Hotelling Trace Criterion as a Figure of Merit for the Optimization of Chromatogram Alignment

Edward J. Soares, Gopal R. Yalla, John B. O'Connor,
Kevin A. Walsh, Amber M. Hupp

Department of Mathematics and Computer Science
Department of Chemistry

College of the Holy Cross

January 10, 2015

Outline

1. **Introduction to Chromatography**

2. Theory and Techniques

3. Experimental Data

4. Results and Future Work

5. Acknowledgements

Outline

Outline

Outline

1 [Introduction to Chromatography](#)

2 [Theory and Techniques](#)

3 [Experimental Data](#)

4 Results and Future Work

5 [Acknowledgements](#)

Outline

Gas Chromatography & Mass Spectrometry

- The gas chromatograph (GC) and mass spectrometer (MS) are two independent instruments which, when combined, create a powerful analytic technique for separating and identifying the components of complex mixtures.

Gas Chromatography & Mass Spectrometry

- The gas chromatograph (GC) and mass spectrometer (MS) are two independent instruments which, when combined, create a powerful analytic technique for separating and identifying the components of complex mixtures.
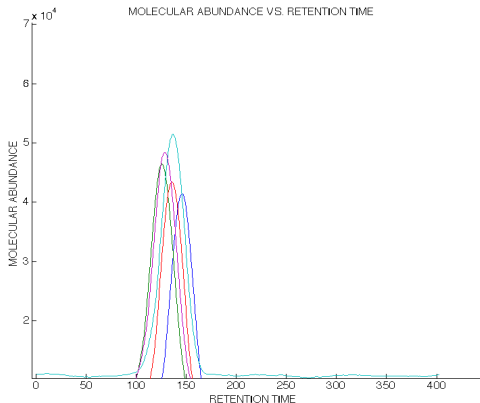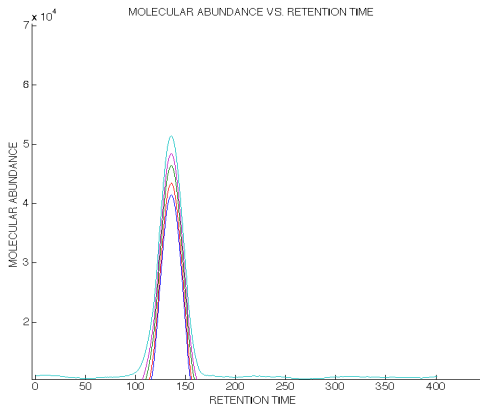
- GC + MS produces chromatograms:
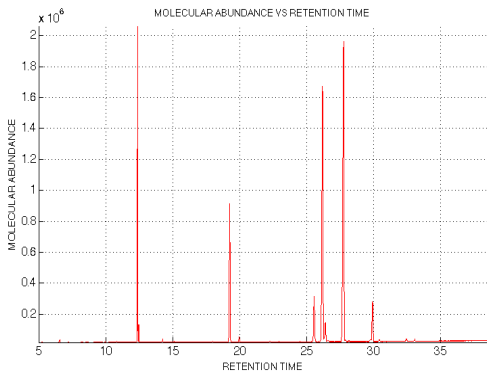
Chromatogram Alignment

- When dealing with multiple samples, fluctuations in peak height and peak location occur.
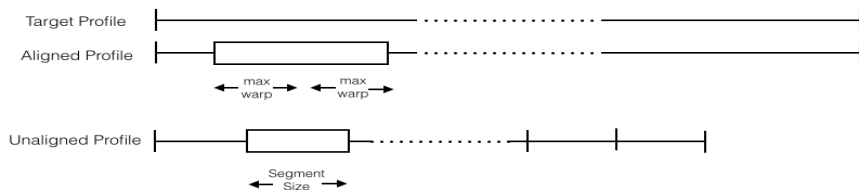
Chromatogram Alignment

- When dealing with multiple samples, fluctuations in peak height and peak location occur.

