An Optimized Retention Time Alignment Method for High-Speed Gas Chromatography

Robert Chaiyo Halvorsen

A thesis

submitted in partial requirement of the

requirements for the degree of

Master of Science

University of Washington

2023

Committee:

Robert E. Synovec

Dan Fu

Program Authorized to Offer Degree: Chemistry

University of Washington

Abstract

An Optimized Retention Time Alignment Method for High-Speed Gas Chromatography

Robert Chaiyo Halvorsen

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

The alignment of chromatograms collected on nominally identical columns using retention time locking (RTL), an instrumental alignment tool, and software-based alignment using correlation optimized warping (COW) is herein analyzed before a more optimized method combining both forms of alignment is described. To analyze the extent of misalignment and the success of corrective methods, three samples were constructed by spiking two sets of analytes into a base test mixture. The three samples were then analyzed by high-speed gas chromatography with four nominally identical columns and identical separation conditions, which due to unavoidable differences from manufacturing imprecision leads to four different sets of retention times. The data is first analyzed without alignment, then using COW alone, then RTL alone, and finally with RTL followed by COW. Principal component analysis (PCA) is used to investigate how well each alignment method allowed for clustering of the chromatograms into the three sample classes in a scores plot without being compromised by the specific column(s) used, where

successful clustering and thus recovery of sample class differences as desired indicated successful alignment. Only the previously unreported method of combining RTL with COW achieved this degree of alignment. To provide a quantitative metric, the degree-of-class separation (*DCS*), measured as the Euclidian distance between the centroids of two clusters in PC space in the scores plot and normalized by their pooled variance, is used. With no alignment, the average *DCS* between sample classes (DCS_{sam}) was 3.0, while the average *DCS* between the four nominally identical columns, i.e., column classes (DCS_{col}) was 76.1 (ideally the DCS_{col} should be 0), indicating the chromatograms were initially classified by the columns used. Using either COW or RTL alone also produced unsatisfactory results, with COW alone incorrectly aligned many peaks, leading to a DCS_{sam} of only 1.9 and DCS_{col} of 1.7, while RTL alone provided a DCS_{sam} 4.7 of and DCS_{col} of 4.2. Finally, using RTL followed by COW alignment, DCS_{sam} increased to 32.5, indicating successful classification by chemical differences between sample classes, while the DCS_{col} decreased to 0.4, indicating virtually no classification due to column-to-column differences, as desired. Thus, RTL provided a "first order" correction of the initial retention mismatch observed for the nominally identical columns, while additional alignment via COW was required to optimize sample classification by PCA.

This thesis expands upon previously published work in Robert C. Halvorsen, Timothy J. Trinklein, Cable G. Warren, Riley D. Rogan and Robert E. Synovec, "Optimizing column-to-column retention time alignment in high-speed gas chromatography by combining retention time locking and correlation optimized warping" *Talanta* 254 (2023) 124173, https://doi.org/10.1016/j.talanta.2022.124173

Table of Contents

List of Figures

List of Tables

ACKNOWLEDGEMENTS

Many thanks to past and present members of the Synovec Group – Riley, Vlada, Paige, Tim, Caitlin, Sonia, Grant, Lina, Cable, Owen, Wenjing, Joe, Meriam, and others – both for their assistance and for making research as inspirational as it is.

My sincere gratitude to Professor Synovec for his mentorship, from devising a way for me to start participating in research in the midst of a lockdown by doing chemometrics work remotely through to the present day.

Finally, I'd like to thank my parents and our cat Noo for always being there for me.

DEDICATION

To my parents, Robert Francis and Pranee Khruasanit Halvorsen, for their love and support through the last 28 years, without which none of this would have been possible.

Chapter 1. Overview of High-Speed Gas Chromatography

1.1 Introduction

Gas chromatography (GC) is an instrumental method of choice for the analysis of volatile and semi-volatile compounds in which volatile analytes are passed through a column either lined with or packed by a stationary phase by means of a carrier gas [1–3]. As the analytes pass through the column, their interactions with the stationary phase, resulting in separation of analytes based on their relevant properties such as boiling points and polarities. The user has a wide variety of stationary phases available allow the user to tune separations to specific properties, and the column is in modern practice often heated to a specific temperature ramp to limit peak broadening while preserving peak separations as desired.

However, one of its main drawbacks can be the relatively slow rate of analysis - many conventional GC methods use run times on the order of 30 min to an hour. Attempts to reduce runtimes while maintaining analytical capability (particularly in terms of peak resolution and capacity) have been ongoing since the advent of capillary columns, as described by Blumberg [4]. He describes several approaches that have been both investigated for and used in reduced GC run time methods, which have historically been described in various ways including ‘Minimum Time Operation’ by Giddings in 1962 [5]. Over time, successful methods that do not require significant modification to the traditional instrumental setup (such as switching to shorter columns with thinner stationary phase films and using only hydrogen as the carrier gas [6]) have become folded into standard practice in gas chromatography, but technological developments primarily in column heating have enabled the development of a subclass of GC systems that take advantage of these to achieve significantly faster temperature programming rates and thus

separation times [7]. These systems, distinct from traditional GC systems by their non-traditional heating systems and corresponding differences in column handling and instrument interfacing, thus form the basis of what are today known as high-speed GC (HS-GC) systems.

In recognition of its utility both in commercial applications and research, several commercial GC manufacturers have offered high-speed GC (HS-GC) systems, primarily using resistive heating techniques. The first commercial HSGC offerings were the "Flash-GC" systems by Thermedics Detection in the 1990s, which were offered both as dedicated instruments as well as upgrade packages [8-11]. This was soon followed in the 2000s by both Agilent's Low Thermal Mass systems [12] and the Gerstel MACH [13], which both served as expansions attached to traditional air oven GC systems – with the LTM, this feature was successfully used to combine a LTM column with a traditional column for 2D GC with heartcutting[14]. With the exception of the revised LTM Series II (which is still offered by Agilent at time of writing), these legacy systems are no longer available.

Current offerings in HSGC systems generally take the form of dedicated instruments with some differences in how the column interfaces with the resistive heating element. With the notable exception of the HyperChrom system which heats the column by a steel capillary which allows the use of traditional columns which are threaded through after being cut to length, this has most often necessitated use of dedicated column modules which are incompatible with traditional columns [15-16]. These include Agilent's Intuvo 9000 and the Teledyne Falcon (known as the CALIDUS prior to Falcon Analytical's acquisition by Teledyne), though with different implementations - the Teledyne's column modules each include their own resistive heating element, while Agilent uses a flat heating element in the GC that a pre-wound column fixed to a metal frame with click connectors is pressed against to allow for reliable interface

between the column and the heating element (Figure 1) [17-18]. With the change from traditional column to column module, both the Intuvo and the Falcon have designed their modules such that they can be installed and removed without the need for trimming the column or fitting ferrules [17-18]. This then significantly reduces the skill level needed for maintenance and thus operation and further through the preservation of column dimensions across installations ensures greater repeatability in results, all of which are significant benefits in routine analysis.

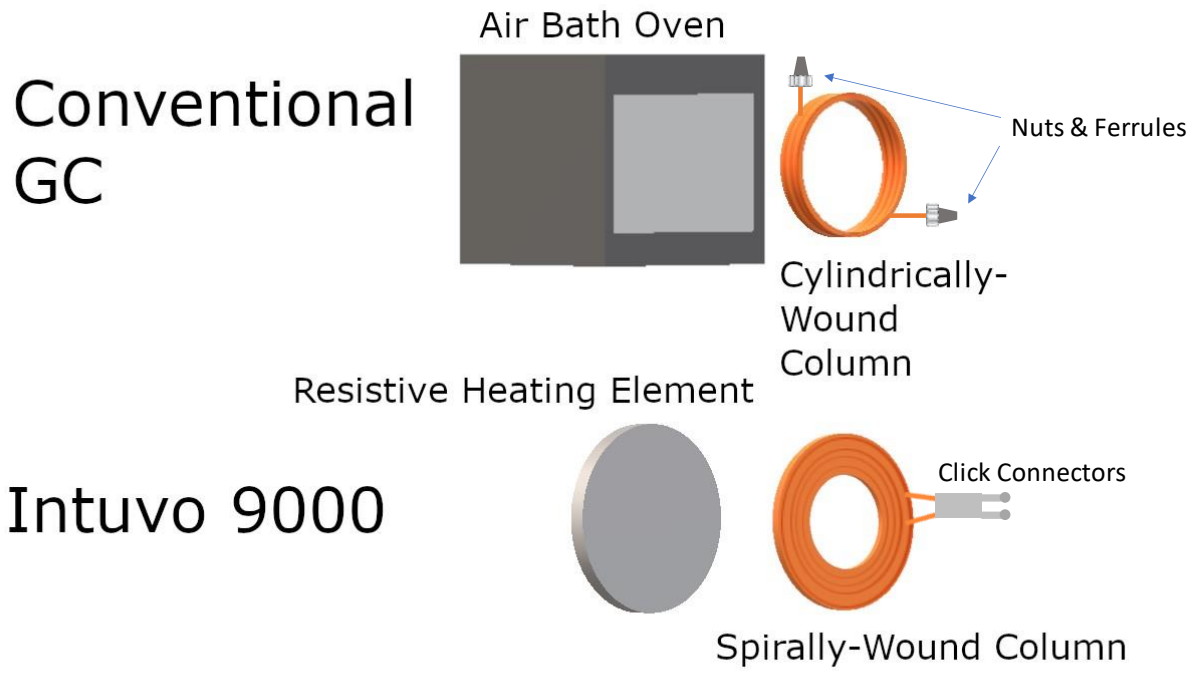

Figure 1. Schematic illustrating differences between column heating in traditional GC systems and in the Intuvo 9000, contrasting the traditional air bath oven with the resistive heating element design as well as the replacement of disposable nuts & ferrules with preinstalled click connectors which are not user-servicable.

1.2 Retention Time Shifting and Alignment in HSGC Systems

Although these developments of HS-GC such as the Intuvo 9000 have been a tremendous enabling technology for gas phase analysis, retention time misalignment remains a major issue as it does in general with all chromatographic instrumentation. In particular, it is important to emphasize the distinction between *run-to-run* misalignment for subsequent separations on a given column, versus *column-to-column* misalignment for chromatograms collected on nominally identical columns. The former arises from minor, low frequency changes in flow and temperature between runs of the same (or similar) sample type on the same instrument and same column. This issue has received extensive attention in the literature [19–21]. The latter issue of column-to-column retention time shifting arises when the analyst replaces a column in an instrument, or when comparing data from interlaboratory studies in which nominally identical instruments and columns are used. If the instrument in one laboratory using the nominally identical column produces instrumental responses (in particular, retention times) different than those obtained on the first instrument, any calibration models (either classical or multivariate) that were applied could produce erroneous results. This issue of column-to-column shifting can be seen to be a specific aspect of the problem of "instrument standardization", or the more general calibration of different instruments to ensure consistent results between them - while this study focuses on the comparison of nominally identical columns used in the same instrument, comparing nominally identical instruments used in different locations (true instrument standardization) is analogous [22]. Significant attention has been dedicated to addressing instrument standardization, particularly for the transfer of calibration models in spectroscopic instruments.

Herein, we investigate to what extent both run-to-run "intra-column" shifting and column-to-column shifting can be addressed by implementing corrective alignment methods. Unlike run-to-run shifting on the same column, the extent and nature of the retention time misalignment between columns is likely to be at least as bad and is often worse, as the original (typically environmental) sources of run-to-run shifting persist while any subtle imperfections in column manufacture may contribute additional shifting. Two popular methods for retention time alignment are retention time-locking (RTL) [23–25] and chemometric alignment [19,20,26-28]. We hypothesize that a combination of RTL and chemometric alignment may be more effective than either method alone to mitigate column-to-column retention shifting.

