



Chromatographic Alignment for Improved Multivariate Analysis

Scott Ramos and Brian Rohrback

Infometrix, Inc.

Bothell, WA

FACSS 2008

Reno, NV

Wednesday, October 1

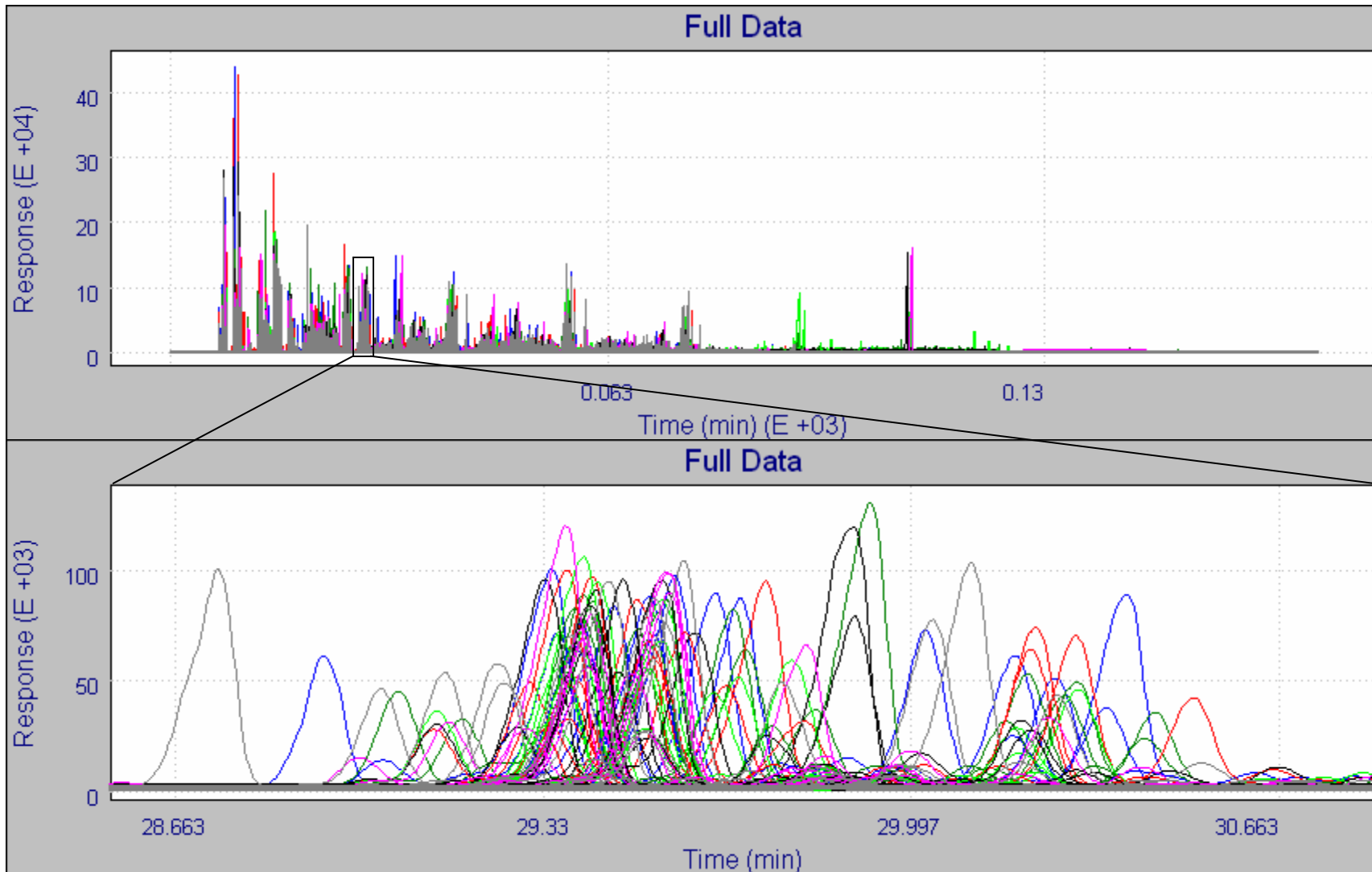


Situation

- Oil companies monitor many processes by chromatography
- Comparisons made within and among refineries
- Historical analyses are frequently compared to recent
- Current practice may requires extensive individual examination
- Result: time and \$\$!!



Naphtha Profiles





Objectives

- Identifying material is not critical
- Characterizing differences within a material type IS important
- Automation could be huge time/cost saver
- Chromatogram database an adjunct to plant historian



Data Description

- Chromatography
 - 100 m dimethylpolysiloxane column, 0.25 mm i.d.
 - Helium carrier, FID
 - 175 minute run (35° for 10 min, 1°/min to 200°), data rate 5 Hz
- Materials included alkylates, naphthas, gasolines
- Data collected over 6-year period
- 5 different GCs, all using same conditions
- At least 1 column change per year per GC



Data Manipulations

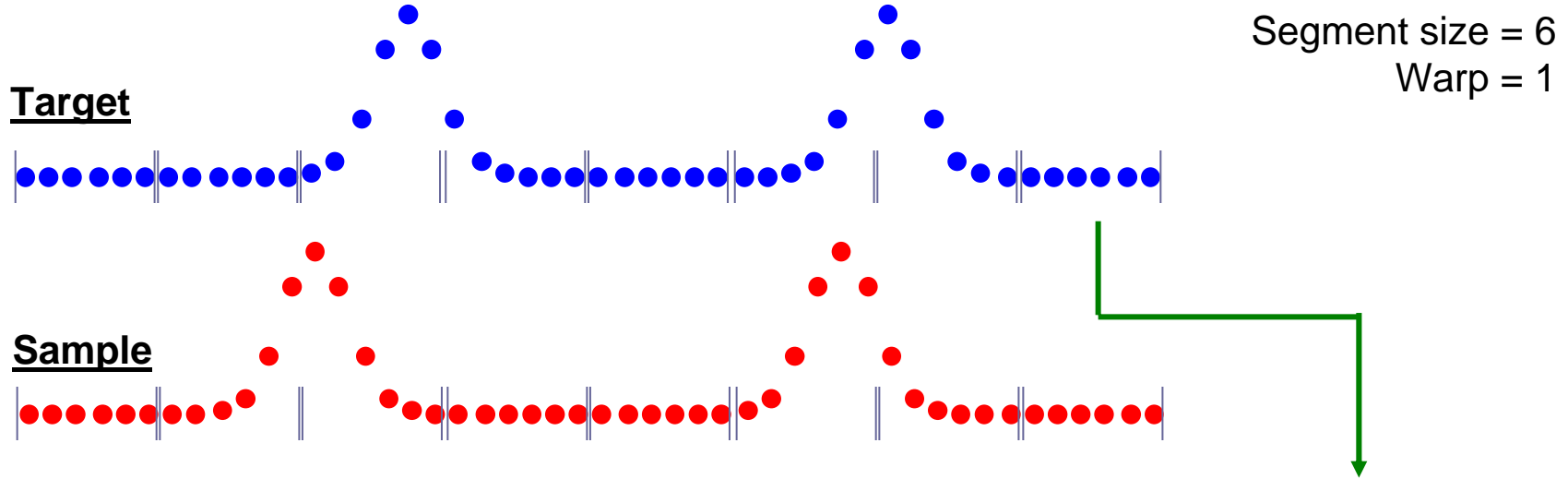
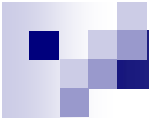
- Chromatogram alignment
- Chromatographic reprocessing
- Chemometrics
- Database archival and retrieval



