EFFECTIVENESS OF GLOBAL, LOW-DEGREE POLYNOMIAL TRANSFORMATIONS FOR GC X GC DATA ALIGNMENT

An Undergraduate Honors Thesis
Submitted in Partial Fulfillment of
A Degree with Distinction for the
College of Arts and Sciences
University of Nebraska-Lincoln

by
Davis Rempe, B.S.
Computer Science, Mathematics

October 24, 2016

Faculty co-advisors:
Stephen Scott, Ph.D., Computer Science and Engineering
Stephen Reichenbach, Ph.D., Computer Science and Engineering

ACKNOWLEDGEMENTS

I would like to thank Chiara Cordero, Wayne E. Rathbun, and Cláudia Alcaraz Zini for collecting all the GCxGC data used as part of these experiments. I also want to thank Stephen Scott and Steve Reichenbach for their time and patience in advising this work.

ABSTRACT

As columns age and differ between systems, retention times for comprehensive two-dimensional gas chromatography (GCxGC) may vary between runs. In order to properly analyze GCxGC chromatograms, it often is desirable to align the retention times of chromatographic features, such as analyte peaks, between chromatograms. Previous work [Reichenbach et al., *Anal. Chem.* 2015, 87, 10056] has shown that global, low-degree polynomial transformation functions – namely affine, second-degree polynomial, and third-degree polynomial – are effective for aligning pairs of two-dimensional chromatograms acquired with dual second columns and detectors (GCx2GC). This work assesses the experimental performance of these global methods on more general GCxGC chromatogram pairs and compares their performance to that of a recent, robust, local alignment algorithm for GCxGC data [Gros et al., *Anal. Chem.* 2012**,** 84, 9033]. Measuring performance with the root-mean-square (RMS) residual differences in retention times for matched peaks suggests that global, low-degree polynomial transformations outperform the local algorithm given a sufficiently large set of alignment points, and are able to improve misalignment by over 95% based on a lower-bound benchmark of inherent variability. However, with small sets of alignment points, the local method demonstrated lower error rates (although with greater computational overhead). For GCxGC chromatogram pairs with only slight initial misalignment, none of the global or local methods performed well. In some cases with initial misalignment near the inherent variability of the system, these methods worsened alignment, suggesting that it may be better not to perform alignment in such cases.

1. INTRODUCTION

This work assesses the performance of global, low-degree polynomial transformations — namely affine, second-degree polynomial, and third-degree polynomial (*1*) — for retention-time alignment between chromatograms obtained by comprehensive two-dimensional gas chromatography (GCxGC). It also compares the performance of these global methods to that of a recent, robust, local alignment algorithm for GCxGC chromatograms proposed by Gros et al. (*2*).

Due to column aging and other run-to-run system variations, retention times may vary between GCxGC chromatograms, even when acquired on the same system. To mitigate this issue, it may be necessary to perform chromatographic alignment by mapping the retention times of one chromatogram to the times of another chromatogram. Alignment methods can be classified as "global or local, i.e., whether the geometric differences between chromatograms are characterized by a single function for the entire chromatogram or by a combination of many functions for different regions of the chromatogram" (*1*).

Previous work (*1*) has investigated global, low-degree polynomial transformation functions for aligning chromatogram pairs acquired by comprehensive two-dimensional gas chromatography with one first-dimension (^{1}D) column and two parallel second-dimension (^{2}D) columns (GCx2GC) (*3-6*). The chromatogram pairs aligned in that work came from the same run with one ^{2}D column to a flame ionization detector (FID) and another ^{2}D column to a mass spectrometer (MS). These chromatogram pairs had significant variations in the ^{2}D retention times. For GCx2GC, low-degree polynomial mapping functions outperformed affine transformations. These polynomial functions were able to approach benchmarks for retention-

time root-mean-square residual error (RMSE) between chromatograms based on consecutive replicate sample runs on the same system and detector.

The present work investigates the performance of these global, low-degree polynomial transformations more generally to align GCxGC chromatograms, i.e.: Does the retention-times RMSE between two chromatograms after alignment approach the noise benchmark? To assess these global methods, chromatograms from three sets of data are used for alignment. Each data set varies an important chromatographic parameter: the date and time, the sample, and the instrument configuration. This allows the alignment methods to be tested across a wide range of situations. The first set of chromatograms was produced from the same diesel sample run over a period of about two and a half years. These chromatograms have moderate initial misalignment due to system and column variations. The second set of chromatograms was produced from samples of three different wine vintages that were run in a period of days. These chromatograms have minimal initial misalignment from run-to-run system variability. The last set of chromatograms was produced from a single cocoa sample, but on systems with two different modulation technologies: flow and thermal. These chromatograms are extremely misaligned due to different system configurations, namely, different modulators, column dimensions, and carrier gas flow.

Additionally, this research compares the performance of these global functions to that of a high-performing local alignment algorithm (*2*). Comparing global and local methods may show whether retention-time differences between GCxGC chromatograms are systemic and therefore well-suited to simple, global functions, or if the differences are too complex and require more sophisticated local methods for alignment. For this, a local alignment method developed by Gros et al. (*2*) is evaluated. Although there are other available local alignment

methods, their work indicates that compared to two other local alignment methods, their robust algorithm "performs the best overall in terms of decreased retention time deviations of matching analytes" (*2*). Gros et al. compared their method to one developed by Pierce et al. (*7*), which was "the first published alignment algorithm for the correction of shifts resulting from uncontrollable variations for whole GC × GC chromatograms" (*2*), and to two-dimensional correlation optimized warping (2-D COW) (*8*) – a multidimensional extension of the original COW algorithm (*9*). As evidenced by these results, the method described by Gros is a high-performing local method.

The experimental methods for testing the various alignment functions follow previous work (*1*). The effectiveness of the alignment methods is measured in terms of the RMSE of the post-alignment retention times for pairs of matched peaks in two GCxGC chromatograms. The error that the alignment methods aim to reach is the benchmark RMSE, computed between pairs of chromatograms from consecutive replicate sample runs on the same system. This benchmark is based on the assumption that the retention-times differences between consecutive replicate sample runs on the same system are unpredictable random noise. Cross-validation experiments are used to evaluate all methods: affine, second-degree polynomial, third-degree polynomial, and the local algorithm from Gros et al. To get an unbiased indicator of performance, these tests use one set of matched peak-pairs to fit (or train) the alignment functions, and a different, disjoint set to measure (or test) the post-alignment RMSE.

2. EXPERIMENTAL SECTION

2.1 Samples

Three different sample types are used to assess performance of the data alignment algorithms. The first is a single distillate diesel sample. The sample was run four different times on the same system over a period of about two and a half years to produce a set of GCxGC chromatograms. Each of these runs were far apart in time, so the chromatograms have moderate misalignments from column differences, such as aging and replacement. The lower-bound benchmark RMSE was determined from a set of four consecutive replicate runs with the same diesel sample on the same system.

The second set of chromatograms came from samples of three different wine vintages. All samples were run within a period of three days as part of a study at the Universidade Federal do Rio Grande do Sul related to the characterization of commercial Merlot wines from the Brazilian Campanha region. All samples were the same Merlot brand, but from different years: 2011, 2012, and 2013. Each sample was run on the system twice consecutively, which provides the replicate runs for determining the alignment benchmark. Because all runs were within a short time period on the same system, the misalignments are relatively small.

The third set of chromatograms came from a single Trinitario cocoa nib sample from Ecuador. The sample was run as part of a study at the Università degli Studi di Torino in Turin, Italy, that focuses on the sensomic characterization of cocoa samples from different botanical and geographical origins. Two chromatograms were first acquired on the system using a reverse-inject differential flow modulator (*10*). The same sample was again run about four months later to acquire three more chromatograms, but this time with a loop-type thermal modulator. The

flow-modulated GCxGC runs were preliminary experiments under unoptimized conditions, making alignment even more difficult. The sample was run consecutively on each modulation platform, so there are replicate runs to determine the alignment benchmarks. Varying the modulation technologies between these sets of runs results in chromatograms with extreme misalignment, much larger than that seen in the diesel chromatograms, particularly in the ^{1}D.

2.2 Instrumentation

For analysis of the diesel sample, all run conditions were in accordance with UOP 990 (*11*), with a modulation period of 8 s and sampling with a flame ionization detector (FID) at 200 Hz, on a LECO GCxGC-FID system (LECO Corp., St. Joseph, MI) with Agilent 6890 GC (Agilent Technologies, Little Falls, DE).

For analysis of the volatile fraction of wine samples, headspace solid-phase micro-extraction (HS-SPME) was performed with one mL of wine, 0.3 g of sodium chloride at 55°C (± 0.9), and a DVB/CAR/PDMS fiber (Supelco, Bellefonte, PA) in 20 mL headspace screw-capped glass vials (*12*). The system was a LECO GCxGC with an Agilent 6890N and time-of-flight mass spectrometric detector (TOFMS). The modulation cycle was 7 s with spectra from 45-450 *m*/*z* acquired at about 100 Hz.

For analysis of the cocoa nib, the GCxGC experimental conditions were different for each modulation technology. The GCx2GC-MS/FID runs with reverse-inject differential flow modulation used an Agilent 7890B GC unit coupled to an Agilent 5977A fast quadrupole MS detector operating in EI mode at 70 eV, and a fast FID. The modulation cycle was 3 s with spectra from 40-240 *m*/*z* acquired at about 35 Hz. The GCxGC-MS runs with thermal modulation used an Agilent 6890 unit with a Zoex loop-type modulator (Zoex

Corp., Lincoln, NE) coupled to an Agilent 5975C MS detector operating in EI mode at 70 eV. The modulation cycle was 3 s with spectra from 40-240 *m*/*z* acquired at about 29 Hz.

Additional details of the instrumental conditions for all systems are included in Appendix A.

2.3 Data Preprocessing

Data preprocessing was performed using GC Image GCxGC Edition Software (R2.6 alpha build) from GC Image, LLC (Lincoln, NE) (*13*). Examples of processed chromatograms for each sample are shown in Appendix C.

For the diesel chromatograms, phase-shifting, baseline correction, and peak detection were performed (*14*). Automated bidirectional peak matching created initial lists of corresponding peaks between all pairs of chromatograms. The lists were edited manually to increase the number and temporal coverage and to ensure correct correspondences, resulting in a total of 112 peaks that were matched across all eight chromatograms (four runs well separated in time and four consecutive replicate runs). Because manual verification was being performed, loose matching criteria were used for creating the initial list to increase the number of prospective peak-pairs and minimize bias. After manual editing, the peaks are well-distributed across the retention times of the chromatograms (Figure C1).

For the wine chromatograms, baseline correction and peak detection were performed. Using the same process as for the diesel sample, a total of 78 peaks were selected and confirmed by MS to correspond across all six chromatograms (Figure C2).

Chromatograms acquired from the cocoa sample yielded fewer corresponding peak-pairs using these peak matching techniques. After baseline correction and peak detection, 33 peaks were confirmed by MS across all five chromatograms (Figures C3 and C4).

2.4 Evaluation Metric

The primary evaluation metric is the RMSE of the post-alignment retention times across the peak sets for pairs of chromatograms. This metric is used in previous work that evaluates the global alignment methods (*1*). For a set containing N_p pairs of corresponding peaks, the RMSE is defined as

$$RMSE = \left(\sqrt{\frac{1}{N_p}\sum_1^{N_p}(x_i - x_i')^2}, \sqrt{\frac{1}{N_p}\sum_1^{N_p}(y_i - y_i')^2}\right) \quad (1)$$

where the retention times for a peak in the target and reference chromatograms are (x_i, y_i) and (x_i', y_i'), respectively, with i indexing the peaks from 1 to N_p.

A blob's retention times indicate its data point with the maximal signal value, i.e., its apex.

2.5 Transformation Models

The global transformation models are identical to those used in previous work (*1*). The first evaluation is with no alignment function applied, i.e. the initial misalignment. This is trivially defined as:

$$f_0(x, y) = (x, y). \quad (2)$$

The affine transformation applies scaling, shearing, and translation:

$$f_1(x,y) = \left(t_x + s_x x + h_x y,\ t_y + h_y x + s_y y\right). \quad (3)$$

where (s_x,s_y) scale in each dimension, (h_x,h_y) shear, and (t_x,t_y) are the translation.

The second and third-degree polynomials simply add extra terms. For both dimensions, the second degree-polynomial adds three more terms:

$$f_2(x,y) = (t_x + s_x x + h_x y + a_x xy + b_x x^2 + c_x y^2, t_y + h_y x + s_y y + a_y xy + b_y x^2 + c_y y^2) \quad (4)$$

The third-degree polynomial adds an additional four terms to the second-degree polynomial:

$$f_3(x,y) = (t_x + s_x x + h_x y + a_x xy + b_x x^2 + c_x y^2 + \alpha_x x^2 y + \beta_x xy^2 + \gamma_x x^3 + \delta_x y^3, t_y + h_y x + s_y y + a_y xy + b_y x^2 + c_y y^2 + \alpha_y x^2 y + \beta_y xy^2 + \gamma_y x^3 + \delta_y y^3) \quad (5)$$

Each global function requires a minimum number of alignment peak-pairs in order to determine the parameters: three peak-pairs for affine, six peak-pairs for second-degree polynomial, and ten peak-pairs for third-degree polynomial. For numbers of peak-pairs larger than the minimum number, the optimal parameters minimize the RMSE of fitted pairs.

The local alignment method of Gros et al. (*2*) also uses corresponding peak-pairs for alignment. These peak-pairs are referred to as alignment points. This algorithm guarantees that these points are perfectly aligned in the final chromatogram produced. Based on these alignment points, displacements for the rest of the data are estimated in both dimensions. In the ^{1}D, displacements are linearly interpolated between alignment points. In the ^{2}D, displacements are estimated using Sibson natural-neighbor interpolation (*15*), based on Voronoi diagrams. For interpolation in the ^{2}D, the algorithm requires the typical peak width (tpw) for both dimensions. This is the number of data-points that make up approximately two standard deviations of a peak (*16*). In the diesel experiments, tpws of 2 and 40 data-points (0.267 min and 0.2 s) were used for the ^{1}D and ^{2}D, respectively. For the wine samples, tpws of 2 and 17 (0.23 min and 0.17 s) data-points were used. For the cocoa samples, tpws of 5 (0.25 min) and 6 data-points (0.17 s for flow

modulation, 0.21 s for thermal) were used. These tpws were roughly determined by visual examination of typical peaks near the center of the chromatogram. This process follows the documentation to users from Gros et al. (*16*). The final step of the algorithm re-interpolates the signal values for all pixels and applies a deformation correction. This part of the algorithm was not executed during the cross-validation testing in this paper, because the focus here is on comparing the retention-times alignments with those of the global methods and not on the separate step of intensity interpolation.

2.6 Evaluation Methodology

The evaluation methodology follows previous work (*1*). Within the alignment points used, the transformations fit the noise as well as the alignment peaks, which is a problem of overfitting. To get an unbiased estimate of a method's performance, a cross-validation technique is employed. The set of corresponding peak-pairs is partitioned into two disjoint sets: a training and testing set. The training set is used as the alignment points for fitting the methods, and the testing set is used to measure their performance. Measuring the error across testing-set peak-pairs after alignment is a good unbiased indicator of the method's performance, as the transformation was not fit to these peak-pairs and their inherent noise.

The experiments are run for every training set size from 3 peak-pairs (the minimum size for affine transformations) to all of the matched peak-pairs, at which point the test set is null. For each training set size, 100 trials are run. The training and testing sets are randomly generated at each trial (and are disjoint complements of the peak-pairs set). Because of the random selection of peak-pairs, the training set may not be well-distributed across the entire chromatogram. The alignment is also done both forward and backward, i.e., peaks from

chromatogram 1 are fit to those in chromatogram 2 and vice versa. The reported RMSE for each training set size is the average RMSE over all 200 trials (with 100 in each direction).