To perform RTL, the inlet pressure of one GC setup is modified in order that the retention times can be matched or "aligned" to those on a similarly configured GC setup. The goal of the RTL method development is to determine what change in inlet pressure is required for alignment. Typically, RTL is performed in a commercial software 'wizard' (a specific type of graphical user interface). First, the analyst chooses a peak whose retention time is measured and which the software aligns to, defined herein as the *target peak*. Chromatograms are collected at varying inlet pressures and the retention time of the target peak in each chromatogram is recorded. This data is used to generate a pressure-retention calibration curve. If alignment is needed, the Wizard uses the information from the calibration to change the inlet pressure the required amount to align the target peak across separations. The new inlet pressure is then incorporated into a revised, or "locked," method [23,29]. The principles of RTL were initially developed to solve the instrument standardization problem such that different instruments with nominally identical columns could be used with the same data processing steps, but RTL is equally applicable to correcting retention time variations on a single instrument resulting from

predictable changes in retention time such as column trimming. It is crucial to note that the retention time correction afforded by RTL is based on a single point of reference (the target peak), so the rest of the separation is not necessarily aligned across chromatograms.

In contrast, chemometric software-based alignment corrects the retention time shifting problem *after* data collection. One of the most popular methods is correlation optimized warping (COW) [19,30–32]. COW subdivides a chromatogram into a series of windows which are iteratively compressed and stretched (i.e., "warped") via linear interpolation to maximize the correlation of the peak positions between the windows of a chromatogram selected for alignment and a target chromatogram. Note that a target *peak*, as used in RTL, is distinct from a target *chromatogram* used in COW. A target chromatogram is the chromatogram that all other chromatograms are aligned to and is usually selected to be the chromatogram most representative of the data set. Unlike RTL, COW strives to align *all* peaks across the chromatograms. In contrast, RTL only aligns the target peak, with the underlying notion that all of the other peaks in the chromatogram have shifted roughly the same amount as the target peak. In contrast, no such assumptions are made by COW.

Herein, we examine the extent to which RTL can be employed as a "first-order correction" to instrumentally align chromatograms which are then more precisely aligned by COW. By using the coarse alignment provided by RTL, COW can be implemented with less risk of peak misalignment and less impact on resulting peak areas. To accomplish this, the clustering of three sample classes (a test mixture at three spike levels) is examined using principal component analysis (PCA) with four levels of alignment: (1) unaligned data, (2) COW only aligned data, (3) RTL only aligned data, and (4) data aligned with RTL and COW together (RTL

+ COW). We hypothesize that this new approach of synergistically using RTL as an instrumental alignment step prior to COW will provide better results than either alignment method alone.

Briefly, PCA projects the chromatographic data, organized in a matrix of samples versus response, into a new set of orthogonal axes which best capture the variance in the data set [33]. If the primary variance captured by PCA is due to differences between sample classes, all chromatograms from a sample will cluster on a PCA scores plot. However, we hypothesize that data collected on nominally identical columns prior to alignment (RTL and/or COW) will not cluster by sample class but will instead cluster by the specific (i.e., "nominally identical") column used. To compare the success of classification for the various levels of alignment in an objective manner, we used the degree of class separation (*DCS*) as a quantitative metric [21,34]. The *DCS* computes the Euclidian distance between the centroids of two clusters in the PCA scores plot normalized by the pooled variance,

$$DCS = \frac{D_{\mathrm{A,B}}}{\sqrt{s_{\mathrm{A}}^2 + s_{\mathrm{B}}^2}} \quad (1)$$

where $D_{\mathrm{A,B}}$ is the Euclidian distance between the two cluster centroids and s_{A}^2 and s_{B}^2 are the variances of each centroid in the PC scores space. Therefore, *DCS* acts as a proxy for classification success. Data can, however, be successfully classified in undesirable ways. We distinguish between *sample classes*, which correspond to differences in the sample (in this case by analyte spike level), and *column classes*, corresponding to differences in the columns used to obtain the chromatograms. Separate *DCS* values are then calculated for classification both by sample class (DCS_{sam}) as desired, and by column class (DCS_{col}), where a high DCS_{col} indicates significant chromatographic variations from column differences. Since the *DCS* calculation produces multiple pairwise values, the averages of DCS_{sam} and DCS_{col} are used for ease of comparison, and future references to DCS_{sam} and DCS_{col} are to these averages. By comparing the

magnitude of DCS_{sam} versus DCS_{col}, we explore to what extent classification is sequentially improved from the unaligned data to the COW only aligned data, to RTL only aligned data and finally to the RTL + COW aligned data.

Chapter 2. Cross-Column Retention Time Shifting in HSGC and Alignment Thereof

2.1. Experimental

2.1.1. Sample Design

To create three sample classes, a 100-component test mixture spanning a broad range of boiling points (54-343°C) (Table 1) was prepared (i.e., referred to as the neat sample, without any spiked analytes) or spiked two-fold with one of two different sets of ten analytes, all of which are present in the original base mixture (Table 1). The base test mixture is referred to as Sample Class 1, with the spiked test mixtures being Sample Classes 2 and 3, respectively. All test mixtures were diluted in HPLC-grade methanol by a factor of 100 to create a nominal concentration of ~1000 ppm for each compound. Five replicates were obtained for each sample class (3 samples) on each column (4 columns) for a total of 60 runs. The spiked analyte peaks were identified by the standard addition method.

Alkanes	Conc (ppm)	**Esters**	Conc (ppm)	**Cyclics**	Conc (ppm)
heptane	592	ethyl formate	600	methylcyclopentane	536
octane	700	methyl decanoate	940	cyclooctane	764
decane	624	methyl caprylate	824	butylcyclohexane	664
undecane	660	methyl salicylate	1056	bicyclohexyl	19.2
dodecane	688	ethyl salicylate	1040	**Halogenated Alkanes**	
tetradecane	736	methyl laurate	776	1,5-dichloropentane	1000

hexadecane	792	methyl caproate	800	1-chlorohexane	928
pristane	724	**Ketones**		1-bromohexane	865
octadecane	808	2-hexanone	724	1-bromoheptane	1052
nonadecane	496	2-heptanone	780	1-bromooctane	1052
eicosane	852	3-heptanone	784	carbon tetrachloride	800
Aromatics		3-octanone	764	1-iodo-2-methylpropane	1440
benzene	748	2-decanone	876	**Organosulfur compounds**	
toluene	832	2-undecanone	728	1-decanethiol	820
3-ethyltoluene	796	2-dodecanone	748	2-thiophenemethylamine	1092
4-ethyltoluene	816	**Alkenes**		2-propylthiophene	936
mesitylene	796	1-hexene	540	2-(methylthio)-pyridine	968
butylbenzene	876	1-undecene	640	1,4-thioxane	1112
tert-butyl benzene	816	**Alkynes**		diphenylsulfide	1088
propylbenzene	776	1-hexyne	516	2-metylthiophene	820
1-ethylnapthelene	904	1-heptyne	716	benzo[b]thiaphene	1908
cyclohexylbenzene	868	1-nonyne	772	1-dodecanethiol	712
diphenylmethane	936	5-decyne	780	2-octanethiol	864
o-xylene	852	**Alcohols**		3-methylthiophene	884
dibutyl phthalate	1264	1-butanol	740	2-methylthianapthene	1480
diethyl phthalate	904	1-pentanol	748	**Organophosphorus compounds**	
2,2'-dimethylbiphenyl	964	2-pentanol	784	dimethyl methyl phosphonate	1204
sec-butylbenzene	772	hexyl alcohol	860	diethyl methyl phosphonate	872
4-phenyltoluene	996	2-heptanol	1072	**Other**	
9-methylanthracene	624	1-octanol	940	(-)-alpha-pinene	872
anthracene	788	1-nonanol	752	o-toluidine	1304
naphthalene	1128	1-tetradecanol	800	(-)-linalool	816
1,2,3,4-tetrahydronapthalene	1752	1-eicosanol	1160	acetophenone	904
phenanthrene	948	2-ethyl-1-hexanol	596	dibenzofuran	1748
hexamethylbenzene	2340	decyl alcohol	860	butyrophenone	1120
		tert-amyl alcohol	676	1,5-dimethyl tetralin	1268

				menthol	540
				3,7-dimethyl-2,6-octadienal	660

Table 1. List of compounds in sample mixtures with concentrations in ppm (all samples diluted in methanol)

2.1.2. Instrumental

All data were collected on an Agilent Intuvo 9000 GC with a flame ionization detector (FID) at 1 kHz. Four nominally identical HP-5MS (5%-diphenyl-methylpolysiloxane) nonpolar columns, 30 m × 250 μm d_c × 0.25 μm d_f, were studied. Instrument control and method adjustments were performed through OpenLab CDS (Agilent Technologies, Palo Alto, CA, USA). Prior to any inlet pressure adjustments from RTL, all runs were performed under constant flow mode at a nominal flow rate of 3 mL/min. An injection volume of 0.5 μL was used at a split ratio of 1:20. The temperature program began at 50 °C for 1 min, followed by a ramp at 150 °C/min to a final temperature of 300 °C, where it was held for 2.33 min for a total run time of 5 min. When RTL was applied the only direct change to the method was a minor alteration in starting inlet pressure. The original method for all columns prior to RTL had an inlet pressure of 17.21 psi (118.64 kPa). Following the application of RTL, with Column 1 staying at the original 17.21 psi (118.64 kPa), the starting inlet pressures for Columns 2, 3, and 4 were 17.47, 16.75, and 16.45 psi (120.47, 115.51, 113.40 kPa) respectively.

2.1.3. Data Processing

Data were processed in MATLAB R2021a (The Mathworks, Inc., Natick, MA, USA). Although the total run time was 5 min, all mixture components (excluding the solvent) eluted in a time window between 105 s to 195 s; therefore, only this time window was used for all subsequent data analyses. Chromatograms were baseline corrected using a rolling ball minimum

algorithm [35], followed by normalizing each chromatogram to the un-spiked component eluting at a t_R of ~150 s having the tallest height in all chromatograms, serving as the internal standard.

Initially, we investigated the scope of the column-to-column and sample-to-sample replicate alignment issues. To do so, the amount of shifting initially present in the unaligned data was assessed using "shift plots." Shift plots were obtained by determining the difference in the retention time for each identifiable peak between each sample/replicate chromatogram and the same analyte peak from the selected reference (or target) chromatogram. Specifically, Replicate 2 from Sample Class 1 using Column 2 was selected as the reference chromatogram for the unaligned data set. Shift plots are similar to the previously reported use of shift functions [38,39]. Unlike shift functions, shift plots do not use interpolation to estimate the retention time shift (difference) between peaks. However, by presenting data for a large number of peaks (~63 peaks), the behavior of retention time shifting as a function of run time can be readily discerned without the need for interpolation. Following the examination of the shift plots for the unaligned data, shift plots were subsequently prepared for the RTL only aligned data and finally to the RTL + COW aligned data. Shift plots for the COW only aligned data were not generated for brevity. For the shift plot generation of aligned data, a chromatogram from Column 2 was selected as the reference/target chromatogram which turned out to be Replicate 1 from Sample Class 1. To quantitatively compare the data in the shift plots, the average shift for each peak was calculated, and then averaged for all peaks in the shift plot to obtain a consensus average and standard deviation of shifting observed on each column in each data set.