Alignment

- Algorithm – COW (correlation optimized warping)
 - Target selection
 - One per material
 - Minimum Mahalanobis distance from PCA of historical set
 - Parameters
 - Optimal alignment
 - Segment size – 100
 - Slack – 5
 - Efficient analyses
 - Pad – 10
 - Global shift – 400
 - Software: LineUp (Infometrix)

Nielsen, N-P. V.; J. M. Carstensen; and J. Smedsgaard, *Aligning of single and multiple wavelength chromatographic profiles for chemometric data analysis using correlation optimised warping*, J. Chromatogr. A, **805**:17-35 (1998).



*Objective:
path of optimal S*

$$S = {}_1c_1 \quad \text{●●●●●} \rightarrow \text{●●●●●}$$

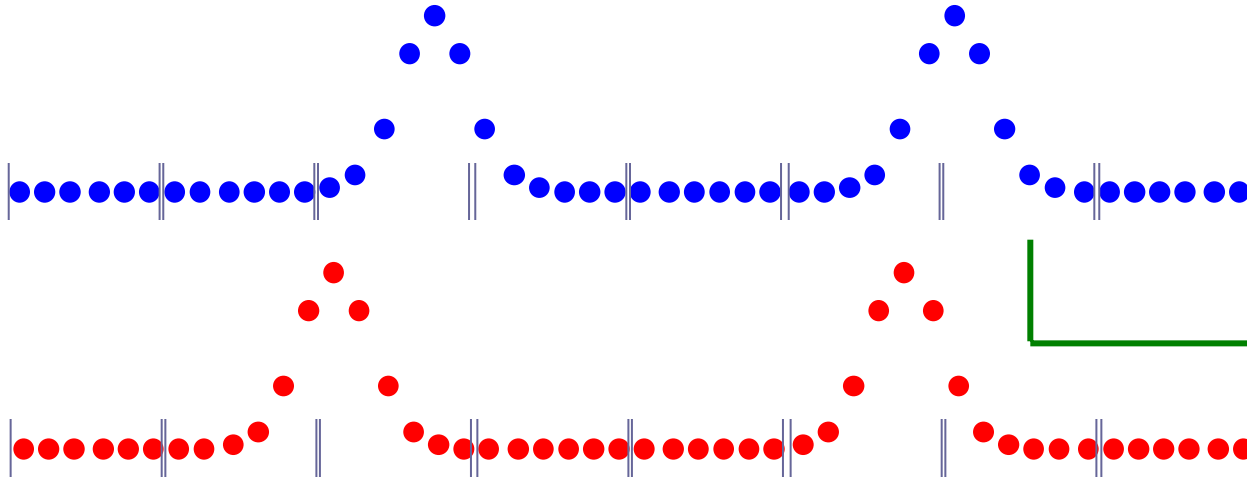
$$S = {}_1c_2 \quad \text{●●●●●} \rightarrow \text{●●●●●}$$

$$S = {}_1c_3 \quad \text{●●●●●} \rightarrow \text{●●●●●}$$

*S is overall Similarity; c is segment correlation; prefix is segment number;
suffix is permutation number; → implies interpolation and comparison to target segment.*

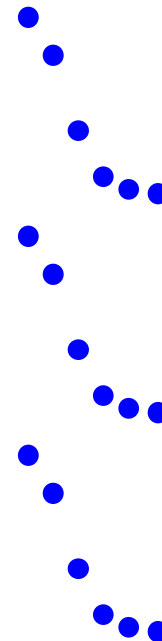


Segment size = 6
Warp = 1



Objective:
path of optimal S

$$S = {}_2C_1 + {}_1C_1$$



$$S = {}_2C_2 + {}_1C_1$$



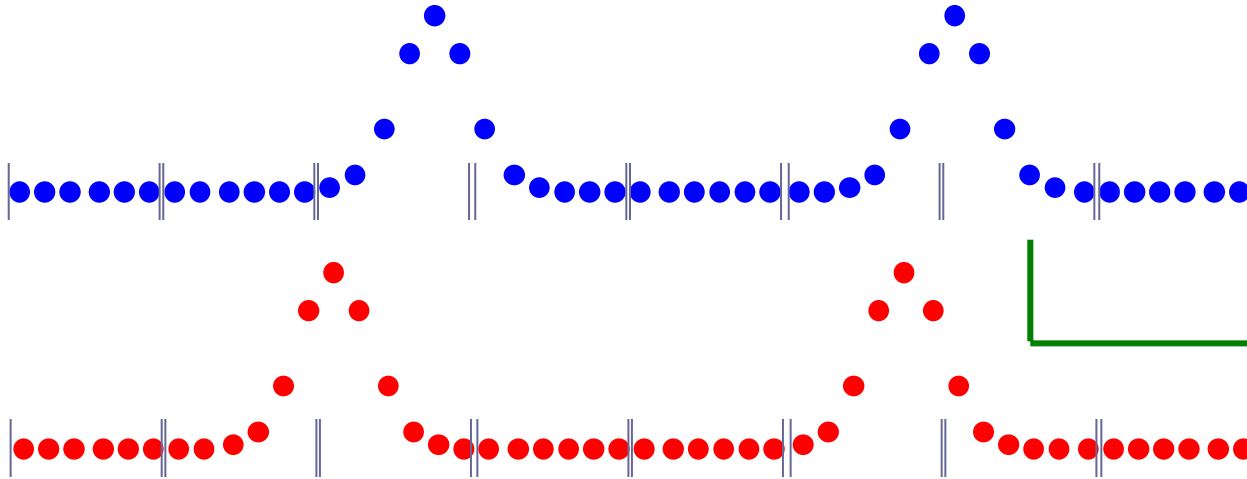
$$S = {}_2C_3 + {}_1C_1$$



S is overall Similarity; c is segment correlation; prefix is segment number;
suffix is permutation number; \rightarrow implies interpolation and comparison to target segment.