2.7 Performance Benchmarks

The global alignment methods are assessed in two ways. First, does the method approach the benchmark error set by the consecutive replicate runs? Second, does the method perform better than the local alignment algorithm? For the first question, the misalignment between consecutive replicate runs can be used as a benchmark indicating the lower bound of alignment performance due to systemic noise. Any misalignment between two replicate chromatograms acquired one after another with the same sample on the same system can be considered the level of random retention-times noise inherent to the system itself.

The degree to which an alignment method approaches the benchmark is measured by its percent improvement I_p. For a specific alignment method, let S be the set of post-alignment average RMSEs for every testing set size and $\min\{s\}, s \in S$, be the best average RMSE achieved for any testing set size. Then that method's percent improvement is defined as

$$I_p = \frac{m_0 - \min_{s \in S}\{s\}}{m_0 - m_b} \times 100 \quad (6)$$

where m_0 is the average testing set RMSE over all trials with no alignment function applied (i.e., the initial misalignment) and m_b is the benchmark RMSE from consecutive replicate runs.

Comparing global performance to the local method is done in multiple ways. If the alignment methods have a RMSE less than that of the local method, they can be said to perform better. The computational overhead (i.e. run-time) of an alignment algorithm is another useful comparison. It is also important to take into consideration how many peak-pairs are required in

order to achieve (or nearly achieve) the method's maximal performance. It may be desired to have a method that can align two chromatograms relatively well using fewer alignment points, rather than one that can achieve a slightly smaller RMSE but which requires more alignment points.

Ideally, the methods should be compared on their performance for specific data sets of interest. For generality, the data sets used here offer a wide range of initial misalignment — from negligible to severely misaligned — so the alignment performance can be considered relative to the initial misalignment. Additionally, each data set varies a different GCxGC chromatogram acquisition parameter. The first varies the analysis over time, the second varies the sample, and the third varies the GCxGC instrument with different modulation platforms.

2.8 Execution Methodology

Experiments were run on the Crane cluster of the Holland Computing Center (*17*) located on the University of Nebraska-Lincoln campus. The cluster has a total of 452 nodes with 64 GB of RAM each. In each of the 16 cores within a single node, there are two Intel Xeon E5-2670 2.60GHz processors.

All alignment methods were implemented in MATLAB. Part of the MATLAB implementation of the local algorithm from Gros et al. was parallelized in order to run much faster across 16 cores on the Crane cluster. Even with the speed boost, and without executing the resampling portion of the algorithm, the local method was more computationally expensive than the simpler global functions.

For the case of 105 peak-pairs for aligning two 1199x1600 diesel chromatograms, fitting the second-degree polynomial to the peak-pairs and computing the transformation for every data

point required 0.1906 s. By comparison, the local algorithm required 8.5971 s to compute the displacements for every data point. Of course, the computation-time difference is smaller if fewer retention times must be transformed (e.g., as would be required to transform a template). However, as these timing results illustrate, the global function requires significantly less computation for larger alignment problems.

3. RESULTS AND DISCUSSION

3.1 Time-Varied Data Results

Chromatograms acquired from the diesel sample were used to test the alignment methods on time-varied data. Tests were performed on chromatograms from four consecutive replicate runs on the same diesel sample to establish a benchmark for the alignment methods. These four chromatograms are labeled runs 17, 18, 19, and 20. The initial misalignment was recorded for consecutive runs: 17 and 18, 18 and 19, and 19 and 20. The results from the cross-validation benchmark tests between runs 18 and 19 are shown in Figure 1. These graphs show the retention-time RMSE for the testing set of peak-pairs for each alignment method as a function of the training-set size, i.e., the number of alignment points used. Each alignment method is represented by a different colored line. The figures for the training sets and additional replicate results can be found in Appendix B (Figure B1).

In both chromatographic dimensions, as the training set size increases, the RMSE of the global functions generally decreases for the testing sets. This makes sense because larger training sets yield better estimates of the global misalignment (because overfitting to noise is reduced), producing the decrease seen in testing-set error.

The RMSE of consecutive replicate runs, which provides our benchmark error, is the blue line in Figure 1. The ^{1}D graph (Figure 1a) shows that none of the alignment algorithms are able to improve upon this initial misalignment.

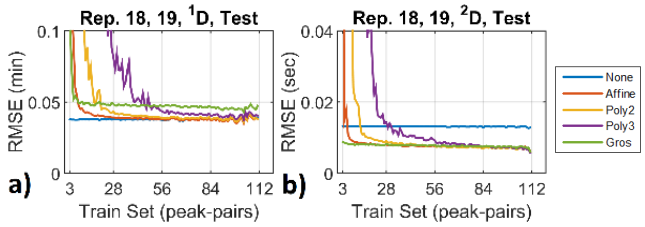


Figure 1. Cross-validation retention-time RMSE results as a function of training set size for consecutive replicate runs of a diesel sample. The RMSE is shown for: a) ^{1}D with the testing set and. b) ^{2}D with the testing set.

In the ^{2}D (Figure 1b), there is only a small improvement of less than 0.01 s. This supports the claim that the initial misalignment of consecutive replicate runs indicates the inherent lower-bound limits on any alignment algorithm.

In the ^{1}D, with no alignment function applied, the RMSE averages 0.0375 min, which is the maximum initial misalignment seen in the ^{1}D across the three replicate tests. In the ^{2}D the initial misalignment is about 0.0131 s. Across all three pairs of replicate runs, the average misalignment in the ^{1}D is 0.0243 min which is less than the modulator sampling noise level of 0.038 min.[i] The peaks in the diesel sample chromatograms are narrow in the ^{1}D, with a tpw of only about 2 modulations, which affects the choice of an alignment benchmark. An alignment method cannot be expected to achieve an RMSE better than the sampling noise, so 0.038 min is the benchmark value in the ^{1}D. Across all three pairs of replicate runs, the average misalignment in the ^{2}D is 0.0125 s. This value is the ^{2}D benchmark RMSE for the alignment methods being tested.

Next, the cross-validation tests were run on every pair-wise combination of four chromatograms acquired over 2.5 years. Due to column aging, these chromatograms exhibit moderate misalignments. The results from one of these pair-wise tests are shown in Figure 2. The names of the

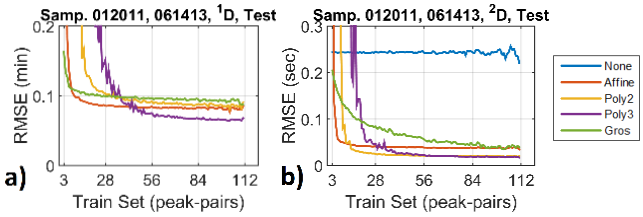


Figure 2. Cross-validation retention-time RMSE results as a function of training set size for chromatograms produced from the same diesel sample about 2.5 years apart. RMSE is shown for: a) ^{1}D with the testing set and b) ^{2}D with the testing set. The names of the samples correspond to the acquisition date (January 20, 2011 and June 14, 2013).

[i] The distillate analyses have a modulation cycle (P_M) of 8 s or $P_M = 0.13$ min. The standard deviation for random uniformly distributed residuals with respect to a single modulation interval is $12^{-1/2} \times P_M$, which is about 0.038 min for these data. This is the RMS retention-time noise level from the sampling effect of modulation and has implications for the benchmark RMSE in the ^{1}D.

samples (January 20, 2011, and June 14, 2013) indicate the dates on which they were run; so, the chromatograms aligned in this figure were acquired about 2.5 years apart. Before any alignment is applied (the blue line in Figure 2), the RMSE is about 0.76 min in the ^{1}D and 0.24 s in the ^{2}D. The "None" function is excluded from plot (2a) to focus on performance of the alignment models.

The testing-set plots in Figure 2 show how the transformations affect peak-pairs that were not used for fitting, for an unbiased evaluation. In both dimensions, significant improvements are seen after applying both the global and local methods to the alignment of chromatograms 012011 and 061413. In the ^{1}D, the third-degree polynomial transformation achieves the smallest RMSE of 0.0641 min compared to the largest RMSE of 0.0871 min for the local algorithm. The (best-performing) third-degree polynomial (0.0641 min) has a percent improvement of $I_p = 96.4\%$ using the benchmark of 0.038 min. Though it has the largest RMSE of the methods tested, resulting in a percent improvement of $I_p = 93.2\%$, Gros' algorithm only requires about 10 peak-pairs to approach its peak performance. This is a smaller training-set size than required for the global methods to reach peak performance.

In the ^{2}D, the third-degree polynomial also achieves the best peak performance, with a minimum RMSE of 0.017 s ($I_p = 98\%$), nearing the 0.0125 s benchmark, compared to 0.0346 s ($I_p = 90.3\%$) for the local method. In the ^{2}D, Gros' algorithm takes much longer to reach its peak performance at around 85 peak-pairs, but it has a lower RMSE than the global functions when the training set size is small.

Training-set data and graphs similar to Figure 2 for all other cross-validation experiments can be found in Appendix B (Figure B2). The patterns discussed with Figure 2 are consistent across most of the experiments. Table 1 summarizes the results from all six cross-validation

Minimum Testing-Set RMSE Reached by Alignment Methods in the ^{1}D (min) and ^{2}D (s) for Diesel Chromatograms										
Chromat-ograms	**None (Avg.)**		**Affine**		**Poly2**		**Poly3**		**Gros et al.**	
	^{1}D	**^{2}D**	**^{1}D**	**^{2}D**	**^{1}D**	**^{2}D**	**^{1}D**	**^{2}D**	**^{1}D**	**^{2}D**
012011-061413	0.7563	0.2414	0.0767	0.0344	0.0806	0.0184	0.0641	0.0170	0.0871	0.0346
012011-090912	0.1024	0.3982	0.0574	0.0147	0.0583	0.0130	0.0592	0.0131	0.0640	0.0435
012011-100412	0.0800	0.0569	0.0502	0.0257	0.0460	0.0225	0.0488	0.0223	0.0612	0.0283
061413-090912	0.8353	0.1819	0.0856	0.0367	0.0868	0.0223	0.0747	0.0221	0.0902	0.0209
061413-100412	0.7940	0.2905	0.0763	0.0558	0.0783	0.0331	0.0511	0.0282	0.0996	0.0558
090912-100412	0.0770	0.4386	0.0631	0.0292	0.0635	0.0247	0.0644	0.0241	0.0671	0.0578
Average	0.4408	0.2679	0.0682	0.0328	0.0689	0.0223	0.0604	0.0211	0.0782	0.0402
Average % Improvement			76.7	87.7	77.8	93.1	77.3	93.6	68.0	86.0

Table 1. Minimum testing-set RMSE for each alignment method in both the first and second chromatographic dimensions for all six experiments with the non-replicate chromatograms from the diesel sample. The "None" columns are the average initial misalignments, not the minimum. The third-degree polynomial function reaches the lowest error on average, and Gros et al. has the highest error on average.

experiments run with the non-replicate diesel chromatograms. Under "None" is the average initial misalignment (m_0 in Eq. (6)). For each experiment, the minimum average testing set RMSE ($\min\{s\}, s \in S$, in Eq. (6)) for each alignment method is shown. The bottom two rows present the averages for the minimum RMSE and percent improvement. Note that the top-performing method in terms of average minimum RMSE may not be the best in terms of average percent improvement (and vice versa), because the average percent improvement depends heavily on initial misalignment. Even if a method averages the smallest RMSE, it may not have the smallest RMSE in cases that the misalignment is very small, which negatively affects its average percent improvement.

On average, all three global alignment methods are able to reach a better peak performance and percent improvement than the local algorithm in both dimensions. The third-degree polynomial averages a 9.3% greater percent improvement than Gros' algorithm in the ^{1}D,

and 7.6% greater in the ^{2}D. In the ^{1}D, the average percent improvement for all alignment methods is noticeably worse than the experiment discussed in Figure 2. For chromatogram pairs with a less significant initial misalignment in the ^{1}D (012011-090912, 012011-100412, and 090912-100412 in Table 1), the alignment methods tend to reach similar minimum RMSE values to the experiments with larger initial misalignments, causing the lower average percent improvements overall. For experiments with large misalignments (> 0.7 min), like in Figure 2, the third-degree polynomial is consistently able to achieve a percent improvement over 95%. In the ^{2}D, both the second and third-degree polynomials average a percent improvement over 93% for all experiments.

There is a clear tradeoff in terms of the number of alignment points used and the minimum RMSE reached for both the local and global methods. If using a very small number of alignment points (~5), it may be preferable to use Gros' algorithm because it starts out at a much lower error than any of the global methods. Though the local method performs relatively well with a small number of alignment points, it is outperformed in both dimensions by the global methods when a larger numbers of alignment points are available. The number of peak-pairs at which the global methods overtake the local method varies between algorithms, and is larger in the ^{1}D than the ^{2}D. With just under 10 points or more, the affine transformation becomes a better choice, attaining a clear performance gain in both dimensions, on average. With around 30 pairs or more, the second-degree polynomial performance overtakes the local method. The third-degree polynomial improves upon the local method when about 50 alignment points or more are available. Though the third-degree polynomial is also able to outperform the second-degree (with ~55 points), the performance gain is small. In terms of percent improvement, the second-degree actually averages better than the third-degree in the ^{1}D, and is within 1% in the ^{2}D. Therefore, for

computational simplicity and because fewer alignment points are required, it may be preferable to use the second-degree function. This result is similar to that seen in previous work for GCx2GC (*1*).

3.2 Sample-Varied Data Results

Chromatograms acquired from the three wine samples were used to test the alignment methods on sample-varied data. The benchmark RMSE for wine sample chromatographic alignment is established with pairs of consecutive replicate runs of the 2011, 2012, and 2013 vintages. The results for the 2011 sample replicate runs are shown in Figure 3. Training-set data and additional replicate runs can be found in Appendix B (Figure B3). The titles of the graphs indicate which two chromatograms were aligned with the year of the sample followed by an "R" and the run number (1 or 2). Figure 3 shows aligned chromatograms for runs 1 and 2 from the 2011 sample. As seen in the testing-set plots, none of the alignment methods are able to improve on the initial misalignment. This indicates there is no systematic retention-time difference between the replicate runs, only retention-time noise. In the ^{1}D (Figure 3a), the RMSE with no alignment is about 0.0368 min, which is the maximum for any of the replicate sample runs. In the ^{2}D (Figure 3b), the initial misalignment is about 0.0137 s. Over the three sets of replicate runs from 2011, 2012, and 2013, the average misalignment in the ^{1}D is 0.03037 min, and in the ^{2}D is 0.01725 s.

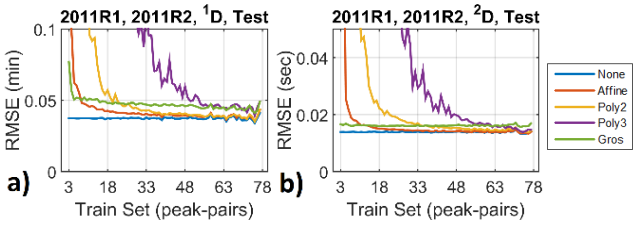


Figure 3. Cross-validation retention-time RMSE results as a function of training set size for consecutive replicate runs of the 2011 wine sample. The names correspond to the vintage year of the wine sample.

The average RMSE in the ^{1}D is less than the modulation sampling noise level of 0.034 min.[ii] The peaks detected in the wine chromatograms are very narrow, with a tpw of about 2 modulations, so the sampling noise must be considered. Therefore, 0.034 min is used as the ^{1}D benchmark RMSE for the alignment of the wine sample chromatograms. The average misalignment in the ^{2}D of 0.01725 s is the other benchmark for the alignment methods.

The chromatograms produced from the second run of each year's sample were tested in every pair-wise combination with the other two years. All these samples were run within a span of three days, so the initial misalignment between them is small, due mainly to run-to-run random variations and sample differences for the different vintages. The results from aligning chromatograms from the 2011 and 2012 samples are shown in Figure 4. The RMSE between the chromatograms without any alignment functions applied is only about 0.0344 min in the ^{1}D (Figure 4a) and 0.02 s in the ^{2}D (Figure 4b). Both these values are just above the benchmark inherent noise threshold in each dimension, suggesting that the alignment methods shouldn't be expected to improve much upon the initial misalignment. This is apparent in both the ^{1}D and ^{2}D testing-set plots which shows that none of the methods are able to improve the alignment more than a few thousandths of a minute and second, respectively. The minimum RMSE reached by Gros' algorithm in the ^{1}D is slightly worse than the initial misalignment.

