can be improved. Overall, β substantially affected $n_{c,2D}$ by influencing retention factors on the secondary column ^2k, and thereby changing the P_M necessary for proper secondary column separations. The necessary changes to P_M modify the modulation ratio, M_R, which affects the primary column peak widths and ^1n_c. Through changes to $^1\beta$, the range of ^2k may be controlled, with subsequent effects to both ^2n_c and ^1n_c. These effects were opposite in direction, such that gains in ^2n_c may result in losses in ^1n_c. Due to the pseudo-isothermal nature of the ^{2}D column separation, there are diminishing returns to extending the ^2n_c at the cost of ^1n_c. The β ratio, β_R, defined as $^1\beta$ divided by $^2\beta$ for the ^{1}D and ^{2}D separations, was introduced as an important parameter for tuning GC×GC separation conditions. β_R can be used to describe the relative ^2k on each column set, as increases in β_R will increase ^2k. It was found that β_R must be considered in conjunction with an appropriate, and not excessive P_M, in order to maximize the $n_{c,2D}$ obtained.

Chapter 4 presented a new method to determine the true modulation ratio, M_R, of a peak in GC×GC, without the requirement for a detector at the end of the primary column as a function of the effective modulation ratio, M_R*, calculated from the effective peak width, ${}^1W_b$*, and the modulation period, P_M. I demonstrated the consequences of modulation period, P_M, selection for a GC×GC separation as it relates to modulator induced band broadening, by mapping from the effective modulation ratio, M_R*, of a modulated ^{1}D peak to the true modulation ratio, M_R. This method was initially developed through the study of simulated GC×GC data, by first defining primary column ^{1}D peak(s) and simulating the modulation process then implementing a new peak measurement method using Gaussian curve fitting to model each modulated ^{2}D peaklet in the unfolded GC×GC data to accurately determine the maxima of the peaklet distribution, followed by Gaussian curve fitting to the maxima to determine the effective ^{1}D peak profile and width, ${}^1W_b$*. The identified relationship between ${}^1W_b$ and ${}^1W_b$* allowed for mapping between the measured M_R* and the premodulation M_R which can also be used to determine loss in peak capacity due to modulation induced band broadening. Experimental validation of the simulated results was provided to span a range of commonly implemented conditions with typical ${}^1W_b$ (2–4.5 s) and P_M (0.25–8 s). Use of $M_R < 2$ significantly broadens the ^{1}D peak ($M_R^* \geq 1.2\ M_R$) corresponding to a loss in ^{1}D peak capacity, ${}^1n_c \geq 20\%$.

Finally, in Chapter 5 the trilinearity of the GC×GC–TOFMS data structure is explored with respect to the quantitative performance of parallel factor analysis, PARAFAC. I demonstrated that retention time shifting from one modulation to the next, $\Delta^2 t_R$, due to temperature programming with GC×GC–TOFMS, which is the principal mechanism of non-trilinearity within a single GC×GC–TOFMS data cube, is not sufficient alone to quantitatively describe the trilinearity of the data cube for the purposes of predicting the performance of

PARAFAC. I demonstrated that the analyte peak width on second dimension separations, ${}^{2}W_{b}$, also impacts trilinearity, along with $\Delta^{2}t_{R}$. The term Trilinearity Deviation Ratio, *TDR*, which is $\Delta^{2}t_{R}$ normalized by ${}^{2}W_{b}$, was introduced as a quantitative metric to assess PARAFAC accuracy with respect to data trilinearity. Consistent with the results from Chapters 3 and 4, the use of a modulation period in the 1 s to 2 s range optimizes the peak capacity for the first dimension separation, with an adequate peak capacity for the second dimension separation concurrent, and an optimized two-dimensional peak capacity, combined with low *TDR* values (0 to 0.05) to facilitate low quantitative errors with PARAFAC. The *TDR* concept in principle is not limited to GC×GC–TOFMS, and should be readily extendable to other comprehensive two-dimensional separation techniques, and can be reduced to bilinearity deviation ratio, *BDR*, for univariate detection methods.

BIBLIOGRAPHY

Adahchour, M., Beens, J., Vreuls, R. J. J., & Brinkman, U. A. T. (2006). Recent developments in comprehensive two-dimensional gas chromatography (GC × GC): IV. Further applications, conclusions and perspectives. *TrAC Trends in Analytical Chemistry*, *25*(8), 821–840. https://doi.org/10.1016/j.trac.2006.03.003

Adahchour, M., van Stee, L. L. P., Beens, J., Vreuls, R. J. J., Batenburg, M. A., & Brinkman, U. A. T. (2003). Comprehensive two-dimensional gas chromatography with time-of-flight mass spectrometric detection for the trace analysis of flavour compounds in food. *Journal of Chromatography A*, *1019*(1–2), 157–172. https://doi.org/10.1016/S0021-9673(03)01131-2