To better visualize the impact of the various alignment methods, PCA was applied. All data were mean centered prior to the application of PCA, which was performed through singular value decomposition in economy mode in MATLAB through the included 'svd' function. The

script to perform PCA was written in-house. Analysis of the data obtained through PCA was performed using the *a priori* information of which sample and column each chromatogram was obtained from and used to label the PC scores plots and calculate *DCS* values in a supervised fashion. Therefore, the primary use of PCA is as a visualization tool and as a space for *DCS* calculation. *DCS* was calculated pairwise across each sample class and column class before averaging the 3 individual DCS_{sam} and 6 individual DCS_{col} values.

The RTL alignment method was performed using Agilent's RTL Wizard [29]. To use the Wizard in OpenLab CDS, a target peak from a reference chromatogram is selected. Three runs are then performed on the sample at varying inlet pressures. The target peak in each corresponding chromatogram is then identified, and the Wizard calculates the inlet pressure adjustment needed to ensure that the target peak is 'locked' to the desired retention time. The base test mixture (un-spiked) component eluting at a t_R of ~150 s (same peak as the internal standard) was selected as the target peak for RTL. The target peak was selected because of its position near the middle of the reference chromatogram.

The code for performing COW was obtained from a publicly available resource maintained by Rasmus Bro and coworkers (Chemometrics Group, Department of Food Science, University of Copenhagen) [36] and is described in the corresponding references [27,30]. When performing COW, the most representative chromatogram for a given data set was selected as the target chromatogram by summing the absolute value of the PCA scores for all chromatograms within the data set as weighted by the amount of variance explained by that PC. This effectively selected the chromatogram closest to the origin on a PCA scores plot. Note that this method of selecting a target chromatogram is very similar to the use of the Mahalanobis Distance but neglects covariance [37]. When COW was applied, automated optimization of segment and slack

was performed, where 'segment' defines the number of sections the chromatogram is divided into and 'slack' limits how much each segment can be warped. For the purpose of optimization only, a 15-point boxcar average of the data was used to facilitate reasonable optimization times on each dataset. This resulted in the average number of data points per peak width-at-base decreasing from 830 points (830 ms) to ~55 datapoints in the averaged dataset, with a visual comparison before and after averaging provided in Fig. 2. The resulting segment and slack lengths were then multiplied by 15 and used for chromatogram alignment.

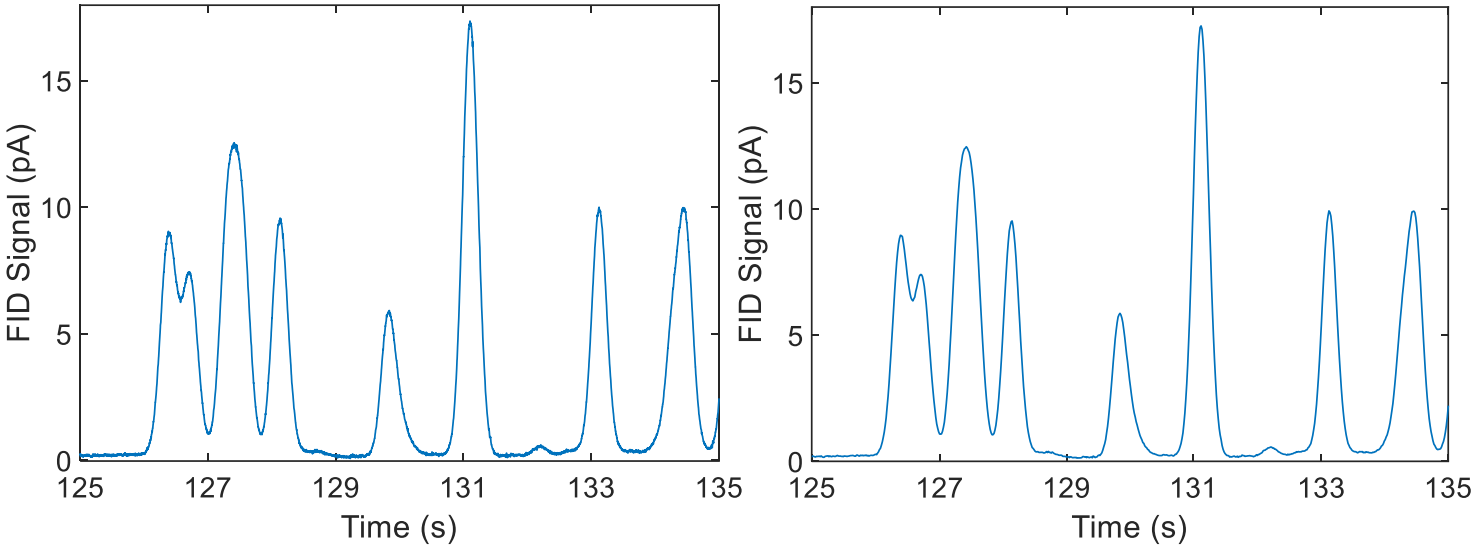

Figure 2. Section of the chromatogram before (A) and after (B) boxcar averaging, showing that while some high-frequency noise is lessened the peak shapes are thoroughly preserved.

2.2. Results and Discussion

A chromatogram of the base test mixture (Sample 1) for a single replicate on Column 1 is shown in Fig. 3A. In Fig. 3B, all 15 runs are provided in an overlay plot, showing how the spike analyte peaks exhibit signal increases for the spiked test mixtures in Sample Classes 2 and 3. A zoom-in view of a section of the three chromatograms containing three spiked analytes is shown in Fig. 3C, with two analytes (3-ethyltoluene and mesitylene) spiked in Sample Class 2 while

tert-butylbenzene was spiked in Sample Class 3. From visual inspection of the signal differences, an analyst could readily classify the three sample classes. Nonetheless, some minor run-to-run shifting is observed. To better visualize the degree of shifting obtained within a single column, Fig. 3D shows an overlay of all fifteen runs of *tert*-butylbenzene on Column 1. For this particular analyte, the retention time shift appears relatively minor. However, focusing on the retention behavior of a single peak does not provide a global understanding of the intra-column shifting since different peaks will shift by different amounts depending on their retention time [38,39]. Later, we will show how this effect can be visualized via shift plots.

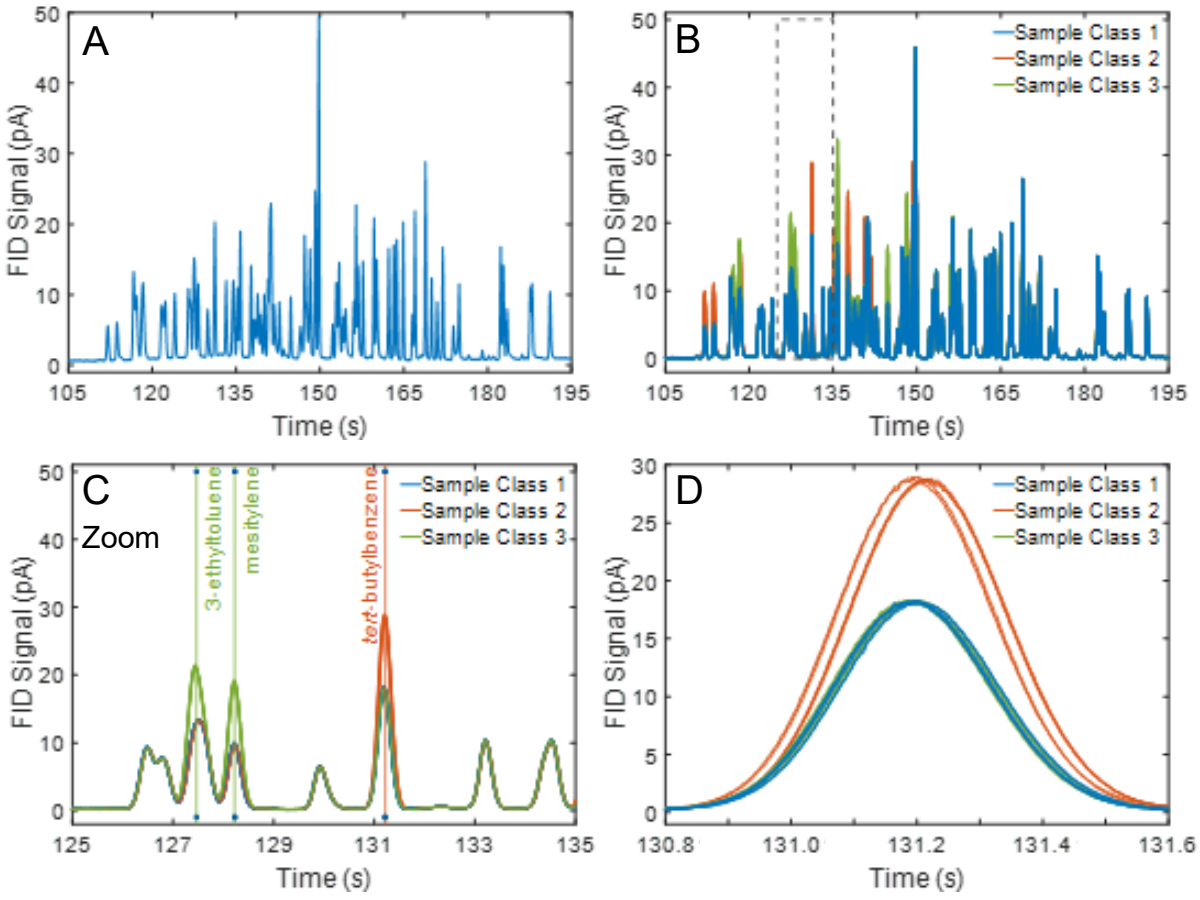

Figure 3. Representative chromatograms from Column 1. (A) A typical chromatogram from the 100 mix/unspiked sample (Class 1). (B) Overlay of all 15 chromatograms collected on Column 1. (C) Zoom-in of (B) at 125-135 s, showing identifiable spike analyte peaks from both sample

classes (labeled). (D) Zoom-in of tert-butylbenzene peak from (C) showing run-to-run peak shifts.

2.2.1. Cross-Column Shifting

Next, the extent of retention time shifting between nominally identical columns was investigated with the hypothesisis that due to unavoidable, subtle differences in column manufacture (e.g., length, inner diameter, and film thickness homogeneity) the magnitude of the retention time shifting between nominally identical columns (column-to-column) might be larger than that of intra-column run-to-run shifting. In the context of HS-GC, a further source of misalignment between separations obtained from different columns could arise from the interface between an individual column and the resistive heating element. Figure 4A shows an overlay of example chromatograms from each column to visually highlight the extent of column-to-column shifting. Note that the amount of column-to-column shifting, illustrated in Fig. 4A, appears greater than the amount of intra-column shifting, given in Fig. 3B.

A zoom-in view between 125-135 s is shown in Fig. 4B. Via this close view, the extent of column-to-column shifting appears to be large enough to complicate traditional quantification using quantification windows. Here, the quantification window is defined as a time interval where integration boundaries would be set to quantify a single analyte peak. In this example, 3-ethyltoluene and mesitylene could represent a problematic pair. For instance, suppose that the quantification window for 3-ethyltoluene was 125-127.5 s and the quantification window for mesitylene was 127.5-128 s. Without retention time correction, the same quantification window could not be used for both columns, i.e., when quantifying 3-ethytoluene, interference for mesitylene will lead to erroneous results, and vice versa. At first glance, it seems plausible that a coarse correction by applying a simple integer shift of the retention times would lead to

alignment. However, this will only lead to successful alignment if all peaks in the separation shift by the same amount from one column to the next.