Segment size = 6
Warp = 1



Objective:
path of optimal S

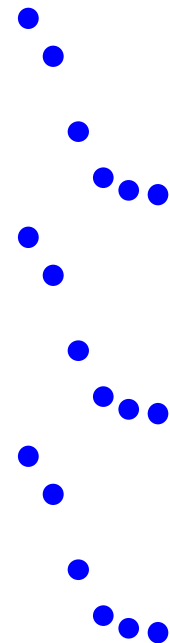
$$S = {}_2C_1 + {}_1C_2$$



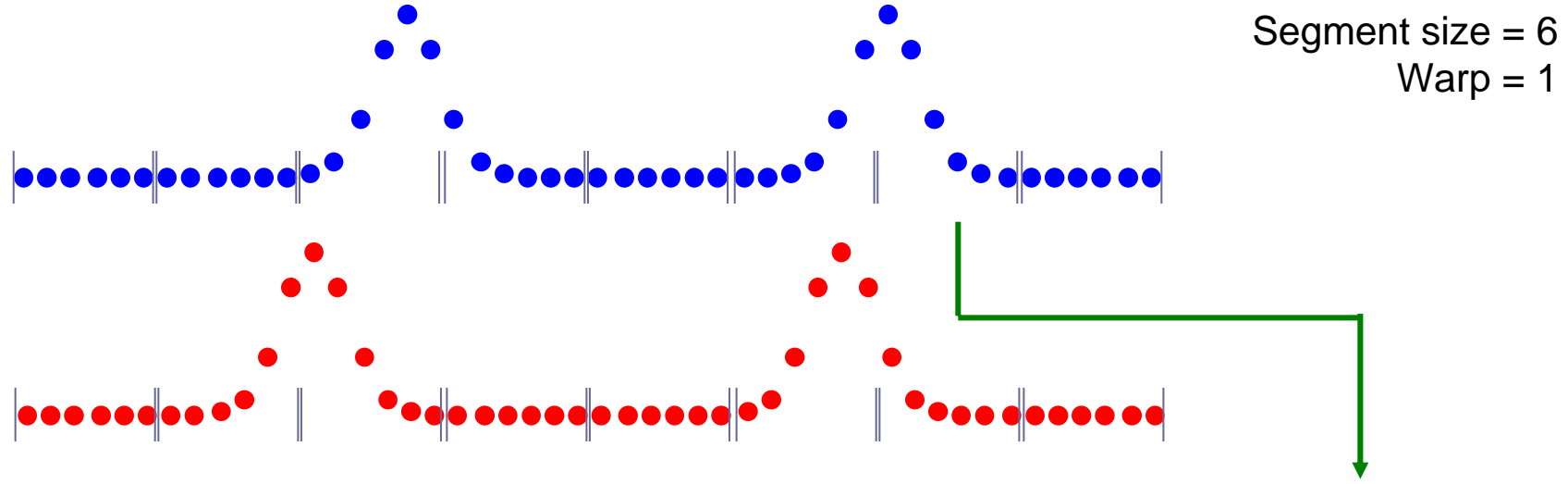
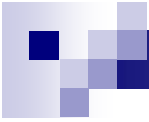
$$S = {}_2C_2 + {}_1C_2$$



$$S = {}_2C_3 + {}_1C_2$$



S is overall Similarity; c is segment correlation; prefix is segment number;
suffix is permutation number; \rightarrow implies interpolation and comparison to target segment.



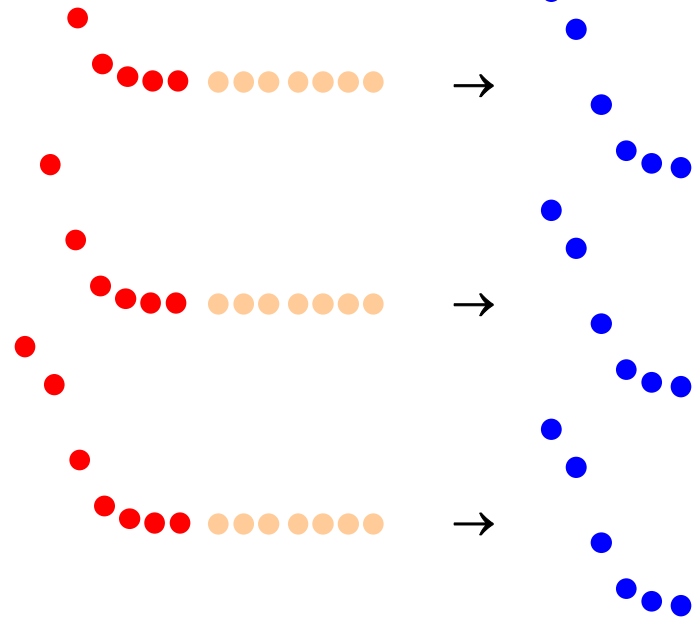
Segment size = 6
Warp = 1

Objective:
path of optimal S

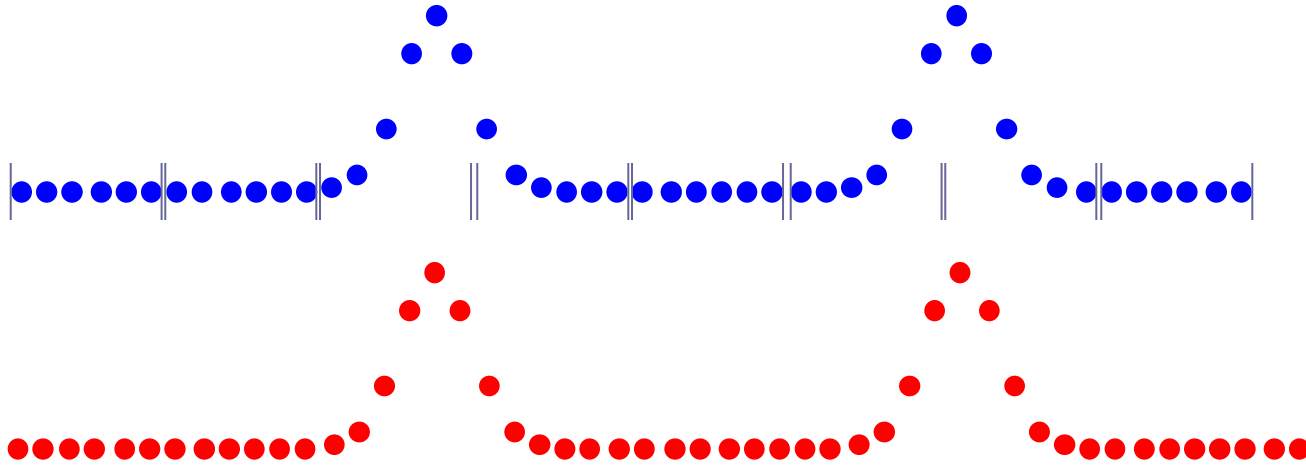
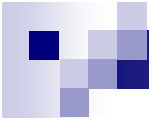
$$S = {}_2C_1 + {}_1C_3$$

$$S = {}_2C_2 + {}_1C_3$$

$$S = {}_2C_3 + {}_1C_3$$



S is overall Similarity; c is segment correlation; prefix is segment number; suffix is permutation number; → implies interpolation and comparison to target segment.