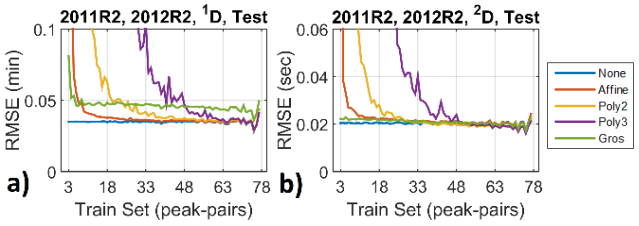


Figure 4. Cross-validation retention-time RMSE results as a function of training set size for alignment of two different wine sample chromatograms. The names correspond to the vintage year of the wine sample.

[ii] The wine analyses have a modulation cycle of 7 s or $P_M = 0.117$ min, so the RMS retention-time noise level from the sampling effect of modulation is 0.034 min.

A table of results and graphs for all other cross-validation experiments can be found in Appendix B (Table B1, Figures B4 and B5). On average, the initial misalignment in both chromatographic dimensions is close to the benchmark values and, as a result, none of the alignment methods achieve notable improvements. The third-degree polynomial even averages a slightly greater minimum value than the initial misalignment in the ^{1}D. These data then suggest that no method, global nor local, is able to perform well. If two chromatograms have only a small initial misalignment, it may be better not to perform any alignment operation at all.

3.3 Instrument-Varied Data Results

Chromatograms acquired from the single cocoa sample were used to test the alignment methods on data obtained with differing instruments. The benchmark RMSE values for the cocoa chromatogram alignments are established using two replicate sample runs with the flow modulator and three replicate runs with the thermal modulator. The results of the second replicate cross-validation experiment with the thermal-modulator chromatograms are shown in Figure 5. Training-set data and additional replicate runs can be found in Appendix B (Figure B6). As expected, only negligible improvements in alignment are seen from any method in either chromatographic dimension. In the ^{1}D (Figure 5a), the average initial misalignment for this experiment is about 0.0438 min which is the maximum seen in any of the replicate experiments. In the ^{2}D (Figure 5b), the initial misalignment is about 0.026 s. Across all three replicate sample run experiments, the average misalignment in

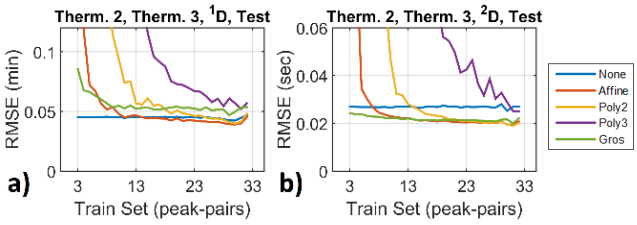


Figure 5. Cross-validation retention-time RMSE results as a function of training set size for consecutive replicate runs of a cocoa sample using a thermal **modulator**.

the ^{1}D is 0.0412 min, which is used as the benchmark. The modulation sampling noise level for these chromatograms[iii] does not greatly affect the benchmark because the peaks detected, with a tpw of about 5 modulations, are wider than those seen from the diesel and wine samples. The average ^{2}D misalignment is 0.0257 s, which is used as the benchmark.

Pairs of chromatograms, one from the two flow-modulator runs and one from the three thermal-modulator runs, were tested in every combination, totaling six experiments. The chromatograms in each experiment were acquired with two different modulators, so the initial misalignment is severe, especially in the ^{1}D because of the constraints posed by the differential flow modulation dynamics to carrier gas volumetric flow. The results from aligning the second flow-modulator chromatogram to the first thermal-modulator chromatogram are shown in Figure 6. The initial misalignment in the ^{1}D (the blue line) is excluded from plot (Figure 6a) because it is so large. In the ^{2}D, the initial misalignment hovers around 0.5 s, and is also excluded from plot (Figure 6b).

In Figure 6, every method offers significant improvement in both dimensions. In the ^{1}D, the affine transformation function reaches the lowest error of 0.488 min (percent improvement $I_p = 98.1\%$), just in front of Gros' algorithm at 0.503 min ($I_p = 98.0\%$). The second and third-degree polynomial transformations are about the same at 0.537 and 0.527 min ($I_p = 97.9\%$), respectively. The percent improvement

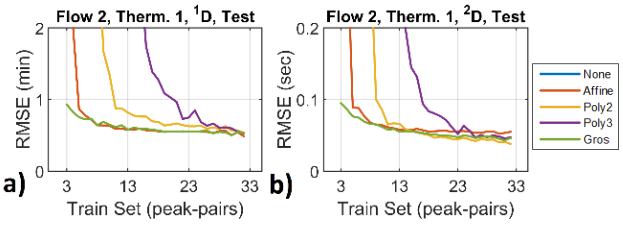


Figure 6. Cross-validation retention-time RMSE results as a function of training set size for chromatograms produced from the same cocoa sample but using two different modulation platforms.

[iii] The cocoa analyses have a modulation cycle of 3 s or $P_M = 0.05$ min, so the RMS noise level from the sampling effect of modulation is 0.0144 min.

from every method is good, even though the benchmark of 0.0412 min is not achieved (perhaps because the initial misalignment is so large).

That the affine transformation performs best suggests that the higher-degree polynomials were not fit with enough alignment points to reach peak performance. Similar to the diesel results, with fewer than 10 alignment points the local method outperforms the global methods in terms of RMSE. Around 10 peak-pairs, though, the affine transformation surpasses Gros' algorithm for a slight performance gain. Because the total number of corresponding peaks across the cocoa chromatograms is only 33, significantly fewer than for the diesel or wine chromatograms, the second and third-degree polynomials do not reach a lower RMSE than the affine transformation or local method. With training sets around 30 peak-pairs, though, they do approach these performances. Figure 6 is a good example of the potential advantages to using Gros' algorithm or the affine transformation when few alignment points are available.

In the ^{2}D, the second-degree polynomial reaches the lowest RMSE of 0.038 s ($I_p = 97.4\%$), followed by Gros' algorithm at 0.043 s ($I_p = 96.3\%$), the third-degree polynomial at 0.046 s ($I_p = 95.7\%$), and the affine transformation at 0.052 s ($I_p = 94.3\%$). Again, every alignment method attains a high percent improvement. The peak RMSE from the second-degree polynomial (0.038 s) also approaches the benchmark set at 0.0257 s. In the ^{2}D, the second-degree polynomial converges to its peak performance with fewer peak-pairs than in the ^{1}D, allowing it to surpass performance of the affine transformation and Gros' algorithm with about 15 alignment points. In terms of percent improvement, this performance gain is small. The third-degree polynomial does not have enough alignment points to be well fit, causing a slightly worse performance than both the second-degree polynomial and local method.

Minimum Testing-Set RMSE Reached by Alignment Methods in the ^{1}D (min) and ^{2}D (s) for Cocoa Chromatograms										
Chromat-ograms	**None (Avg.)**		**Affine**		**Poly2**		**Poly3**		**Gros et al.**	
	^{1}D	**^{2}D**	**^{1}D**	**^{2}D**	**^{1}D**	**^{2}D**	**^{1}D**	**^{2}D**	**^{1}D**	**^{2}D**
Flow 1-Thermal 1	23.2438	0.5204	0.4897	0.0491	0.5417	0.0332	0.5385	0.0358	0.5159	0.0408
Flow 1-Thermal 2	23.2396	0.5094	0.4822	0.0383	0.5265	0.0265	0.5159	0.0324	0.5268	0.0321
Flow 1-Thermal 3	23.2261	0.5271	0.4783	0.0455	0.5352	0.0273	0.5378	0.0362	0.5195	0.0378
Flow 2-Thermal 1	23.2566	0.4952	0.4879	0.0524	0.5367	0.0379	0.5274	0.0457	0.5025	0.0431
Flow 2-Thermal 2	23.2523	0.4842	0.4801	0.0420	0.5226	0.0316	0.5102	0.0427	0.5139	0.0355
Flow 2-Thermal 3	23.2389	0.5018	0.4757	0.0481	0.5304	0.0308	0.5276	0.0450	0.5057	0.0404
Average	23.2429	0.5064	0.4823	0.0459	0.5322	0.0312	0.5262	0.0396	0.5141	0.0383
Average % Improvement			98.1	95.8	97.9	98.8	97.9	97.1	98.0	97.4

Table 2. Minimum testing set RMSE reached by each alignment method in both the first and second chromatographic dimensions for all six experiments run with the chromatograms from the cocoa sample. The "None" columns are the average initial misalignments, not the minimum. All methods perform well as indicated by the high percent improvements.

Table 2 shows a summary of the results from all six cross-validation experiments. Graphs from the other experiments are in Appendix B (Figure B7). The average case performance of the global and local methods closely mirrors the performances discussed with Figure 6. Although a global function was able to, on average, outperform the local method (affine in ^{1}D and second-degree polynomial in ^{2}D), the performance gain is minimal in terms of percent improvement. All methods perform well, averaging a percent improvement over 95%. In line with conclusions from the diesel alignment results, it may be preferable to use Gros' algorithm if very few alignment points are available, affine transformation when more than a few alignment points are available, and polynomial transformation when 30 or more alignment points are available.

4. CONCLUSIONS

This work indicates that low-degree polynomial transformation functions will, on average, outperform the local alignment method developed by Gros et al., if given a sufficient number of alignment points for a good fit. Looking at cross-validation tests run on diesel chromatograms, which were acquired at varying times, the global methods consistently achieve a lower peak RMSE than the local method. The cross-validation experiments run with the cocoa sample chromatograms, acquired with differing instrument configurations, support this conclusion, although the local algorithm still averaged a percent improvement of over 97% in both dimensions. In general, although the third-degree polynomial transformation consistently reaches the lowest minimum RMSE with sufficient fitting (requiring about 55 alignment points), the performance gain over the second-degree polynomial is not significant and may not be worth the extra computational cost.

The tests run on GCxGC chromatograms acquired from varying wine samples indicate that no alignment method, global or local, is able to significantly improve alignment when initial misalignment is close to the retention-times noise level. The third-degree polynomial and local method actually made the alignment slightly worse in several cases, suggesting that when misalignment is very small, it may be better not to apply any alignment operation.

This research suggests that for the purpose of chromatographic alignment between two GCxGC chromatograms, it may be preferable to use global, low-degree transformation functions such as second-degree polynomials rather than local methods when a sufficient number of alignment points are available. These global transformations show a better average performance and incur less computational overhead. However, if working with fewer than 10 alignment points, it may be better to use Gros' algorithm. In order to outperform Gros' algorithm, the

affine transformation needed as many as 10 alignment points and the second-degree polynomial needed around 30 points.

The training set size at which the alignment methods reach their peak performance may be affected by how the alignment points are chosen. In the experiments presented here, these peak-pairs were chosen randomly from a large, well-distributed set, but choosing a subset that is better distributed across the range of retention times in a chromatogram may reduce the training set size required to approach peak performance. For the global methods, more distributed alignment points would allow the systemic misalignment trends to be modeled with fewer points. Although this would also help the local method, it is already approaching peak performance with very few points in most cases, suggesting that better distributed alignment points might reduce the set size at which the performance of the global methods overtakes the local method.

A final consideration is the problem of incorrect alignment points. With a local method, the associated error is localized but larger; whereas with a global method the associated error is smaller but global. If alignment point errors are possible, a global method with many alignment points to regularize the fit may be preferred.

REFERENCES

1. Reichenbach, S. E.; Rempe, D. W.; Tao, Q.; Bressanello, D.; Liberto, E.; Bicchi, C.; Balducci, S.; Cordero, C. *Anal. Chem.* **2015,** *87*, 10056-10063.
2. Gros, J.; Nabi, D.; Dimitriou, P.; Rutler, R.; Arey, J. *Anal. Chem.* **2012,** *84*, 9033-9040.
3. Seeley, J.; Kramp, F.; Sharpe, K. *J. Sep. Sci.* **2001,** *24*, 444-450.
4. Seeley, J.; Kramp, F.; Sharpe, K.; Seeley, S. *J. Sep. Sci.* **2002,** *25*, 53-59.
5. Nicolotti, L.; Cordero, C.; Bressanello, D.; Cagliero, C.; Liberto, E.; Magagna, F.; Rubiolo, P.; Sgorbini, B.; Bicchi, C. *J. Chromatogr., A* **2014,** *1360*, 264-275.
6. Bressanello, D.; Liberto, E.; Collino, M.; Reichenbach, S.; Benetti, E.; Chiazza, F.; Bicchi, C.; Cordero, C. *J. Chromatogr., A* **2014,** *1361*, 265-276.
7. Pierce, K. M.; Wood, L. F.; Wright, Bob W.; Synovec, R. E. *Anal. Chem.* ***2005,*** *77*, 7735-7743.
8. Zhang, D.; Huang, X.; Regnier, F. E.; Zhang, M. *Anal. Chem.* **2008**, *80*, 2664−2671.
9. Nielsen, N.-P. V.; Carstensen, J. M.; Smedsgaard, J. *J. Chromatogr., A* **1998**, *805*, 17−35.
10. Cordero C, Rubiolo P, Cobelli L, Stani G, Miliazza A, Giardina M, Firor R, Bicchi C. *J. Chromatogr., A.* **2015**, *1417*, 79-95
11. UOP LLC. *UOP 990-11: Organic Analysis of Distillate by Comprehensive Two-Dimensional Gas Chromatography with Flame Ionization Detection*; 2011.
12 Welke, J. E.; Zanus, M.; Lazarotto, M.; Schmitta, K. G.; Zini, C. A. *J. Braz. Chem. Soc.* **2012**, *23*, 678-687.
13. GC Image, LLC. *GC Image GCxGC Software;* GC Image, LLC: Lincoln NE, 2015.
14. Reichenbach, S. Data acquisition, visualization, and analysis. In *Comprehensive Two Dimensional Gas Chromatography;* Ramos, L., Ed.; Elsevier, 2009; pp 77-106.
15. Sibson, R. A Brief Description of Natural Neighbor Interpolation. In *Interpreting Multivariate Data;* Barnett, V., Ed.; John Wiley & Sons: New York, 1981; pp 21–36.
16. Gros, J.; Arey, J. *Documentation for the Gros-Arey code to align GCxGC chromatograms as implemented in MATLAB*; EPFL, Switzerland, November, 2014.
17. Holland Computing Center. HCC Documentation. https://hcc-docs.unl.edu/display/HCCDOC/HCC+Documentation (accessed April 12, 2016).

APPENDIX A. EXPERIMENTAL CONDITIONS

A.1 Instrumentation for Diesel Sample Runs

For the analysis of the diesel sample, all run conditions were in accordance with UOP 990, with a modulation period of 8 s and sampling with a flame ionization detector (FID) at 200 Hz. Diesel sample runs used a LECO GC x GC FID system equipped with an Agilent 6890 GC and LECO GC x GC accessories (modulator and secondary oven).

A.2 Instrumentation for Wine Sample Runs

Wine samples (750 mL each) were protected from direct light and stored in a cool place. After opening the bottles, smaller volumes of each wine were placed in 200 mL screw-capped dark glass flasks and were frozen (-18°C) in order to avoid loss of volatiles until chromatographic analyses. Headspace microextraction (HS-SPME) was performed with one mL of wine, 0.3 g of sodium chloride at 55°C (± 0.9), and a DVB/CAR/PDMS fiber (Supelco, Bellefonte, PA) in 20 mL headspace screw-capped glass vials. SPME fibers were previously conditioned according to manufacturer´s instructions. The system employed for GC × GC was an Agilent 6890N (Agilent Technologies, Palo Alto, CA) with a time-of-flight mass spectrometric detector (TOFMS) equipped with a CombiPAL autosampler (CTC Analytics, Zwingen, Switzerland), a secondary oven for the second chromatographic column, and a quadjet cryogenic modulator (two cold and two hot) where cold jets were supported by nitrogen gas cooled with liquid nitrogen. Desorption took place at 250°C, in the injection port, where the fiber was kept for five (5) minutes. Other parameters employed were: modulation period of 7 s, oven temperature offset of 10°C, transfer line temperature of 300°C, detector temperature 240°C, ionization energy of 70 eV, detector of voltage 1500 V, mass range 45 to 450 m/z, and data

acquisition rate of 100 Hz. Carrier gas was helium (purity 5.0, White Martins, Pinhais, Brazil) and its linear velocity was 1.0 mL min-1. Stationary phase of the first dimension column (^{1}D) was a DB-WAX (30 m × 0.25 mm × 0.25 μm) and a DB-17ms (1.70 m × 0.18 mm × 0.18 μm) in the second dimension (^{2}D).