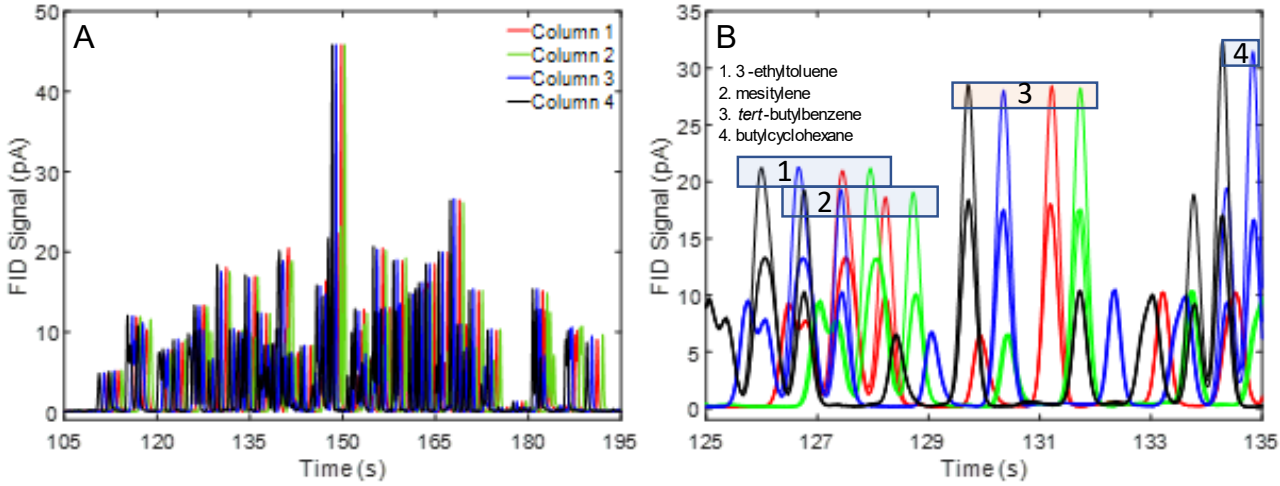

Figure 4. (A) Overlay of representative chromatograms from all four nominally identical columns prior to any alignment, with a single chromatogram shown from each column. (B) Overlay of one representative chromatogram from each sample class per column at 125-135 seconds, further illustrating the extent of column-to-column shifting.

To determine if this was the case, the appearance of shift plots as given in Fig. 5 was examined to compare the overall native shifting from one column to the next. For this purpose, Column 2 was used as the reference column, with the reference chromatogram obtained from Replicate 2, Sample Class 1. The shift plot for Column 1 is given in Fig. 5A and shows a noticeable variation as a function of retention time. Only Column 2 (Fig. 5B) has a (relatively) flat pattern in shifting, which is a consequence of the reference chromatogram originating from Column 2. Note that while Fig. 5B shows the extent of intra-column shifting using Column 2, in contrast Figs. 5A, C, and D highlight the level of column-to-column shifting for Columns 1, 3, and 4, relative to Column 2, respectively. For shift plots reflecting column-to-column shifting, simple integer shifts, equivalent to moving the points across the y-axis, would be ineffective in

aligning more than one peak and a few other peaks with close retention times. Relative to Column 2 as the reference column, Column 1 had an average shift of -577 ms with a standard deviation of 77.9 ms, Column 2 had an average shift of -18.4 ms and a standard deviation of 16.4 ms (again reflective of the reference chromatogram being selected from Column 2), Column 3 had an average shift of -2233 ms and a standard deviation of 99.9 ms, and Column 4 had an average shift of -3880 ms and a standard deviation of 109 ms. With an average peak width-at-base of ~830 ms, this represents significant shifting. A full comparison of the average shifting and the standard deviations of shifting for each column in each data set are in Tables 2 and 3 at the end of the analysis (p. 30) as supplementary information providing more insight into shifting patterns.

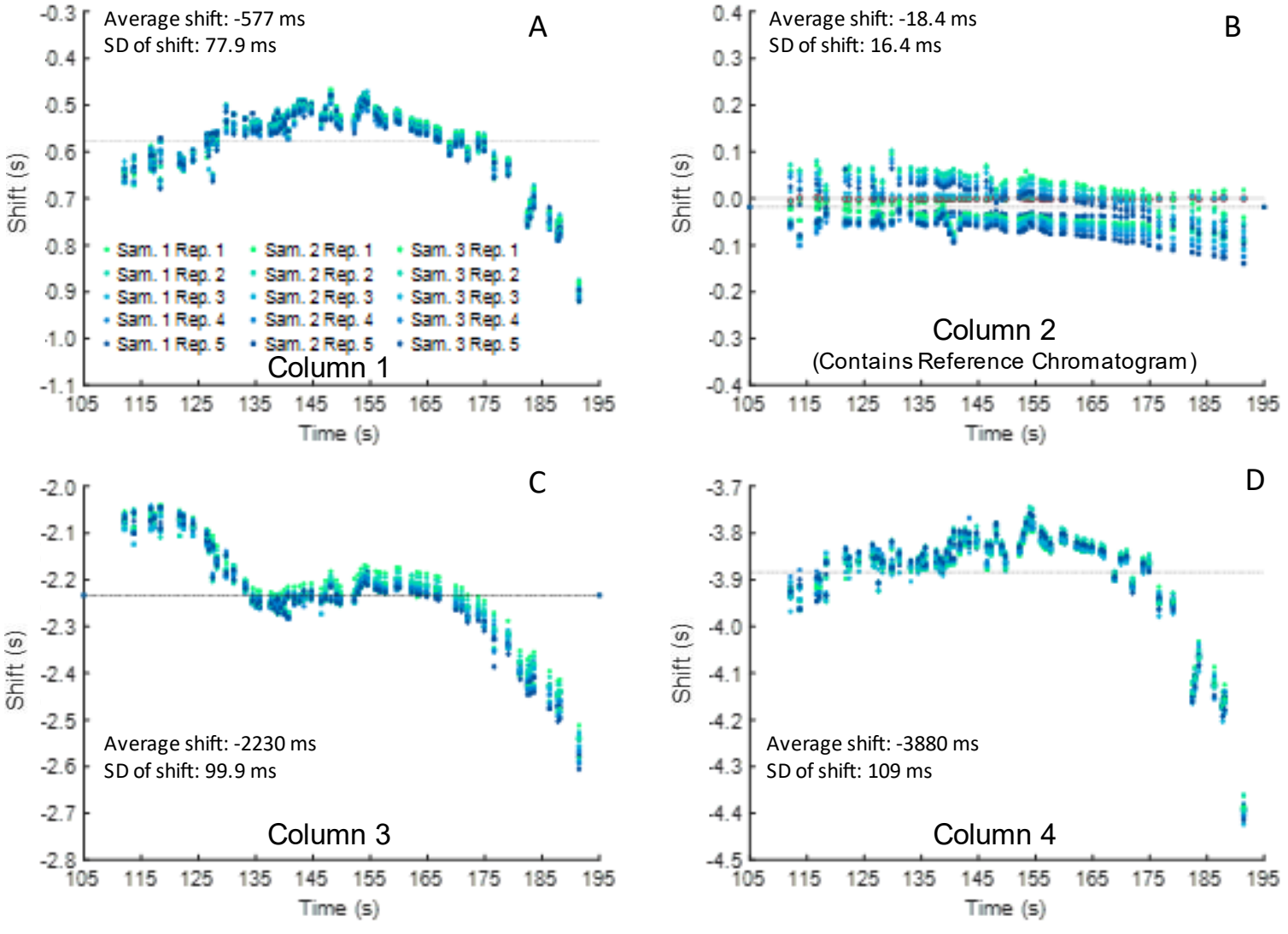

Figure 5. Shift plots for the unaligned data set. All shift plots show a dashed line at the average shift value, with Column 2 also having a solid line at 0 shift. The Reference Chromatogram in Column 2 is highlighted in dark red. (A) Column 1 shift plot. (B) Column 2 shift plot (reference column), illustrating the extent of run-to-run shifting. (C) Column 3 shift plot. (D) Column 4 shift plot.

Figure 6A shows the PCA scores plot obtained for the unaligned data. Clearly, PCA alone does not cluster by sample class but rather by the column class. The degree of shifting across columns leads to a significant amount of variation among chromatograms. The clustering by sample class relative to the clustering by column class was quantified by the *DCS* metric, via Eq. (1), and then averaged across each pairwise value. Specifically, we compare the magnitude of DCS_{col} and DCS_{sam}, where DCS_{col} is the *DCS* metric of the column classes whereas DCS_{sam} is calculated for the three sample classes. If chromatograms cluster according to the sample classes rather than the column classes, the DCS_{sam} should be much greater than DCS_{col} (indeed, DCS_{col} should ideally be 0). Without alignment, we observed a DCS_{col} of 76.1, markedly higher than the DCS_{sam} of 3.0 (Table 4). This indicates that the samples are being classified by column class, rather than sample class. The source of clustering can be visualized in the loadings for PC1 and PC2 (Fig. 6B) which appear to show contributions from all component peaks (~100), indicating that the retention time differences between each column cause classification according to the

column used, leading to the four column classes. The derivative-like appearance of profile loadings is consistent with that observed due to retention time shifting [40].

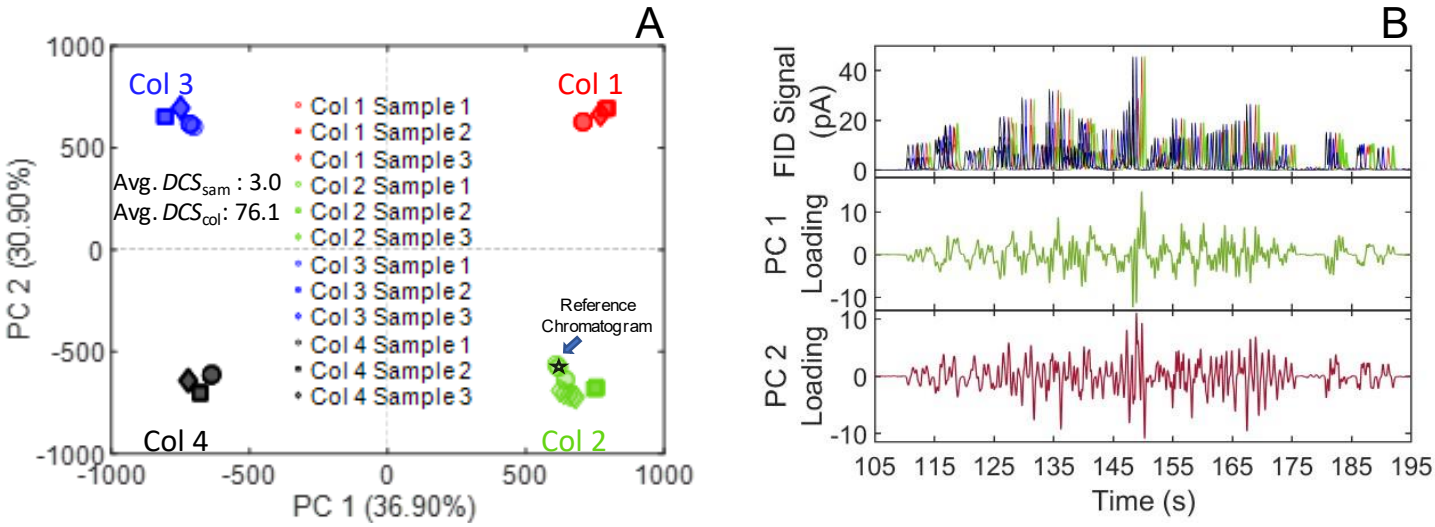

Figure 6. PCA results for the unaligned data set. (A) The scores plot of PC1 and PC2, showing clustering primarily by column (color coded) and not by sample class (coded by shape). (B) An overlay of all 60 chromatograms in the unaligned data set from all four columns, top, followed by the PC1 and PC2 loadings, middle and bottom. PC1 and PC2 both show loading contributions throughout the chromatogram except for locations with negligible to no signal, supporting the conclusion that PC1 and PC2 both correspond with misalignment and not sample differences.

