Separation & Purification Reviews
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/lspr20

A Review of Chemometrics Applied to Comprehensive Two-dimensional Separations from 2008-2010

Karisa M. Pierce [a] & Rachel E. Mohler [b]
[a] Department of Chemistry and Biochemistry, Seattle Pacific University, Seattle, WA, USA
[b] Chevron Energy Technology Co., Richmond, CA, USA

Available online: 13 Oct 2011

To cite this article: Karisa M. Pierce & Rachel E. Mohler (2012): A Review of Chemometrics Applied to Comprehensive Two-dimensional Separations from 2008-2010, Separation & Purification Reviews, 41:2, 143-168

To link to this article: http://dx.doi.org/10.1080/15422119.2011.591868

Separation & Purification Reviews, 41:143–168, 2012

ISSN: 1542-2119 print/1542-2127 online
DOI: 10.1080/15422119.2011.591868

Taylor & Francis
Taylor & Francis Group

A Review of Chemometrics Applied to Comprehensive Two-dimensional Separations from 2008–2010

KARISA M. PIERCE[1] and RACHEL E. MOHLER[2]
[1]*Department of Chemistry and Biochemistry, Seattle Pacific University, Seattle, WA*
[2]*Chevron Energy Technology Co., Richmond, CA*

This review article covers developments in multidimensional separations combined with chemometrics that were published in 2008 through 2010, specifically for multidimensional gas chromatography, liquid chromatography, and electrophoresis. Although different instrumentation is used to generate multidimensional separations data, many similar data processing options and chemometrics can be applied in order to objectively distill the data into useful knowledge while reducing manual analysis and preserving data integrity. This review article describes the chemometrics employed in the referenced studies in terms of unsupervised, supervised, preprocessing, resolution, and image analysis algorithms. Other factors that affect converting data into useful knowledge are the structure of the data and the format of the data submitted to the analysis methods, so the studies are also described in terms of data dimensionality and data format (i.e., whether peak tables or raw data points were analyzed).

KEYWORDS Comprehensive two-dimensional gas chromatography, two-dimensional liquid chromatography, two-dimensional gel electrophoresis, chemometrics

Received February 28, 2011; Accepted May 27, 2011
Address correspondence to Karisa M. Pierce, Assistant Professor, Department of Chemistry and Biochemistry, Seattle Pacific University, 3307 Third Avenue West, Suite 205, Seattle, WA 98119-1950. E-mail: pierck1@spu.edu

INTRODUCTION

This review article covers developments in chemometrics applied to comprehensive two-dimensional (2D) separations that were published in 2008 through 2010, where chemometrics is defined as algorithmic and mathematical techniques for extracting information from chemical data. This article is designed to act as an update to the early 2008 review article in the *Journal of Chromatography A* covering the same topic but over the time period 2002 through 2007 (1). Since the 2008 review, there have been other reviews published describing quantitative analysis of comprehensive two-dimensional gas chromatography and in 2009, Amigo et al. wrote an excellent review article describing chemometrics and data analysis methods for one-dimensional (1D) and multidimensional chromatography, but we will limit our scope to chemometrics with multidimensional chromatography and multidimensional electrophoresis (2–4).

Analysts are increasingly turning to multidimensional separations to gain an improved understanding of complex samples. Typical multidimensional separations include comprehensive two-dimensional gas chromatography (GCxGC), two-dimensional liquid chromatography (LCxLC), and multidimensional electrophoresis (4–11). Thus, this review article is broken into three main sections based on the analytical instrumentation. Common issues arise among these three types of instruments when the analyst attempts to distill the data into useful knowledge and these issues are addressed by the chemometric techniques discussed herein. There are generally two approaches to converting the data into useful knowledge; one approach is to analyze peak data tables generated by the instrument software and the other approach is to analyze the raw data points.

The latter approach forces the analyst to manage massive volumes of data, so each of the three main sections is organized by separating the projects that processed raw data from the projects that processed peak tables provided by the instrument software. Another issue when handling multidimensional data is managing the structure of the data. For example, a univariate detector, e.g., flame ionization detector (FID), on a 2D chromatographic instrument continually collects signal data as a function of time, yielding a 1D data vector as shown in Figure 1A. This vector is composed of consecutive second dimension separations that can be reshaped into a 2D matrix as in Figure 1B where the peak "slices" that elute at a similar first dimension separation time and identical second dimension separation time are considered to be a single compound. Thus, it is necessary to combine multiple 1D peak "slices" into a single 2D peak with an integrated volume which, ideally, is proportional to chemical concentration. These 2D chromatograms are frequently depicted as a surface plot (Figure 1C) or as a contour plot (Figure 1D) where the contours represent signal magnitude.

Data analysis options are related to whether or not the data has been reshaped into a 2D matrix. An instrument with two separation dimensions

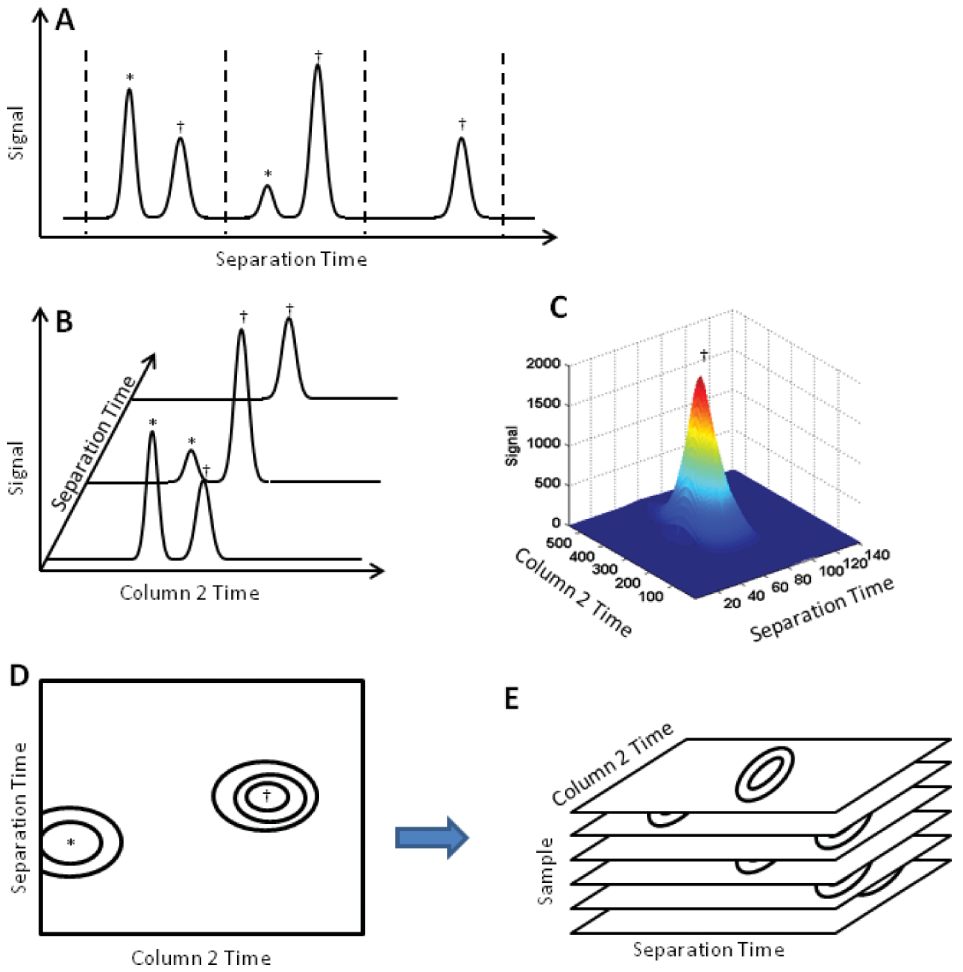

FIGURE 1 A univariate detector on a 2D chromatographic instrument collects a 1D data vector composed of consecutive second dimension separations (A) that can then be reshaped into a 2D matrix (B) where the peak "slices" that elute at a similar separation time and identical column 2 time are considered to be a single compound. Thus, multiple 1D peak "slices" are actually a single peak with an integrated volume that is proportional to chemical concentration. These 2D chromatograms are frequently depicted as surface plots (C) or as a contour plot (D) where the contours represent signal magnitude. Two chemicals represented by the asterisk and the cross are shown in these figures. When multiple 2D chromatograms are combined into a single 3-way array, then another dimension is added to the data (E). (color figure available online.)

and a univariate detector, e.g., GCxGC-FID, will generate 2D data, but an instrument with two separation dimensions combined with a multivariate detector, e.g., time-of-flight mass spectrometer (TOFMS), will generate three-dimensional (3D) data. When multiple 2D or 3D chromatograms are combined into a single 3-way or 4-way array, then another dimension is added to the data (Figure 1E) and this added dimensionality is yet another

factor that affects data analysis options. Thus, the projects in each main section are also organized by increasing data dimensionality.