Adcock, J. L., Adams, M., Mitrevski, B. S., & Marriott, P. J. (2009). Peak Modeling Approach to Accurate Assignment of First-Dimension Retention Times in Comprehensive Two-Dimensional Chromatography. *Analytical Chemistry*, *81*(16), 6797–6804. https://doi.org/10.1021/ac900960n

Bailey, H. P., & Rutan, S. C. (2011). Chemometric resolution and quantification of four-way data arising from comprehensive 2D-LC-DAD analysis of human urine. *Chemometrics and Intelligent Laboratory Systems*, *106*(1), 131–141. https://doi.org/10.1016/j.chemolab.2010.07.008

Bailey, H. P., Rutan, S. C., & Carr, P. W. (2011). Factors that affect quantification of diode array data in comprehensive two-dimensional liquid chromatography using chemometric data analysis. *Journal of Chromatography A*, *1218*(46), 8411–8422. https://doi.org/10.1016/j.chroma.2011.09.057

Beens, J., Adahchour, M., Vreuls, R. J. J., van Altena, K., & Th. Brinkman, U. A. (2001). Simple, non-moving modulation interface for comprehensive two-dimensional gas chromatography. *Journal of Chromatography A*, *919*(1), 127–132. https://doi.org/10.1016/S0021-9673(01)00785-3

Blumberg, L. M. (1999). Theory of Fast Capillary Gas Chromatography Part 4: Column Performance vs.Liquid Film Thickness. *Journal of High Resolution Chromatography*, *22*(9), 501–508. https://doi.org/10.1002/(SICI)1521-4168(19990901)22:9<501::AID-JHRC501>3.0.CO;2-N

Blumberg, L. M. (2003). Comprehensive two-dimensional gas chromatography: metrics, potentials, limits. *Journal of Chromatography A*, *985*(1–2), 29–38. https://doi.org/10.1016/S0021-9673(02)01416-4

Blumberg, L. M. (2008). Accumulating resampling (modulation) in comprehensive two-dimensional capillary GC (GC×GC). *Journal of Separation Science*, *31*(19), 3358–3365. https://doi.org/10.1002/jssc.200800424

Blumberg, L. M. (2011). Multidimensional Gas Chromatography: Theoretical Considerations. In L. Mondello (Ed.), *Comprehensive Chromatography in Combination with Mass Spectrometry* (pp. 13–63). John Wiley & Sons, Inc. Retrieved from http://onlinelibrary.wiley.com/doi/10.1002/9781118003466.ch2/summary

Blumberg, L. M., David, F., Klee, M. S., & Sandra, P. (2008). Comparison of one-dimensional and comprehensive two-dimensional separations by gas chromatography. *Journal of Chromatography A*, *1188*(1), 2–16. https://doi.org/10.1016/j.chroma.2008.02.044

Blumberg, L. M., & Klee, M. S. (2001). Elution parameters in constant-pressure, single-ramp temperature-programmed gas chromatography. *Journal of Chromatography A*, *918*(1), 113–120. https://doi.org/10.1016/S0021-9673(01)00659-8

Booksh, K. S., & Kowalski, B. R. (1994). Theory of Analytical Chemistry. *Analytical Chemistry*, *66*(15), 782A–791A. https://doi.org/10.1021/ac00087a718

Bro, R. (1997). PARAFAC. Tutorial and applications. *Chemometrics and Intelligent Laboratory Systems*, *38*(2), 149–171. https://doi.org/10.1016/S0169-7439(97)00032-4

Bro, R., Andersson, C. A., & Kiers, H. A. L. (1999). PARAFAC2—Part II. Modeling chromatographic data with retention time shifts. *Journal of Chemometrics*, *13*(3-4), 295–309. https://doi.org/10.1002/(SICI)1099-128X(199905/08)13:3/4<295::AID-CEM547>3.0.CO;2-Y

Bruckner, C. A., Prazen, B. J., & Synovec, R. E. (1998). Comprehensive Two-Dimensional High-Speed Gas Chromatography with Chemometric Analysis. *Analytical Chemistry*, *70*(14), 2796–2804. https://doi.org/10.1021/ac980164m

Bushey, M. M., & Jorgenson, J. W. (1990). Automated instrumentation for comprehensive two-dimensional high-performance liquid chromatography of proteins. *Analytical Chemistry*, *62*(2), 161–167. https://doi.org/10.1021/ac00201a015

Carroll, J. D., & Chang, J.-J. (1970). Analysis of individual differences in multidimensional scaling via an n-way generalization of "Eckart-Young" decomposition. *Psychometrika*, *35*(3), 283–319. https://doi.org/10.1007/BF02310791

Carr, P. W., Stoll, D. R., & Wang, X. (2011). Perspectives on Recent Advances in the Speed of High-Performance Liquid Chromatography. *Analytical Chemistry*, *83*(6), 1890–1900. https://doi.org/10.1021/ac102570t

Comas, E., Gimeno, R. A., Ferré, J., Marcé, R. M., Borrull, F., & Rius, F. X. (2004). Quantification from highly drifted and overlapped chromatographic peaks using second-order calibration methods. *Journal of Chromatography A*, *1035*(2), 195–202. https://doi.org/10.1016/j.chroma.2004.02.069