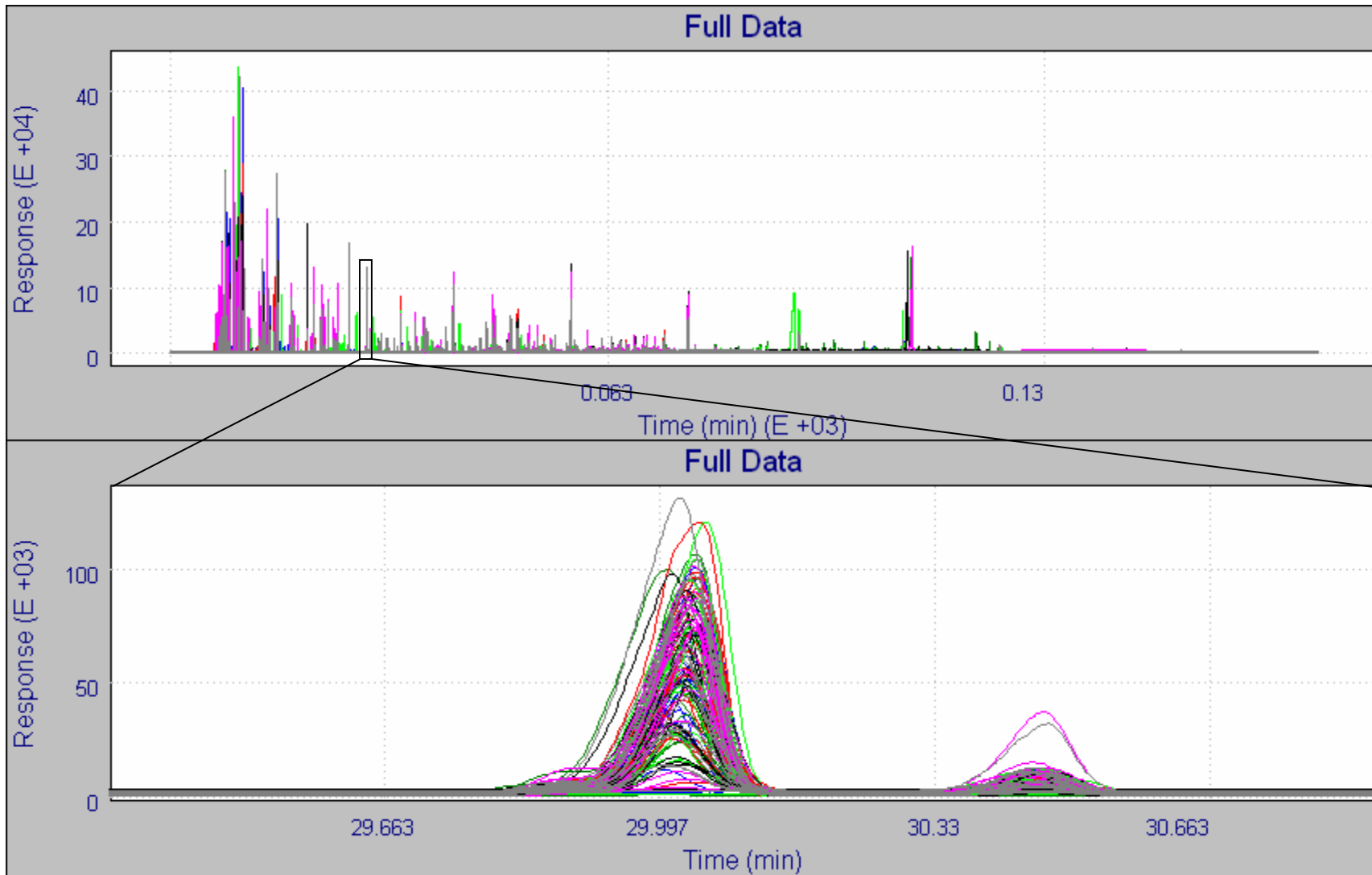


*Objective:
path of optimal S*

$$S = {}_8C_2 \cdots {}_2C_3 + {}_1C_3$$

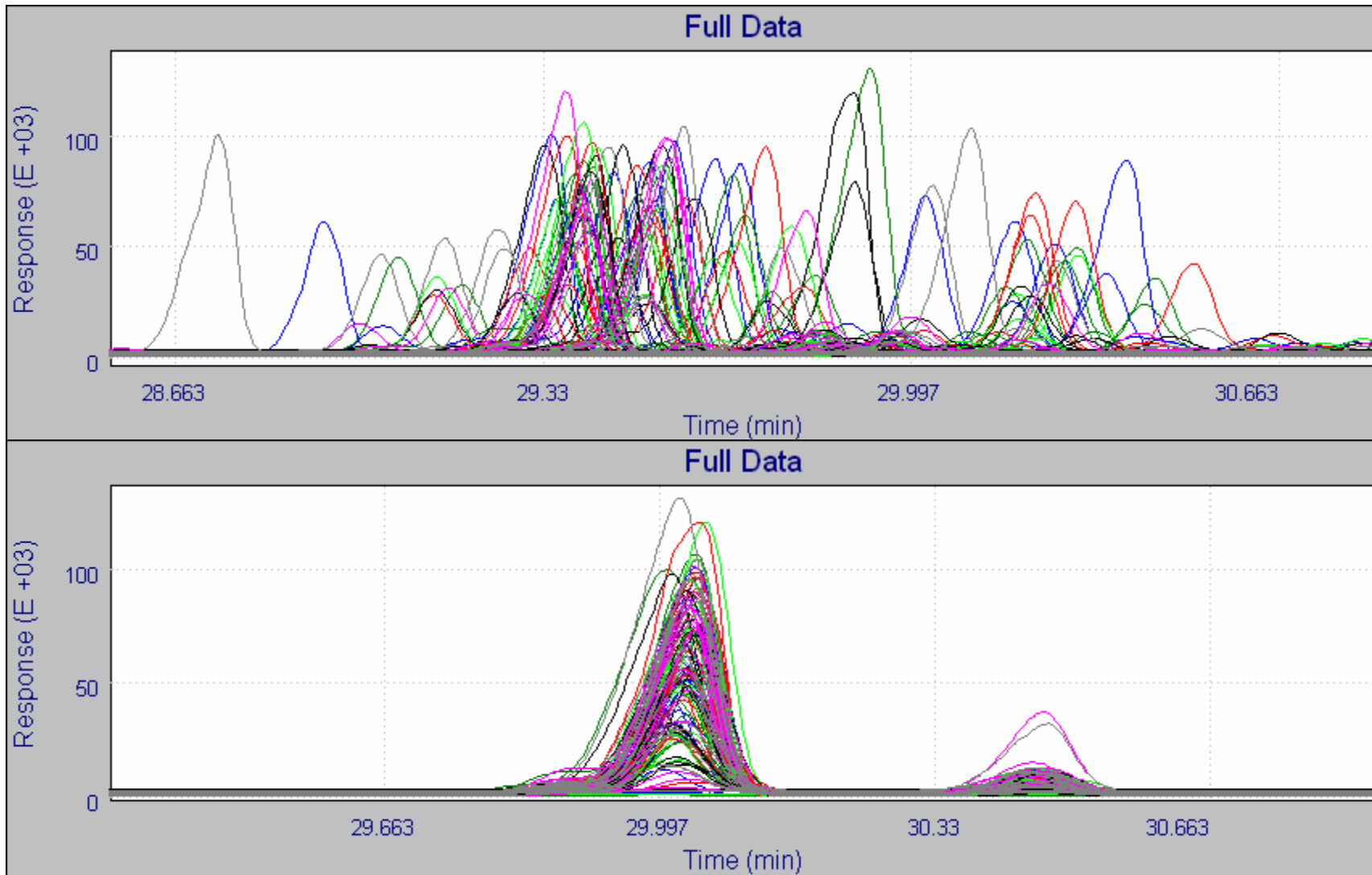