This review focuses on analytical methods that use multidimensional chromatography combined with chemometrics that are objective, reduce manual analysis, and preserve data integrity (i.e., preserve resolution and precision throughout data exportation, compression, or reduction steps). Many chemometric data mining methods discussed herein can be loosely grouped into two categories: unsupervised or supervised. Unsupervised techniques are helpful when the user desires to discover the class of a complex sample, so no knowledge of class membership is required as an input for the chemometric algorithm. Principal components analysis (PCA) and hierarchical clustering analysis (HCA) are two common unsupervised techniques.

Supervised techniques are powerful for determining the complex sample components that distinguish given sample classes, which must be input by the user. Partial least squares analysis (PLS), linear discriminant analysis (LDA), or other methods based on the Fisher criterion are some of the common supervised techniques. Each chemometric technique will be briefly described when it is first referenced in the remainder of this text. Data preprocessing methods like baseline correction, normalization, alignment, and mathematical resolution are important steps in chemometric analysis because the preprocessing methods reduce variations that are unrelated to chemical variations.

Correlation optimized warping (COW) and wavelet transforms are two common techniques that are useful for preprocessing data. Mathematical resolution like parallel factor analysis (PARAFAC) is also a unique and important category of chemometrics. Thus, each main section in this review article is loosely organized by the categories (unsupervised, supervised, preprocessing, or resolution) of data analysis options that were prominently employed in the referenced project (though many of the projects could fit into more than one of these categories). Finally, another commonality evident among the projects discussed herein is a fifth data analysis option: automated 2D image analysis algorithms that supplant manual approaches. Again, each chemometric technique will be briefly described when it is first referenced in the remainder of this text. Detailed descriptions of the chemometric algorithms can be found in excellent textbooks written by Beebe et al., Brereton, and Sharaf et al. (12–14).

MULTIDIMENSIONAL GAS CHROMATOGRAPHY

Unsupervised Classification of Raw Data

The following projects focused on unsupervised classification techniques applied to 2D GCxGC data. The authors of these projects used

unsupervised chemometrics to classify their samples and discover chemical components that distinguish the classes. Groger et al. obtained comprehensive 2D gas chromatography with flame ionization detection (GCxGC-FID) and comprehensive 2D gas chromatography with total ion current mass spectrometry (GCxGC-TIC) chromatograms of heroin and cannabis samples (15).

The first step to processing data is to remove irrelevant variations prior to submitting the chromatograms to unsupervised and supervised analysis. For the heroin and cannabis study, Groger et al. applied baseline correction, normalization and alignment algorithms to their data. Baseline correction is an important preprocessing step for chromatographic data because the detector signal can have a constant offset or drift due to column bleed or background ionization and this can obscure important chemical variations across samples.

Normalization is also an important preprocessing step for chromatograms, because sample preparation can introduce bias and injectors are notorious for providing poor injection volume precision. Proper normalization reduces these sources of variation. Alignment is also an important preprocessing step because retention time variations between chromatograms occur due to uncontrollable pressure and temperature fluctuations as well as column evolution. Retention time alignment algorithms are designed to objectively improve retention time precision by shifting peak positions so each chemical has a constant reproducible retention time while preserving the accuracy of peak volumes.

Groger et al. used a dynamic warping algorithm described by Tomasi et al. known as COW (16). COW is a popular retention time alignment algorithm that works by subdividing the data into local regions that are iteratively stretched and compressed by interpolation until the correlation between the sample and target chromatograms is maximized. Following preprocessing of the data, HCA was used for unsupervised classification of the illicit drugs, followed by supervised Fisher criterion to locate chemicals that significantly distinguished the drugs.

HCA is one of many clustering algorithms that generally work by calculating the distance between samples in variable-space, where distance is often defined as the Euclidean or Mahalanobis distance among all samples from the centroid or origin. Classification is achieved by considering close samples to be similar to each other while increasingly more distant samples increasingly differ from each other. The Fisher criterion is a statistic that calculates the ratio of signal variance between classes to signal variance within each class as a function of an independent variable, in this case, retention time. Retention times with large Fisher ratios reveal the chemicals with statistically significant variations that differentiate the sample classes. The groups discovered in HCA can be used to identify the sample classes for the Fisher criterion, since the Fisher criterion requires class information.

Vial et al. obtained GCxGC-TIC chromatograms of tobacco plant extracts and they used PCA and a related chemometric technique called independent component analysis (ICA) to discriminate three types of tobacco plants (17).

PCA works by highly loading the chromatographic signals that have the greatest variation across all samples, and reducing the chromatograms from n-dimensional variable space into a much lower dimensional principal component (PC) space wherein the PCs are orthogonal, nested, and ordered based on eigenvalue or variance captured. In the lower dimensional PC space, similarities and differences among samples can be visualized and quantified because noise which may have obscured important chemical variations in the higher-dimensional variable space is removed in the truncated PC space.

ICA is designed to identify and extract independent pure component signals from a complex sample profile. Prior to PCA and ICA, Vial et al. also aligned their chromatograms using the COW algorithm (16). Zhang et al. reported an adaptation of the COW alignment algorithm for GCxGC-TIC and selected ion monitoring (SIM) chromatograms (18). Zhang et al.'s COW adaptation uses SIM information to help match peaks during alignment. Alignment algorithms that utilize mass spectral information for matching peaks are increasingly important as the data set of chromatograms moves from a set of similar samples (homogeneous data set) to a set composed of a wider variety of samples (heterogeneous data set).

Supervised Classification of Raw Data

The following projects focused on supervised classification methods applied to 2D and 3D GCxGC data wherein the authors were specifically authenticating blend composition. Pierce and Schale obtained GCxGC-TIC and GCxGC-TOFMS chromatograms of biodiesels blended with conventional diesels. The chromatograms were exported out of the ChemStation (Agilent, Santa Clara, CA) and ChromaTOF (LECO Corp., St. Joseph, MI) instrument software as .txt or .cdf files. These were imported into MATLAB (MATLAB, Natick, MA) where PLS and N-PLS were used to model and predict biodiesel blend percent compositions of training sets and independent test sets. The PLS algorithm was from the PLS Toolbox by Eigenvector Research Inc. (Eigenvector Research Inc., Wenatchee, WA). The N-PLS algorithm was from the N-Way Toolbox by Rasmus Bro (www.models.life.ku.dk/source/nwaytoolbox/). To remove chemically-irrelevant variations, the chromatograms underwent baseline correction and normalization prior to chemometric analysis (19).

PLS works by finding linear combinations of highly loaded chromatographic signals that co-vary with given quantitative information, thus creating a model that can be used to predict that quantitative property for new incoming chromatograms. N-PLS is an extension of PLS into multiple dimensions.

Importing chromatograms into MATLAB to increase data analysis options is another commonality among many of the references discussed herein. Pedroso et al. obtained GCxGC-FID chromatograms of adulterated gasolines. The chromatograms were exported out of the instrument software as ASCII files and then imported into MATLAB where N-PLS was used to model and predict adulteration of the gasolines (20).

De Godoy et al. obtained GCxGC-FID chromatograms of kerosene and gasoline blends (21). They used N-PLS, PARAFAC and PARAFAC2 to quantify the kerosene in the blends. PARAFAC works by using the fact that multidimensional chromatographic signals of pure chemicals are 3D, trilinear (defined by the outer product of three vectors), and additive in the presence of co-eluting components of selected chromatographic subregions. As long there is some physical separation (selectivity) of each of the co-eluting components on two of the three dimensions, then PARAFAC can mathematically resolve the component signals. This yields a prediction of the pure component concentrations and spectral profiles in the absence of a given pure standard. This mathematical resolution of overlapping peaks is called peak deconvolution. PARAFAC2 is a derivative of PARAFAC that is robust against retention time shifting. The authors concluded for their study that N-PLS resulted in the lowest RMSECV followed by PARAFAC2 and PARAFAC (21).

Mathematical Peak Resolution of Raw Data

The following projects focused on mathematical resolution (peak deconvolution) for quantification using raw 3D data generated by a GCxGC with a multivariate detector, as shown in Figure 2A. Hoggard and Synovec et al. obtained GCxGC-TOFMS chromatograms of forensic chemical warfare precursor samples. They used PARAFAC to quantify impurities in the samples and these quantified contaminants allowed successful determination of the sample sources (22). Snyder et al. also applied this same PARAFAC algorithm to select subregions of GCxGC-TOFMS chromatograms of human and mouse brain tissue to detect and quantify betamethylamino-alanine (BMAA), revealing evidence that it crosses the blood-brain barrier when mice are fed BMAA (23).

