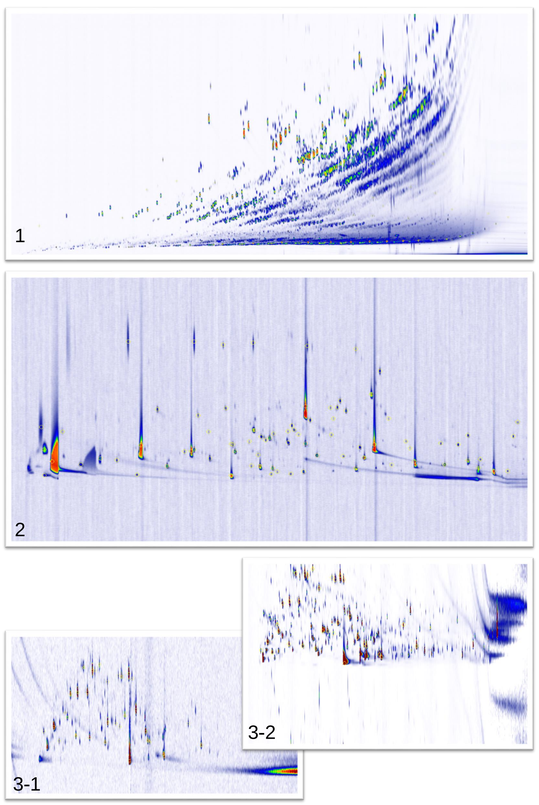
Alignment for Comprehensive Two-Dimensional Gas Chromatography (GCxGC) with Global, Low-Order Polynomial Transformations

Davis Rempe, Stephen Reichenbach, Stephen Scott

Introduction

As columns age and differ between systems, **retention times** for GC x GC may **vary between runs**. In order to properly analyze chromatograms, it is often desirable to align chromatographic features between chromatograms. This **alignment** can be characterized by a **mapping of retention times from one chromatogram to the retention 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. Previous work has shown that global, low-degree polynomial transformation functions – namely **affine, second-degree polynomial, and third-degree polynomial** – are effective for aligning pairs of chromatograms acquired through GCxGC with dual secondary columns and detectors (GC x 2GC). This work assesses the experimental performance of these same methods on more general GC x GC chromatogram pairs and **compares their performance to that of a recent, robust, local alignment algorithm** for GC x GC data [Gros et al*., Anal. Chem.* 2012, 84, 9033].

Data Samples



1: Single diesel sample
2: Three wine vintage samples
3-1: Cocoa sample with flow modulator
3-2: Cocoa sample with thermal modulator

Methods

Preprocessing: Data preprocessing used GC Image GCxGC Edition Software (R2.6 alpha build) from GC Image, LLC (Lincoln NE, USA). For each set of chromatograms, a list of corresponding peaks – peaks which could be located in all chromatograms – was compiled to be used as alignment points. The diesel sample chromatograms produces **112 corresponding peaks**, the wine samples have **78 peaks**, and the cocoa sample has only **33 peaks**.

Evaluation Metric: The primary evaluation metric is the root-mean-square error (RMSE) of the post-alignment retention times across the peak sets for pairs of chromatograms. If each peak in the first chromatogram has the retention-times (x_i, y_i) and the corresponding peak in the other chromatogram has the retention times (x_i', y_i'), then for a set of N_p corresponding peak pairs the two-dimensional RMSE is

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

Evaluation: A cross-validation technique is used to avoid overfitting and provide an unbiased estimator of alignment performance. In **cross-validation**, a peak-pairs set is partitioned into a **training set**, which is used to determine the transform, and a **testing set**, which is used to independently evaluate performance. To account for variability, results over multiple rounds of cross-validation are evaluated. Each cross-validation result is computed across random partitions, for each transformation method (including no transformation), for both the training set and the testing set, at each training set size from 3 peak-pairs (the minimum size for the affine transformation) to the total number of peak-pairs, and for both directions (i.e., switching the target and reference chromatograms). For each training set size, 100 cross-validation trials are run. The RMSE for a training set size is the average across all these trials.

Performance Benchmarks: The performance of the global transformation models are assessed in two ways. First, they are **compared to a benchmark** computed as the RMSE between corresponding peaks for contiguous runs using the same sample, detector, and system. This benchmark is the system's inherent variability or noise and is regarded as the lower bound on alignment performance. Secondly, the peak RMSE is **compared to that of the local algorithm**, as well as the number of alignment points needed to reach this peak RMSE.

Alignment Models

There are **three global**, polynomial transformation alignment methods being tested, along with the **one local** algorithm from Gros et al. On each test, the identity transformation function is also tested which provides the average initial misalignment between the pair of chromatograms.