Naphtha Profiles, Aligned



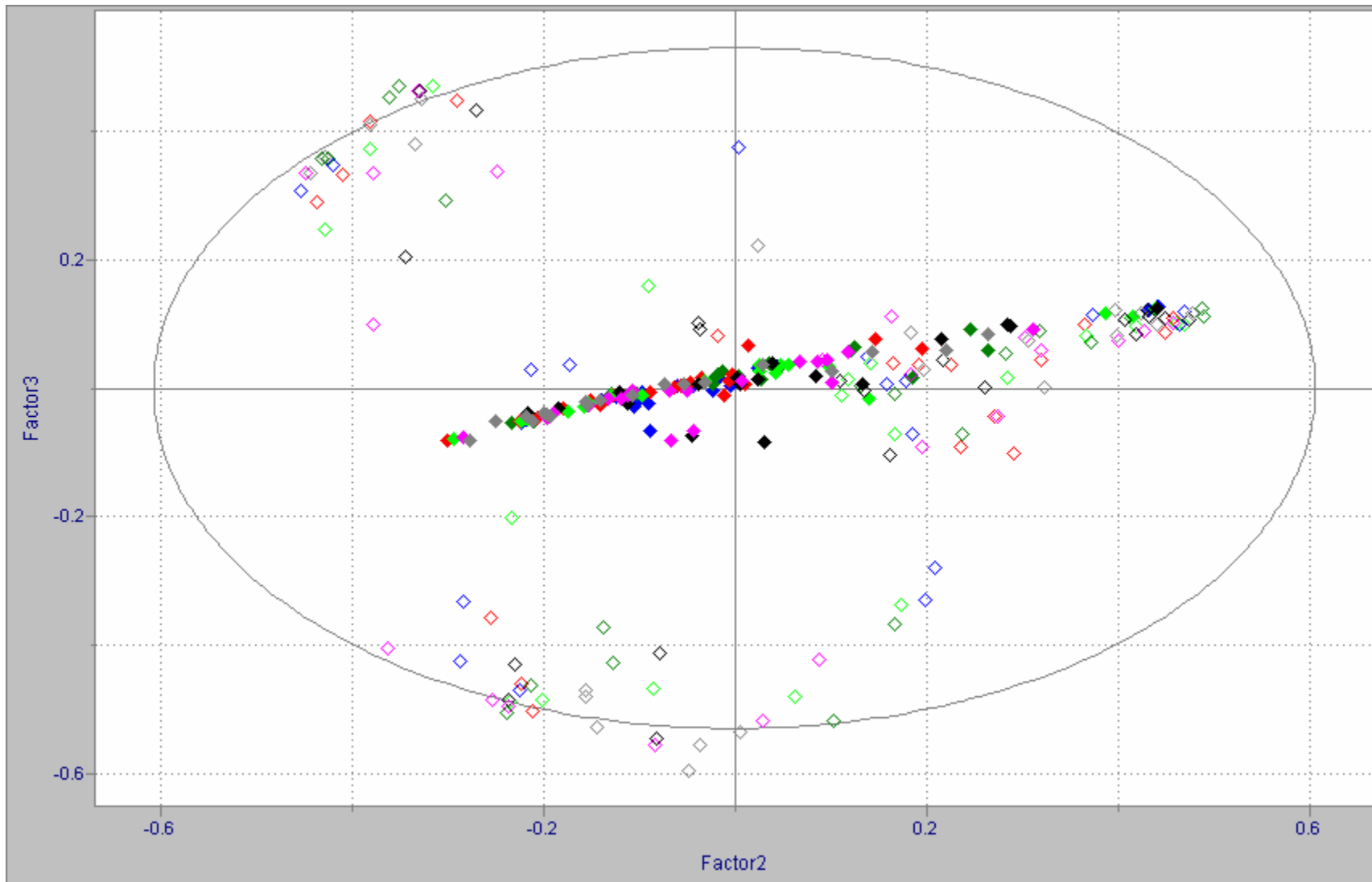


Naphthas, Before & After Alignment



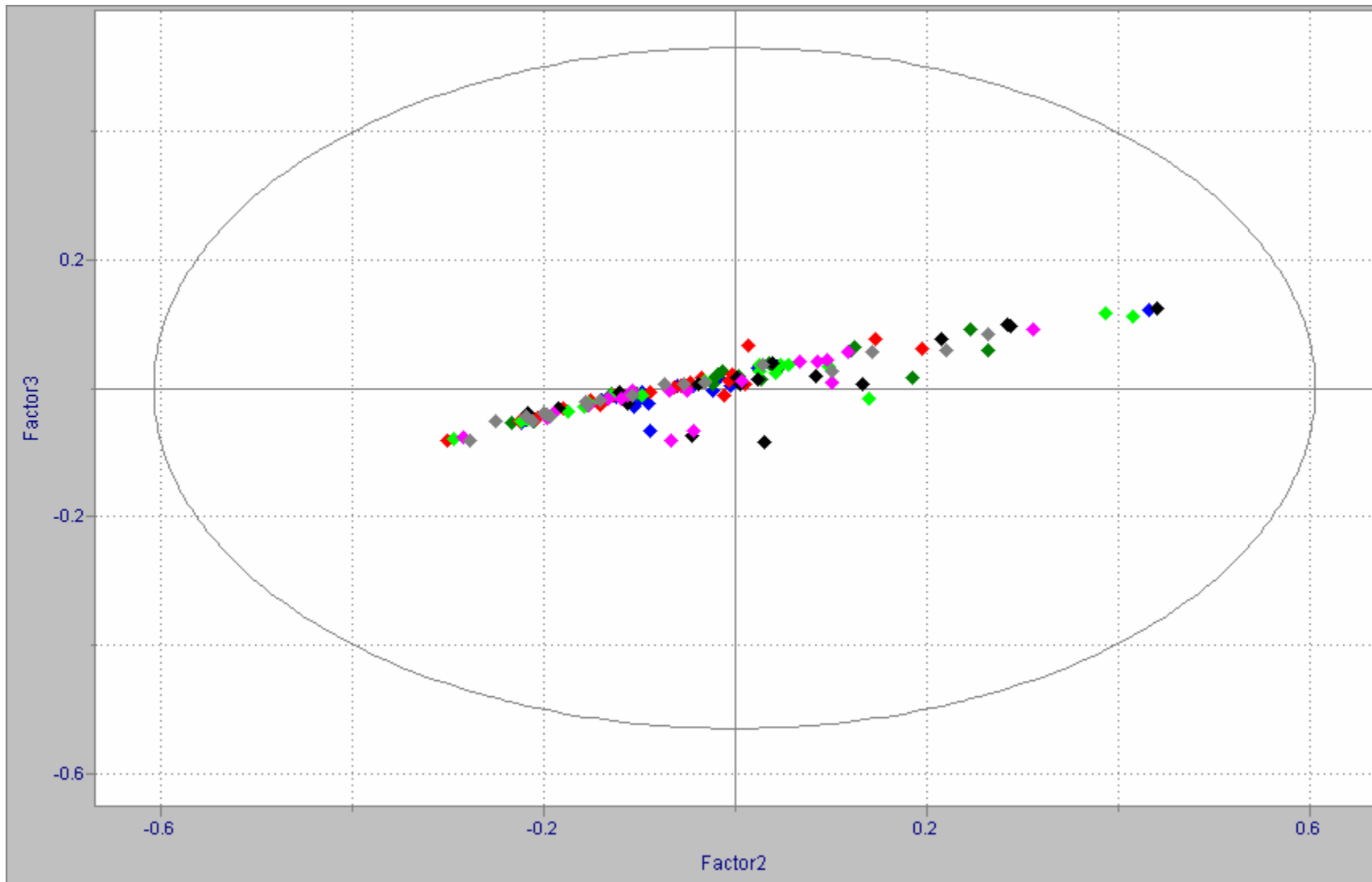