2.2.2. COW Alignment

We first investigated if COW alone was sufficient to correct column-to-column retention time shifting. Figure 7A shows an overlay of all 60 chromatograms after applying COW. Despite excellent alignment for the middle of the chromatograms, there is significant misalignment at the beginning and end of the chromatograms, as shown in the zoom-in view of the end of the chromatogram in Fig. 7B.The severity of misalignment results in poor classification by sample class via PCA (Fig. 7C), with a DCS_{sam} of 1.9 and a DCS_{col} of 1.7. The locations of the misaligned peaks are clearly displayed in the PC1 loadings, with PC2 recovering spike analyte locations outside of those eluting in the misaligned section near ~115 s, which have negligible

loading contributions (Fig. 7D). Note that the failure of COW to succesfully align this data set likely results from the shifting observed between different columns dominating over the run-to-run shifting for a given column alone.

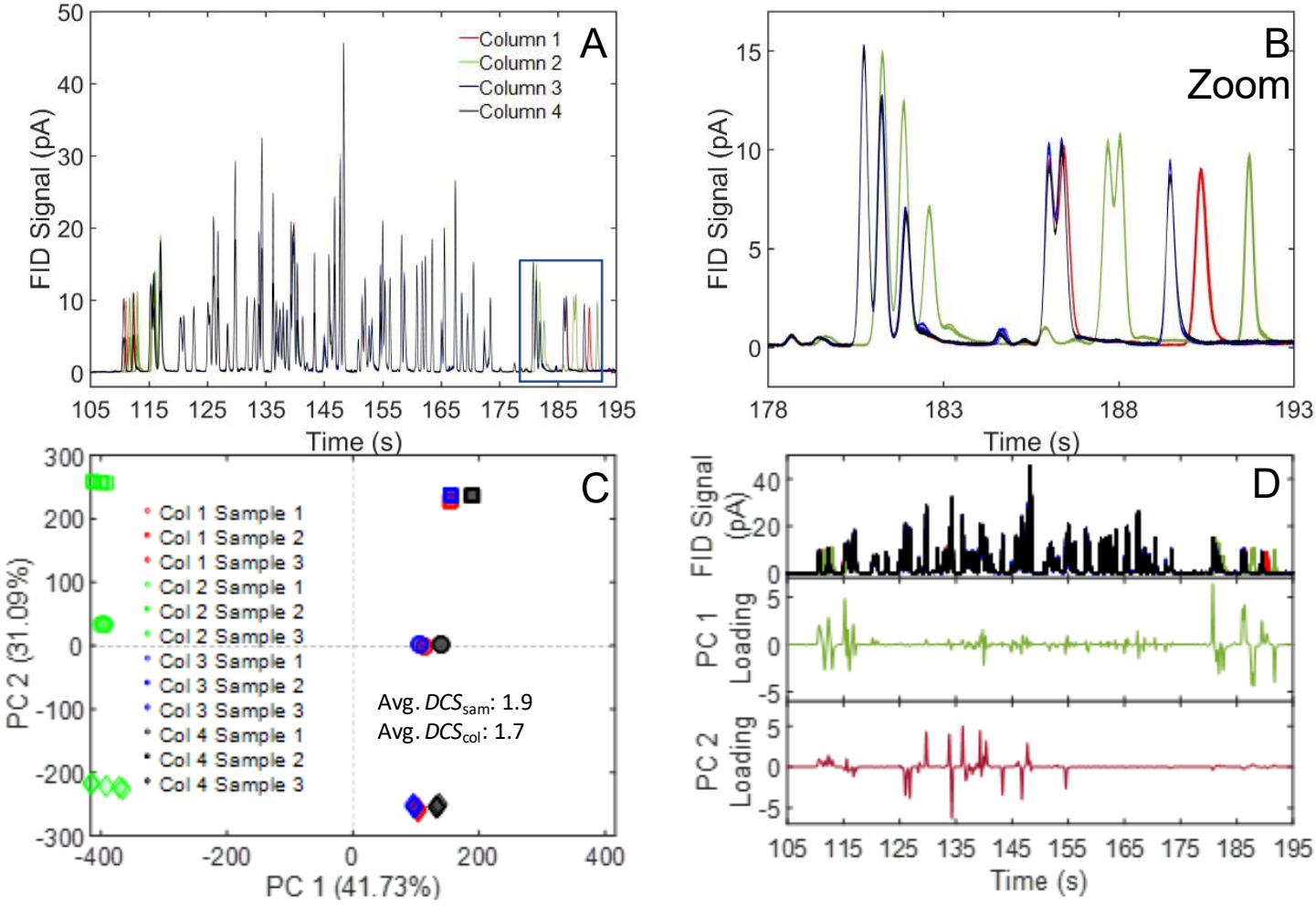

Figure 7. (A) Overlay of all 60 chromatograms after direct COW alignment of the unaligned dataset, showing presence of misaligned peaks near the beginning and end of the separations. (B) Zoom-in showing misaligned peaks near the end of the chromatographic window. (C) PCA scores plot, showing classification by column differences on PC1 and classification by sample class on PC2. (D) Overlay of all 60 COW-only aligned chromatograms (top). PC1 loadings (middle) appear to capture the misalignment at the beginning and end of the chromatograms, while PC2 loadings (bottom) correspond with spike analyte peaks.

2.2.3. RTL Alignment

Next, the extent to which RTL alone corrected retention time shifting was investigated. An overlay of a typical chromatogram from each of the four columns with the RTL target peak indicated is shown in Fig. 8A. From visual inspection, considerable improvement in alignment is apparent in Fig. 8A compared to Fig. 4A, where no alignment was performed. Three zoom-in views of different sections of the chromatogram overlay are shown in Figs. 8B-D to visualize the success of RTL correction at different regions of the chromatogram. Fig. 8B shows a zoom-in view of Section 1 in the chromatogram in Fig. 8A. The three spiked analytes 3-ethyltoluene, mesitylene, and *tert*-butylbenzene are indicated. Although RTL correction has been applied, appreciable shifting is readily discernable. Conversely, in Section 2 in the chromatogram, shown in Fig. 8C, the amount of shifting appears minor and largely corrected by RTL. This is a result of Section 2 containing the RTL target peak. Finally, Fig. 8D shows Section 3 of the chromatogram

later in the run, equally removed from the RTL target peak as Section 1 - like Fig. 7C, retention time shifting appears largely corrected near the end of the chromatogram in Section 3.

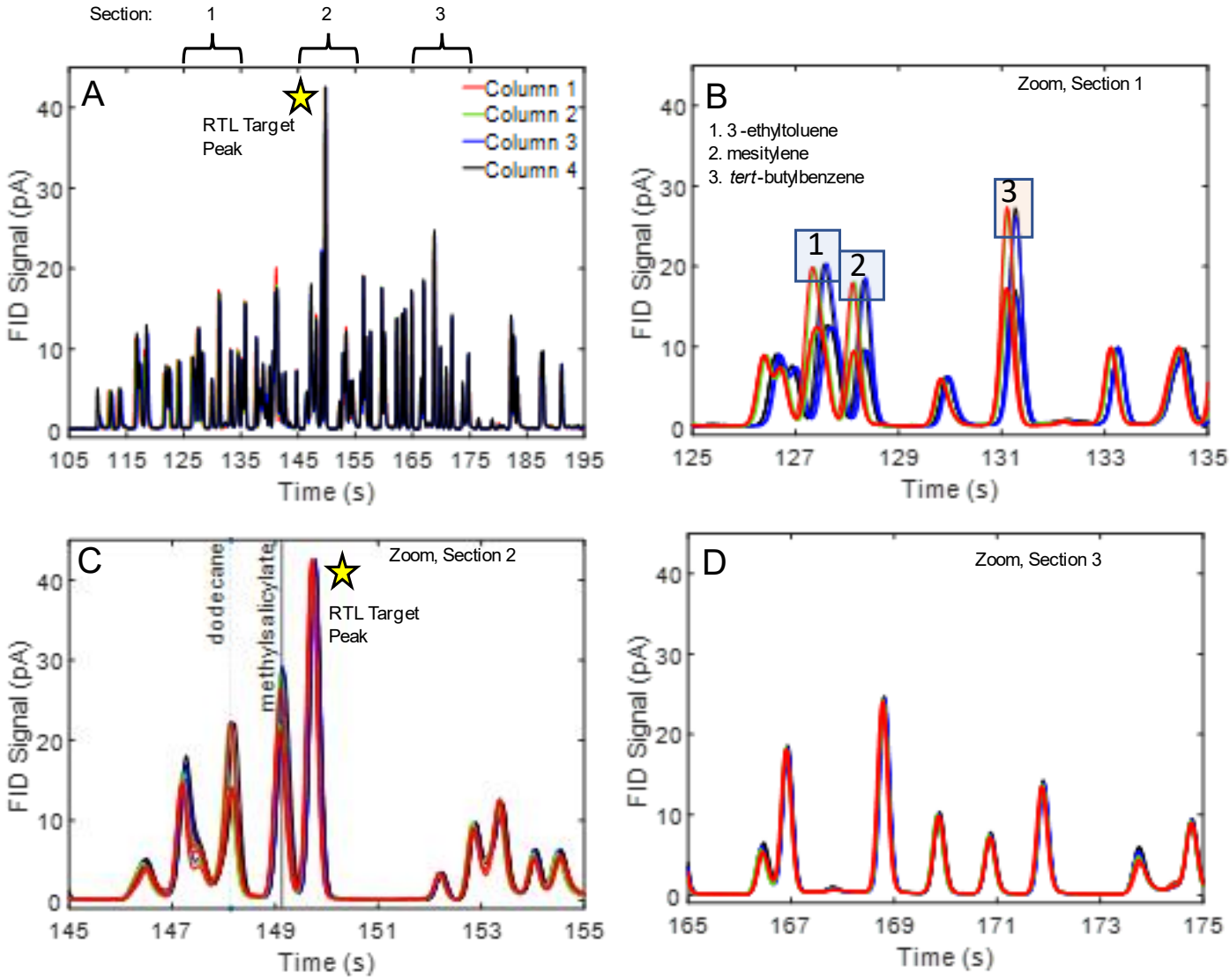

Figure 8. Overlay of all chromatograms from the RTL-aligned data set. (A) The full target window, showing significant improvement over the unaligned data set. (B) Zoom-in of (A) at the 125-135 s window (Section 1), showing appreciable shifting. (C) Zoom-in of (A) for peaks eluting in the middle of the target window, including the RTL target peak (Section 2) at 150 s, showing minor shifting. (D) Zoom-in of (A) from 165-175 s (Section 3). Note that here peaks appear to have negligible remaining misalignment, which was generally true of other peaks that eluted later than the RTL target peak.