Furthermore, this analysis method combined with the instrumental method resulted in a 0.7 ppb LOD while the commercial method to detect BMAA has a 25 ppb LOD. Quantifying the peak volumes by PARAFAC greatly improved the signal to noise ratio (S/N) by removing noise as an independent component separate from the chemical signal. To produce these sophisticated results, researchers must be able to export raw data at the data point level (Figure 2B) from native instrument software and import it into a powerful data platform like MATLAB while preserving data integrity.

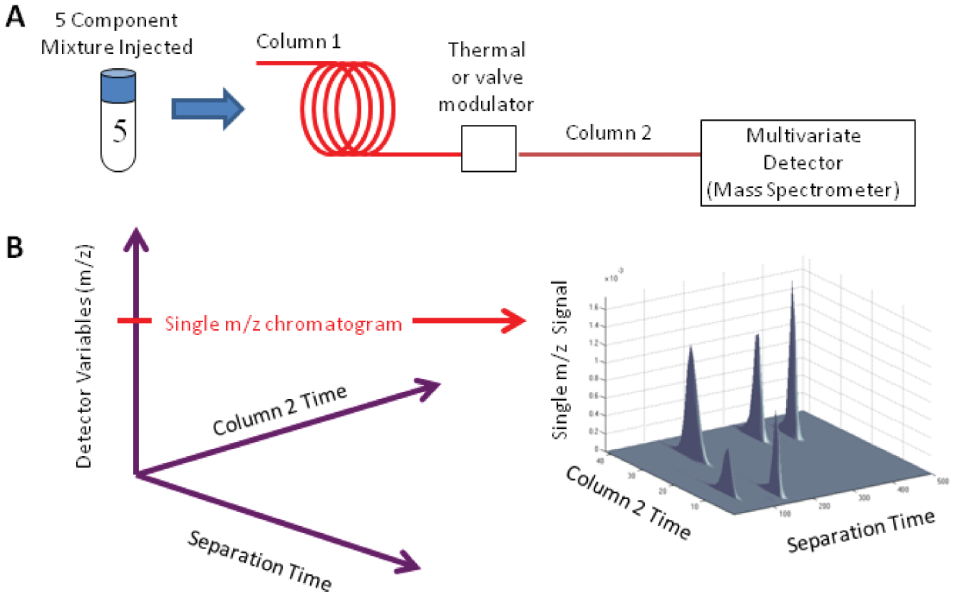

FIGURE 2 A comprehensive 2D gas chromatograph has a thermal or valve modulator that periodically injects effluent from a long nonpolar first column onto a short polar second column (A). When the 2D gas chromatograph has a multivariate detector, it produces 3D chromatograms (B) with high data density. Some researchers analyze the raw data at this data point level, while others analyze the peak lists output by the instrument software. (color figure available online.)

This can be a substantial and important hurdle to doing chemometric analysis. In related work, Hoggard and Synovec presented an algorithmic method to automatically select an appropriate number of factors for PARAFAC models applied to select subregions in GCxGC-TOFMS chromatograms, thus reducing by one the number of user inputs required for PARAFAC analysis (24, 25). This is an important advancement in converting PARAFAC into an objective method that requires fewer user inputs than prior PARAFAC algorithms. Furthermore, Hoggard converted this PARAFAC algorithm into an automated comprehensive PARAFAC algorithm that automatically resolves all peaks in an *entire* GCxGC-TOFMS chromatogram (26).

PARAFAC was already a powerful chemometric method in that it improves S/N so well for select subregions, but being converted into an algorithm that can process entire multidimensional chromatograms with fewer user inputs makes it even more powerful. In the same laboratory, Watson et al. built a comprehensive three-dimensional gas chromatography system with flame ionization detector (GCxGCxGC-FID). They applied PARAFAC to their chromatograms to add mathematical resolution to the three dimensions of chromatographic resolution already achieved (27). For PARAFAC to be accurate, the data must be trilinear, but PARAFAC2 can be successfully

applied to data that deviates from trilinearity due to retention time shifting. Skov et al. applied PARAFAC2 to GCxGC-TOFMS data for the first time (28). Their algorithm corrected within-run retention time shifting using the mass spectral information to restore trilinearity to the data and to ultimately provide accurate quantitative information.

When an analyst comprehensively compares multiple chromatograms from third-order instrumentation, then a 4-way array of data must be processed, and the following articles report techniques for analyzing such large volumes of four dimensional (4D) data. Mohler et al. obtained GCxGC-TOFMS chromatograms of metabolites from yeast extracts (29). They imported the chromatograms into MATLAB and determined whether metabolites were cycling as a function of cellular respiration using an S-ratio algorithm they developed in MATLAB. The S-ratio algorithm was built to analyze the entire 4-way array of multiple GCxGC-TOFMS chromatograms (both separation dimensions, all m/z, and all samples) and find complex sample components that cyclically vary in concentration.

Cycling concentrations were confirmed using PARAFAC quantification. Humston et al. also used the S-ratio algorithm applied to GCxGC-TOFMS chromatograms to identify cycling metabolites (30). They also confirmed the cyclical concentration variations using PARAFAC. Furthermore, they applied PCA to optimize sampling time. In related work, Humston et al. obtained GCxGC-TOFMS chromatograms of Cacao beans with various degrees of moisture damage (31).

They applied PCA, classification and regression trees (CART), random forests (RF), PARAFAC (24), and the F-ratio algorithm (which has built-in baseline correction and normalization options) (32), all implemented in MATLAB, with the goal of building a predictive model for quality of Cacao beans using the entire 4D data array. The tree algorithms search through each variable to find the value of a single variable that best splits the data into two groups, based on minimizing the mean square error of the model.

This splitting of the data continues until the specified criterion is met, thus resulting in a tree-like structure. RF is similar to CART with the main difference being that RF is an ensemble of trees. The authors confirmed that the analytes of interest found using the F-ratio algorithm did indeed have significant peak volume differences among the classes according to the peak list output by the native instrument software (31). In related work, Humston et al. also obtained GCxGC-TOFMS chromatograms of mixtures of naphthalene and its isotopic derivative as a standard with dozens of other similar compounds that chromatographically overlap the naphthalenes.

They imported the chromatograms into Matlab and used an isotopic dilution standardization method to quantify the unlabeled target analyte using only a single chromatogram based on the $^{13}C/^{12}C$ ratio. They used PARAFAC to mathematically resolve the naphthalenes and then a classical least squares (CLS) model for quantification (33). The CLS algorithm

mathematically resolves the ^{12}C contribution to the signal from the ^{13}C contribution to the signal since the pure component signals are additive and linearly related to concentration. These PARAFAC and/or CLS methods could be useful for the third order GCxGC isotope ratio mass spectrometry instrument developed and reported by Tobias and Sacks et al. (34).

Classification and Regression of Processed Data

Many of the references discussed heretofore used MATLAB as the main analysis platform, but researchers do import their chromatograms into other platforms such as image analysis programs, and other software packages which provide PCA, PLS, and other chemometrics (35, 36). For instance, Li et al. used the SIMCA platform (Umetrics, Umea, Sweden) to analyze peak lists by partial least squares discriminant analysis (PLSDA) with orthogonal signal correction (OSC), resulting in the discovery of five plasma biomarkers for diabetes (37).

The definition of PLSDA is essentially equal to the PLS definition provided above, but some authors make the distinction that PLS is specifically the main algorithm by which the data model is generated, while the term PLSDA indicates the PLS model was used for determining the sample components that co-vary with the given quantitative information. OSC is a data filter applied prior to PLS analysis which removes variance from the independent variables (e.g., chromatographic signals) that is orthogonal to the given information (e.g., quantitative information or classifications).

Thus, the important variations in independent variables that correlate with given information have a better chance of being captured and modeled by primary and secondary latent variables. Cordero et al. used GC Image software (GC Image, Lincoln, NE), to directly compare GCxGC-qMS chromatographic images based on a peak matching algorithm that utilized retention times and mass spectral information for accurate matching (38).

The result was identification of components that differentiated volatile fractions of coffee and plant leaves from a variety of geographic sources and roasting processes. In related work, Cordero et al. applied a similar method to headspace solid phase micro extraction (HS-SPME) GC×GC-qMS chromatograms and they developed a set of criteria for assessing the source and quality of volatile fractions of hazelnut products (39).