Cook, D. W., Rutan, S. C., Stoll, D. R., & Carr, P. W. (2015). Two dimensional assisted liquid chromatography – a chemometric approach to improve accuracy and precision of quantitation in liquid chromatography using 2D separation, dual detectors, and multivariate curve resolution. *Analytica Chimica Acta*, *859*, 87–95. https://doi.org/10.1016/j.aca.2014.12.009

Davis, J. M. (1991). Statistical theory of spot overlap in two-dimensional separations. *Analytical Chemistry*, *63*(19), 2141–2152. https://doi.org/10.1021/ac00019a014

Davis, J. M., & Giddings, J. C. (1983). Statistical theory of component overlap in multicomponent chromatograms. *Analytical Chemistry*, *55*(3), 418–424. https://doi.org/10.1021/ac00254a003

Davis, J. M., & Giddings, J. C. (1985). Statistical method for estimation of number of components from single complex chromatograms: theory, computer-based testing, and analysis of errors. *Analytical Chemistry*, *57*(12), 2168–2177. https://doi.org/10.1021/ac00289a002

Davis, J. M., Stoll, D. R., & Carr, P. W. (2008). Dependence of Effective Peak Capacity in Comprehensive Two-Dimensional Separations on the Distribution of Peak Capacity between the Two Dimensions. *Analytical Chemistry*, *80*(21), 8122–8134. https://doi.org/10.1021/ac800933z

de Juan, A., & Tauler, R. (2001). Comparison of three-way resolution methods for non-trilinear chemical data sets. *Journal of Chemometrics*, *15*(10), 749–771. https://doi.org/10.1002/cem.662

del Olmo, M., González-Casado, A., Navas, N. A., & Vilchez, J. L. (1997). Determination of bisphenol A (BPA) in water by gas chromatography-mass spectrometry. *Analytica Chimica Acta*, *346*(1), 87–92. https://doi.org/10.1016/S0003-2670(97)00182-7

Faber, N. M., Buydens, L. M. C., & Kateman, G. (1994). Generalized rank annihilation method. I: Derivation of eigenvalue problems. *Journal of Chemometrics*, *8*(2), 147–154. https://doi.org/10.1002/cem.1180080206

Fitz, B. D., Reaser, B. C., Pinkerton, D. K., Hoggard, J. C., Skogerboe, K. J., & Synovec, R. E. (2014). Enhancing Gas Chromatography–Time of Flight Mass Spectrometry Data Analysis Using Two-Dimensional Mass Channel Cluster Plots. *Analytical Chemistry*, *86*(8), 3973–3979. https://doi.org/10.1021/ac5004344

Fitz, B. D., & Synovec, R. E. (2016). Extension of the two-dimensional mass channel cluster plot method to fast separations utilizing low thermal mass gas chromatography with time-of-flight mass spectrometry. *Analytica Chimica Acta*, *913*, 160–170. https://doi.org/10.1016/j.aca.2016.01.045

Fitz, B. D., Wilson, R. B., Parsons, B. A., Hoggard, J. C., & Synovec, R. E. (2012). Fast, high peak capacity separations in comprehensive two-dimensional gas chromatography with time-of-flight mass spectrometry. *Journal of Chromatography A*, *1266*, 116–123. https://doi.org/10.1016/j.chroma.2012.09.096

Fraga, C. G., Pérez Acosta, G. A., Crenshaw, M. D., Wallace, K., Mong, G. M., & Colburn, H. A. (2011). Impurity Profiling to Match a Nerve Agent to Its Precursor Source for Chemical Forensics Applications. *Analytical Chemistry*, *83*(24), 9564–9572. https://doi.org/10.1021/ac202340u

Giddings, J. C. (1984). Two-dimensional separations: concept and promise. *Analytical Chemistry*, *56*(12), 1258A–1270A. https://doi.org/10.1021/ac00276a003

Giddings, J. C. (1991). *Unified separation science*. New York: Wiley. Retrieved from http://catalog.hathitrust.org/api/volumes/oclc/21764363.html

Gross, G. M., Prazen, B. J., Grate, J. W., & Synovec, R. E. (2004). High-Speed Gas Chromatography Using Synchronized Dual-Valve Injection. *Analytical Chemistry*, *76*(13), 3517–3524. https://doi.org/10.1021/ac049909g

Harshman, R. A. (1970). University microfilms. Michigan: Ann Arbor. Michigan: Ann Arbor.