- Without peak location alignment, trends determined by chemometric methods will be skewed or meaningless.
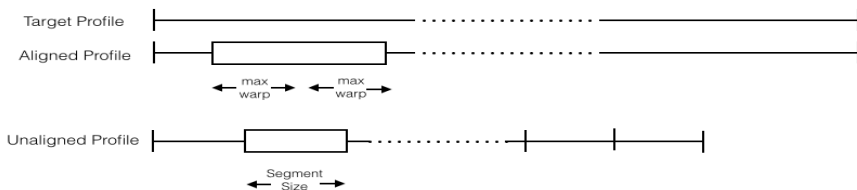
Alignement Techniques

Correlation Optimized Warping (**COW**): Given two parameters segment size and max warp, a chromatogram P is aligned to a target chromatogram T.

Alignement Techniques

Correlation Optimized Warping (**COW**): Given two parameters segment size and max warp, a chromatogram P is aligned to a target chromatogram T.
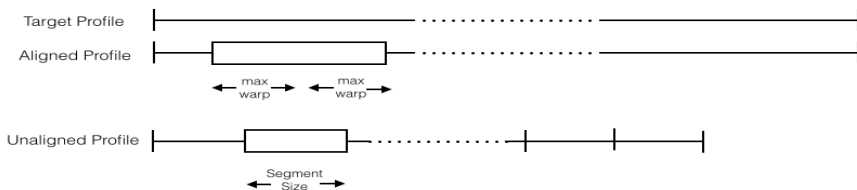

Developed our own..

→ Faster version of COW.

→ Only Align on Peaks

→ Includes a Peak Detection Feature

Alignement Techniques

Correlation Optimized Warping (**COW**): Given two parameters segment size and max warp, a chromatogram P is aligned to a target chromatogram T.

Naturally named, **MOO-COW**:

$\rightarrow$ Faster version of COW.

$\rightarrow$ Only Align on Peaks

$\rightarrow$ Includes a Peak Detection Feature

Alignement Techniques

Correlation Optimized Warping (**COW**): Given two parameters segment size and max warp, a chromatogram P is aligned to a target chromatogram T.
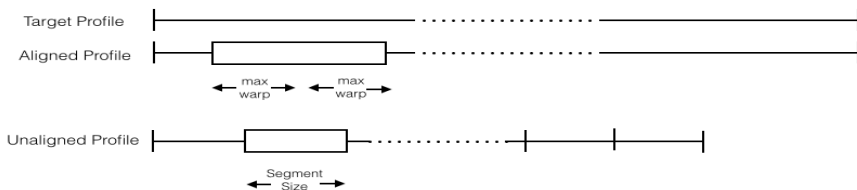

What is the optimal choice of COW parameters?

Alignment Metrics

Warping Effect = Simplicity + Peak Factor

- **Simplicity**: *How close is data to rank 1 matrix*

$$\text{simplicity} = \sum_{r=1}^{R} \left(\text{SVD} \left(\mathbf{X} / \sqrt{\sum_{k=1}^{K} \sum_{n=1}^{N_k} \sum_{m=1}^{M} x_{knm}^2} \right) \right)^4$$

Alignment Metrics

Warping Effect = Simplicity + Peak Factor

- **Simplicity**: *How close is data to rank 1 matrix*

$$\text{simplicity} = \sum_{r=1}^{R} \left(\text{SVD} \left(\mathbf{X} \Big/ \sqrt{\sum_{k=1}^{K} \sum_{n=1}^{N_k} \sum_{m=1}^{M} x_{knm}^2} \right) \right)^4$$

- **Peak Factor**: *How much the shape and peak area of chromatograms have been changed by warping*

$$\text{peak factor} = \frac{1}{N} \sum_{k=1}^{K} \sum_{n=1}^{N_k} (1 - \min(c_{kn}, 1)^2)$$

Alignment Metrics (con't)

- Incorporates both **within class** and **between class** variation in the data set.
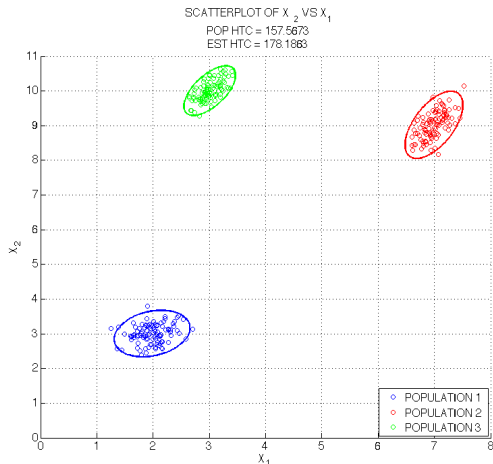
$$HTC = \text{trace}(S_2^{-1} S_1)$$