PCA Scores, Unaligned + Aligned



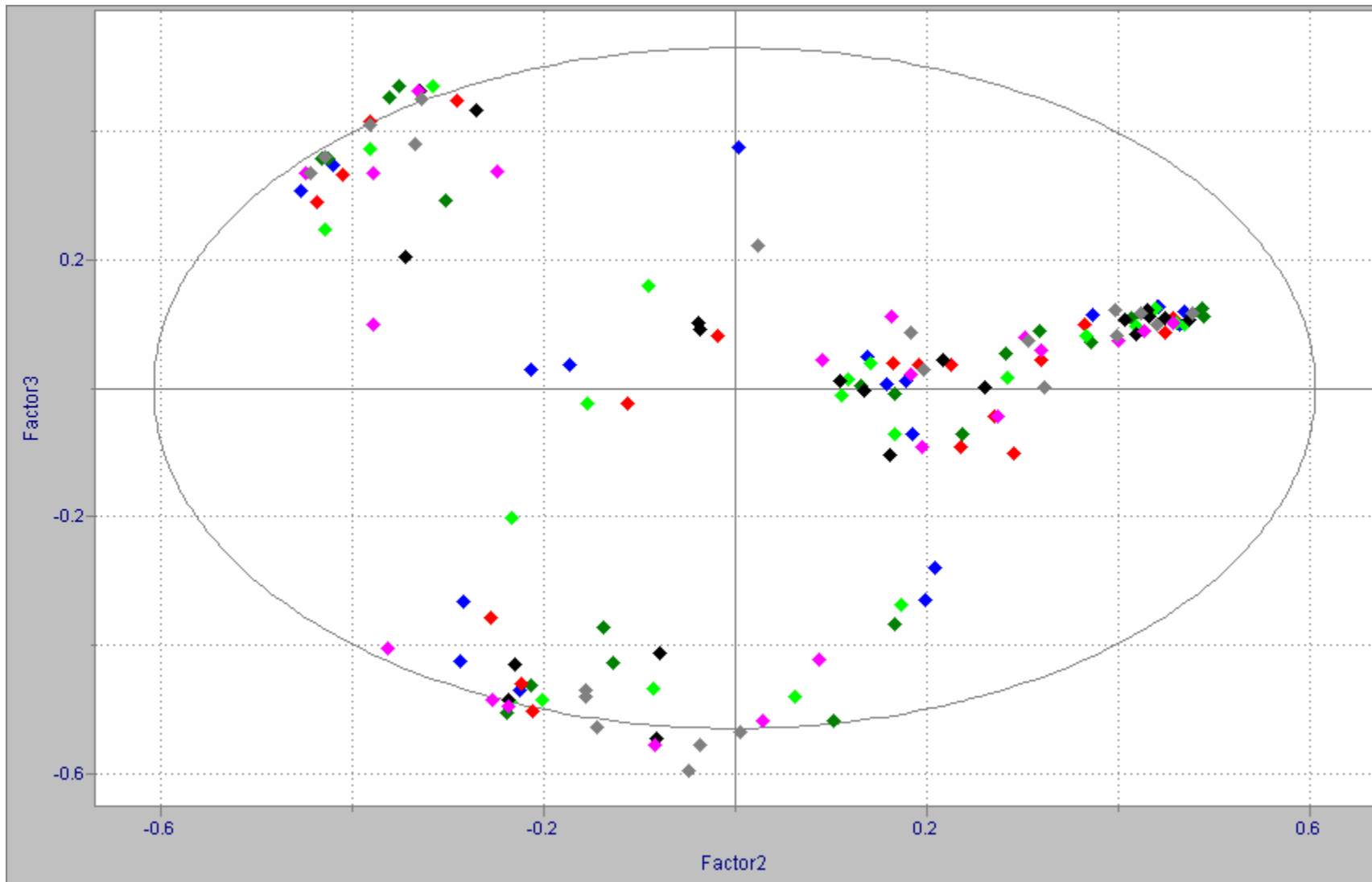


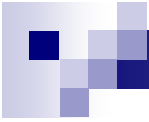
Naphtha Scores, Aligned Profiles