Hoggard, J. C. (2011). peg2mat3p8; http://depts.washington.edu/synlab/software/. Retrieved from http://depts.washington.edu/synlab/software/

Hoggard, J. C., Siegler, W. C., & Synovec, R. E. (2009). Toward automated peak resolution in complete GC × GC–TOFMS chromatograms by PARAFAC. *Journal of Chemometrics*, *23*(7-8), 421–431. https://doi.org/10.1002/cem.1239

Hoggard, J. C., & Synovec, R. E. (2007). Parallel Factor Analysis (PARAFAC) of Target Analytes in GC × GC−TOFMS Data: Automated Selection of a Model with an Appropriate Number of Factors. *Analytical Chemistry*, *79*(4), 1611–1619. https://doi.org/10.1021/ac061710b

Hoggard, J. C., & Synovec, R. E. (2008). Automated Resolution of Nontarget Analyte Signals in GC × GC-TOFMS Data Using Parallel Factor Analysis. *Analytical Chemistry*, *80*(17), 6677–6688. https://doi.org/10.1021/ac800624e

Hoggard, J. C., Wahl, J. H., Synovec, R. E., Mong, G. M., & Fraga, C. G. (2010). Impurity Profiling of a Chemical Weapon Precursor for Possible Forensic Signatures by Comprehensive Two-Dimensional Gas Chromatography/Mass Spectrometry and Chemometrics. *Analytical Chemistry*, *82*(2), 689–698. https://doi.org/10.1021/ac902247x

Jalali-Heravi, M., Parastar, H., Kamalzadeh, M., Tauler, R., & Jaumot, J. (2010). MCRC software: A tool for chemometric analysis of two-way chromatographic data. *Chemometrics and Intelligent Laboratory Systems*, *104*(2), 155–171. https://doi.org/10.1016/j.chemolab.2010.08.002

Johnson, K. J., Wright, B. W., Jarman, K. H., & Synovec, R. E. (2003). High-speed peak matching algorithm for retention time alignment of gas chromatographic data for chemometric analysis. *Journal of Chromatography A*, *996*(1–2), 141–155. https://doi.org/10.1016/S0021-9673(03)00616-2

Juan, A. de, Jaumot, J., & Tauler, R. (2014). Multivariate Curve Resolution (MCR). Solving the mixture analysis problem. *Analytical Methods*, *6*(14), 4964–4976. https://doi.org/10.1039/C4AY00571F

Khummueng, W., Harynuk, J., & Marriott, P. J. (2006). Modulation Ratio in Comprehensive Two-dimensional Gas Chromatography. *Analytical Chemistry*, *78*(13), 4578–4587. https://doi.org/10.1021/ac052270b

Kinghorn, R. M., & Marriott, P. J. (1998). Comprehensive Two-Dimensional Gas Chromatography Using a Modulating Cryogenic Trap. *Journal of High Resolution Chromatography*, *21*(11), 620–622. https://doi.org/10.1002/(SICI)1521-4168(19981101)21:11<620::AID-JHRC620>3.0.CO;2-#

Klee, M. S. (2005). *GC Inlets An Introduction* (Second). Wilmington, DE USA: Agilent Technogies, Inc. Retrieved from http://www.agilent.com/cs/library/usermanuals/public/5958-9468_041007.pdf

Klee, M. S., & Blumberg, L. M. (2010). Measurement of retention in comprehensive two-dimensional gas chromatography using flow modulation with methane dopant. *Journal of Chromatography A*, *1217*(11), 1830–1837. https://doi.org/10.1016/j.chroma.2010.01.027

Klee, M. S., Cochran, J., Merrick, M., & Blumberg, L. M. (2015). Evaluation of conditions of comprehensive two-dimensional gas chromatography that yield a near-theoretical maximum in peak capacity gain. *Journal of Chromatography A*, *1383*, 151–159. https://doi.org/10.1016/j.chroma.2015.01.031

Koek, M. M., Muilwijk, B., van Stee, L. L. P., & Hankemeier, T. (2008). Higher mass loadability in comprehensive two-dimensional gas chromatography–mass spectrometry for improved analytical performance in metabolomics analysis. *Journal of Chromatography A*, *1186*(1–2), 420–429. https://doi.org/10.1016/j.chroma.2007.11.107

Krakowska, B., Stanimirova, I., Orzel, J., Daszykowski, M., Grabowski, I., Zaleszczyk, G., & Sznajder, M. (2014). Detection of discoloration in diesel fuel based on gas chromatographic fingerprints. *Analytical and Bioanalytical Chemistry*, *407*(4), 1159–1170. https://doi.org/10.1007/s00216-014-8332-4

Kuligowski, J., Quintás, G., Tauler, R., Lendl, B., & de la Guardia, M. (2011). Background Correction and Multivariate Curve Resolution of Online Liquid Chromatography with Infrared Spectrometric Detection. *Analytical Chemistry*, *83*(12), 4855–4862. https://doi.org/10.1021/ac2004407

Lisec, J., Hoffmann, F., Schmitt, C., & Jaeger, C. (2016). Extending the Dynamic Range in Metabolomics Experiments by Automatic Correction of Peaks Exceeding the Detection Limit. *Analytical Chemistry*, *88*(15), 7487–7492. https://doi.org/10.1021/acs.analchem.6b02515

Li, S., Hamilton, J. C., & Gemperline, P. J. (1992). Generalized rank annihilation method using similarity transformations. *Analytical Chemistry*, *64*(6), 599–607. https://doi.org/10.1021/ac00030a007