A.3 Instrumentation for Cocoa Sample Runs

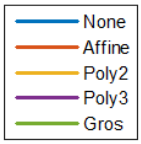
For the analysis of the volatile fraction of cocoa samples, headspace solid-phase micro-extraction (HS-SPME) was performed on 1.00 g of cocoa nibs finely milled with liquid nitrogen at 45°C (± 0.9) for 40 minutes. A DVB/CAR/PDMS fiber (Supelco, Bellefonte, PA) was used in 20 mL headspace screw-capped glass vials. The GC x 2GC-MS/FID runs with reverse-inject differential flow modulation used an Agilent 7890B GC unit coupled to an Agilent 5977A fast quadrupole MS detector (Agilent, Little Falls, DE) operating in EI mode at 70 eV, and a fast FID. The GC transfer line was set at 270°C. A scan range of 40-240 m/z with a scanning rate of 20,000 amu/s was used, and the spectra generation frequency was 35 Hz. The FID base temperature was 280°C, with H_2 flow of 40 mL/min, air flow of 240 mL/min, and make-up (N_2) of 450 mL/min, at a sampling frequency of 150 Hz. The system was equipped with reverse-inject differential flow consisting of one CFT plate connected to a three-way solenoid valve that receives a controlled supply of carrier gas (helium) from an auxiliary electronic pressure control module (EPC). Pulse time was set at 200 ms and modulation period of 3 s. The ^{1}D used a SolGel-Wax column (100% polyethylene glycol)(30 m × 0.25 mm d_c, 0.25 μm d_f) from SGE Analytical Science (Ringwood, Australia) coupled with a ^{2}D OV1701 column (86% polydimethylsiloxane, 7% phenyl, 7% cyanopropyl) (5 m × 0.25 mm d_c, 0.25 μm d_f) from Mega (Legnano, Milan, Italy). Cocoa volatiles extracted by HS-SPME were thermally desorbed into

the GC split/splitless injector port in split mode, with split ratio 1:20, and injector temperature 250°C. The carrier gas was helium at a constant flow of 0.3 mL/min in the ^{1}D and 20 mL/min in the ^{2}D. The temperature program went from 50°C (0.5 min) to 250°C at 2°C/min (5 min). Connection between the ^{2}D column and the two parallel detectors was by a three-way unpurged splitter (G3181B, Agilent, Little Falls, DE). The deactivated capillary to the MS detector was 0.17 m long with 0.1 mm d_c, and to the FID detector was 1.3 m long with 0.45 mm d_c. Split ratio was 25:75 (MS:FID).

The GC x GC-MS runs with thermal modulation used an Agilent 6890 unit coupled to an Agilent 5975C MS detector (Agilent, Little Falls, DE) operating in EI mode at 70eV. The GC transfer line was set at 270°C with scan range 40-240 m/z and a scanning rate of 12,500 amu/s. The spectra generation frequency was 29 Hz. The system was equipped with a two-stage KT 2004 loop-type thermal modulator (Zoex Corporation, Houston, TX) cooled with liquid nitrogen. The hot jet pulse time was set at 250 ms and used a modulation period of 3 s. The fused silica capillary loop dimensions were 1.0 x 0.1 mm (inner diameter). The ^{1}D used a SolGel-Wax column (100% polyethylene glycol)(30 m × 0.25 mm d_c, 0.25 μm d_f) from SGE Analytical Science (Ringwood, Australia) coupled with a ^{2}D OV1701 column (86% polydimethylsiloxane, 7% phenyl, 7% cyanopropyl) (1 m × 0.1 mm d_c c, 0.10 μm d_f) from Mega (Legnano, Milan, Italy). Cocoa volatiles extracted by HS-SPME were thermally desorbed into the GC split/splitless injector port in split mode, with split ratio 1:20, and injector temperature 250°C. The carrier gas was helium at a constant flow of 1.8 mL/min. Temperature program was from 40°C (1 min) to 200°C at 3°C/min and to 250°C at 10°C/min (5 min).

APPENDIX B. ADDITIONAL RESULTS

All figures in this appendix use the same legend as the main body of the paper, shown below.



B.1 Additional Results for Time-Varied Data

Figure B1 shows the results for the alignment of two additional pairs of consecutive replicate diesel sample runs, along with additional training-set plots for the chromatogram pair presented in the paper. The misalignment between consecutive replicate runs indicates a benchmark for the lower bound of alignment performance due to systemic noise. Chromatogram pair 18 and 19 were discussed in the paper. For consecutive replicate diesel sample runs 17 and 18, the ^{1}D misalignment is about 0.0176 min, and the ^{2}D is about 0.0156 s. For replicate runs 19 and 20 the ^{1}D misalignment averages 0.0177 min, and the ^{2}D is about 0.0089 s. These ^{1}D values are less than the modulator sampling noise level for the diesel sample chromatograms (calculated in the paper). The ^{2}D misalignments are in line with the benchmark used in the paper.

Figure B2 shows the performance of the global and local algorithms for the alignment of all six pairings of diesel sample chromatograms acquired over various periods of time. (Testing-set figures for pairing 012011 and 061413 are presented in the paper.) In each test, every method offers significant improvements in alignment for both chromatographic dimensions. In the ^{1}D, the initial misalignment of the chromatogram pairs ranges from about 0.07 min to over 0.83 min, many times greater than the benchmark. The third-order polynomial tends to reach around 0.06

min whereas the affine and second-order polynomial reach just under 0.07 min. The local algorithm averages just under 0.08 min.

In the ^{2}D, the initial misalignment ranges from 0.06 s to about 0.44 s. The third-order polynomial reaches the lowest RMSE in all but one of the results from Figure S4, averaging about 0.02 s. The second-order polynomial is about the same. The affine and local methods still improve the initial misalignment, but only get to about 0.03 and 0.04 s, respectively. These results are consistent with those presented in the paper.

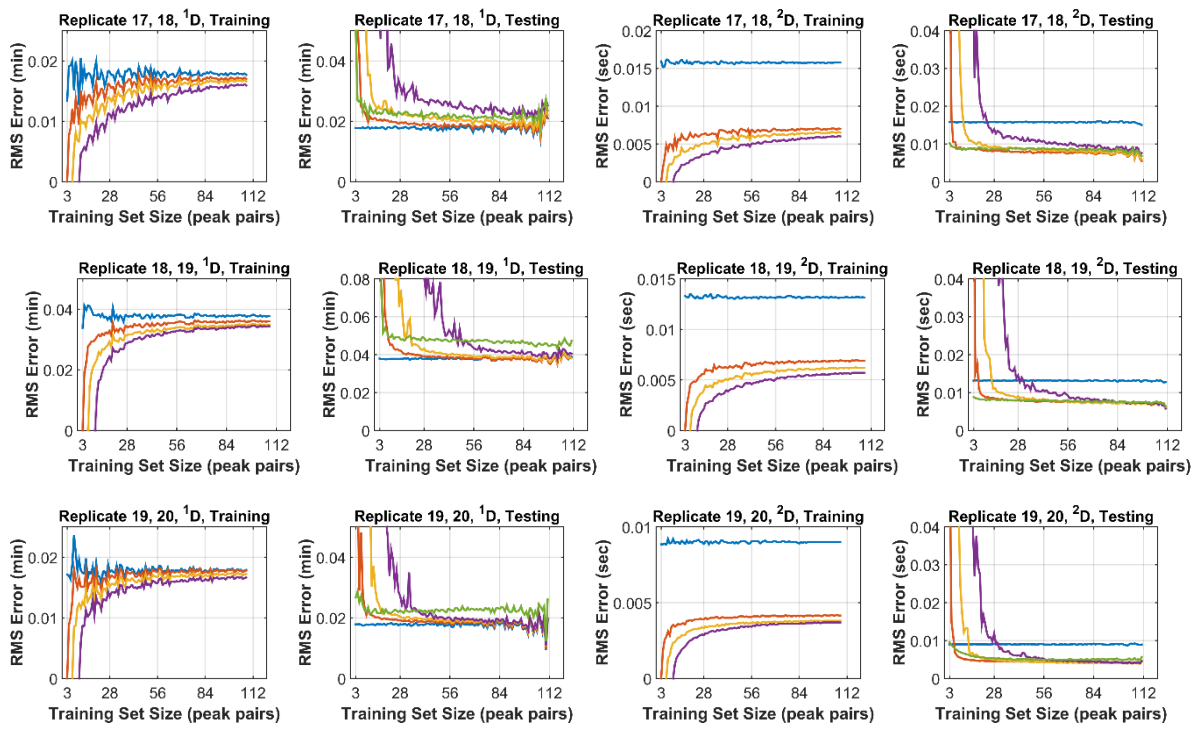


Figure B1. Cross-validation retention-time RMSE results as a function of training set size for consecutive replicate runs of a diesel sample. From left to right, the RMSE is shown for the ^{1}D with the training set, ^{1}D with the testing set, ^{2}D with the training set, and ^{2}D with the testing set. The performance of the local algorithm from Gros et al. is only shown in the testing plots because it is guaranteed to perfectly align the training set. The top row is for chromatograms from diesel runs #17 and #18, the middle row is for runs #18 and #19, and the bottom row is for runs #19 and #20.

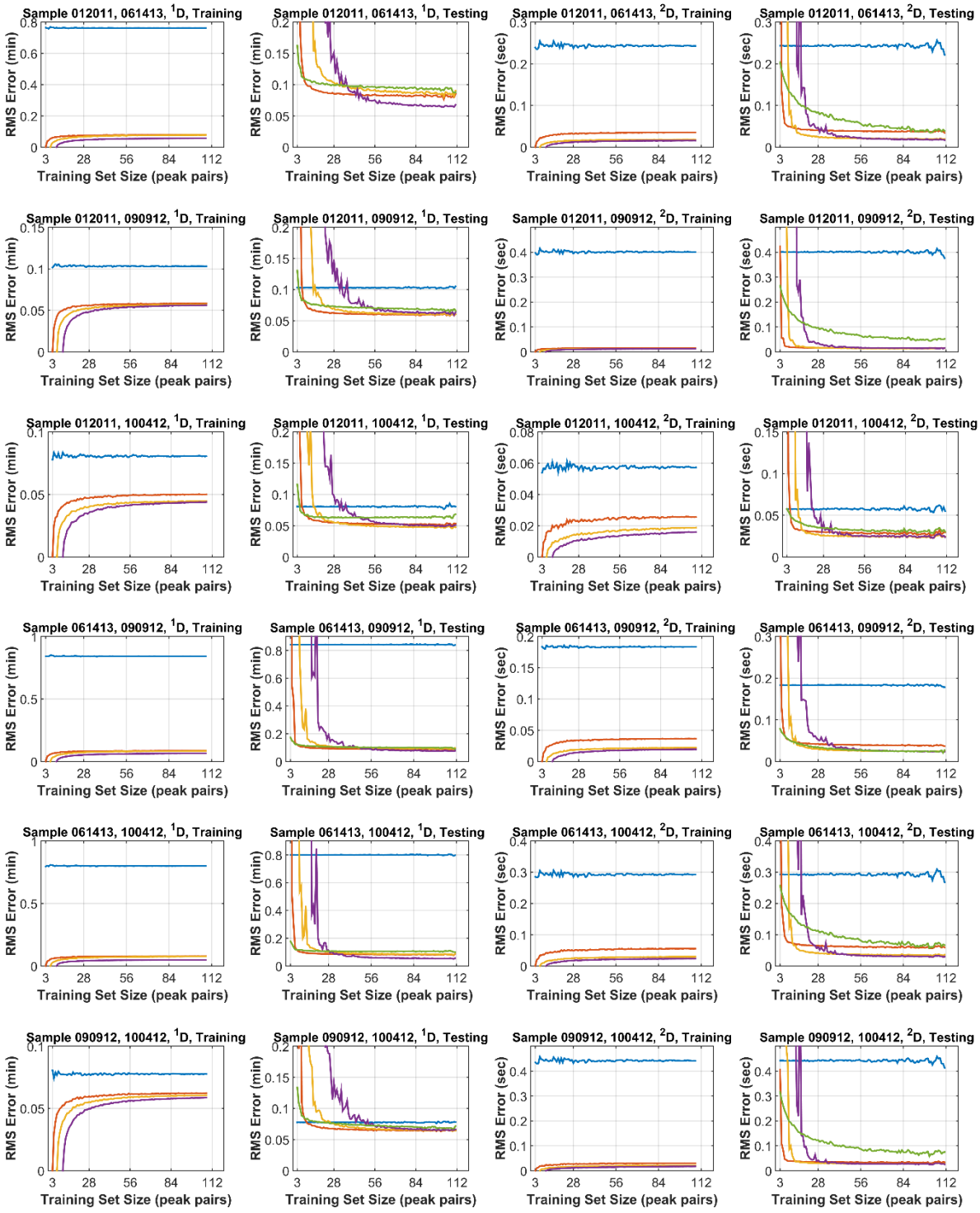


Figure B2. Cross-validation retention-time RMSE results as a function of training set size for chromatograms produced from the same diesel sample. From left to right, the RMSE is shown for the ^{1}D with the training set, ^{1}D with the testing set, ^{2}D with the training set, and ^{2}D with the testing set. The names of the samples correspond to the acquisition date (i.e. for the top row January 20, 2011 and September 9, 2012). Each row is for a different chromatogram pair.

B.2 Additional Results for Sample-Varied Data

Figure B3 shows the results for alignment of two additional pairs of consecutive replicate wine sample runs, along with additional training-set figures for the pair presented in the paper. The 2011 pair is discussed in the paper. For the ^{1}D, the benchmarks from both additional pairs are less than the modulation sampling noise level of 0.034 min, like the one presented in the paper. In the ^{2}D, the benchmark for both pairs is just over 0.015 s, right around the benchmark used in the paper of 0.01725 s.