Schmarr and Bernhardt exported their GCxGC-qMS chromatograms of fruit volatiles out of the instrument software as .csv files, converted them into gray scale TIFF images, and warped them using Delta2d v. 4.02 (DECODON, Greifswald, Germany) software for alignment, PCA, and HCA (40). Vaz-Freire et al. obtained GCxGC-TOFMS chromatograms of three olive oils and then converted the ChromaTOF software contour plots into .jpg images using ImageJ 1.37v (Wayne Rasband, NIH, USA) software that

converts .jpg images into gray scale 8 bit images. The gray scale images were submitted to PCA using Statistica 6.0 software (Statsoft Inc) to classify the oils (41). Reichenbach et al. used their GC Image © GCxGC Software R2.1 and R2.2 to analyze 1.7 terabytes of 3D chromatographic data (http://www.gcimage.com) (42).

Their data set consisted of comprehensive two-dimensional gas chromatography with high-resolution mass spectrometry (GCxGC-HRMS) separations of breast cancer tumor samples representing three grades of the cancer. The GC Image software provided baseline correction, alignment, blob detection, and differential analysis for classification and biomarker detection. The GC Image software was capable of comprehensively processing the huge chromatograms because it stores the integer-mass chromatograms in random access memory (RAM) and then it accesses the high resolution information stored on the hard disk only when necessary.

The GC Image software can also export the chromatograms into universally available file formats, which is extremely important for users interested in applying chemometrics offered by other data analysis platforms. Reichenbach's report of successfully dealing with terabytes of data is significant considering that most of the projects described so far dealt with entire chromatograms and huge volumes of data, rather than dealing with the significantly smaller peak lists. Analyzing entire chromatograms should return results that are equal in accuracy to analyzing peak lists as long as appropriate preprocessing methods are used in both cases and as long as the peak list is exhaustive.

Qiu et al.'s project is one example of a method that generates and uses peak lists for chemometric analysis (35). The peak lists were generated by importing the .csv format GCxGC-FID chromatograms into Excel (Microsoft Corp., Redmond, WA) where the peaks were algorithmically matched and quantified. Then the peak lists were submitted to PCA and PLSDA in SIMCA.

Again, most of the projects described so far dealt with comprehensively analyzing entire chromatograms instead of peak lists. As long as appropriate preprocessing methods are used and as long as data integrity is maintained when a raw data file in its native format is reformatted into a compatible and exportable format, then analyzing entire chromatograms will return results that are equal in accuracy to analyzing peak lists. The following projects focused on methods for processing peak lists that are generated by the native instrument software, many of which were quantified peak lists provided by the LECO Pegasus GCxGC-TOFMS (LECO Corp., St. Joseph, MI) with ChromaTOF software.

The ChromaTOF quantification method uses mathematical peak resolution causing improved S/N that is ideal for trace-analysis applications like analyzing breath samples from cardiac surgery patients, detecting contaminants in wine and grape extracts, mathematically resolving and quantifying target steroids in nutritional supplements, and differentiating metabolite

profiles from central carbon metabolic cycles of *Methylobacterium extorquens* grown on two different carbon sources (43–46). The ChromaTOF peak lists contain retention time information on both columns, peak volume information, and mass spectral information. Thus, a peak list is always smaller than its corresponding raw chromatogram (in terms of data file size), and peak lists are more easily exported out of instrument software.

Indrasti et al. obtained ChromaTOF peak lists for glycerolysis products from animal and plant oils. They imported the peak lists into SAS version 6.12 (SAS Cary, NC) to normalize the data and submit it to analysis of variance (ANOVA) applications in order to determine the sources of variation in the data (47). The peak lists were also submitted to PCA in Unscrambler version 9.6 (CAMO Software AS, Oslo, Norway). The highly loaded variables in PCA were confirmed to be monoglycerides and diglycerides that differentiated the sample classes. Gaquerel et al. obtained ChromaTOF peak lists for the volatile emissions of plants that were provoked by herbivore interaction.

The peak lists were submitted to ANOVA for feature selection, and the features were submitted to HCA and PCA to classify plant emissions (48). Kempa et al. obtained ChromaTOF's text format peak lists for green algae that were differentially labeled with ^{13}C. The ChromaTOF peak lists with mass spectra were processed by novel software called MetMax which aligned the peak lists and assembled each into a matrix with other peak information. These matrices were then submitted to PCA and ICA in MATLAB (49).

Stanimirova et al. obtained ChromaTOF peak lists for honeys from two different sources. The peak lists were classified and compared by linear discriminant analysis (LDA) and PLSDA with support vector machines and Pearson VII universal kernel (50). LDA is a multivariate method that is related to the univariate Fisher criterion described earlier, so LDA seeks variables that separate classes. LDA works by finding linear combinations of chromatographic retention times that have signals with maximum Fisher ratios as long as the classes are normally distributed. Chin et al. obtained GCxGC-TOFMS chromatograms of fatty acid methyl esters from six different animals. They submitted the ChromaTOF peak list volumes to PCA in an effort to model biological variability of FAMES among individual animals (51).

The following projects focused on methods for processing ChromaTOF peak lists (LECO Corp., St. Joseph, MI) specifically applied to metabolomics studies. Metabolomics is a research field that focuses on comprehensively studying the small molecule metabolites produced in cellular processes. Ma et al. obtained ChromaTOF peak lists for metabolic extracts from two lines of transgenic *Artemisia annua L.* The peak lists were imported into SIMCA and submitted to PLSDA with analysis of variable importance in projection scores for feature selection and to determine metabolites that differentiated the classes (52).

Wang et al. also obtained ChromaTOF peak lists for metabolite samples. They developed an algorithm called distance and spectrum correlation optimization alignment (DISCO) which merges multiple entries of the same peak appearing in a ChromaTOF list for a single chromatogram, makes an alignment template, and aligns peak lists using mass spectral information (53). Ralston-Hooper et al. developed software called Mssort, which also merges multiple entries of the same peak appearing in a ChromaTOF peak list and aligns peak lists. The aligned peak lists were applied to PCA to reveal metabolic differences among the *Diporeia* samples grown in different environments (54).

Oh et al. also developed a peak sorting and alignment algorithm for the ChromaTOF peak lists and demonstrated that the algorithm aligns both chromatographic dimensions and matches mass spectra among peak lists (55). Almstetter et al. also algorithmically aligned ChromaTOF peak lists of metabolites, then they manually reduced the lists to a subset of features and submitted these to PCA to compare two classes (56). Li et al. obtained GCxGC-TOFMS chromatograms of metabolite samples from three genetic types of *E. coli* and they exported both the ChromaTOF peak lists and the m/z 73 chromatograms as .csv files. The peak lists were processed to merge split peaks and these were submitted to PCA and PLSDA. PLS results were used to optimize the parameters for fuzzy c-means clustering (57).

Fuzzy c-means clustering is similar to HCA in terms of using Euclidean or Mahalanobis distances in variable space to determine similarity of samples, but fuzzy c-means is more flexible because a single sample is assigned to multiple classes to differing degrees and final classification is determined using some convergence criterion, like minimizing an objective function or reaching a given sensitivity threshold for minimal difference between one model and the next iteration. Other clustering techniques can be used prior to fuzzy c-means clustering in order to estimate the number of clusters expected in the data set

These algorithms that match mass spectra for alignment purposes are related to target analyte analysis algorithms that are designed to match observed mass spectra with library mass spectra for identification. The automated mass spectral deconvolution and identification system (AMDIS) and the open mass spectrometry search algorithm (OMSSA) are target analyte analysis algorithms that were applied to 2D separations data during the 2008–2010 time period (58–61). The target analyte alogorithms are widely used in the tandem mass spectrometry (MS/MS) community.

MULTIDIMENSIONAL LIQUID CHROMATOGRAPHY

There were plenty of publications in the field of GCxGC combined with chemometrics during the 2008 through 2010 period, but there were also

a number of developments in LCxLC combined with chemometrics during the same time period. Developing an LCxLC instrument includes challenges that are different than the obstacles to developing a GCxGC instrument, but the data analysis challenges are often similar, including preprocessing, data exportation, and reduction of chemically irrelevant variations, all for the purpose of improving the interpretation of samples. An example is van der Klift et al. who adapted GCxGC software for visualizing and quantifying their LCxLC data in MATLAB (62).

Another example of the crossover between GCxGC and LCxLC is Reichenbach et al's. GC Image© software for both GC×GC and/or LC×LC, which is designed to identify and precisely quantify 2D peaks (http://www.gcimage.com) (63). Furthermore, Reichenbach et al. obtained LCxLC-UV chromatograms and applied the GC Image software to achieve sample classification via a sophisticated baseline correction algorithm, peak detection algorithm, and K-nearest neighbors (KNN) classification (64). In related work, Reichenbach et al. demonstrated their template pattern matching software called *Smart Templates* for LCxLC chromatograms.