Liu, Z., & Phillips, J. B. (1991). Comprehensive Two-Dimensional Gas Chromatography using an On-Column Thermal Modulator Interface. *Journal of Chromatographic Science*, *29*(6), 227–231. https://doi.org/10.1093/chromsci/29.6.227

Li, X., Stoll, D. R., & Carr, P. W. (2009). Equation for Peak Capacity Estimation in Two-Dimensional Liquid Chromatography. *Analytical Chemistry*, *81*(2), 845–850. https://doi.org/10.1021/ac801772u

Mahé, L., Courtiade, M., Dartiguelongue, C., Ponthus, J., Souchon, V., & Thiébaut, D. (2012). Overcoming the high-temperature two-dimensional gas chromatography limits to elute heavy compounds. *Journal of Chromatography A*, *1229*, 298–301. https://doi.org/10.1016/j.chroma.2012.01.030

Marriott, P. J., Wu, Z., & Schoenmakers, P. (2003). Nomenclature and Conventions in Comprehensive Multidimensional Chromatography. *LC-GC Europe*, *16*(6), 335–339.

Mohler, R. E., Dombek, K. M., Hoggard, J. C., Pierce, K. M., Young, E. T., & Synovec, R. E. (2007). Comprehensive analysis of yeast metabolite GC×GC–TOFMS data: combining discovery-mode and deconvolution chemometric software. *Analyst*, *132*(8), 756–767. https://doi.org/10.1039/B700061H

Mohler, R. E., Dombek, K. M., Hoggard, J. C., Young, E. T., & Synovec, R. E. (2006). Comprehensive Two-Dimensional Gas Chromatography Time-of-Flight Mass Spectrometry Analysis of Metabolites in Fermenting and Respiring Yeast Cells. *Analytical Chemistry*, *78*(8), 2700–2709. https://doi.org/10.1021/ac052106o

Moler, C., & Stewart, G. (1973). An Algorithm for Generalized Matrix Eigenvalue Problems. *SIAM Journal on Numerical Analysis*, *10*(2), 241–256. https://doi.org/10.1137/0710024

Mostafa, A., Edwards, M., & Górecki, T. (2012). Optimization aspects of comprehensive two-dimensional gas chromatography. *Journal of Chromatography A*, *1255*, 38–55. https://doi.org/10.1016/j.chroma.2012.02.064

Murphy, R. E., Schure, M. R., & Foley, J. P. (1998). Effect of Sampling Rate on Resolution in Comprehensive Two-Dimensional Liquid Chromatography. *Analytical Chemistry*, *70*(8), 1585–1594. https://doi.org/10.1021/ac971184b

Murray, J. A. (2012). Qualitative and quantitative approaches in comprehensive two-dimensional gas chromatography. *Journal of Chromatography A*, *1261*, 58–68. https://doi.org/10.1016/j.chroma.2012.05.012

Nadeau, J. S., Wilson, R. B., Hoggard, J. C., Wright, B. W., & Synovec, R. E. (2011). Study of the interdependency of the data sampling ratio with retention time alignment and principal component analysis for gas chromatography. *Journal of Chromatography A*, *1218*(50), 9091–9101. https://doi.org/10.1016/j.chroma.2011.10.031

Oca, M. L., Ortiz, M. C., Herrero, A., & Sarabia, L. A. (2013). Optimization of a GC/MS procedure that uses parallel factor analysis for the determination of bisphenols and their

diglycidyl ethers after migration from polycarbonate tableware. *Talanta*, *106*, 266–280. https://doi.org/10.1016/j.talanta.2012.10.086

Parastar, H., Radović, J. R., Jalali-Heravi, M., Diez, S., Bayona, J. M., & Tauler, R. (2011). Resolution and Quantification of Complex Mixtures of Polycyclic Aromatic Hydrocarbons in Heavy Fuel Oil Sample by Means of GC × GC-TOFMS Combined to Multivariate Curve Resolution. *Analytical Chemistry*, *83*(24), 9289–9297. https://doi.org/10.1021/ac201799r

Parastar, H., & Tauler, R. (2014). Multivariate Curve Resolution of Hyphenated and Multidimensional Chromatographic Measurements: A New Insight to Address Current Chromatographic Challenges. *Analytical Chemistry*, *86*(1), 286–297. https://doi.org/10.1021/ac402377d

Parsons, B. A., Pinkerton, D. K., & Synovec, R. E. (2017). Implications of Phase Ratio for Maximizing Peak Capacity in Comprehensive Two-Dimensional Gas Chromatography Time-of-Flight Mass Spectrometry. *Journal of Chromatography A*, Manuscript JCA-16-2350, *currently submitted*

Parsons, B. A., Pinkerton, D. K., Wright, B. W., & Synovec, R. E. (2016). Chemical characterization of the acid alteration of diesel fuel: Non-targeted analysis by two-dimensional gas chromatography coupled with time-of-flight mass spectrometry with tile-based Fisher ratio and combinatorial threshold determination. *Journal of Chromatography A*, *1440*, 179–190. https://doi.org/10.1016/j.chroma.2016.02.067