A more comprehesive visualization of the shifting for the RTL-aligned data is provided in the shift plots in Fig. 9. Similar to the shift plots in Fig. 5 but now showing the impact of RTL,

these shift plots present the difference in the retention time for each identifiable peak between each sample/replicate RTL-aligned chromatogram and the same analyte peak from the selected reference/target RTL-aligned chromatogram (which has shift differences of zero by definition). Therefore, instead of showing the amount of "natural" shifting in the data, the intent of Fig. 9 is to show how much shifting remains after RTL was applied. Note that while the RTL target peak has a 0 ms average shift in the target peak for Column 2 and only -4, 42, and 66 ms average shift in Columns 1, 3, and 4 respectively, much of the rest of the chromatogram is not appreciably aligned. Furthermore, the column-to-column shifting observed is similar to that of the data set without applying the RTL method. Because the reference chromatogram was obtained from Column 2, its shift plot in Fig. 9B shows only intra-column shifting. Figs. C and D reflect column-to-column shifting for Columns 3 and 4, respectively. Note that in all four shift plots, the magnitude of shift in Sections 2 and 3 is low. The minor shift near the RTL target peak (Section 2) is anticipated as this is the area most accurately aligned using RTL, while the minor amount of shifting in Section 3 may be coincidental. In contrast, note that the magnitude of the shifts in Section 1 are high, which illustrates why the observed shifting in Fig. 8B appears high. Note that the minimal remaining misalignment for the RTL target peak eliminates the offsets seen in the unaligned shift plots (Fig. 3), as seen in the shift plots for the RTL aligned data set in Fig. 9. This is reflected in reduced average shift values, which for the RTL aligned data set are -29.3 ms with a standard deviation of 46.5 ms for Column 1, -12.2 ms with a standard deviation of 18.9 for Column 2, 72.0 ms with a standard deviation of 101 ms for Column 3, and 62.4 ms with a standard deviation of 68.8 ms for Column 4. The new shift plots emphasize that RTL correction

primarily addresses linear retention time shifting, whereas the shifting in real chromatograms is inherently non-linear, and is thus not fully addressed by RTL alone [39].

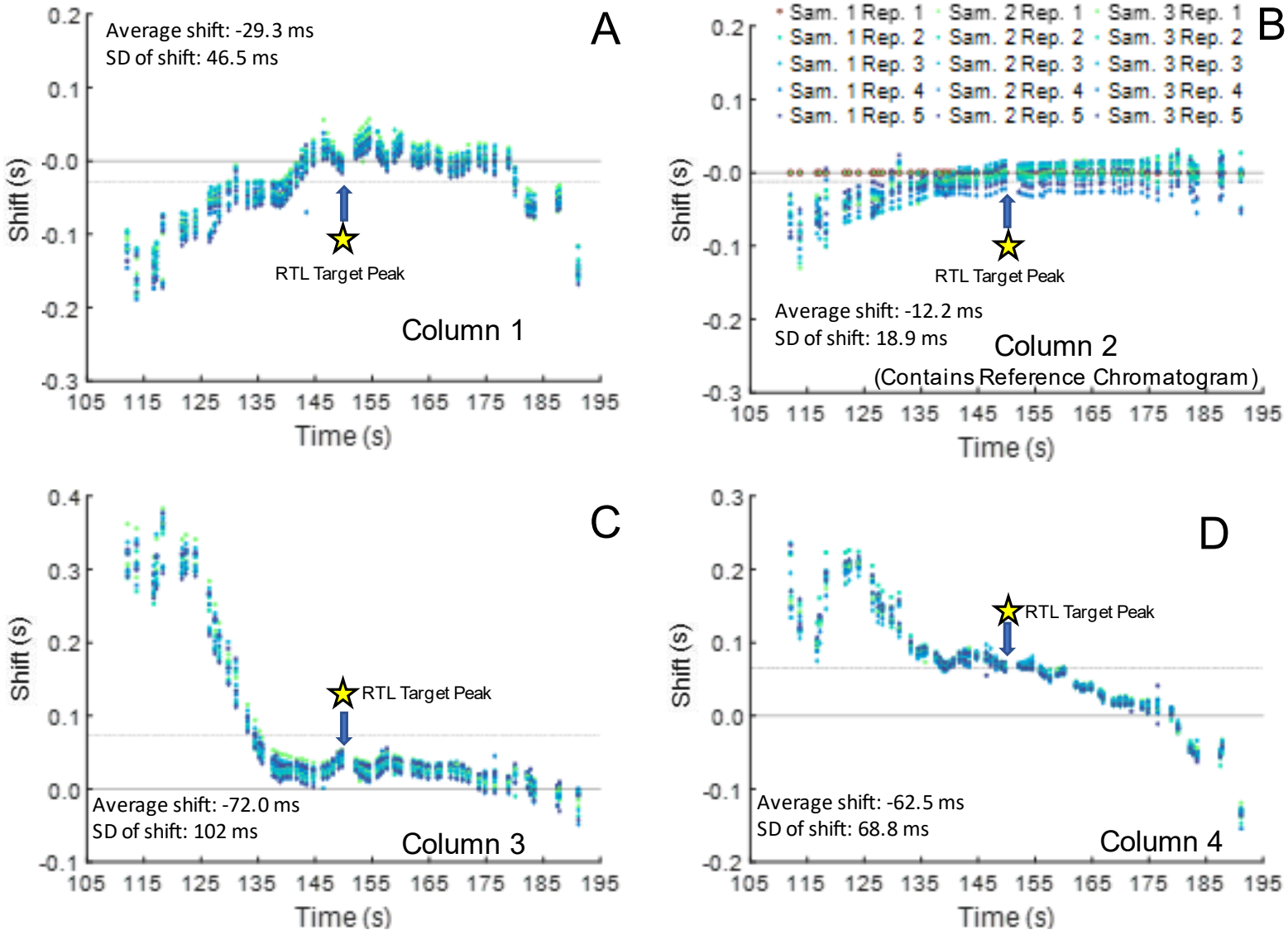

Figure 9. Shift plots for the RTL-aligned data set. As in Fig. 3, dashed lines indicate average shift with solid lines at 0 shift for reference, and the reference chromatogram is highlighted in red. (A) Column 1 shift function, illustrating run-to-run shifting. (B) Column 2 shift plot, illustrating intra-column shifting. (C) Column 3 shift plot. (D) Column 4 shift plot.

The extent to which RTL alignment improved sample clustering by PCA was examined next. The scores plot is shown in Fig. 10A. Note that DCS_{sam} has approximately doubled to 4.7

from 3.0 for the unaligned data set, whereas DCS_{col} has been greatly reduced from 76.1 in the unaligned data set to 4.2 with RTL. PC2 now appears to primarily reflect the differences in sample class rather than column class, as its scores cluster by the sample class. However, PC1 still appears to reflect variations in alignment caused by column differences. Note that Columns 1 and 2 are "split" from Columns 3 and 4 across PC1, reflecting that only two of the four known column classes are captured. This is further exemplified in the loadings plot, given in Fig. 10B. The PC1 loadings have contributions from a majority of the peaks, concurrent with the derivative-like shape effect that is consistent with misalignment [40]. Note that the magnitude of the peaks in the PC1 loading decrease after the RTL target peak, which is consistent with conclusions from the interpretation of Fig. 9, where the magnitude of shifting is generally lower after the RTL target peak. In contrast, the PC2 loading appears to capture the spiked analyte changes in the three sample classes, which are indicated with vertical lines, with solid lines used for spike analytes in Class 2 and dashed lines used for spike analytes in Class 3. All of the Class 2 spike analytes appear as positive peaks in the PC2 loadings, whereas the the Class 3 spike analytes appear as negative peaks in the PC2 loadings. Next, Fig. 10C plots the PC2 scores for all replicates to illustrate that PC2 corresponds strongly with sample classes and correctly categorizes the chromatograms by sample class. To quantify this, the same *DCS* formula was used on the one-dimensional distances between PC2 scores to obtain one-dimensional *DCS* values. With a one-dimensional DCS_{sam} of 288.0, more than three times greater than the one-dimensional DCS_{col} of 74.8, the succesful classification by sample class on PC2 alone is confirmed. While the scores plot and corresponding *DCS* values show significant improvement

over the unaligned data, we predict that additional alignment using the COW algorithm should provide optimal classification by sample class.

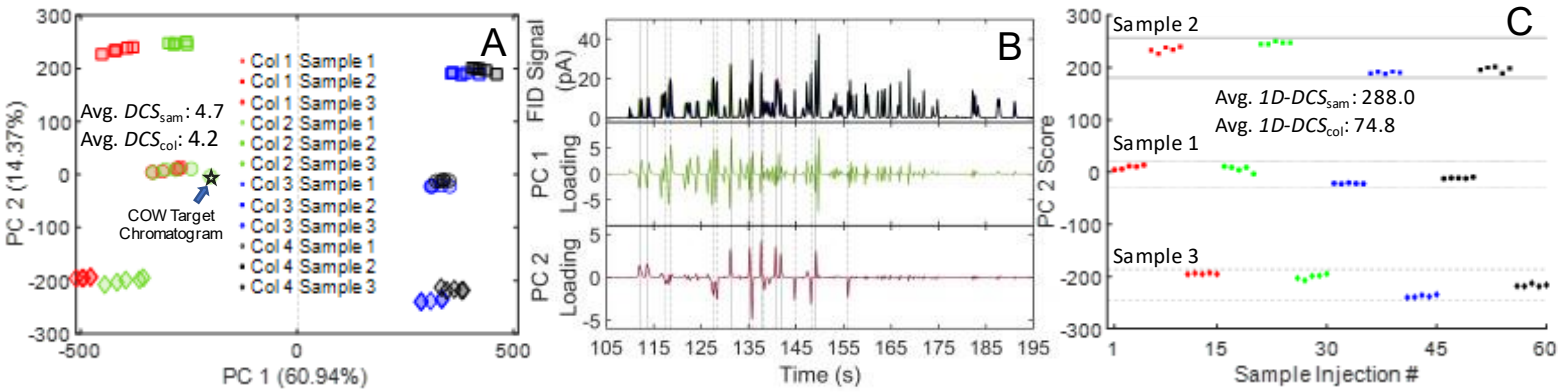

Figure 10. The PCA results of the RTL-aligned data. (A) PCA scores plot showing clustering by both sample class on PC2 and by the nominal column used on PC1, resulting in a low DCS both for sample class and column. The target chromatogram for subsequent COW alignment is indicated. (B) Overlay of all 60 RTL-aligned chromatograms (top), PC1 loadings (middle), and PC2 loadings (bottom). The PC1 loadings appear to capture misalignment, while the PC2 loadings appear to capture the sample class differences because the peaks correspond to the locations of the spike analytes. (C) Plot of PC2 scores across injections, showing clear groupings by sample.

2.3.4. RTL + COW Alignment Method

Finally, we investigated how alignment could be improved by applying COW to the data set experimentally obtained using RTL (RTL + COW alignment). For this purpose the target chromatogram for COW alignment was selected from the RTL aligned data set, as indicated in Fig. 10A. Figure 11A shows an overlay of all 60 chromatograms after RTL + COW alignment, with virtually no shifting visually apparent. Figure 11B shows a zoom-in of Section 2, previously shown for RTL alignment only in Fig. 8B. While significant shifting was apparent in this section with RTL alignment alone, the additional application of COW appears to have eliminated shifting in Section 2.

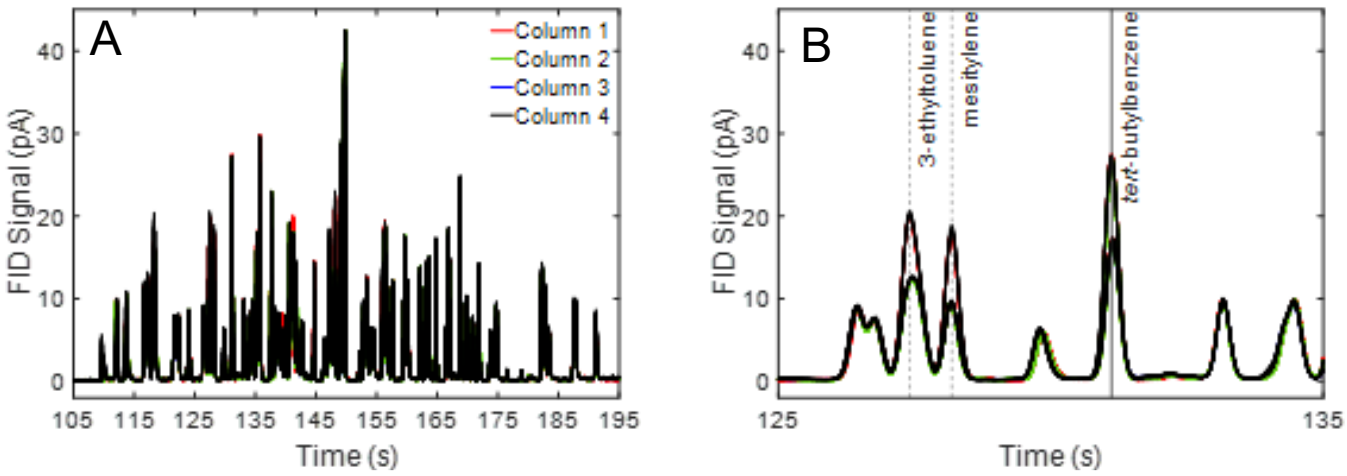

Figure 11. (A) Chromatographic overlay of all 60 RTL + COW aligned chromatograms. (B) Zoom-in of A between 125-135 s. Note that this is the same section shown in Fig. 5B, now with markedly less misalignment due to the subsequent application of COW.