$S_1 =$ Between Class Covariance Matrix

$S_2 =$ Within Class Covariance Matrix

Hotelling Trace Criterion

- Incorporates both **within class** and **between class** variation in the data set.
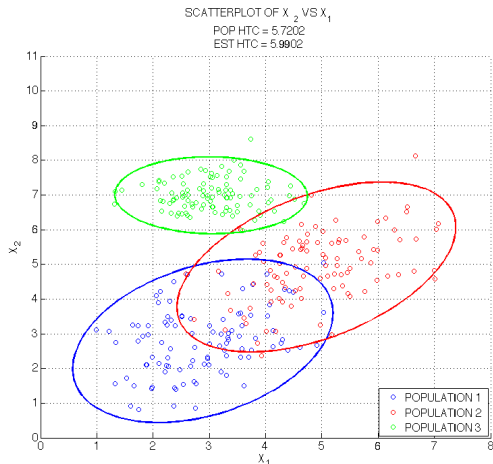

↑ **HTC**

Alignment Metrics (con't)

Hotelling Trace Criterion

- Incorporates both **within class** and **between class** variation in the data set.

↓ **HTC**

Data

5 Classes of Biodiesel:

Data

5 Classes of Biodiesel:

Soy (6 different samples)

Data

5 Classes of Biodiesel:

Soy (6 different samples)
Canola (3 different samples)

Data

5 Classes of Biodiesel:

Soy (6 different samples)
Canola (3 different samples)
Tallow (3 different samples)

Data

5 Classes of Biodiesel:

Soy (6 different samples)
Canola (3 different samples)
Tallow (3 different samples)
Waste Grease (2 different samples)

Data

5 Classes of Biodiesel:

Soy (6 different samples)

Canola (3 different samples)

Tallow (3 different samples)

Waste Grease (2 different samples)

Hybrid (1 sample)

Data

5 Classes of Biodiesel:

Soy (6 different samples)
Canola (3 different samples)
Tallow (3 different samples)
Waste Grease (2 different samples)
Hybrid (1 sample)

}

- Each sample tested
  3 different runs

Data

5 Classes of Biodiesel:

Soy (6 different samples)
Canola (3 different samples)
Tallow (3 different samples)
Waste Grease (2 different samples)
Hybrid (1 sample)

- Each sample tested 3 different runs

- 45 Total Chromatograms

Data

5 Classes of Biodiesel:

Soy (6 different samples)
Canola (3 different samples)
Tallow (3 different samples)
Waste Grease (2 different samples)
Hybrid (1 sample)

- Each sample tested 3 different runs

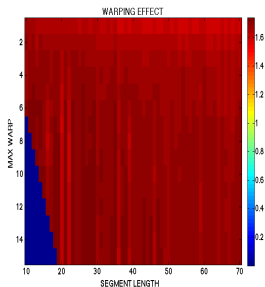- 45 Total Chromatograms

Data Preprocessing:

Baseline Corrected
Aligned (COW)
Normalized
PC Transformed
Computed Metric

Data

5 Classes of Biodiesel:

Soy (6 different samples)
Canola (3 different samples)
Tallow (3 different samples)
Waste Grease (2 different samples)
Hybrid (1 sample)

- Each sample tested 3 different runs
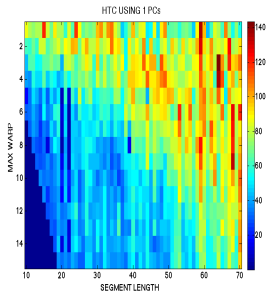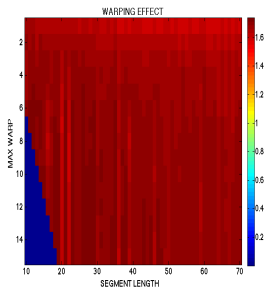
- 45 Total Chromatograms