Peroni, D., & Janssen, H.-G. (2014). Comprehensive two-dimensional gas chromatography under high outlet pressure conditions: A new approach to correct the flow-mismatch issue in the two dimensions. *Journal of Chromatography A*, *1332*, 57–63. https://doi.org/10.1016/j.chroma.2014.01.051

Peroni, D., Sampat, A. A. S., van Egmond, W., de Koning, S., Cochran, J., Lautamo, R., & Janssen, H.-G. (2013). Comprehensive two-dimensional gas chromatography with a multi-capillary second dimension: A new column-set format for simultaneous optimum linear velocity operation. *Journal of Chromatography A*. https://doi.org/10.1016/j.chroma.2013.07.097

Phillips, J. B., & Beens, J. (1999). Comprehensive two-dimensional gas chromatography: a hyphenated method with strong coupling between the two dimensions. *Journal of Chromatography A*, *856*(1–2), 331–347. https://doi.org/10.1016/S0021-9673(99)00815-8

Pierce, K. M., Hoggard, J. C., Hope, J. L., Rainey, P. M., Hoofnagle, A. N., Jack, R. M., … Synovec, R. E. (2006). Fisher Ratio Method Applied to Third-Order Separation Data To Identify Significant Chemical Components of Metabolite Extracts. *Analytical Chemistry*, *78*(14), 5068–5075. https://doi.org/10.1021/ac0602625

Pinkerton, D. K., Parsons, B. A., Anderson, T. J., & Synovec, R. E. (2015). Trilinearity deviation ratio: A new metric for chemometric analysis of comprehensive two-dimensional gas chromatography time-of-flight mass spectrometry data. *Analytica Chimica Acta*, *871*, 66–76. https://doi.org/10.1016/j.aca.2015.02.040

Pinkerton, D. K., Parsons, B. A., & Synovec, R. E. (2016). Method to determine the true modulation ratio for comprehensive two-dimensional gas chromatography. *Journal of Chromatography A*, *1476*, 114–123. https://doi.org/10.1016/j.chroma.2016.11.015

Pinkerton, D. K., Pierce, K. M., & Synovec, R. E. (2016). Chemometric Resolution of Complex Higher Order Chromatographic Data with Spectral Detection. In *Resolving Spectral Mixtures* (pp. 333-352). Cambridge, MA: Elsevier

Płotka, J. M., Morrison, C., Adam, D., & Biziuk, M. (2012). Chiral Analysis of Chloro Intermediates of Methylamphetamine by One-Dimensional and Multidimentional NMR and GC/MS. *Analytical Chemistry*, *84*(13), 5625–5632. https://doi.org/10.1021/ac300503g

Poe, R. B., & Rutan, S. C. (1993). Effects of resolution, peak ratio and sampling frequency in diode-array fluorescence detection in liquid chromatography. *Analytica Chimica Acta*, *283*(2), 845–853. https://doi.org/10.1016/0003-2670(93)85298-X

Pool, W. G., de Leeuw, J. W., & van de Graaf, B. (1996). A rapid routine to correct for skewing in gas chromatography/mass spectrometry. *Journal of Mass Spectrometry*, *31*(2), 213–215. https://doi.org/10.1002/(SICI)1096-9888(199602)31:2<213::AID-JMS284>3.0.CO;2-6

Porter, S. E. G., Stoll, D. R., Rutan, S. C., Carr, P. W., & Cohen, J. D. (2006). Analysis of Four-Way Two-Dimensional Liquid Chromatography-Diode Array Data: Application to Metabolomics. *Analytical Chemistry*, *78*(15), 5559–5569. https://doi.org/10.1021/ac0606195

Prazen, B. J., Bruckner, C. A., Synovec, R. E., & Kowalski, B. R. (1999). Second-order chemometric standardization for high-speed hyphenated gas chromatography: Analysis of GC/MS and comprehensive GC×GC data. *Journal of Microcolumn Separations*, *11*(2), 97–107. https://doi.org/10.1002/(SICI)1520-667X(1999)11:2<97::AID-MCS2>3.0.CO;2-Z

Prazen, B. J., Synovec, R. E., & Kowalski, B. R. (1998). Standardization of Second-Order Chromatographic/Spectroscopic Data for Optimum Chemical Analysis. *Analytical Chemistry*, *70*(2), 218–225. https://doi.org/10.1021/ac9706335

Prebihalo, S., Brockman, A., Cochran, J., & Dorman, F. L. (2015). Determination of emerging contaminants in wastewater utilizing comprehensive two-dimensional gas-chromatography coupled with time-of-flight mass spectrometry. *Journal of Chromatography A*, *1419*, 109–115. https://doi.org/10.1016/j.chroma.2015.09.080

Reaser, B. C., Yang, S., Fitz, B. D., Parsons, B. A., Lidstrom, M. E., & Synovec, R. E. (2016). Non-targeted determination of 13C-labeling in the Methylobacterium extorquens AM1 metabolome using the two-dimensional mass cluster method and principal component analysis. *Journal of Chromatography A*, *1432*, 111–121. https://doi.org/10.1016/j.chroma.2015.12.088

Reid, V. R., & Synovec, R. E. (2008). High-speed gas chromatography: The importance of instrumentation optimization and the elimination of extra-column band broadening. *Talanta*, *76*(4), 703–717. https://doi.org/10.1016/j.talanta.2008.05.012

Robards, K., Haddad, P. R., & Jackson, P. E. (2004). *Principles and Practice of Modern Chromatographic Methods*. Elsevier, Ltd.