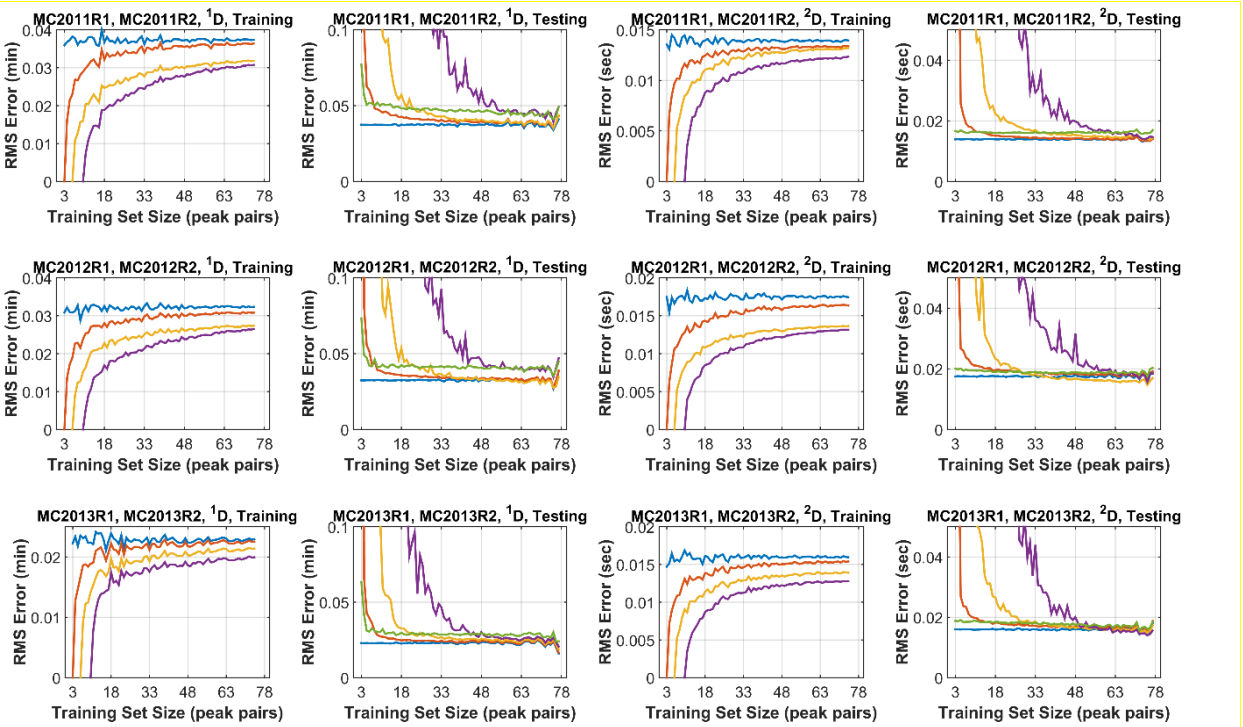


Figure B3. Cross-validation retention-time RMSE results as a function of training set size for consecutive replicate runs of the various wine samples. The names correspond to the vintage year of the wine sample. The top row is for chromatograms from vintage year 2011, runs #1 and #2. The middle row is for chromatograms from vintage year 2012, runs #1 and #2. The bottom row is for chromatograms from vintage year 2013, runs #1 and #2.

Figure B4 shows the cross-validation performance of the alignment methods for all three pairs of chromatograms from different wine samples run in a very short period of time. The

2011, 2012 pair is discussed in the paper. In the ^{1}D, the initial misalignments are just barely greater than the benchmark RMSE. Because of this, no method is able to improve on the initial misalignment in either test. The initial misalignment between pairs in the ^{2}D is around 0.02 s, also just above the benchmark value. So, there is little improvement on the alignment from any method. These results are in line with those found in the paper.

Table B1 summarizes the results of all three cross-validation experiments run on the wine chromatograms. It shows the minimum testing set RMSE reached for all alignment methods in both dimensions along with the average initial misalignment. The cells marked red are RMSE values greater than the average initial misalignment for that experiment.

Figure B5 visualizes the wine alignment results by plotting the minimum RMSE reached by each method against the average initial misalignment. The red dot-dashed line shows the RMSE benchmarks. The black dashed line shows the identity function – where the initial misalignment and minimum RMSE would be equal. A point above this line indicates that a method's resulting alignment is worse than it initially was, and one below offers an improvement. Each alignment method is represented by a different colored point. Between the two dimensions, several points from the third-order polynomial and local method slightly worsen the initial misalignment (as shown in data points above the dashed line). When the points fall under the identity function, it is not by much, showing negligible improvement on the wine chromatogram alignment. These data support the idea that if two chromatograms have only a small initial misalignment, it may be better not to perform any alignment operation at all.

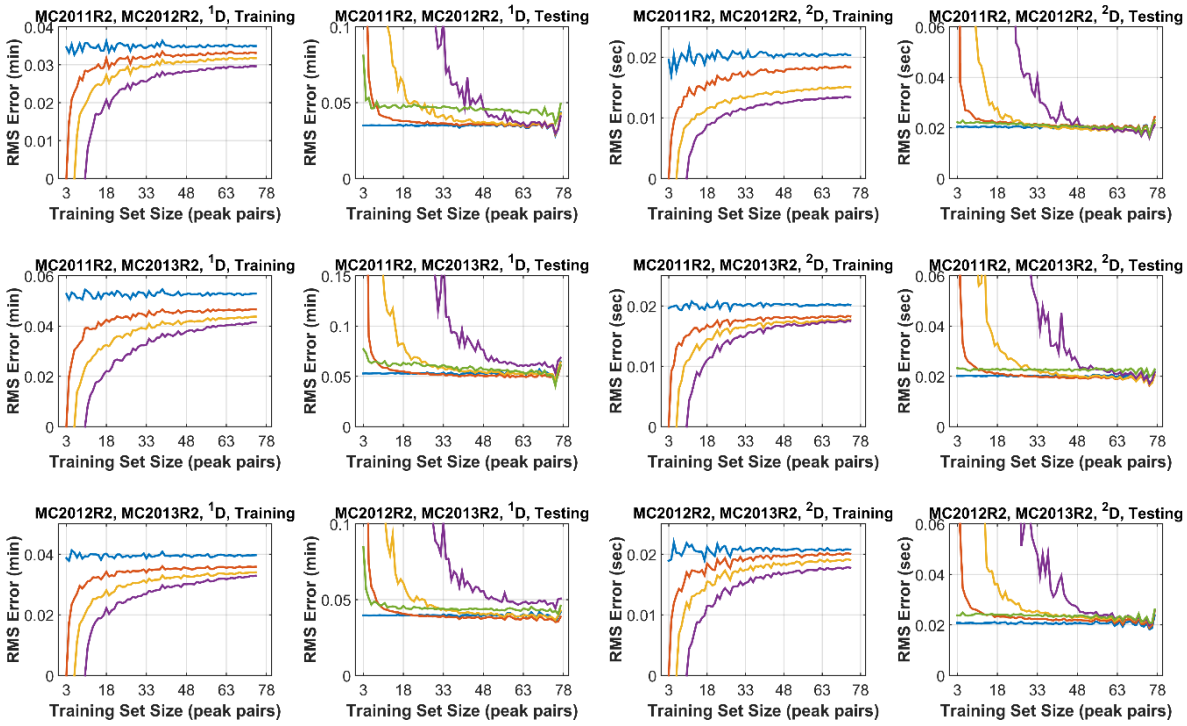


Figure B4. Cross-validation retention-time RMSE results as a function of training set size for alignment of two different wine sample chromatograms. The names correspond to the vintage year of the wine sample. For example, the top row is for chromatograms from the second runs of the 2011 and 2013 samples.

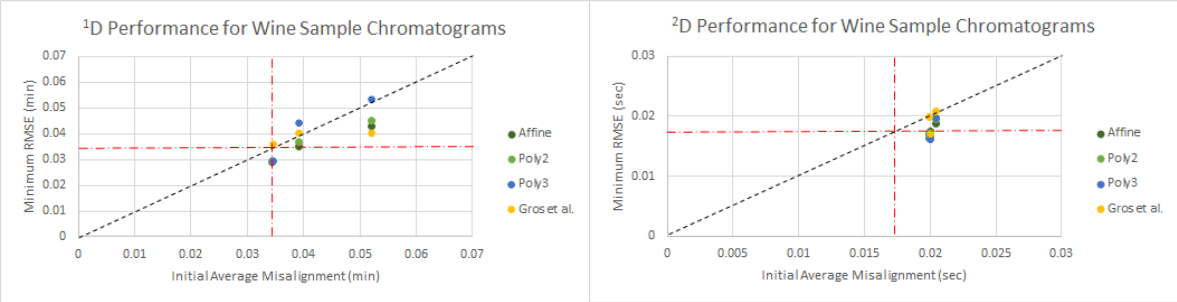


Figure B5. Minimum testing-set RMSE reached by the alignment methods on the wine sample chromatograms relative to the average initial misalignment. The red dot-dashed line shows the benchmark RMSE values (0.034 min and 0.01725 sec). The black dashed line shows the identity function – where the initial misalignment and minimum RMSE would be equal. A point above this line indicates that a method's resulting alignment is worse than it initially was, and one below offers an improvement.

Minimum RMSE Reached by Alignment Methods in the ^{1}D (min) and ^{2}D (sec) for Wine Chromatograms										
Chromat-ograms	**None (Avg.)**		**Affine**		**Poly2**		**Poly3**		**Gros et al.**	
	^{1}D	**^{2}D**	**^{1}D**	**^{2}D**	**^{1}D**	**^{2}D**	**^{1}D**	**^{2}D**	**^{1}D**	**^{2}D**
2011-2012	0.0344	0.0200	0.0289	0.0175	0.0294	0.0167	0.0297	0.0161	0.0362	0.0171
2011-2013	0.0520	0.0199	0.0427	0.0164	0.0452	0.0166	0.0533	0.0171	0.0406	0.0198
2012-2013	0.0391	0.0204	0.0352	0.0188	0.0367	0.0195	0.0442	0.0197	0.0405	0.0208
Average	0.0418	0.0201	0.0356	0.0176	0.0371	0.0176	0.0424	0.0176	0.0391	0.0192

Table B1. Minimum testing-set RMSE reached by each alignment method in both the first and second chromatographic dimensions for all three experiments run with the non-replicate chromatograms from the wine samples. The "None" columns are the average initial misalignments, not the minimum. The red boxes indicate where the initial misalignment was made worse by a method. On average, no method was able to improve upon the initial alignment significantly in either dimension.

B.3 Additional Results for Instrument-Varied Data

Figure B6 shows the results for the alignment of two additional pairs of consecutive replicate cocoa sample runs, along with additional training-set figures for the pair presented in the paper. The top row aligns replicate sample runs performed on a flow modulation platform, and the bottom two rows were performed on a thermal modulation platform. The Thermal 2, 3 pair is discussed in the paper. For the other pairs in the ^{1}D, the benchmarks are about 0.037 min and 0.043 min, respectively. These are consistent with the 0.0412 min benchmark used in the paper. For the ^{2}D, the average initial misalignment are about 0.03 s and 0.022 s, respectively, right around the paper benchmark of 0.0257 s. All methods did improve the alignment of the replicate flow-modulated chromatograms, indicating that there was a systematic misalignment between them. This is due to a small phase-roll affected by the alignment algorithms.

Figure B7 shows the cross-validation performance of the alignment methods on all six pairs of chromatograms acquired on different modulation platforms. The Flow 2, Thermal 1 pair is discussed in the paper, but additional training-set plots are presented here. All methods significantly improved alignment in both dimensions. In the ^{1}D, the initial misalignments are consistently around 23.2 min, well above the benchmark. Across these six experiments, the

affine transformation is able to reach about 0.48 min, the local algorithm from Gros et al. is about 0.51 min, and the second and third-order polynomials reach about 0.53 min. As observed in the paper, the higher-degree polynomials might require more peak-pairs for maximal performance.

In the ^{2}D, the initial misalignment is around 0.5 s. The second-order polynomial reaches a minimum RMSE of about 0.031 s on average, the third-order polynomial and local algorithm are both around 0.038 s, and the affine transformation averages around 0.046 s. These values are consistent with the example presented in the paper. All are effective, achieving between about 96% and 99% improvement.

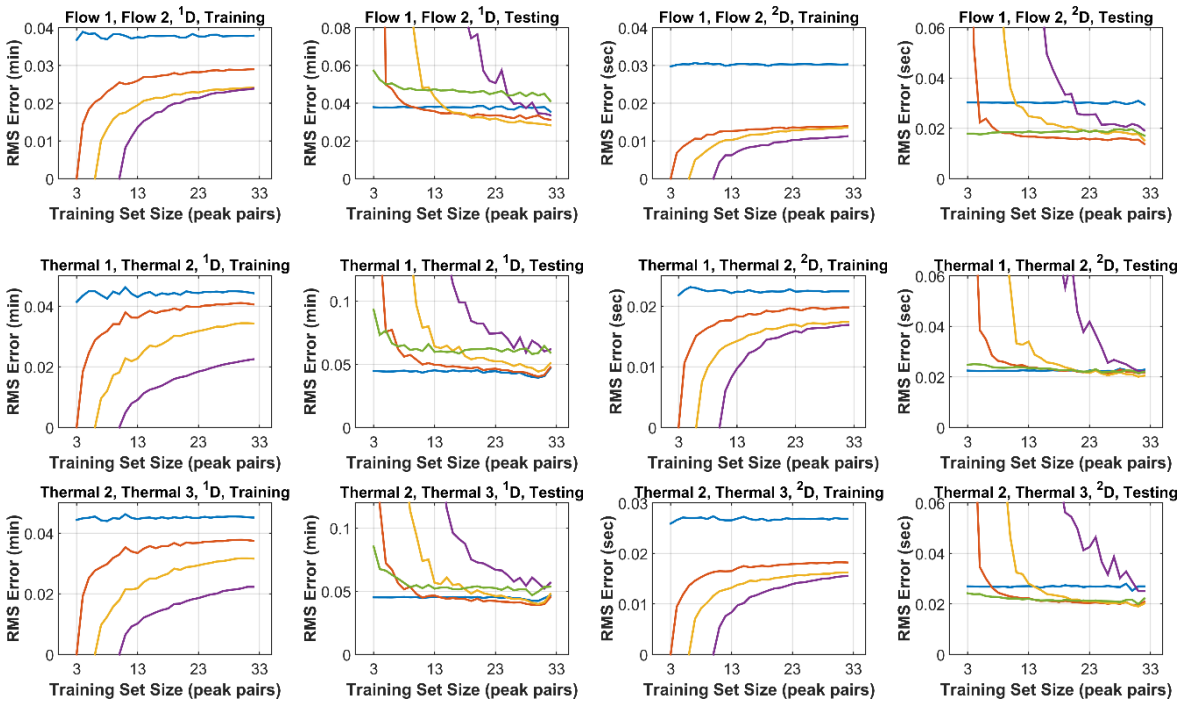


Figure B6. Cross-validation retention-time RMSE results as a function of training set size for consecutive replicate runs of a cocoa sample using different modulation platforms. The top row is for chromatograms from runs #1 and #2 using a flow modulator. The middle row is for chromatograms from runs #1 and #2 using a thermal modulator. The bottom row is for chromatograms from runs #2 and #3 using a thermal modulator.