Data Preprocessing:

Baseline Corrected
Aligned (COW)
Normalized
PC Transformed
Computed Metric

Data

5 Classes of Biodiesel:

Soy (6 different samples)
Canola (3 different samples)
Tallow (3 different samples)
Waste Grease (2 different samples)
Hybrid (1 sample)

- Each sample tested 3 different runs
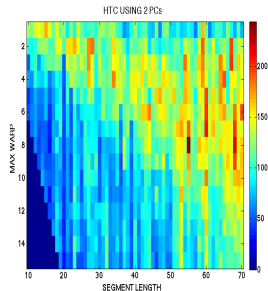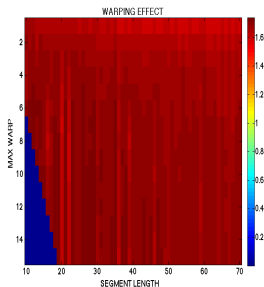
- 45 Total Chromatograms

Data Preprocessing:

Baseline Corrected
Aligned (COW)
Normalized
PC Transformed
Computed Metric

Data

5 Classes of Biodiesel:

Soy (6 different samples)
Canola (3 different samples)
Tallow (3 different samples)
Waste Grease (2 different samples)
Hybrid (1 sample)

- Each sample tested 3 different runs
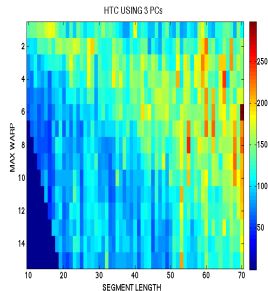
- 45 Total Chromatograms

Data Preprocessing:

Baseline Corrected
Aligned (COW)
Normalized
PC Transformed
Computed Metric

Data

5 Classes of Biodiesel:

Soy (6 different samples)
Canola (3 different samples)
Tallow (3 different samples)
Waste Grease (2 different samples)
Hybrid (1 sample)

- Each sample tested 3 different runs
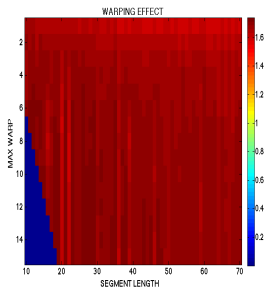
- 45 Total Chromatograms

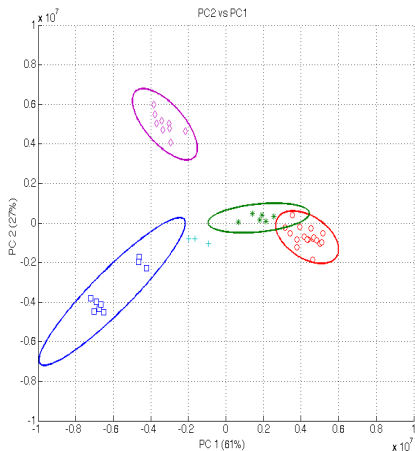## Data Preprocessing:

Baseline Corrected
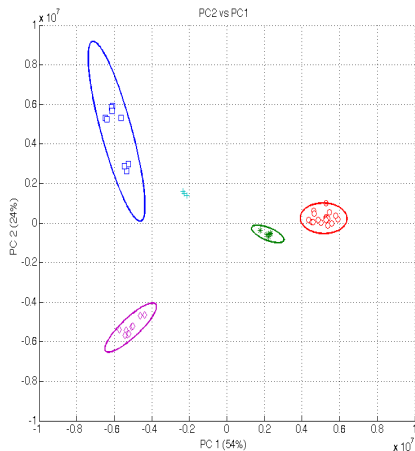Aligned (COW)
Normalized
PC Transformed
Computed Metric