Rodier, C., Vandenabeele-Trambouze, O., Sternberg, R., Coscia, D., Coll, P., Szopa, C., … Despois, D. (2001). Detection of martian amino acids by chemical derivatization coupled to gas chromatography: In situ and laboratory analysis. *Advances in Space Research*, *27*(2), 195–199. https://doi.org/10.1016/S0273-1177(01)00047-3

Rutledge, D. N., & Jouan-Rimbaud Bouveresse, D. (2007). Multi-way analysis of outer product arrays using PARAFAC. *Chemometrics and Intelligent Laboratory Systems*, *85*(2), 170–178. https://doi.org/10.1016/j.chemolab.2006.06.011

Sadoughi, N., Schmidtke, L. M., Antalick, G., Blackman, J. W., & Steel, C. C. (2015). Gas Chromatography–Mass Spectrometry Method Optimized Using Response Surface Modeling for the Quantitation of Fungal Off-Flavors in Grapes and Wine. *Journal of Agricultural and Food Chemistry*. https://doi.org/10.1021/jf505444r

Sampat, A., Lopatka, M., Sjerps, M., Vivo-Truyols, G., Schoenmakers, P., & van Asten, A. (2016). Forensic potential of comprehensive two-dimensional gas chromatography. *TrAC Trends in Analytical Chemistry*, *80*, 345–363. https://doi.org/10.1016/j.trac.2015.10.011

Sanchez, E., & Kowalski, B. R. (1986). Generalized rank annihilation factor analysis. *Analytical Chemistry*, *58*(2), 496–499. https://doi.org/10.1021/ac00293a054

Sanchez, E., Scott Ramos, L., & Kowalski, B. R. (1987). Generalized rank annihilation method. *Journal of Chromatography A*, *385*, 151–164. https://doi.org/10.1016/S0021-9673(01)94629-1

Sasaki, T., Koshi, E., Take, H., Michihata, T., Maruya, M., & Enomoto, T. (2017). Characterisation of odorants in roasted stem tea using gas chromatography–mass spectrometry and gas chromatography-olfactometry analysis. *Food Chemistry*, *220*, 177–183. https://doi.org/10.1016/j.foodchem.2016.09.208

Seeley, J. V. (2002). Theoretical study of incomplete sampling of the first dimension in comprehensive two-dimensional chromatography. *Journal of Chromatography A*, *962*(1–2), 21–27. https://doi.org/10.1016/S0021-9673(02)00461-2

Seeley, J. V., Kramp, F., & Hicks, C. J. (2000). Comprehensive Two-Dimensional Gas Chromatography via Differential Flow Modulation. *Analytical Chemistry*, *72*(18), 4346–4352. https://doi.org/10.1021/ac000249z

Seeley, J. V., Micyus, N. J., Bandurski, S. V., Seeley, S. K., & McCurry, J. D. (2007). Microfluidic Deans Switch for Comprehensive Two-Dimensional Gas Chromatography. *Analytical Chemistry*, *79*(5), 1840–1847. https://doi.org/10.1021/ac061881g

Seeley, J. V., & Seeley, S. K. (2013). Multidimensional Gas Chromatography: Fundamental Advances and New Applications. *Analytical Chemistry*, *85*(2), 557–578. https://doi.org/10.1021/ac303195u

Siegler, W. C., Fitz, B. D., Hoggard, J. C., & Synovec, R. E. (2011). Experimental Study of the Quantitative Precision for Valve-Based Comprehensive Two-Dimensional Gas Chromatography. *Analytical Chemistry*, *83*(13), 5190–5196. https://doi.org/10.1021/ac200302b

Skov, T., Hoggard, J. C., Bro, R., & Synovec, R. E. (2009). Handling within run retention time shifts in two-dimensional chromatography data using shift correction and modeling. *Journal of Chromatography A*, *1216*(18), 4020–4029. https://doi.org/10.1016/j.chroma.2009.02.049

Stein, S. E. (1999). An integrated method for spectrum extraction and compound identification from gas chromatography/mass spectrometry data. *Journal of the American Society for Mass Spectrometry*, *10*(8), 770–781. https://doi.org/10.1016/S1044-0305(99)00047-1

Stoll, D. R., Cohen, J. D., & Carr, P. W. (2006). Fast, comprehensive online two-dimensional high performance liquid chromatography through the use of high temperature ultra-fast gradient elution reversed-phase liquid chromatography. *Journal of Chromatography A*, *1122*(1–2), 123–137. https://doi.org/10.1016/j.chroma.2006.04.058

Tauler, R. (1995). Multivariate curve resolution applied to second order data. *Chemometrics and Intelligent Laboratory Systems*, *30*(1), 133–146. https://doi.org/10.1016/0169-7439(95)00047-X

Thekkudan, D. F., Rutan, S. C., & Carr, P. W. (2010). A study of the precision and accuracy of peak quantification in comprehensive two-dimensional liquid chromatography in time.