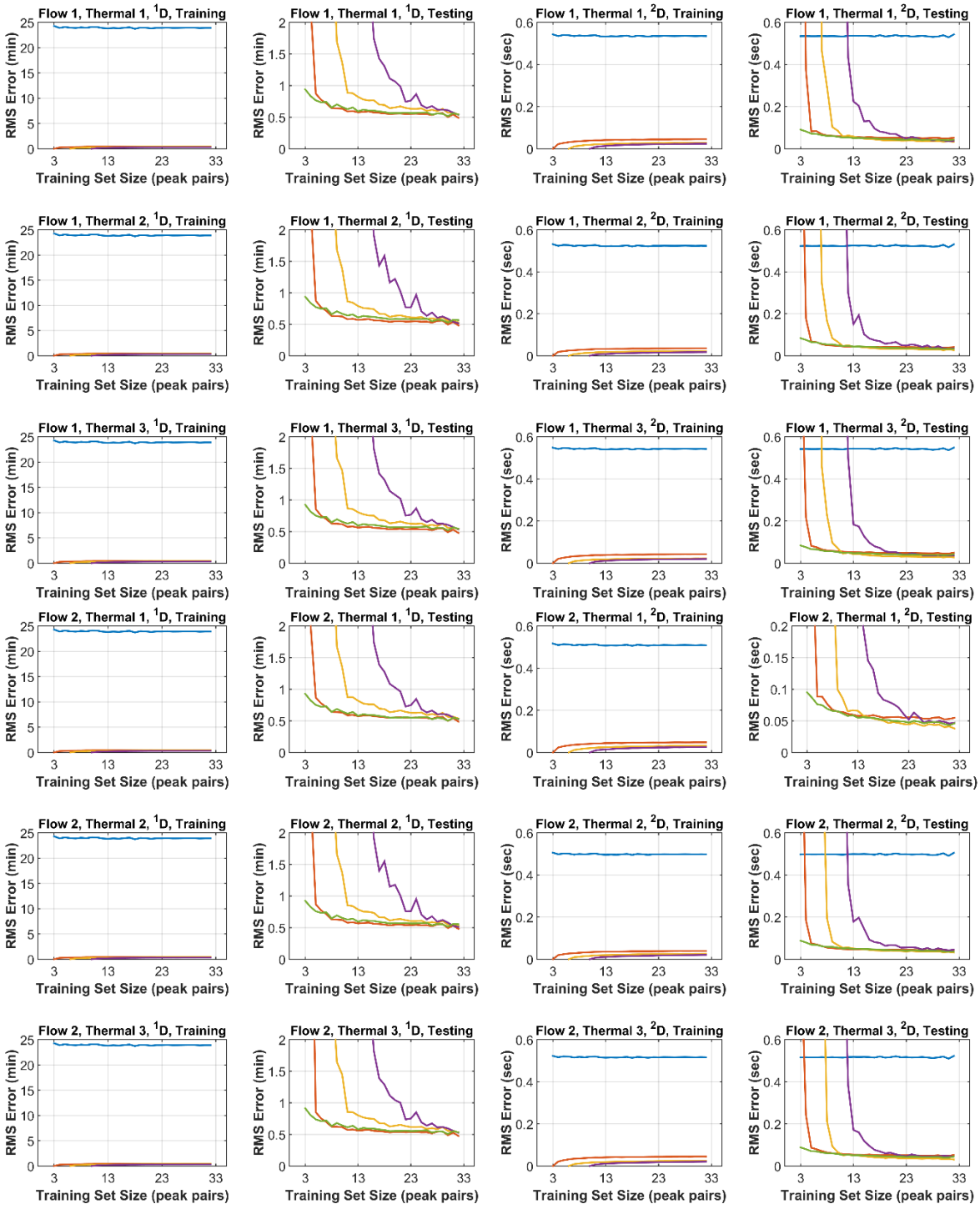


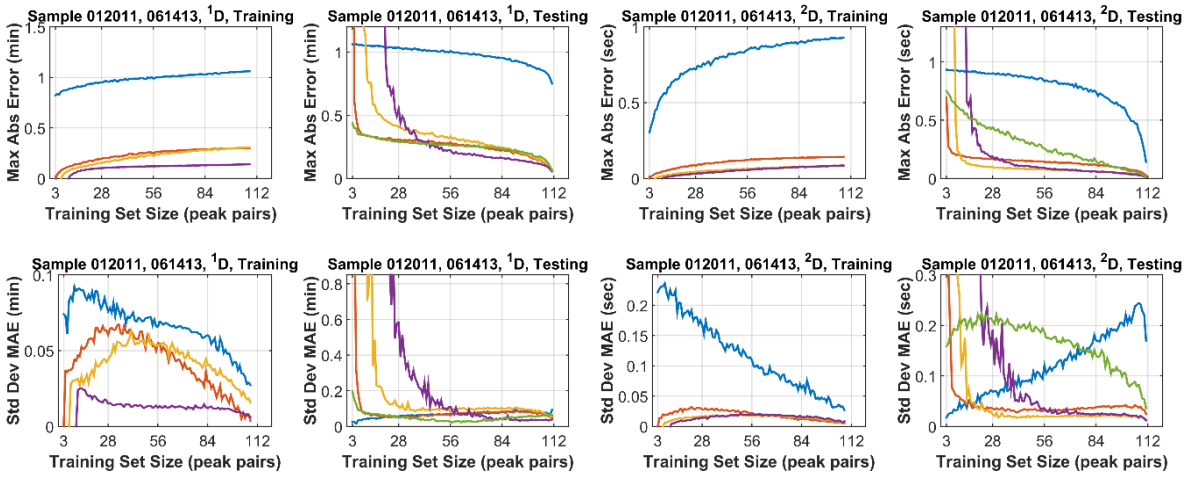
Figure B7. Cross-validation retention-time RMSE results as a function of training set size for chromatograms produced from the same cocoa sample but using two different modulation platforms. Each row is a pair of chromatograms from two different runs. For example the top row is for chromatograms from run #1 on the flow modulator, and run #1 on the thermal modulator.

B.4 Maximum Alignment Error

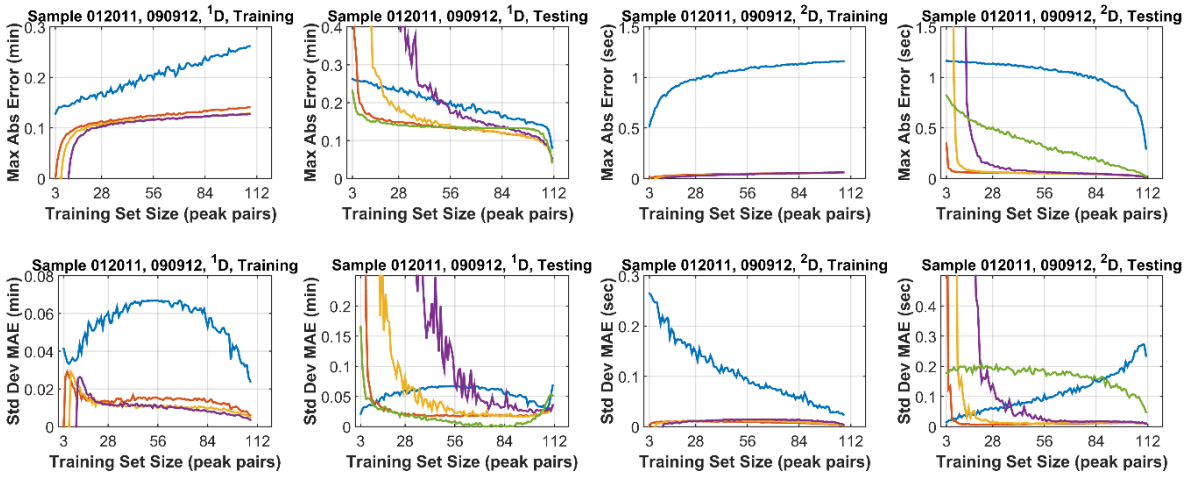
All figures presented so far have shown the root-mean-square-error (RMSE) with respect to retention times of matched peaks in a chromatogram pair. This metric indicates average-case performance of an alignment method, but the worst-case scenario must also be considered. Figures B8, B9, and B10 show the average maximum absolute alignment error (MAE) across all trials run for each training set size. The standard deviation of this MAE is also shown. Figure B8 is for the diesel runs, Figure B9 is for the wine runs, and Figure B10 is for the cocoa runs.

Across all experiments, behavior in both chromatographic dimensions is similar. If the training set has enough peak pairs, all methods reach a similar MAE for the testing set. The number of peak-pairs required to reach this convergent value differs between the different alignment methods. The local method from Gros et al. and the affine transformation require a much smaller training set than the second and third-degree polynomials in order to reach a lower MAE. This is consistent with the corresponding RMSE behavior. For the diesel runs in the ^{2}D, the local algorithm tends to have a higher standard deviation than the other methods, but this trend does not hold in the wine and cocoa experiments.

A. Runs 012011 and 061413.



B. Runs 012011 and 090912.



C. Runs 012011 and 100412.

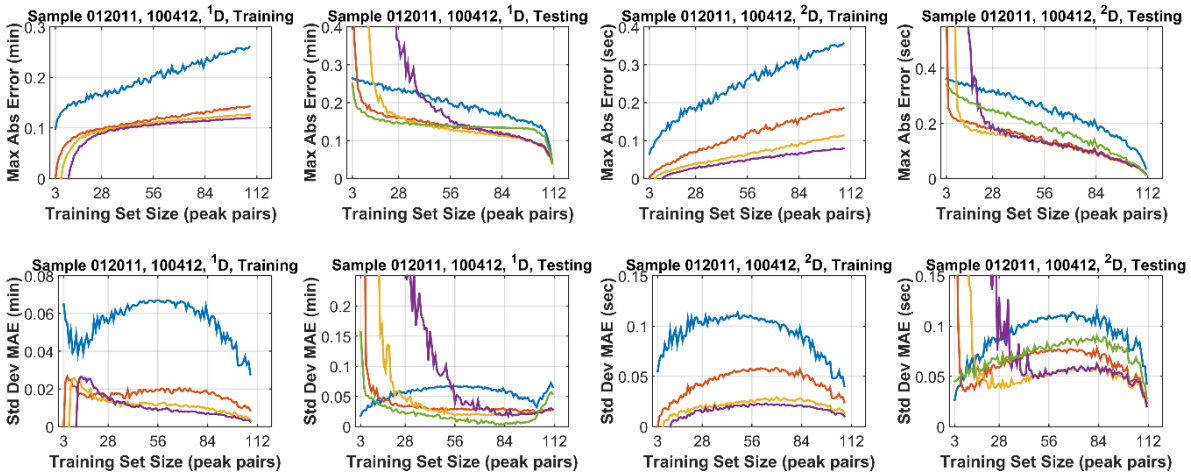
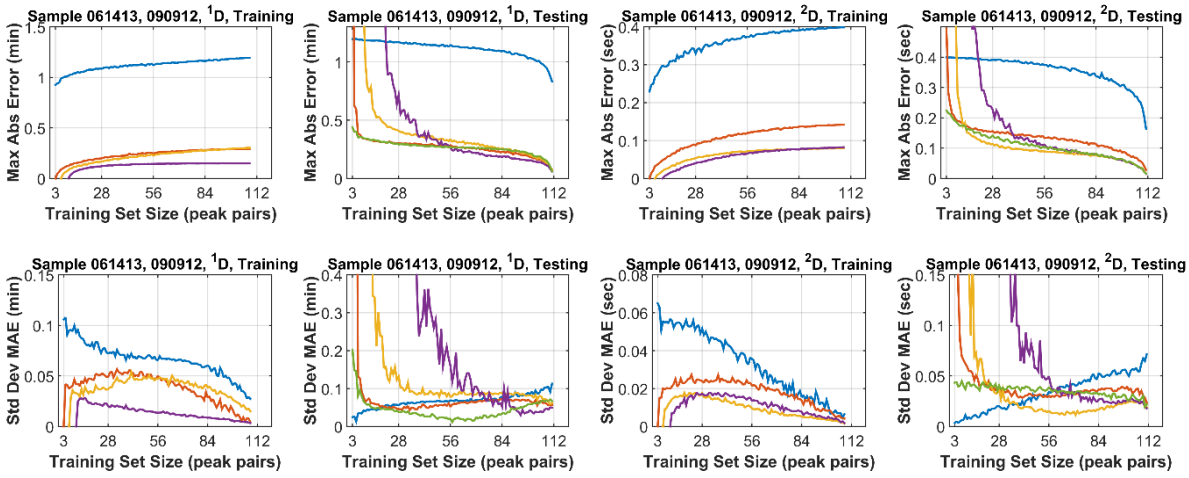
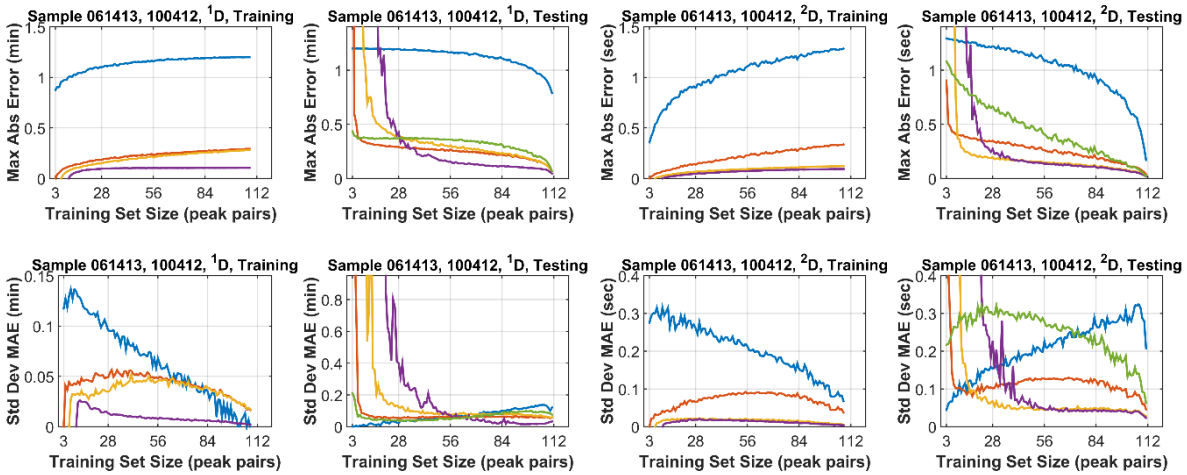


Figure S8. Maximum absolute error as a function of the training set size for diesel sample chromatograms. Sets of rows with maximum absolute error on the top of each set and the standard deviation of maximum absolute error on the bottom of each set are for: A. Runs 012011 and 061413, B. Runs 012011 and 090912, and C. Runs 012011 and 100412.

D. Runs 061413 and 090912.



E. Runs 061413 and 100412.



F. Runs 090912 and 100412.

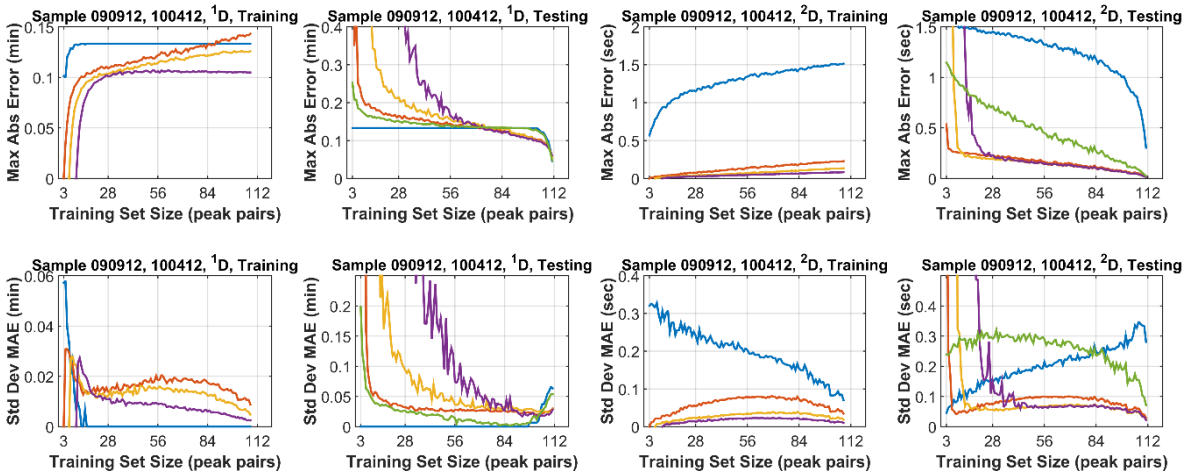
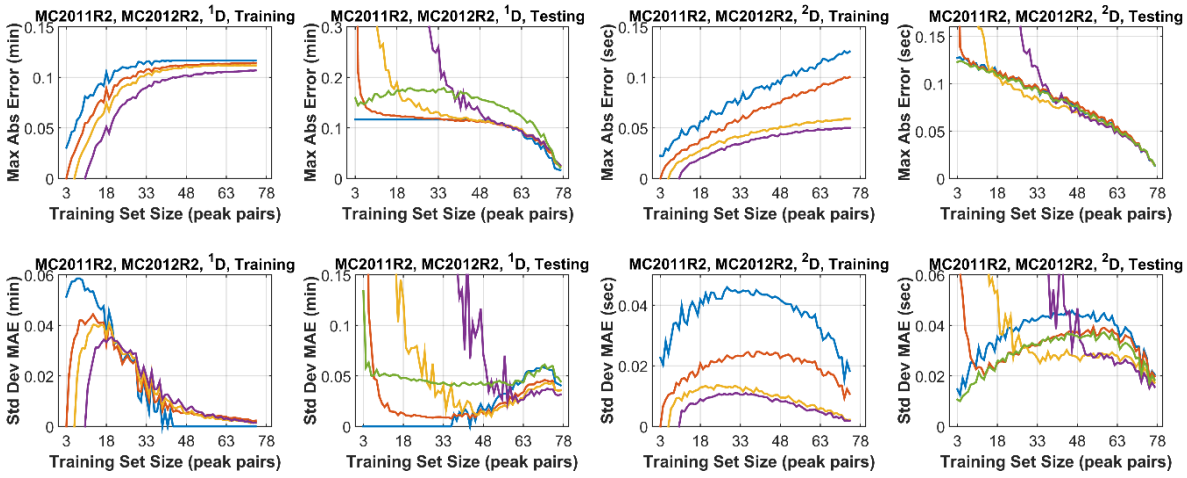
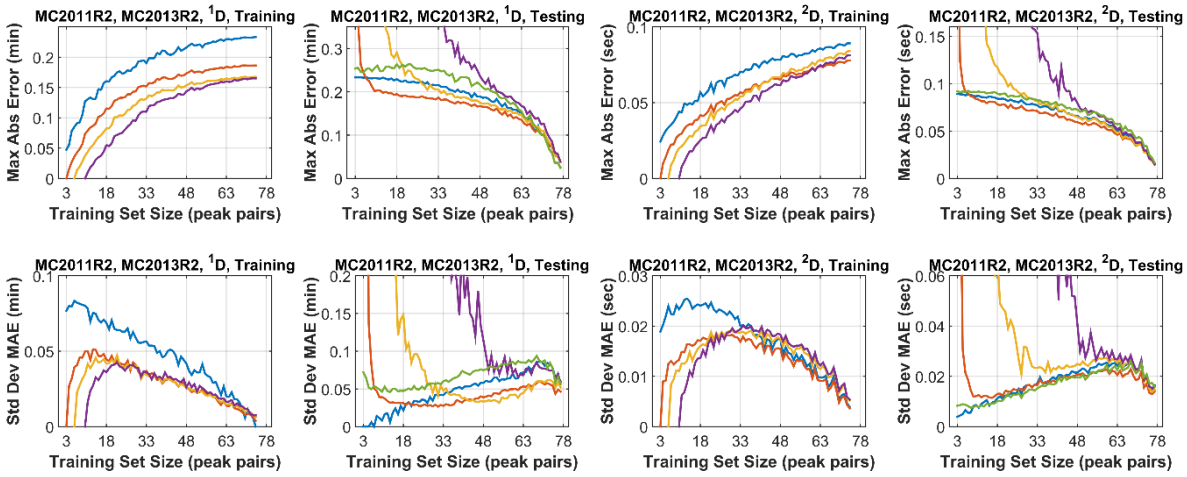