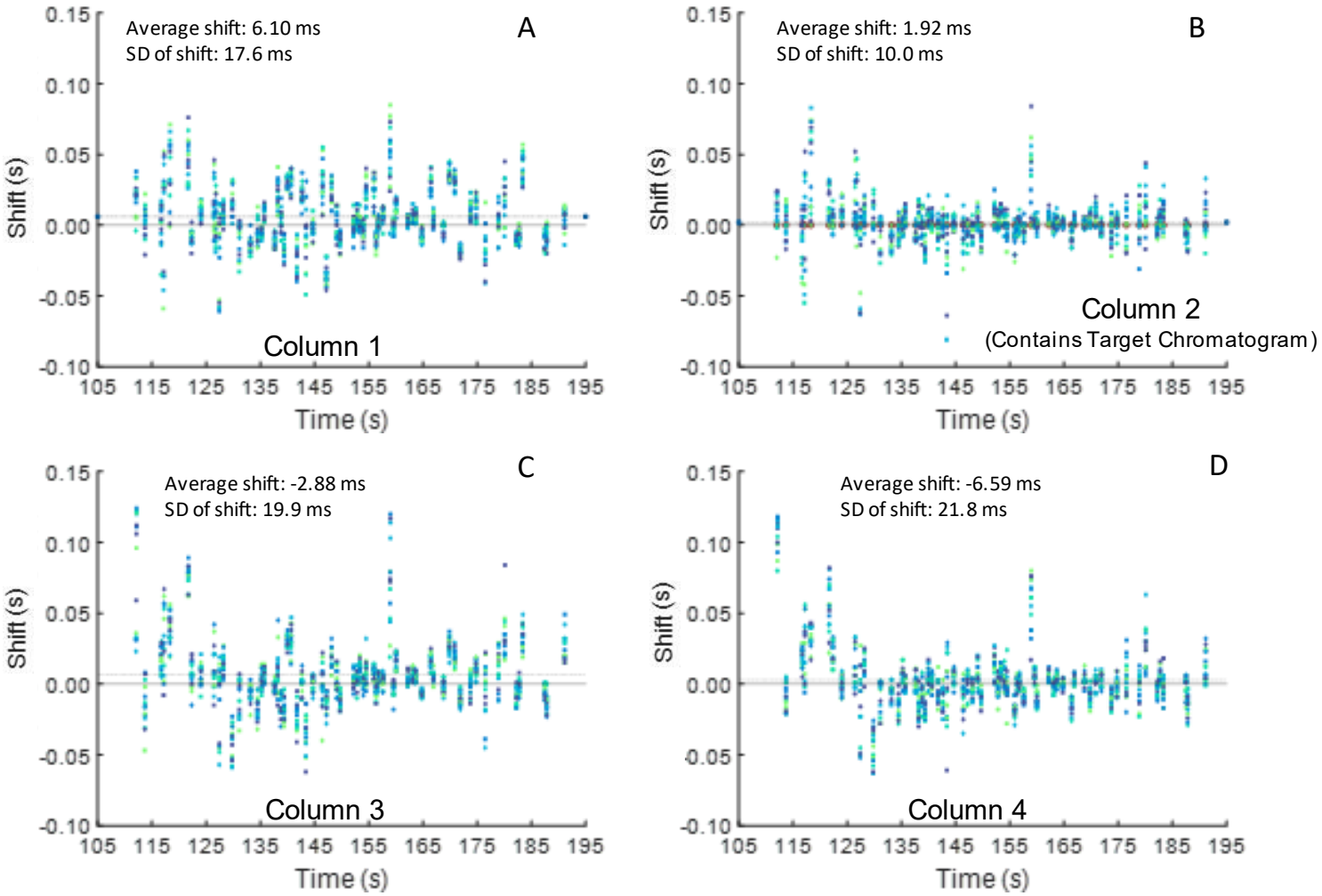

Figure 12. Shift plots for remaining shifting after RTL + COW alignment, using the retention time in the COW-aligned chromatogram and the retention time of the target chromatogram. (A)

Column 1 shift plot. (B) Column 2 shift plot, reflecting remaining within-column shifting. (C) Column 3 shift plot. (D) Column 4 shift plot.

The alignment is further confirmed via the shift plots, given in Fig. 12, which show the remaining shifting post COW + RTL alignment between each chromatogram and the retention times in the target chromatogram used in the COW alignment. The remaining average shifts and standard deviations for the RTL + COW data set are 6.10 ms and 17.6 ms for Column 1, 1.91 ms and 10.0 ms for Column 2, 2.88 ms and 19.9 ms for Column 3, and 6.59 and 21.8 ms for Column 4. PCA was perfomed on the RTL + COW aligned data set to provide another quantitative assessment of the alignment correction.

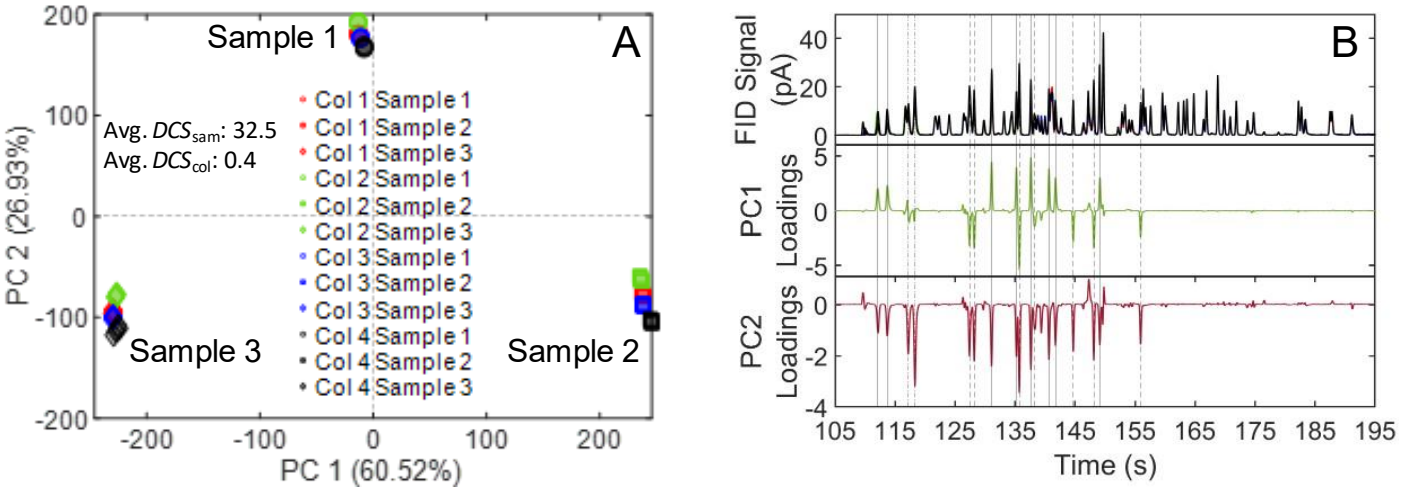

Figure 13. PCA results of the RTL + COW aligned data set. (A) PCA scores plot, illustrating definitive clustering by true sample class. (B) Overlay of all 60 RTL + COW aligned chromatograms (top). PC1 (middle) and PC2 loadings (bottom). Both PC loadings appear to capture variance resulting from the sample classes instead of the column classes.

Figure 13 shows the PCA scores plot of RTL + COW aligned data. Visually, excellent clustering by sample class is observed on PC1 and PC2. This is further confirmed by the *DCS* metric, with a DCS_{sam} of 32.5 and a DCS_{col} of only 0.4, a marked improvement over the RTL-only aligned data *DCS* values of 3.0 and 76.1 respectively. Analysis of the loadings plots indicates that PC1 and PC2 both show contributions solely from spike analyte peaks.

Dataset	Column #	Average shift (ms)	SD of shift (ms)
Unaligned	1	-577	77.9
	2	-18.4	16.4
	3	-2230	99.9
	4	-3880	109
RTL Only	1	-23.9	46.5
	2	-12.2	18.9
	3	72.0	102
	4	62.5	68.8
RTL + COW	1	6.10	17.6
	2	1.92	10.0
	3	2.88	19.9
	4	6.59	21.8

Table 2. The average shift of each peak was used to reduce the data so that the average and standard deviation (SD) of the shifts on a column were obtained. This was done for each column and data set to quantify reductions in retention time shifting with alignment methods.

Dataset	Column #	Average σ_{t_R}(ms)	σ_{t_R} SD (ms)
Unaligned	1	12.4	6.00
	2	38.4	7.00
	3	15.2	6.10
	4	13.1	6.04
RTL Only	1	12.3	5.20
	2	13.7	5.27
	3	10.8	5.84
	4	6.53	5.66
RTL + COW	1	8.85	6.62
	2	8.96	6.89
	3	8.16	5.01
	4	8.98	6.97

Table 3. In addition to the average shift for each peak being used to quantify shifting, the standard deviation of shifting for each peak ($\boldsymbol{\sigma_{t_R}}$) was calculated similarly to quantify consistency in peak retention times.

2.3. Conclusion

Retention time shifting remains a significant challenge to the chemical analysis using chromatography especially when chemometrics is being implemented. This challenge was critically examined in this report for HSGC with the Intuvo GC instrument. Specifically in this report, we determined that COW alone was unable to align across the column-to-column data and analyzed the nature of shifting across nominally identical Intuvo GC columns with no alignment, RTL-alignment, and RTL + COW alignment. We have introduced the concept of performing coarse instrumental alignment with RTL, followed by finer alignment via COW (RTL + COW), thus combining the pre-analysis method adjustment provided by RTL with the post-analysis chemometric alignment provided by COW. The coarse alignment by RTL, while insufficient alone, sufficiently reduced retention time shifting to prevent the misalignment of peaks by COW that occurred when COW was applied without RTL. This method provided excellent classification of the sample classes, which as seen in the comparison of DCS values in Table 4 was significantly greater than that achieved through either RTL or COW alignment alone.

Alignment type	$\boldsymbol{DCS_{sam}}$	$\boldsymbol{DCS_{col}}$
Unaligned	3.0	76.1
COW	1.9	1.7
RTL	4.7	4.2
RTL + COW	32.5	0.4

Table 4. A comparison of the averaged *DCS* values after PCA of the unaligned, COW only aligned, RTL only aligned, and COW + RTL aligned data sets. Note the increase in DCS_{sam} and the decrease in DCS_{col} as overall alignment improves.

Sample classes featuring more heterogenous samples where analytes are present in some and absent in others (as opposed to the entirely native spikes used here) would be at greater risk of misalignment from COW, and as such in greater need of coarse alignment. Although demonstrated in the context of HSGC, the concepts introduced herein should be generalizable to alignment in other chromatographic methods.