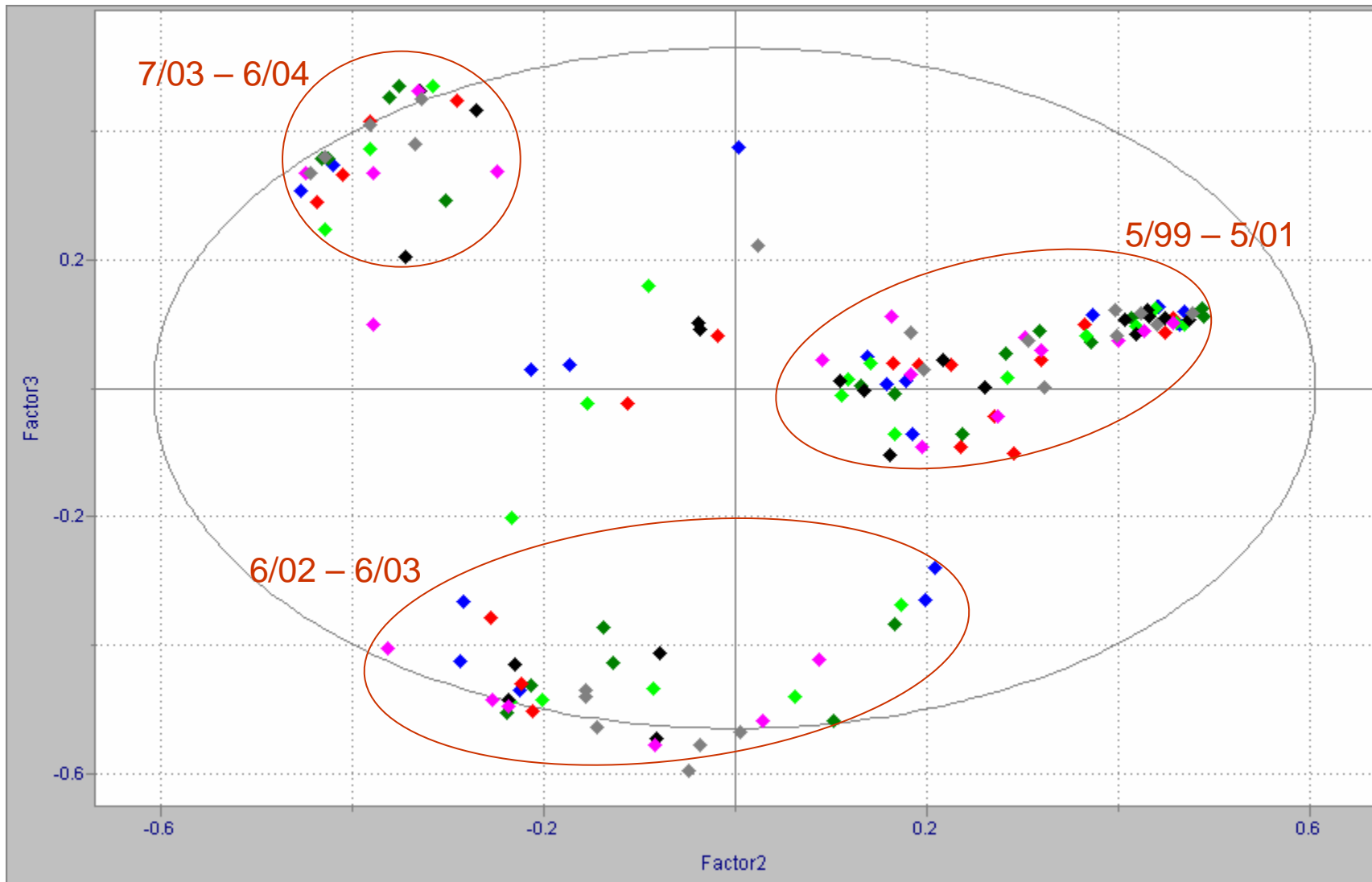


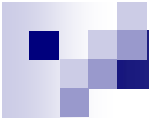
Naphtha Scores, Unaligned Profiles



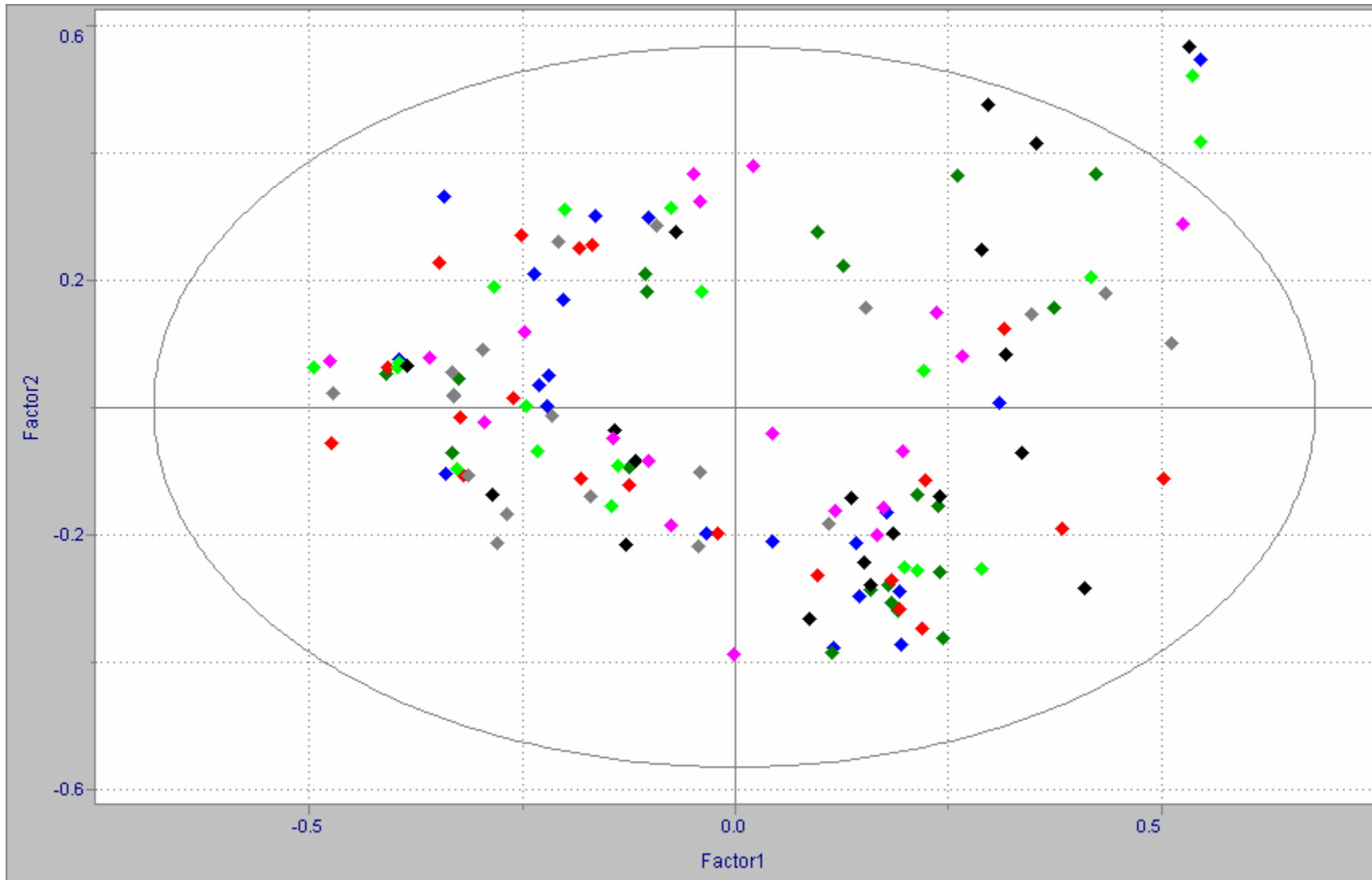