Figure B8 continued. D. Runs 061413 and 090912, E. Runs 061413 and 100412, and F. Runs 090912 and 100412.

A. Samples 2011, 2012, Runs #2.



B. Samples 2011, 2013, Runs #2.



C. Samples 2012, 2013, Runs #2.

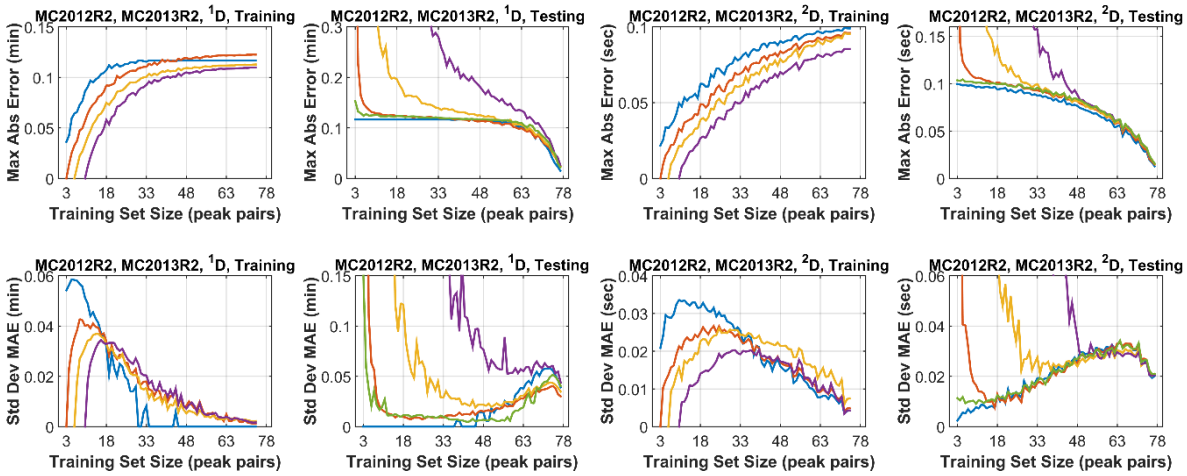
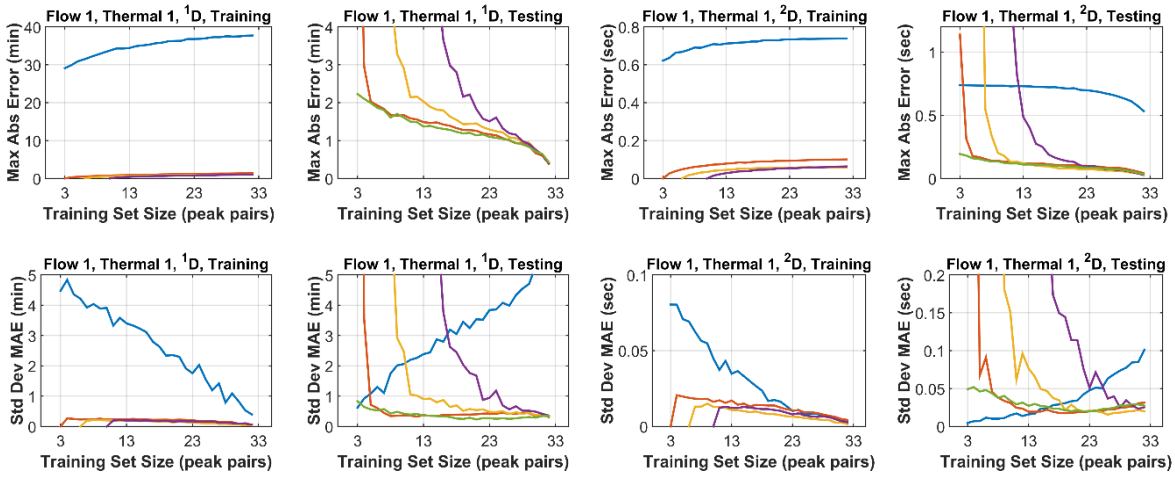
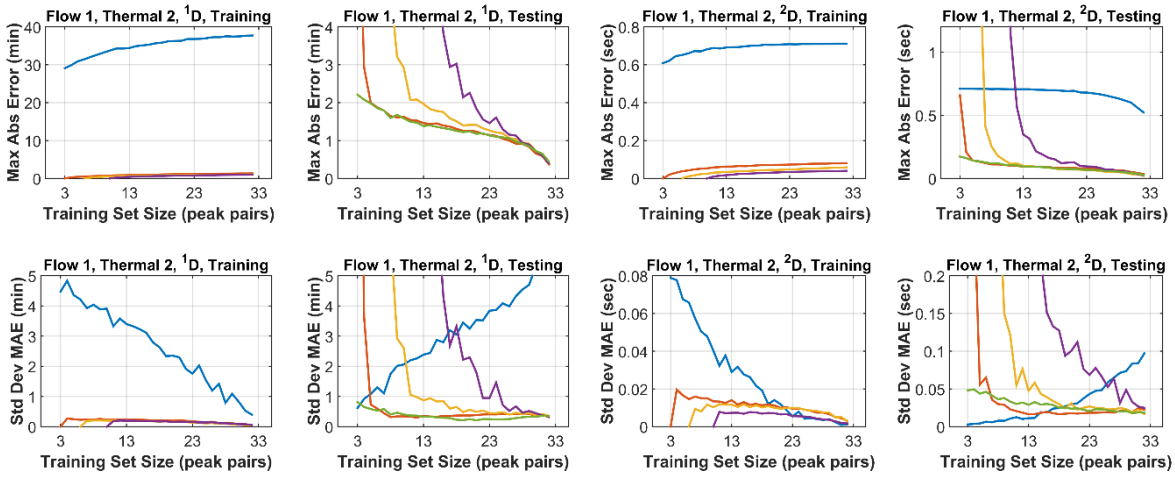


Figure B9. Maximum absolute error as a function of the training set size for wine sample Sets of rows with maximum absolute error on the top of each set and the standard deviation of maximum absolute error on the bottom of each set are for: A. Samples 2011, 2012, Runs #2, B. Samples 2011, 2012, Runs #2, and C. Samples 2012, 2013, Runs #2.

A. Runs #1 and #1.



B. Runs #1 and #2.



C. Runs #1 and #3.

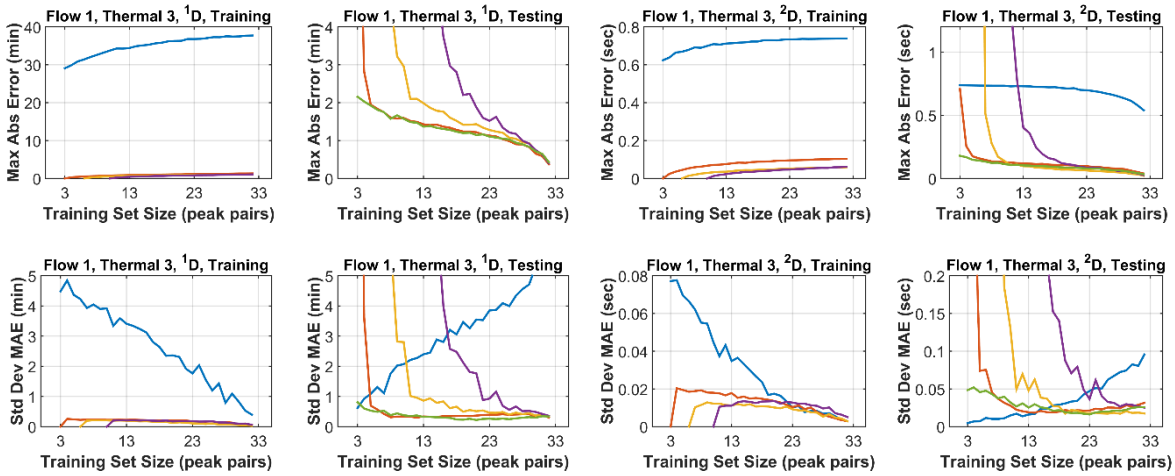
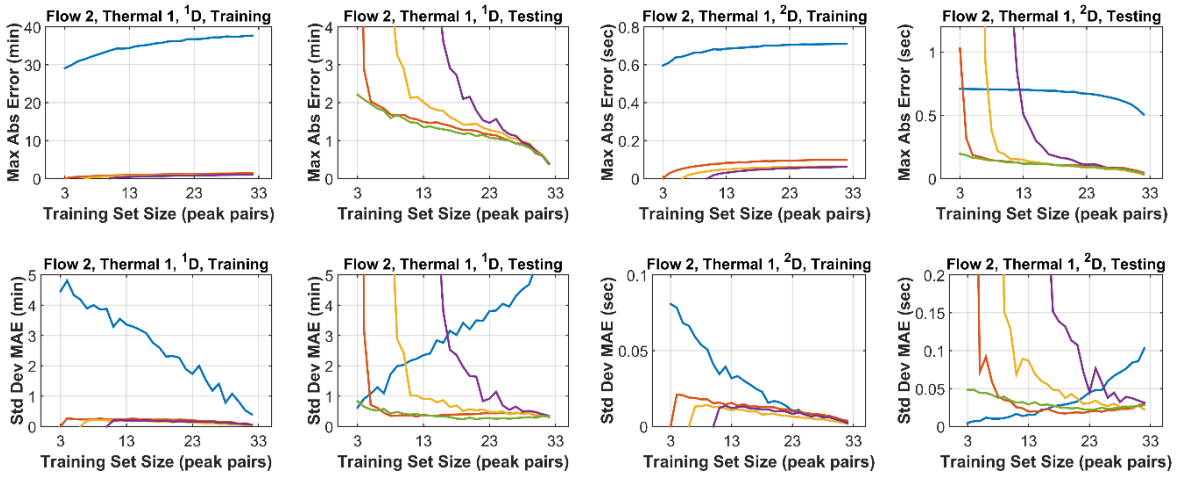
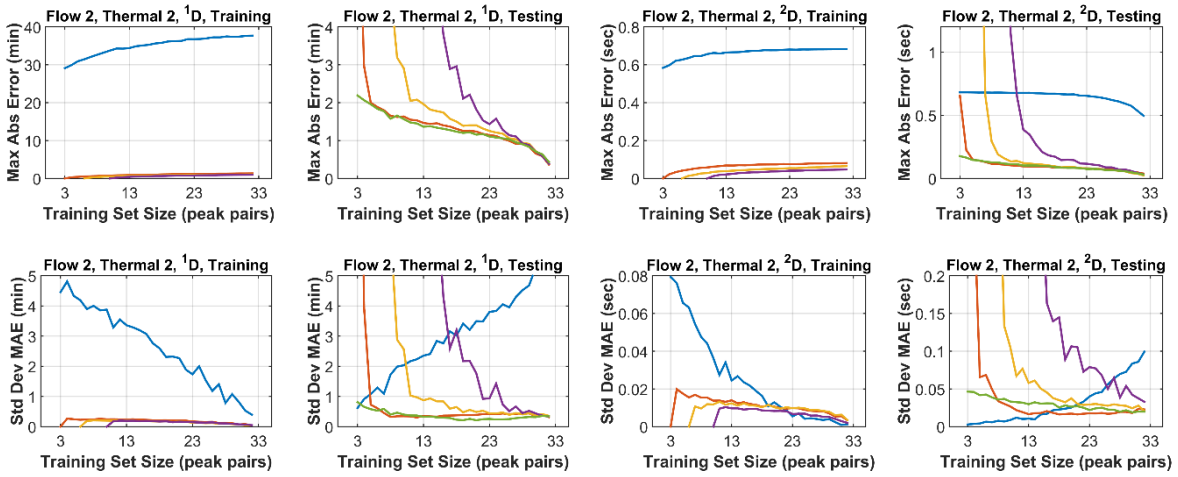


Figure B10. Maximum absolute error as a function of the training set size for cocoa sample chromatograms with different modulation platforms. Sets of rows with maximum absolute error on the top of each set and the standard deviation of maximum absolute error on the bottom of each set are for: A. Runs #1 and #1, B. Runs #1 and #2, and C. Runs #1 and #3.

D. Runs #2 and #1.



E. Runs #2 and #2.



F. Runs #2 and #3.

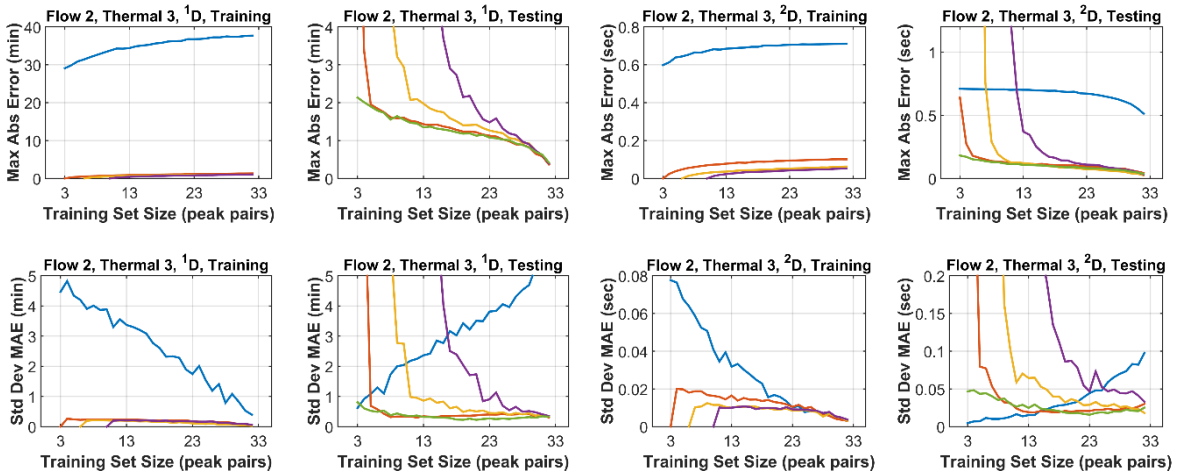


Figure B10 continued. D. Runs #2 and #1, E. Runs #2 and #2, and D. Runs #2 and #3.