The software takes chromatograms and creates a typical pattern of peaks with retention time data and chemical identities or sample classes. This typical pattern is used with rules, constraints, and mass spectral information to match peaks among chromatograms (65). Sherma used office scanner technology, Proquant image software, and ImageQuant for analyzing LCxLC data (66). Sherma's coupling of office imaging technology and LCxLC instrumentation might not fit in the chemometrics category as precisely as Reichenbach's GCImage software, but Sherma's work is an example of a creative crossover between separation science and image analysis techniques. Another LCxLC software system that was reported is called Chrom^{square} ver. 1.0 software (Chromaleont, Messina, Italy) which was used by both Dugo et al. and Mondello et al. for baseline correction, internal standard normalization, peak-finding and peak integration (8, 67).

Thekkudan et al. reported applying a Gaussian fitting algorithm to LCxLC chromatograms to quantify peak volumes (68). Stevenson et al. demonstrated their algorithm written in Wolfram Mathematica 7 (Hearn Scientific Software, Melbourne, VIC, Australia) which automates peak finding and 2D plotting for 2D HPLC chromatograms. The algorithm is based on using the derivative of raw chromatograms to find peak maxima, interpolate the chromatographic signals, and merge first dimension peak slices (69). The term "slice" is defined in Figure 1. Cantwell et al. simulated LCxLC data in order to demonstrate the Messick, Kalivas, and Lang (MKL) multivariate selectivity metric, which quantifies the selectivity of each peak with respect to all other peaks in the chromatogram using the individual component profiles resolved by PARAFAC (70–73). This was an interesting development because the best precision with which the popular PARAFAC algorithm can resolve a component is related to this multivariate selectivity value (70).

The selectivity and resolution of each peak is related to the orthogonality of the stationary phases and Dazykowski et al. developed a PCA and HCA method for determining the best pairs of coupled thin layer chromatography (TLC) phases for optimal 2D TLC (74). Kallio et al. developed software in MATLAB (The MathWorks, Natick, MA, USA) for visualizing and interpreting comprehensive LCxLC (or GCxGC) data once .csv format chromatograms are imported into MATLAB. The programs allows the user to visualize 2D and 3D plots, compare two 2D plots, and determine retention times, peak heights, and volumes (75).

MULTIDIMENSIONAL ELECTROPHORESIS

The 2D gel electrophoresis community has data processing methods and issues in common with the GCxGC and LCxLC communities since a 2D electropherogram has the same dimensionality as a GCxGC or LCxLC chromatogram generated by a univariate detector, as depicted in Figure 3. Quintana et al. obtained 2D gel electrophoresis separations of salivary proteins from healthy individuals as a function of sampling time. They imported the images into Statistica software (StatSoft, Tulsa, OK) and applied ANOVA (ratioing between-class variance to within-class variance), PCA, and HCA. The authors concluded there was no statistical significance in time of sampling due to variability of protein patterns, but individuals could be identified by PCA (76).

Plymoth et al. also obtained 2D gel electrophoresis separations of salivary protein samples from subjects that were smokers and nonsmokers. They imported the images into SIMCA-P software (Umetrics AB, Umea, Sweden) and applied PCA, HCA, and PLSDA. A correlation was found between smoker profiles and diagnosis of chronic obstructive pulmonary disease (77). Dowsey et al. used images of 2D gel electrophoresis separations to

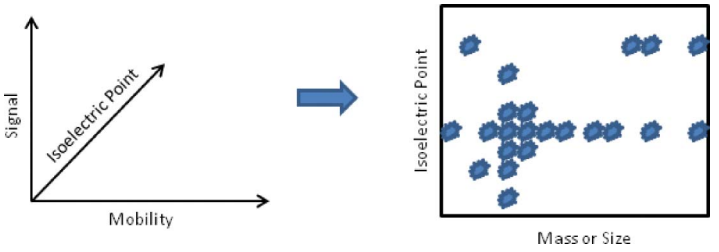

FIGURE 3 In 2D electrophoresis, a mixture of macro-molecules is often separated by isoelectric focusing in one dimension and mobility through a cross-linked gel matrix in a second physically perpendicular dimension. The 2D electropherograms generated by a univariate detector have the same dimensionality as a GCxGC or LCxLC chromatogram generated by a univariate detector, so many data analysis methods are common to all three types of instruments. (color figure available online.)

demonstrate their automated image processing algorithm that interpolates, normalizes, and aligns the images to do differential analysis of proteomics samples. The processing algorithm was developed using GPU technology on Nvidia (http://www.nvidia.com) consumer graphics card hardware with the Cg programming language (78).

Daszykowski et al. used images of 2D gel electrophoresis separations to demonstrate application of a fuzzy warping algorithm from the MATLAB toolbox (79). In related work, Daszykowski et al. developed a method for automatically preprocessing electrophoretic images at the pixel level, rather than the "spot" level, to avoid problems with missing elements in lists of spot data. Working with the pixels requires the user to input fewer optimal parameters so the user is less likely to inadvertently introduce variability into the data. The authors used robust orthogonal regression for baseline correction that works well even for undetected overlapping spots (80).

Marengo et al. used 2D gel electrophoresis separations of proteomics samples to demonstrate ranking PCA via forward search variable selection for discovering biomarkers (81). Ranking PCA is supervised PCA coupled to a forward search variable selection algorithm. Given class information, PCA is calculated iteratively where at each iteration, a new variable is added to the data set submitted to PCA. Variables that improve the clustering and separation between two classes are kept and variables that diminish the clustering are discarded. This observation results in discovering all variables that are possibly relevant to classification.

Soggiu et al. used 2D gel electrophoresis separations of serum and plasma samples to demonstrate a computational approach for detecting low abundance proteins. The traditional approach of unobscuring low abundance proteins is to chemically reduce the concentration of highly abundant proteins. Soggiu et al's. computational approach is to use wavelet denoising applied with the commercial imaging software ImageMaster 2D Platinum 6.0.1 (GE Healthcare) to transform images from a noisy function domain to a less noisy wavelet coefficients domain, and then reconstruct the less noisy image (82).

Rye et al. used images of entire 2D gel electrophoresis separations of animal tissue extracts to apply image morphology preprocessing that reduces streaks, corrects nonuniform backgrounds, and aligns entire images at the pixel level in MATLAB. The preprocessed images were then submitted to PCA and PLS for classification and modeling time after slaughter (83). Apraiz et al. used 2D isoelectric focusing and SDS-PAGE to obtain separations of proteins in mussels after an oil spill. The images were submitted to PCA using UMAX Image Scanner (GE Healthcare) (84).

Barbas et al. obtained 2D capillary electrophoresis-UV separations of samples from controls and diabetics that had been treated with antioxidants. The electropherograms were exported as ASCII files and preprocessed with Needelman-Wunsch dynamic programming for alignment in SEQSEE

(computational chemistry program), baseline correction, normalization, and multidimensional scaling. The preprocessed electropherograms were submitted to differential analysis, such as subtracting pairs of profiles to find differences between the two classes (85).

This review covers developments in 2D image analysis during the 2008 through 2010 period because our previous review article covered the 2002 through 2007 time period, but our previous review neglected to report a 2003 article by Luhn et al., which describes automated analysis of images of 2D gel electropherograms (86). Luhn et al. developed a method of using standard positions and image fusion to create proteome maps from 2D gel electrophoresis images in Delta2d (DECODON, Greifswald, Germany) software.

Finally, Skinner described a new instrument that coupled liquid chromatography and capillary array electrophoresis that produced 2D separation profiles. The chromatograms were imported into MATLAB where 2D Gaussian fitting was used to obtain peak widths for peak capacity evaluations (87). Skinner's coupling of LC and electrophoresis instrumentation is an example of the creative crossovers among GCxGC, LCxLC, and 2D electrophoresis projects in the separation science communities.

CONCLUSION

We reviewed the developments of chemometrics applied to multidimensional separations combined with chemometrics during the 2008–2010 time period, and we hope we did not inadvertently omit publications that were within the scope of this review article. Table 1 contains a categorized list of the chemometrics and software that were referenced in this review. It has repeatedly been shown that the advanced processing techniques can offer improved insight into the samples, both qualitatively and quantitatively.

The chemometric data processing tools are becoming increasingly mature and easy to use, and in some cases, as with PARAFAC and alignment algorithms, are evolving to require fewer user inputs and/or be applicable to higher-order data. The instrumentation and stand alone software are becoming more sophisticated allowing improved data processing options. Although, there has been significant advancement in the chemometric methods applied to multidimensional data, one area for improvement is in algorithms that utilize the entire data structure and find information that is only contained in higher-order data.