The **affine transformation** is linear scaling and shearing plus translation:

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

where (s_x,s_y) are the scale parameters, (h_x,h_y) are the shear parameters, and (t_x,t_y) are the translation parameters. This requires at least 3 alignment peak-pairs. The **second-degree polynomial** adds three additional terms in each dimension and requires at least 6 peak-pairs:

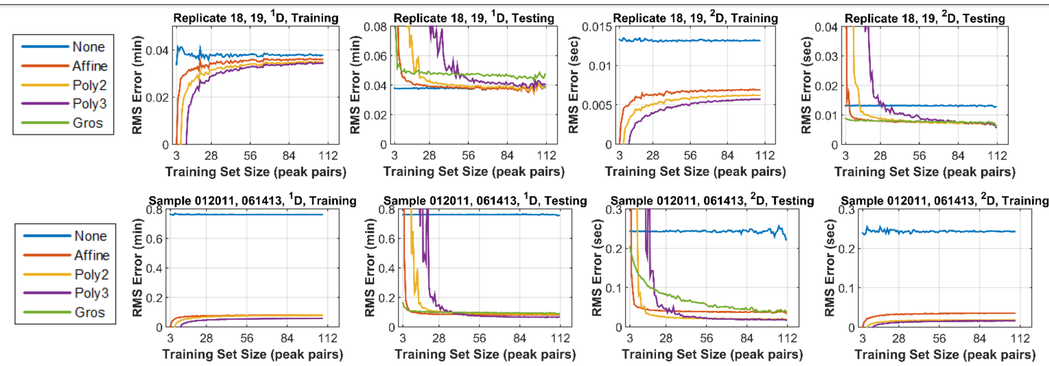
$$f_2(x,y) = (t_x + s_x x + h_x y + a_x xy + b_x x^2 + c_x y^2, t_y + h_y x + s_y y + a_y xy + b_y x^2 + c_y y^2).$$

The **third-degree polynomial** adds four additional terms in each dimension and requires at least 10 peak-pairs:

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

The **local algorithm** guarantees that alignment points (i.e., training set peak-pairs) 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 first dimension, displacements are linearly interpolated between alignment points. In the second, displacements are estimated using Sibson natural-neighbor interpolation, based on Voronoi diagrams.

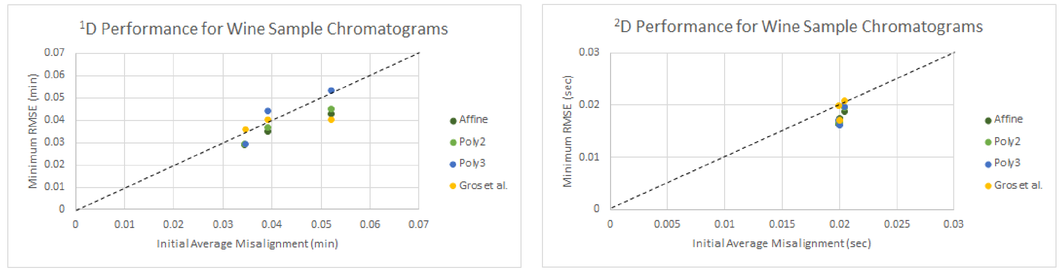
Results



Minimum RMSE Reached by Alignment Methods in the ^{1}D (min) and ^{2}D (sec) for Diesel Chromatograms

Chromatograms	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	0.017	0.0871	0.0346
012011-090912	0.1024	0.3982	0.0574	0.0147	0.0583	0.013	0.0592	0.0131	0.064	0.0435
012011-100412	0.08	0.0569	0.0502	0.0257	0.046	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	0.794	0.2905	0.0763	0.0558	0.0783	0.0331	0.0511	0.0282	0.0996	0.0558
090912-100412	0.077	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

Results from a diesel replicate and sample run, summary table.



Summary of all results from wine sample alignments. These figures show the minimum testing RMSE reached by each of the methods as a function of the initial misalignment. The dashed line is the identity function – where the RMSE reached is the same as the initial misalignment.

Conclusions

This work indicates that **low-order polynomial transformation functions** will, on average, **out-perform that of the local alignment method** developed by Gros et al. if given a **sufficient amount of alignment points** to optimally fit the functions. Although the third-degree polynomial transformation consistently reaches the lowest minimum RMSE when optimally parameterized (around 55 alignment points), the performance gain over the second-order polynomial is not significant and may not be worth the extra computational cost.

The tests run on GC x GC chromatograms from wine samples also indicate that **no alignment method**, global or local, is able to **perform well when initial misalignment is minimal.**

In order to out-perform the affine transformation and Gros' algorithm, the second-order polynomial needed around 30 alignment points. If working with a smaller number of points**, it may be preferable to use the method from Gros et al.** due to its ability to approach its peak RMSE **with fewer than 10 points**, though this peak will not be as low as is possible with the global transformations.

References

Reichenbach, S. E.; Rempe, D. W.; Tao, Q.; Bressanello, D.; Liberto, E.; Bicchi, C.; Balducci, S.; Cordero, C. Alignment for Comprehensive Two-Dimensional Gas Chromatography with Dual Secondary Columns and Detectors. *Anal. Chem.* **2015,** *87* (19), 10056-10063.

Gros, J.; Nabi, D.; Dimitriou, P.; Rutler, R.; Arey, J. Robust Algorithm for Aligning Two-Dimensional Chromatograms. *Analytical Chemistry* **2012,** *84,* 9033-9040.

GC Image, LLC. *GC Image GCxGC Software;* GC Image, LLC: Lincoln NE, 2015.

The University of Nebraska–Lincoln is an equal opportunity educator and employer.
© 2014 The Board of Regents of the University of Nebraska. All rights reserved.