Journal of Chromatography A, *1217*(26), 4313–4327. https://doi.org/10.1016/j.chroma.2010.04.039

Toraman, H. E., Franz, K., Ronsse, F., Van Geem, K. M., & Marin, G. B. (2016). Quantitative analysis of nitrogen containing compounds in microalgae based bio-oils using comprehensive two-dimensional gas-chromatography coupled to nitrogen chemiluminescence detector and time of flight mass spectrometer. *Journal of Chromatography A*, *1460*, 135–146. https://doi.org/10.1016/j.chroma.2016.07.009

Tranchida, P. Q., Donato, P., Cacciola, F., Beccaria, M., Dugo, P., & Mondello, L. (2013). Potential of comprehensive chromatography in food analysis. *TrAC Trends in Analytical Chemistry*, *52*, 186–205. https://doi.org/10.1016/j.trac.2013.07.008

Tranchida, P. Q., Maimone, M., Franchina, F. A., Bjerk, T. R., Zini, C. A., Purcaro, G., & Mondello, L. (2016). Four-stage (low-)flow modulation comprehensive gas chromatography–quadrupole mass spectrometry for the determination of recently-highlighted cosmetic allergens. *Journal of Chromatography A*, *1439*, 144–151. https://doi.org/10.1016/j.chroma.2015.12.002

Vasquez, N. P., Crosnier de bellaistre-Bonose, M., Lévêque, N., Thioulouse, E., Doummar, D., Billette de Villemeur, T., … Moussa, F. (2015). Advances in the metabolic profiling of acidic compounds in children's urines achieved by comprehensive two-dimensional gas chromatography. *Journal of Chromatography B*, *1002*, 130–138. https://doi.org/10.1016/j.jchromb.2015.08.006

Vezzani, S., Castello, G., & Pierani, D. (1998). Measurement and prediction of dead times and column diameter in capillary gas chromatography by using air, methane and some solvents. *Journal of Chromatography A*, *811*(1–2), 85–96. https://doi.org/10.1016/S0021-9673(98)00215-5

Vosough, M., Bayat, M., & Salemi, A. (2010). Matrix-free analysis of aflatoxins in pistachio nuts using parallel factor modeling of liquid chromatography diode-array detection data. *Analytica Chimica Acta*, *663*(1), 11–18. https://doi.org/10.1016/j.aca.2010.01.039

Wilson, R. B., Hoggard, J. C., & Synovec, R. E. (2012). Fast, high peak capacity separations in gas chromatography-time-of-flight mass spectrometry. *Analytical Chemistry*, *84*(9), 4167–4173. https://doi.org/10.1021/ac300481k

Wilson, R. B., Hoggard, J. C., & Synovec, R. E. (2013). High throughput analysis of atmospheric volatile organic compounds by thermal injection – isothermal gas chromatography – time-of-flight mass spectrometry. *Talanta*, *103*, 95–102. https://doi.org/10.1016/j.talanta.2012.10.013

Wilson, R. B., Siegler, W. C., Hoggard, J. C., Fitz, B. D., Nadeau, J. S., & Synovec, R. E. (2011). Achieving high peak capacity production for gas chromatography and comprehensive two-dimensional gas chromatography by minimizing off-column peak broadening. *Journal of Chromatography A*, *1218*(21), 3130–3139. https://doi.org/10.1016/j.chroma.2010.12.108

Wong, Y. F., West, R. N., Chin, S.-T., & Marriott, P. J. (2015). Evaluation of fast enantioselective multidimensional gas chromatography methods for monoterpenic compounds: Authenticity control of Australian tea tree oil. *Journal of Chromatography A*, *1406*, 307–315. https://doi.org/10.1016/j.chroma.2015.06.036

Yang, S., Nadeau, J. S., Humston-Fulmer, E. M., Hoggard, J. C., Lidstrom, M. E., & Synovec, R. E. (2012). Gas chromatography–mass spectrometry with chemometric analysis for

determining 12C and 13C labeled contributions in metabolomics and 13C flux analysis. *Journal of Chromatography A, 1240*, 156–164. https://doi.org/10.1016/j.chroma.2012.03.072

Yang, S., Sadilek, M., Synovec, R. E., & Lidstrom, M. E. (2009). Liquid chromatography–tandem quadrupole mass spectrometry and comprehensive two-dimensional gas chromatography–time-of-flight mass spectrometry measurement of targeted metabolites of Methylobacterium extorquens AM1 grown on two different carbon sources. *Journal of Chromatography A, 1216*(15), 3280–3289. https://doi.org/10.1016/j.chroma.2009.02.030