Processing peak lists utilizes the entire data structure when locating and quantifying peaks, but the instances where raw data (pixel-based data) is used, it is possible that higher-order algorithms could yield better results than lower-order algorithms, in some cases.

TABLE 1 Categorized List of the Chemometrics and Software that Were Referenced in This Review

Category	Software	Analyze Data Points (pixel level analysis)	Analyze Peak Tables (peak list analysis)	2D Instrument	3D Instrument
Preprocessing	baseline correction	8, 15, 19, 32, 42, 64, 67, 80, 83, 85		8, 15, 19, 80, 83, 85	19, 32, 42, 64, 67, 80
	normalization	8, 15, 19, 32, 67, 78, 85	47, 56	8, 15, 19, 78, 85	19, 32, 47, 56, 67
	wavelet noise reduction	82		82	
	alignment	15, 17, 18, 42, 49, 78, 79, 83, 85	49, 53, 54, 55, 56	15, 17, 18, 28, 78, 79, 83, 85	18, 42, 49, 53, 54, 55
	COW alignment (16)	15, 17, 18		15, 17, 18	18
	DISCO alignment		53		53
	image fusion	86		86	
	MSsort		54		54
Unsupervised	MetMax		49		49
	HCA	15, 40, 74, 76, 77	48	15, 74, 76, 77	40, 48
	Fuzzy c-means clustering		57		57
	KNN	64			64
	PCA	17, 31, 35, 36, 40, 41, 74, 76, 77, 83, 84	47, 48, 49, 51, 54, 56, 57	17, 35, 36, 74, 76, 77, 83, 84	30, 31, 35, 36, 40, 41, 47, 48, 49, 51, 54, 56, 57
Supervised	Ranking PCA	81		81	
	ICA	17	49	17	49
	OSC		37		37
	LDA		50		50
	Fisher criterion	15, 32		15	32
	differential analysis	42, 75, 78, 85		78, 85	42, 75
	ANOVA	76	47, 48	76	47, 48
	N-PLS	19, 20, 21, 83		19, 20, 21, 83	19
	PLSDA	35, 37, 77	35, 50, 52, 57	35, 77	35, 37, 50, 52, 57
	S-ratio algorithm	29, 30			29, 30
	CART	31			31
	Random Forests	31			31

Resolution	PARAFAC	21, 22, 23, 24, 25, 26, 27, 29, 30, 33, 70		21	22, 23, 24, 25, 26, 27, 29, 30, 33, 70
	PARAFAC2	21, 22, 28		21	22, 28
	CLS	33			33
Peak Metrics	Gaussian fitting algorithm	68, 87		87	68
	SAXICAB	34			34
	MKL multivariate selectivity (73)	70, 72			70, 72
	Peak-picking and separation performance (69)	69		69	
Target	AMDIS (59)	58			58
	OMSSA target analysis (61)	60			60
Software for Visualizing Data, Includes Multiple Analysis Functions	ChromaTOF (includes resolution, quantification, target analysis)		43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57		43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57
	GC Image (includes alignment, quantification)	38, 39, 42, 63, 64			38, 39, 42, 63, 64
	Delta2D (includes alignment)	40, 86		86	40
	ImageJ 1.37v	41			41
	Statistica	41, 76		76	41
	Unscrambler		47		47
	UMAX Image Scanner (84)	84		84	
	Smart Templates (65)	65			65
	ProQuant	66			66
	ImageQuant	66			66
	Chrom^{square} ver. 1.0	8, 67			8, 67

REFERENCES

1. Pierce, K., Hoggard, J., Mohler, R., and Synovec, R. (2008) Recent advancements in comprehensive two-dimensional separations with chemometrics. *Journal of Chromatography A*, 1184: 341–352.
2. Amigo, J.M., Skov, T., and Bro, R. (2009) ChroMATHography: Solving chromatographic issues with mathematical models and intuitive graphics. *Chemical Reviews*, DOI: 10.1021/cr900394n.
3. Amador-Muñoz, O. and Marriott, P.J. (2008) Quantification in comprehensive two-dimensional gas chromatography and a model of quantification based on selected summed modulated peaks. *Journal of Chromatography A*, 1184: 323–340.
4. Cortes, H.J., Winniford, B., Luong, J., and Pursch, M. (2009) Comprehensive two-dimensional gas chromatography review. *Journal of Separation Science*, 32: 883–904.
5. Stoll, D.R. (2010) Recent progress in online, comprehensive two-dimensional high-performance liquid chromatography for non-proteomic applications. *Analytical and Bioanalytical Chemistry*, 397: 979–986.
6. Janssen, H.-G., Steenbergen, H., and de Koning, S. (2009) The role of comprehensive chromatography in the characterization of edible oils and fats. *European Journal of Lipid Science and Technology*, 111: 1171–1184.
7. Hamilton, J.F. (2010) Using comprehensive two-dimensional gas chromatography to study the atmosphere. *Journal of Chromatographic Science*, 48: 274–282.
8. Mondello, L., Herrero, M., Kumm, T., Dugo, P., Cortes, H., and Dugo, G. (2008) Quantification in comprehensive two-dimensional liquid chromatography. *Analytical Chemistry*, 80: 5418–5424.
9. Mondello, L., Tranchida, P.Q., Dugo, P., and Dugo, G. (2008) Comprehensive two-dimensional gas chromatography-mass spectrometry: A review. *Mass Spectrometry Reviews*, 27: 101–124.
10. Wang, Y., Chen, Q., Norwood, D.L., and McCaffrey, J. (2010) Recent developments in the applications of comprehensive two-dimensional gas chromatography. *Journal of Liquid Chromatography and Related Technologies*, 33: 1082–1115.
11. François, I., Sandra, K., and Sandra, P. (2009) Comprehensive liquid chromatography: Fundamental aspects and practical considerations – A review. *Analytica Chimica Acta*, 641: 14–31.
12. Brereton, R.G. (2003) *Chemometrics: Data Analysis for the Laboratory and Chemical Plant*; Wiley: New York, United States.
13. Beebe, K.R., Pell, R.J., and Seasholtz, M.B. (1998) *Chemometrics: A Practical Guide*; Wiley-Interscience: New York.
14. Sharaf, M.A., Illman, D.L., and Kowalski, B.R. (1986) *Chemometrics*. Wiley: New York, United States.
15. Groger, T., Schaffer, M., Putz, M., Ahrens, B., Drew, K., Eschner, M., and Zimmermann, R. (2008) Application of two-dimensional gas chromatography combined with pixel-based chemometric processing for the chemical profiling of illicit drug samples. *Journal of Chromatography A*, 1200: 8–16.

16. Tomasi, G., Berg, F.v.d., and Andersson, C. (2004) Correlation optimized warping and dynamic time warping as preprocessing methods for chromatographic data. *Journal of Chemometrics*, 18: 231–241.
17. Vial, J., Noçairi, H., Sassiat, P., Mallipatu, S., Cognon, G., Thiébaut, D., Teillet, B., and Rutledge, D.N. (2009) Combination of dynamic time warping and multivariate analysis for the comparison of comprehensive two-dimensional gas chromatogramsApplication to plant extracts. *Journal of Chromatography A*, 1216: 2866–2872.
18. Zhang, D., Huang, X., Regnier, F.E., and Zhang, M. (2008) Two-Dimensional correlation optimized warping algorithm for aligning GCxGC-MS data. *Analytical Chemistry*, 80: 2664–2671.
19. Pierce, K.M. and Schale, S.P. (2010) Predicting percent composition of blends of biodiesel and conventional diesel using gas chromatography–mass spectrometry,comprehensive two-dimensional gas chromatography–mass spectrometry,and partial least squares analysis. *Talanta*, doi:10.1016/j.talanta.2010.07.084:
20. Pedroso, M., Degodoy, L., Ferreira, E., Poppi, R., and Augusto, F. (2008) Identification of gasoline adulteration using comprehensive two-dimensional gas chromatography combined to multivariate data processing. *Journal of Chromatography A*, 1201: 176–182.
21. de Godoy, L.A., Ferreira, E.C., Pedroso, M.P., Fidelis, C.H.d., Augusto, F., and Poppi, R.J. (2008) Quantification of kerosene in gasoline by comprehensive two-dimensional gas chromatography and N-Way Multivariate Analysis. *Analytical Letters*, 41: 1603–1614.
22. Hoggard, J.C., Wahl, J.H., Synovec, R.E., Mong, G.M., and 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: 689–698.
23. Snyder, L.R., Hoggard, J.C., Montine, T.J., and Synovec, R.E. (2010) 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. *Journal of Chromatography A*, 1217: 4639–4647.
24. Hoggard, J.C. and Synovec, R.E. (2007) Parallel Factor Analysis (PARAFAC) of Target Analytes in GCxGC-TOFMS Data: Automated Selection of a Model with an Appropriate Number of Factors. *Analytical Chemistry*, 79: 1611–1619.
25. Hoggard, J.C., and Synovec, R.E. (2008) Automated resolution of nontarget analyte signals in GC × GC-TOFMS data using parallel factor analysis. *Analytical Chemistry*, 80: 6677–6688.
26. Hoggard, J.C., Siegler, W.C., and Synovec, R.E. (2009) Toward automated peak resolution in complete GC × GC-TOFMS chromatograms by PARAFAC. *Journal of Chemometrics*, 23: 421–431.
27. Watson, N.E., Siegler, W.C., Hoggard, J.C., and Synovec, R.E. (2007) Comprehensive three-dimensional gas chromatography with parallel factor analysis. *Analytical Chemistry*, 79: 8270–8280.
28. Skov, T., Hoggard, J.C., Bro, R., and 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: 4020–4029.