Chapter 3. Conclusions

This thesis has discussed the issue of retention time shifting and how two well-known methods that have been widely used to correct limited retention time shifting can be combined to align more severely shifted data. As retention time shifting can be seen to be one specific facet of instrument standardization more generally, this is part of broader work towards ensuring that data across experiments, instruments, and labs can be directly compared and analyzed. This then helps further scientific collaboration and enables greater reproducibility in studies despite differences in environment or in instrumentation. Our work so far has forwarded a promising method for the correction of retention time shifts that are more reflective of cross-instrument differences, as well as new tools such as shift plots which can be used to analyze the specific details of retention time shifting. Both are widely applicable to enable future work, and additional studies on alignment of different samples for other chemometric analysis methods such as PLS are underway.

References

[1] M.R. Ras, R.M. Marcé, F. Borrull, Volatile organic compounds in air at urban and industrial areas in the Tarragona region by thermal desorption and gas chromatography–mass spectrometry, Environ. Monit. Assess. 161 (2010) 389–402. https://doi.org/10.1007/s10661-009-0755-6.

[2] B.K. Lavine, N. Mirjankar, R. LeBouf, A. Rossner, Prediction of mold contamination from microbial volatile organic compound profiles using solid phase microextraction and gas chromatography/mass spectrometry, Microchem. J. 103 (2012) 37–41. https://doi.org/10.1016/j.microc.2012.01.002.

[3] A. Ribes, G. Carrera, E. Gallego, X. Roca, M.J. Berenguer, X. Guardino, Development and validation of a method for air-quality and nuisance odors monitoring of volatile organic compounds using multi-sorbent adsorption and gas chromatography/mass spectrometry thermal desorption system, J. Chromatogr. A 1140 (2007) 44–55. https://doi.org/10.1016/j.chroma.2006.11.062.

[4] M.S. Klee, L.M. Blumberg, Theoretical and practical aspects of fast gas chromatography and method translation, J. Chromatogr. Sci. 40 (2002) 234–247. https://doi.org/10.1093/chromsci/40.5.234.

[5] J.C. Giddings, Theory of Minimum Time Operation in Gas Chromatography, Anal. Chem. 34 (1962) 314–319. https://doi.org/10.1021/ac60183a005.

[6] Supelco, Fast GC, (2015). https://www.sigmaaldrich.com/deepweb/assets/sigmaaldrich/product/documents/166/314/t407096.pdf.

[7] M.R. Jacobs, E.F. Hilder, R.A. Shellie, Applications of resistive heating in gas chromatography: A review, Anal. Chim. Acta 803 (2013) 2–14. https://doi.org/10.1016/j.aca.2013.04.063.

[8] K. Maštovská, J. Hajšlová, M. Godula, J. Křivánková, V. Kocourek, Fast temperature programming in routine analysis of multiple pesticide residues in food matrices. J. Chromatogr. A 907 (2001) 235-245. https://doi.org/10.1016/S0021-9673(00)01045-1.

[9] M. van Lieshout, R. Derks, H-G Janssen, C. A. Cramers, Fast Capillary Gas Chromatography: Comparison of Different Approaches. J. High Resol. Chromatogr. 1998, 21, (11) 583-586. https://doi.org/10.1002/(SICI)1521-4168(19981101)21:11<583::AID-JHRC583>3.0.CO;2-4.

[10] A. Wang, H. D. Tolley, M. L. Lee, Gas chromatography using resistive heating technology. J. Chromatogr. A, 1261 (2012) 46-57. https://doi.org/10.1016/j.chroma.2012.05.021.

[11] T. A. Williams, M. Riddle, S. L. Morgan, W. E. Brewer, Rapid Gas Chromatographic Analysis of Drugs of Forensic Interest, J. Chromatogr. Sci. 37 (6) (1999) https://doi.org/10.1093/chromsci/37.6.210.

[12] B. D. Fitz, B. C. Mannion, K. To, T. Hoac, R. E. Synovec, Evaluation of injection methods for fast, high peak capacity separations with low thermal mass gas chromatography, J. Chromatogr. A, 1392 (2015) 82-90. https://doi.org/10.1016/j.chroma.2015.03.009.

[13] K. Sasamoto, N. Ochiai, H. Kanda, Dual low thermal mass gas chromatography-mass spectrometry for fast dual-column separation of pesticides in complex sample, Talanta 72 (2007) 1637–1643. https://doi.org/10.1016/j.talanta.2007.03.020.

[14] J. Luong, R. Gras, G. Yang, H. Cortes, R. Mustacich, Multidimensional gas chromatography with capillary flow technology and LTM-GC, J. Sep. Sci. 31 (2008) 3385–3394. https://doi.org/10.1002/jssc.200800163.

[15] P. Boeker, J. Leppert, Flow field thermal gradient gas chromatography, Anal. Chem. 87 (2015) 9033–9041. https://doi.org/10.1021/acs.analchem.5b02227.

[16] J. Leppert, M. Härtel, T.M. Klapötke, P. Boeker, Hyperfast Flow-Field Thermal Gradient GC/MS of Explosives with Reduced Elution Temperatures, Anal. Chem. 90 (2018) 8404–8411. https://doi.org/10.1021/acs.analchem.8b00900.

[17] K. Ramos, A. Riddell, H. Tsiagras, A.M. Hupp, Analysis of biodiesel-diesel blends: Does ultrafast gas chromatography provide for similar separation in a fraction of the time?, J. Chromatogr. A 1667 (2022) 462903. https://doi.org/10.1016/j.chroma.2022.462903.

[18] Agilent Technologies, Intuvo 9000 GC System, (2022). https://www.agilent.com/en/product/gas-chromatography/gc-systems/intuvo-9000-gc-system.

[19] N.-P. Vest Nielsen, J.M. Carstensen, J. Smedsgaard, Aligning of single and multiple wavelength chromatographic profiles for chemometric data analysis using correlation optimised warping, J. Chromatogr. A 805 (1998) 17-35. https://doi.org/10.1016/S0021-9673(98)00021-1.

[20] 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. https://doi.org/10.1016/S0021-9673(03)00616-2.

[21] K.M. Pierce, B.W. Wright, R.E. Synovec, Unsupervised parameter optimization for automated retention time alignment of severely shifted gas chromatographic data using the piecewise alignment algorithm, J. Chromatogr. A 1141 (2007) 106–116. https://doi.org/10.1016/j.chroma.2006.11.101.

[22] Y. Wang, D.J. Veltkamp, B.R. Kowalski, Multivariate instrument standardization, Anal. Chem. 63 (1991) 2750–2756. https://doi.org/10.1021/ac00023a016.

[23] N. Etxebarria, O. Zuloaga, M. Olivares, L.J. Bartolomé, P. Navarro, Retention-time locked methods in gas chromatography, J. Chromatogr. A 1216 (2009) 1624–1629. https://doi.org/10.1016/j.chroma.2008.12.038.

[24] A. Kende, Z. Csizmazia, T. Rikker, V. Angyal, K. Torkos, Combination of stir bar sorptive extraction-retention time locked gas chromatography-mass spectrometry and automated mass spectral deconvolution for pesticide identification in fruits and vegetables, Microchem. J. 84 (2006) 63–69. https://doi.org/10.1016/j.microc.2006.04.015.

[25] Y. Li, Q. Ruan, Y. Li, G. Ye, X. Lu, X. Lin, G. Xu, A novel approach to transforming a non-targeted metabolic profiling method to a pseudo-targeted method using the retention time locking gas chromatography/mass spectrometry-selected ions monitoring, J. Chromatogr. A 1255 (2012) 228–236. https://doi.org/10.1016/j.chroma.2012.01.076.

[26] J.H. Christensen, G. Tomasi, A.B. Hansen, Chemical fingerprinting of petroleum biomakers using time warping and PCA, Environ. Sci. Technol. 39 (2005) 255–260. https://doi.org/10.1021/es049832d.

[27] T. Skov, F. Van Den Berg, G. Tomasi, R. Bro, Automated alignment of chromatographic data, J. Chemom. 20 (2006) 484–497. https://doi.org/10.1002/cem.1031.

[28] 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. https://doi.org/10.1016/j.chroma.2009.02.049.

[29] I. Eisele, Retention Time Locking of Organochlorine Pesticides on an Agilent 8860 GC System Using the OpenLab Retention Time Locking Wizard, Agil. Appl. Note. (2019).

[30] G. Tomasi, F. Van Den Berg, C. Andersson, Correlation optimized warping and dynamic time warping as preprocessing methods for chromatographic data, J. Chemom. 18 (2004) 231–241. https://doi.org/10.1002/cem.859.

[31] L. Zheng, D.G. Watson, B.F. Johnston, R.L. Clark, R. Edrada-Ebel, W. Elseheri, A chemometric study of chromatograms of tea extracts by correlation optimization warping in conjunction with PCA, support vector machines and random forest data modeling, Anal. Chim. Acta 642 (2009) 257–265. https://doi.org/10.1016/j.aca.2008.12.015.

[32] D. Jacobson, A.R. Monforte, A.C.S. Ferreira, Untangling the Chemistry of Port Wine Aging with the Use of GC-FID, Multivariate Statistics, and Network Reconstruction, J. Agric. Food Chem. 61 (2013) 2513–2521. https://doi.org/10.1021/jf3046544.

[32] E.H. Ebrahimabadi, S.M. Ghoreishi, S. Masoum, A.H. Ebrahimabadi, Combination of GC/FID/Mass spectrometry fingerprints and multivariate calibration techniques for recognition of antimicrobial constituents of Myrtus communis L. essential oil, J. Chromatogr. B 1008 (2016) 50–57. https://doi.org/10.1016/j.jchromb.2015.11.010.

[33] R. Bro, A.K. Smilde, Principal component analysis, Anal. Methods 6 (2014) 2812–2831. https://doi.org/10.1039/c3ay41907j.

[34] P.E. Sudol, D. V. Gough, S.E. Prebihalo, R.E. Synovec, Impact of data bin size on the classification of diesel fuels using comprehensive two-dimensional gas chromatography with

principal component analysis, Talanta 206 (2020) 120239. https://doi.org/10.1016/j.talanta.2019.120239.

[35] C.N. Cain, S. Schöneich, R.E. Synovec, Development of an enhanced total ion current chromatogram algorithm to improve untargeted peak detection, Anal. Chem. 16 (2020) 11365-11373. https://doi.org/10.1021/acs.analchem.0c02136.

[36] G. Tomasi, T. Skov, F. van den Berg, Dynamic Time Warping (DTW) and Correlation Optimized Warping (COW). http://www.models.kvl.dk/DTW_COW (accessed November 11, 2022).

[37] R.G. Brereton, G.R. Lloyd, Re-evaluating the role of the Mahalanobis distance measure, J. Chemom. 30 (2016) 134–143. https://doi.org/10.1002/cem.2779.

[38] J.S. Nadeau, B.W. Wright, R.E. Synovec, Chemometric analysis of gas chromatography-mass spectrometry data using fast retention time alignment via a total ion current shift function, Talanta 81 (2010) 120–128. https://doi.org/10.1016/j.talanta.2009.11.046.

[39] M.D. Krebs, R.D. Tingley, J.E. Zeskind, M.E. Holmboe, J.M. Kang, C.E. Davis, Alignment of gas chromatography-mass spectrometry data by landmark selection from complex chemical mixtures, Chemom. Intell. Lab. Syst. 81 (2006) 74–81. https://doi.org/10.1016/j.chemolab.2005.10.001.

[40] G. Malmquist, R. Danielsson, Alignment of chromatographic profiles for principal component analysis: a prerequisite for fingerprinting methods, J. Chromatogr. A 687 (1994) 71–88. https://doi.org/10.1016/0021-9673(94)00726-8.