WARPING EFFECT

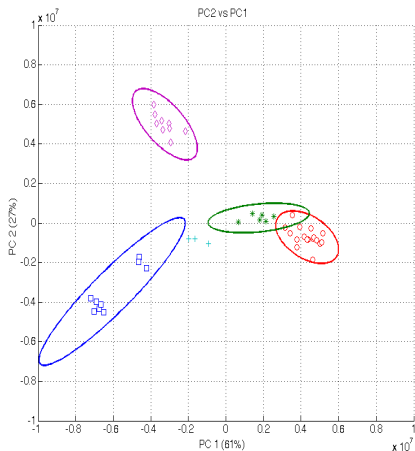HTC USING 3 PCs

Sample Results

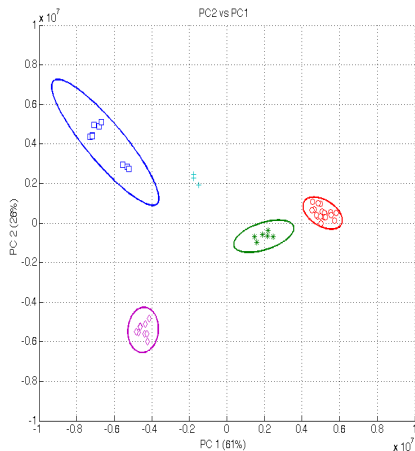**Max Warp Effect**

Max HTC (1 PC)

soy ($\circ$), canola ($\diamond$), tallow ($\square$), waste grease ($*$), hybrid ($+$).
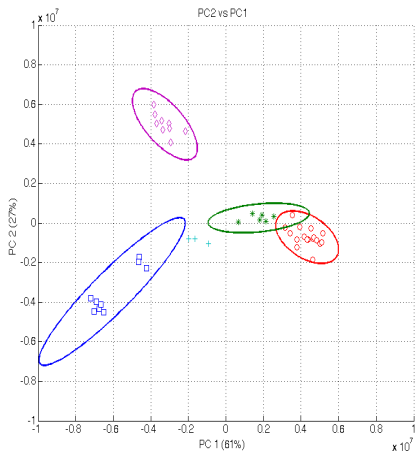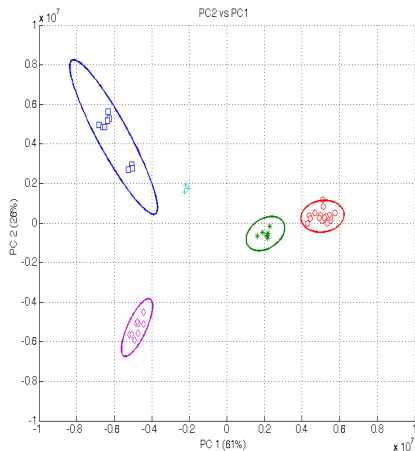
Sample Results

Max Warp Effect

Max HTC (2 PC)

soy (○), canola (◇), tallow (□), waste grease (∗), hybrid (+).

Max Warp Effect

Max HTC (3 PC)

soy (○), canola (◇), tallow (□), waste grease (∗), hybrid (+).

Results

Summary:

- HTC leads to better alignment than warping effect

Results

Summary:

- HTC leads to better alignment than warping effect

- Shown From:

  $\rightarrow$ Superior clustering in principal component space

Results

Summary:

- HTC leads to better alignment than warping effect
- Shown From:
  - $\rightarrow$ Superior clustering in principal component space
  - $\rightarrow$ Greater Euclidean Distance between class means

Results

Summary:

- HTC leads to better alignment than warping effect
- Shown From:
    - $\rightarrow$ Superior clustering in principal component space
    - $\rightarrow$ Greater Euclidean Distance between class means
    - $\rightarrow$ Smaller within class variation.

Results

Summary:

- HTC leads to better alignment than warping effect

- Shown From:

  - $\rightarrow$ Superior clustering in principal component space

  - $\rightarrow$ Greater Euclidean Distance between class means

  - $\rightarrow$ Smaller within class variation.

Future Work:

- Improve upon COW and other alignment algorithms.

- Build classification scheme for unknown biofuels with similar chemical makeup as a given training set by using HTC as a figure of merit.

Acknowledgements

- *Journal of Chemometrics*

Acknowledgements

- *Journal of Chemometrics*

- University Syringe Program Grant from Hamilton Company (AMH).

- Robert L. Ardizzone Fund for Junior Faculty Excellence (AMH).

- College of the Holy Cross.

Acknowledgements