29. Mohler, R., Tu, B., Dombek, K., Hoggard, J., Young, E., and Synovec, R. (2008) Identification and evaluation of cycling yeast metabolites in two-dimensional comprehensive gas chromatography–time-of-flight-mass spectrometry data. *Journal of Chromatography A*, 1186: 401–411.
30. Humston, E.M., Dombek, K.M., Hoggard, J.C., Young, E.T., and Synovec, R.E. (2008) Time-Dependent profiling of metabolites from Snf1 mutant and wild type yeast cells. *Analytical Chemistry*, 80: 8002–8011.
31. Humston, E.M., Knowles, J.D., McShea, A., and Synovec, R.E. (2010) Quantitative assessment of moisture damage for cacao bean quality using two-dimensional gas chromatography combined with time-of-flight mass spectrometry and chemometrics. *Journal of Chromatography A*, 1217: 1963–1970.
32. Pierce, K.M., Hoggard, J.C., Hope, J.L., Rainey, P.M., Hoofnagle, A.N., Jack, R.M., Wright, B.W., and Synovec, R.E. (2006) Fisher ratio method applied to third-order separation data to identify significant chemical components of metabolite extracts. *Analytical Chemistry*, 78: 5068–5075.
33. Humston, E.M., Hoggard, J.C., and Synovec, R.E. (2010) Utilizing the third order advantage with isotope dilution mass spectrometry. *Analytical Chemistry*, 82: 41–43.
34. Tobias, H.J., Sacks, G.L., Zhang, Y., and Brenna, J.T. (2008) Comprehensive Two-dimensional gas chromatography combustion isotope ratio mass spectrometry. *Analytical Chemistry*, 80: 8613–8621.
35. Qiu, Y., Lu, X., Pang, T., Ma, C., Li, X., and Xu, G. (2008) Determination of radix ginseng volatile oils at different ages by comprehensive two-dimensional gas chromatography/time-of-flight mass spectrometry. *Journal of Separation Science*, 31: 3451–3457.
36. Qiu, Y., Lu, X., Pang, T., Zhu, S., Kong, H., and Xu, G. (2007) Study of traditional Chinese medicine volatile oils from different geographical origins by comprehensive two-dimensional gas chromatography–time-of-flight mass spectrometry (GC×GC–TOFMS) in combination with multivariate analysis. *Journal of Pharmaceutical and Biomedical Analysis*, 43: 1721–1727.
37. Li, X., Xu, Z., Lu, X., Yang, X., Yin, P., Kong, H., Yu, Y., and Xu, G. (2009) Comprehensive two-dimensional gas chromatography/time-of-flight mass spectrometry for metabonomics: Biomarker discovery for diabetes mellitus. *Analytica Chimica Acta*, 633: 257–262.
38. Cordero, C., Liberto, E., Bicchi, C., Rubiolo, P., Reichenbach, S.E., Tian, X., and Tao, Q. (2010) Targeted and non-targeted approaches for complex natural sample profiling by GCxGC-qMS. *Journal of Chromatographic Science*, 48: 251–261.
39. Cordero, C., Liberto, E., Bicchi, C., Rubiolo, P., Schieberle, P., Reichenbach, S.E., and Taod, Q. (2010) Profiling food volatiles by comprehensive two-dimensional gas chromatography coupled with mass spectrometry: Advanced fingerprinting approaches for comparative analysis of the volatile fraction of roasted hazelnuts (*Corylus avellana L.*) from different origins. *Journal of Chromatography A*, 1217: 5848–5858.
40. Schmarr, H.-G. and Bernhardt, J. (2010) Profiling analysis of volatile compounds from fruits using comprehensive two-dimensional gas chromatography and image processing techniques. *Journal of Chromatography A*, 1217: 565–574.

41. Vaz-Freire, L., Dasilva, M., and Freitas, A. (2009) Comprehensive two-dimensional gas chromatography for fingerprint pattern recognition in olive oils produced by two different techniques in Portuguese olive varieties Galega Vulgar, Cobrançosa e Carrasquenha. *Analytica Chimica Acta*, 633: 263–270.
42. Reichenbach, S.E., Tian, X., Tao, Q., Ledford, E.B., Wu, Z., and Fiehn, O. (2010) Informatics for cross-sample analysis with comprehensive two-dimensional gas chromatography and high-resolution mass spectrometry (GCxGC-HRMS). *Talanta*, doi: 10.1016/j.talanta.2010.09.057
43. Dasgupta, S., Banerjee, K., Patil, S.H., Ghaste, M., Dhumal, K.N., and Adsule, P.G. (2010) Optimization of two-dimensional gas chromatography time-of-flight mass spectrometry for separation and estimation of the residues of 160 pesticides and 25 persistent organic pollutants in grape and wine. *Journal of Chromatography A*, 1217: 3881–3889.
44. Mieth, M., Schubert, J.K., Groger, T., Sabel, B., Kischkel, S., Fuchs, P., Hein, D., Zimmermann, R., and Miekisch, W. (2010) Automated needle trap heart-cut GC/MS and needle trap comprehensive two-dimensional GC/TOF-MS for breath gas analysis in the clinical environment. *Analytical Chemistry*, 82: 2541–2551.
45. Yang, S., Sadilek, M., Synovec, R.E., and 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: 3280–3289.
46. Stepan, R., Cuhra, P., and Barsova, S. (2008) Comprehensive two-dimensional gas chromatography with time-of-flight mass spectrometric detection for the determination of anabolic steroids and related compounds in nutritional supplements. *Food Additives and Contaminants*, 25: 557–565.
47. Indrasti, D., Che Man, Y.B., Chin, S.T., Mustafa, S., Mat Hashim, D., and Abdul Manaf, M. (2010) Regiospecific Analysis of Mono- and Diglycerides in Glycerolysis Products by GC × GC-TOF-MS. *Journal of the American Oil Chemists' Society*, DOI 10.1007/s11746-010-1614–x
48. Gaquerel, E., Weinhold, A., and Baldwin, I.T. (2009) Molecular interactions between the specialist herbivore *Manduca sexta* (Lepidoptera, Sphigidae) and Its natural host *Nicotiana attenuata*. VIII. An unbiased GC x GC-ToFMS analysis of the plant's elicited volatile emissions. *Plant Physiology*, 149: 1408–1423.
49. Kempa, S., Hummel, J., Schwemmer, T., Pietzke, M., Strehmel, N., Wienkoop, S., Kopka, J., and Weckwerth, W. (2009) An automated GCxGC-TOFMS protocol for batch-wise extraction and alignment of mass isotopomer matrixes from differential ^{13}C-labelling experiments: a case study for photoautotrophic-mixotrophic grown *Chlamydomonas reinhardtii* cells. *Journal of Basic Microbiology*, 49: 82–91.
50. Stanimirova, I., Üstün, B., Cajka, T., Riddelova, K., Hajslova, J., Buydens, L.M.C., and Walczak, B. (2010) Tracing the geographical origin of honeys based on volatile compounds profiles assessment using pattern recognition techniques. *Food Chemistry*, 118: 171–176.