APPENDIX C. SAMPLE CHROMATOGRAMS

Figures C1 through C4 show examples of the sample chromatograms and peaks used for alignment experiments. Figure C1 shows a diesel sample chromatogram acquired on June 14, 2013. All diesel chromatograms, including the replicate runs, look very similar to this one. The yellow circles show the 112 peaks that correspond across all diesel chromatograms and were used as alignment points.

Figure C2 shows a wine sample chromatogram acquired from the second run of the 2011 vintage sample. Because misalignment is so minimal between wine chromatograms, they all look nearly identical to Figure C2. The yellow circles show the 78 peaks used for alignment of the wine chromatograms.

Figures C3 and C4 show a cocoa sample chromatogram acquired on a system using a thermal and flow modulator, respectively. The two other thermal modulated chromatograms used in experiments look very similar to C3, and the one other flow modulator chromatogram resembles Figure C4. The yellow circles show the 33 peaks used for alignment of cocoa sample chromatogram pairs.

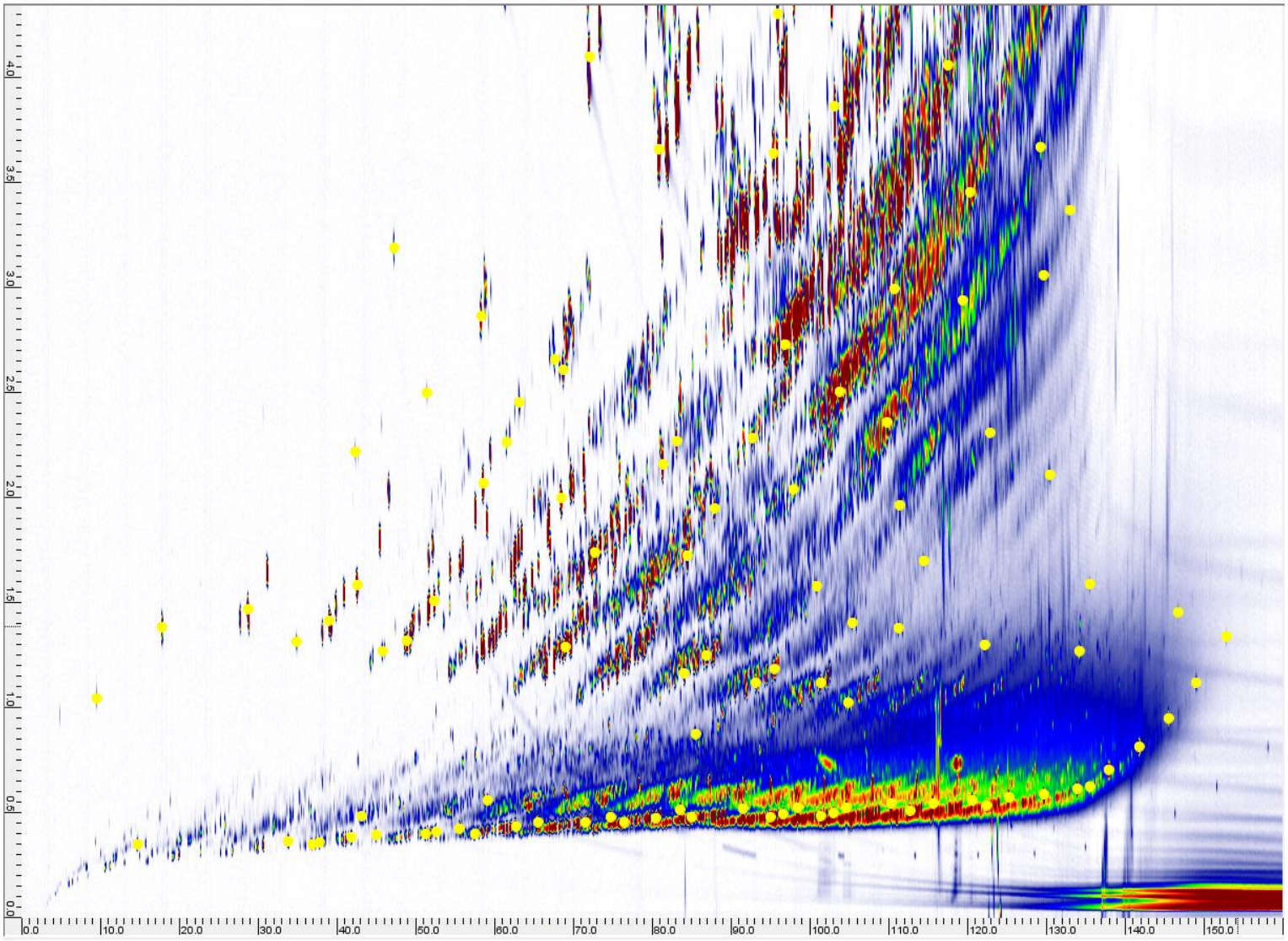


Figure C1. Diesel chromatogram 061413. Closed yellow circles represent peaks that were matched across all diesel chromatograms (112 peaks) and used as alignment points.

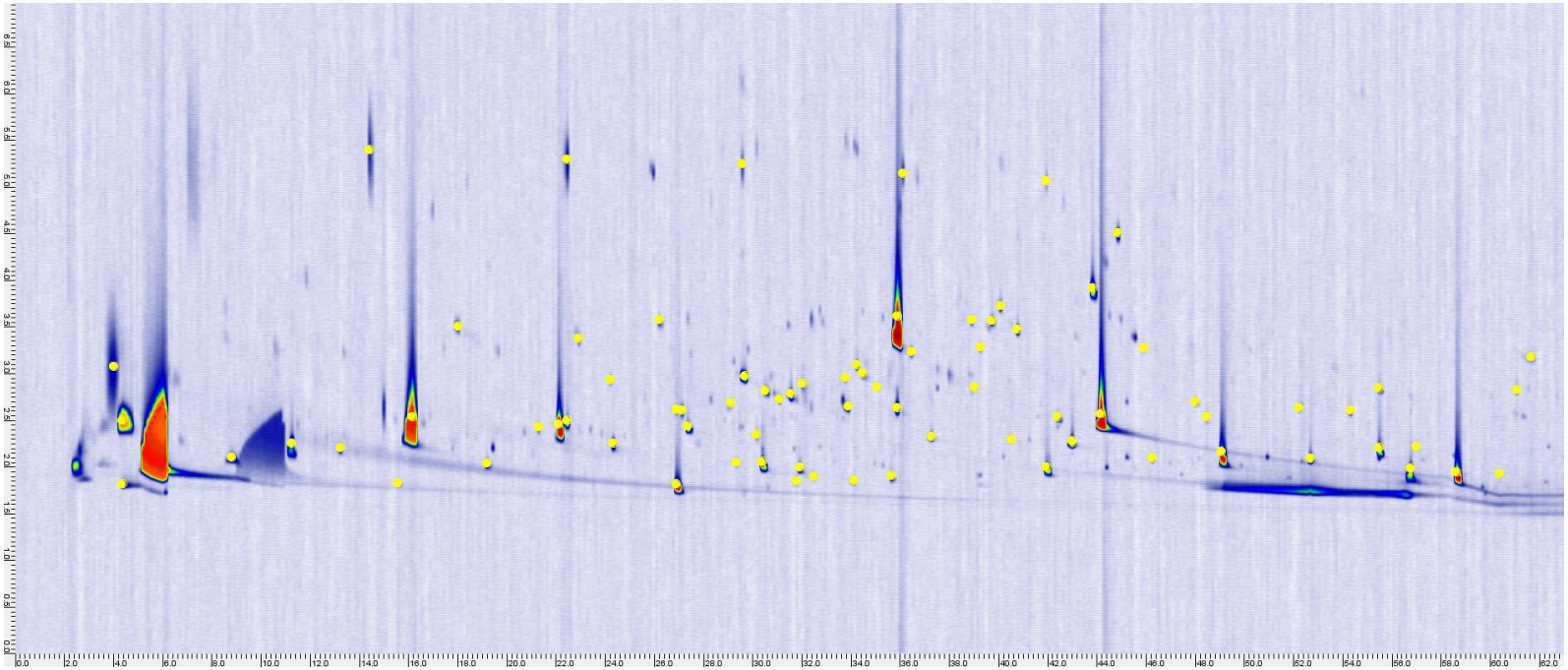


Figure C2. Wine chromatogram MC2011R2. Closed yellow circles represent peaks that were matched across all wine chromatograms (78 peaks) and used as alignment points.

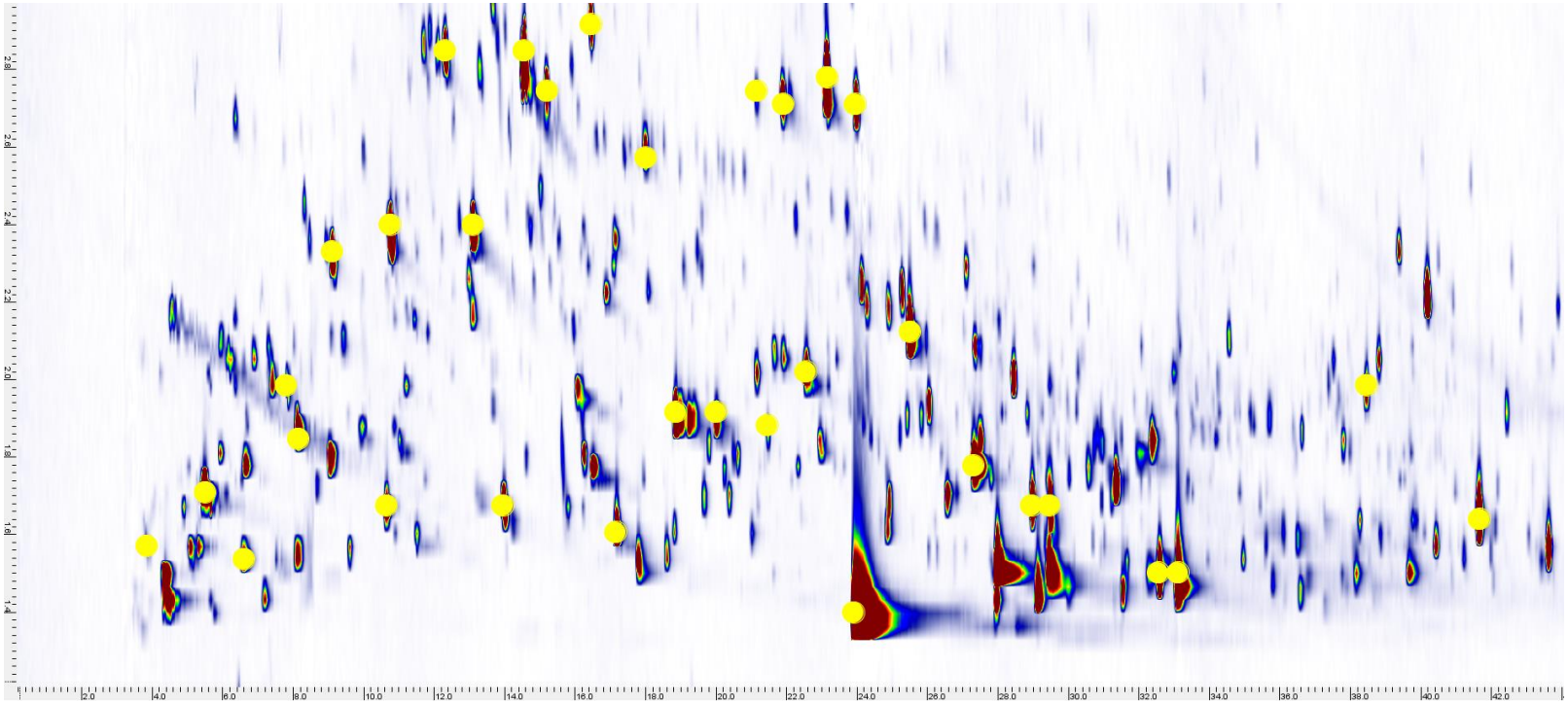


Figure C3. Cocoa sample chromatogram Thermal 1. Closed yellow circles represent peaks that were matched across all cocoa chromatograms (33 peaks) and used as alignment points.

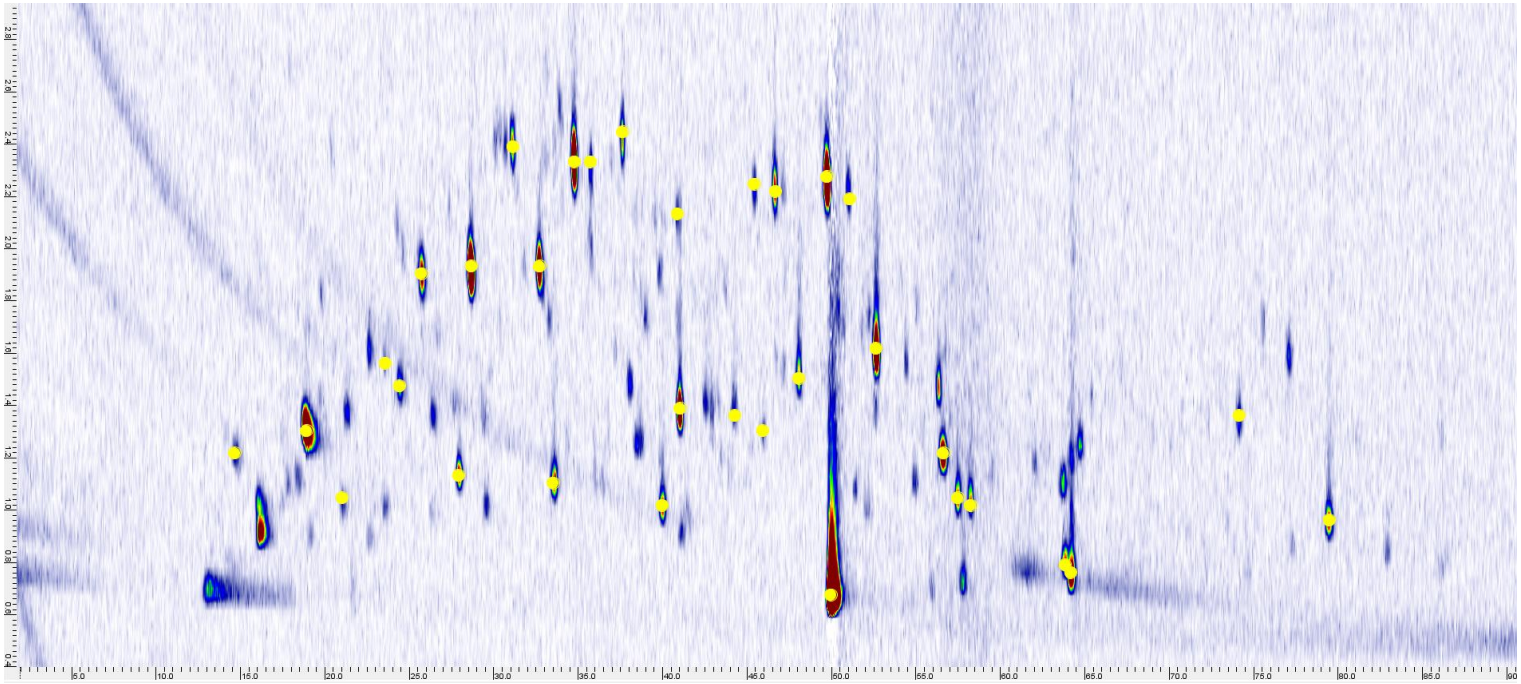


Figure C4. Cocoa sample chromatogram Flow 1. Closed yellow circles represent peaks that were matched across all cocoa chromatograms (33 peaks) and used as alignment points.