51. Chin, S.-T., Che Man, Y.B., Tan, C.P., and Hashim, D.M. (2009) Rapid profiling of animal-derived fatty acids using fast GC × GC coupled to time-of-flight mass spectrometry. *Journal of the American Oil Chemists' Society*, 86: 949–958.
52. Ma, C., Wang, H., Lu, X., Wang, H., Xu, G., and Liu, B. (2009) Terpenoid metabolic profiling analysis of transgenic Artemisia annua L. by comprehensive two-dimensional gas chromatography time-of-flight mass spectrometry. *Metabolomics*, 5: 497–506.
53. Wang, B., Fang, A., Heim, J., Bogdanov, B., Pugh, S., Libardoni, M., and Zhang, X. (2010) DISCO: Distance and Spectrum correlation optimization alignment for two-dimensional gas chromatography time-of-flight mass spectrometry-based metabolomics. *Analytical Chemistry*, 82: 5069–5081.
54. Ralston-Hooper, K., Hopf, A., Oh, C., Zhang, X., Adamec, J., and Sepulveda, M. (2008) Development of GCxGC/TOF-MS metabolomics for use in ecotoxicological studies with invertebrates. *Aquatic Toxicology*, 88: 48–52.
55. Oh, C., Huang, X., Regnier, F., Buck, C., and Zhang, X. (2008) Comprehensive two-dimensional gas chromatography/time-of-flight mass spectrometry peak sorting algorithm. *Journal of Chromatography A*, 1179: 205–215.
56. Almstetter, M.F., Appel, I.J., Gruber, M.A., Lottaz, C., Timischl, B., Spang, R., Dettmer, K., and Oefner, P.J. (2009) Integrative normalization and comparative analysis for metabolic fingerprinting by comprehensive two-dimensional gas chromatography-time-of-flight mass spectrometry. *Analytical Chemistry*, 81: 5731–5739.
57. Li, X., Lu, X., Tian, J., Gao, P., Kong, H., and Xu, G. (2009) Application of Fuzzy c-Means clustering in data analysis of metabolomics. *Analytical Chemistry*, 81: 4468–4475.
58. Ma, Y. and Hays, M.D. (2008) Thermal extraction–two-dimensional gas chromatography–mass spectrometry with heart-cutting for nitrogen heterocyclics in biomass burning aerosols. *Journal of Chromatography A*, 1200: 228–234.
59. Stein, S.E. (1999) An integrated method for spectrum extraction and compound identification from gas chromatography/ mass spectrometry data. *Journal of the American Society of Mass Spectrometry*, 10: 770–781.
60. Fang, X., Balgley, B.M., Wang, W., Park, D.M., and Lee, C.S. (2009) Comparison of multidimensional shotgun technologies targeting tissue proteomics. *Electrophoresis*, 30: 4063–4070.
61. Geer, L.Y., Markey, S.P., Kowalak, J.A., Wagner, L., Xu, M., Maynard, D.M., Yang, X., Shi, W., and Bryant, S.H. (2004) Open mass spectrometry search algorithm. *Journal of Proteome Research*, 3: 958–964.
62. Klift, E.J.C.v.d., Viv´o-Truyols, G., Claassen, F.W., Holthoon, F.L.v., and Beek, T.A.v. (2008) Comprehensive two-dimensional liquid chromatography with ultraviolet, evaporative light scattering and mass spectrometric detection of triacylglycerols in corn oil. *Journal of Chromatography A*, 1178: 43–55.
63. Reichenbach, S.E. (2009) Quantification in comprehensive two-dimensional liquid chromatography. *Analytical Chemistry*, 81: 5099–5101.
64. Reichenbach, S.E., Tian, X., Tao, Q., Stoll, D.R., and Carr, P.W. (2010) Comprehensive feature analysis for sample classification with comprehensive two-dimensional LC. *Journal of Separation Science*, 33: 1365–1374.

65. Reichenbach, S.E., Carr, P.W., Stoll, D.R., and Tao, Q. (2009) Smart Templates for peak pattern matching with comprehensive two-dimensional liquid chromatography. *Journal of Chromatography A*, 1216: 3458–3466.
66. Sherma, J. (2010) Planar chromatography. *Analytical Chemistry*, 82: 4895–4910.
67. Dugo, P., Cacciola, F., Donato, P., Airado-Rodríguez, D., Herrero, M., and Mondello, L. (2009) Comprehensive two-dimensional liquid chromatography to quantify polyphenols in red wines. *Journal of Chromatography A*, 1216: 7483–7487.
68. Thekkudan, D.F., Rutan, S.C., and 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: 4313–4327.
69. Stevenson, P.G., Mnatsakanyan, M., Guiochon, G., and Shalliker, R.A. (2010) Peak picking and the assessment of separation performance in two-dimensional high performance liquid chromatography. *Analyst*, 135: 1541–1550.
70. Cantwell, M.T., Porter, S.E.G., and Rutan, S.C. (2007) Evaluation of the multivariate selectivity of multi-way liquid chromatography methods. *Journal of Chemometrics*, 21: 335–345.
71. Olivieri, A.C. (2005) Computing sensitivity and selectivity in parallel factor analysis and related multiway techniques: The need for further developments in net analyte signal theory. *Analytical Chemistry*, 77: 4936–4946.
72. Olivieri, A.C. (2008) Analytical advantages of multivariate data processing. One, two, three, Infinity? *Analytical Chemistry*, 80: 5713–5720.
73. Messick, N.J., Kalivas, J.H., and Lang, P.M. (1996) Selectivity and related measures for nth-order data. *Analytical Chemistry*, 68: 1572–1579.
74. Daszykowski, M., Hawrył, M., Waksmundzka-Hajnos, M., and Walczak, B. (2008) Identification of similar and orthogonal chromatographic thin-layer systems for two-dimensional separations of flavonoids and their analogues. *Acta Chromatographica*, 20: 283–307.
75. Kallio, M., Kivilompolo, M., Varjo, S., Jussila, M., and Hyötyläinen, T. (2009) Data analysis programs for comprehensive two-dimensional chromatography. *Journal of Chromatography A*, 1216: 2923–2927.
76. Quintana, M., Palicki, O., Lucchi, G., Ducoroy, P., Chambon, C., Salles, C., and Morzel, M. (2009) Inter-individual variability of protein patterns in saliva of healthy adults. *Journal of Proteomics*, 72: 822–830.
77. Plymoth, A., Lofdahl, C.G., Ekberg-Jansson, A., Dahlback, M., Broberg, P., Foster, M., Fehniger, T.E., and Marko-Varga, G. (2007) Protein expression patterns associated with progression of chronic obstructive pulmonary disease in bronchoalveolar lavage of smokers. *Clinical Chemistry*, 53: 636–644.
78. Dowsey, A.W., Dunn, M.J., and Yang, G.Z. (2008) Automated image alignment for 2D gel electrophoresis in a high-throughput proteomics pipeline. *Bioinformatics*, 24: 950–957.
79. Daszykowski, M., Færgestad, E.M., Grove, H., Martens, H., and Walczak, B. (2009) Matching 2D gel electrophoresis images with MATLAB 'Image Processing Toolbox'. *Chemometrics and Intelligent Laboratory Systems*, 96: 188–195.

80. Daszykowski, M., Wróbel, M.S., Bierczynska-Krzysik, A., Silberring, J., Lubec, G., and Walczak, B. (2009) Automatic preprocessing of electrophoretic images. *Chemometrics and Intelligent Laboratory Systems*, 97: 132–140.
81. Marengo, E., Robotti, E., Bobba, M., and Gosetti, F. (2010) The principle of exhaustiveness versus the principle of parsimony: A new approach for the identification of biomarkers from proteomic spot volume datasets based on principal component analysis. *Analytical and Bioanalytical Chemistry*, 397: 25–41.
82. Soggiu, A., Marullo, O., Roncada, P., and Capobianco, E. (2009) Empowering spot detection in 2DE Images by wavelet denoising. *In Silico Biology*, 9: 125–133.
83. Rye, M.B., Faergestad, E.M., Martens, H., Wold, J.P., and Alsberg, B.K. (2008) An improved pixel-based approach for analyzing images in two-dimensional gel electrophoresis. *Electrophoresis*, 29: 1382–1393.
84. Apraiz, I., Cajaraville, M.P., and Cristobal, S. (2009) Peroxisomal proteomics: Biomonitoring in mussels after the Prestige's oil spill. *Marine Pollution Bulletin*, 58: 1815–1826.
85. Barbas, C., Vallejo, M., Garcia, A., Barlow, D., and Hannabrown, M. (2008) Capillary electrophoresis as a metabolomic tool in antioxidant therapy studies. *Journal of Pharmaceutical and Biomedical Analysis*, 47: 388–398.
86. Luhn, S., Berth, M., Hecker, M., and Bernhardt, J. (2003) Using standard positions and image fusion to create proteome maps from collections of two-dimensional gel electrophoresis images. *Proteomics*, 3: 1117–1127.
87. Skinner, C.D. (2010) A liquid chromatography to capillary array electrophoresis interface for two-dimensional separations. *Analyst*, 135: 358–367.