Naphtha Scores, Unaligned Profiles



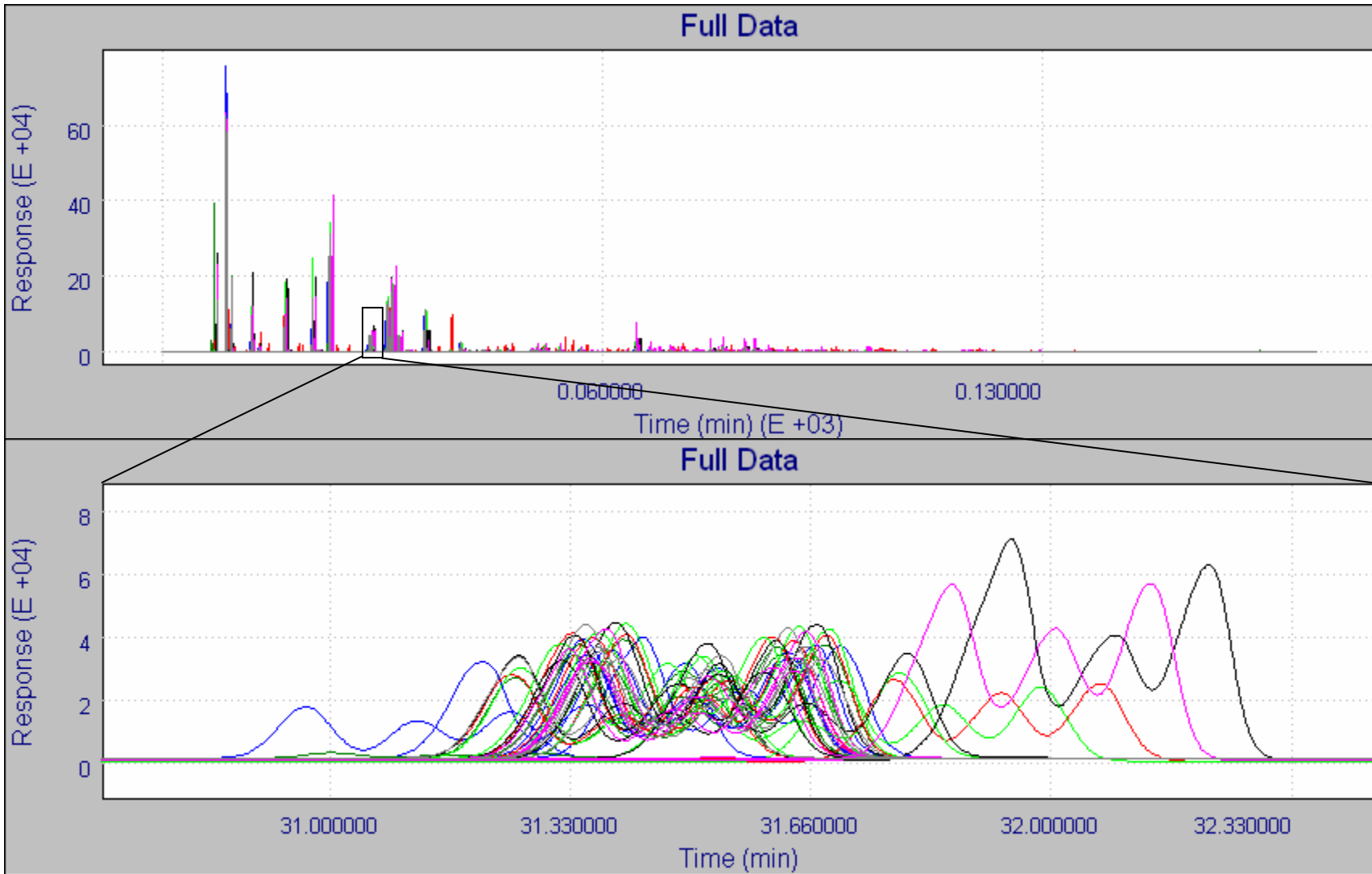


Naphtha Scores, Aligned Profiles



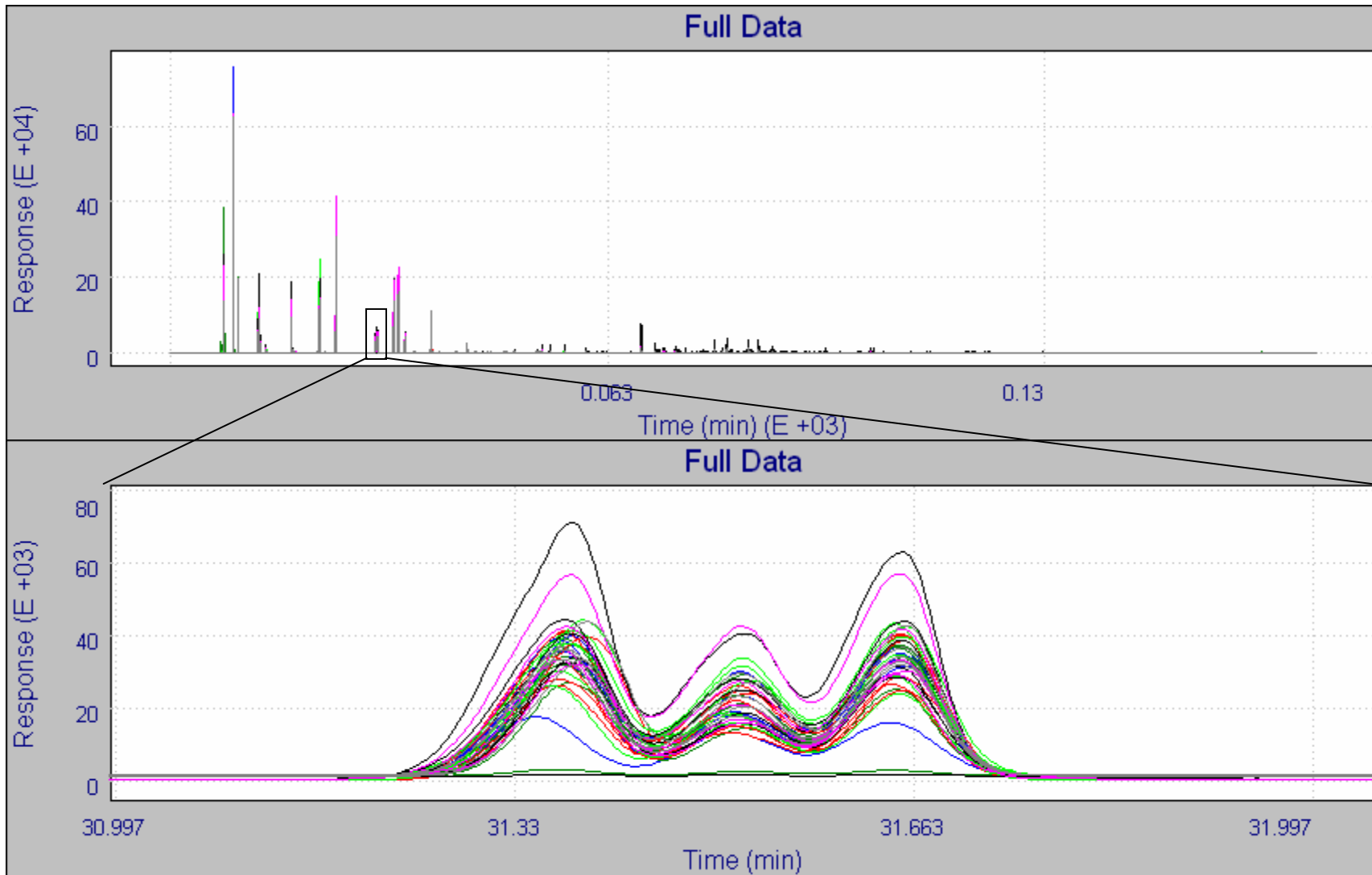


Alkylate Profiles



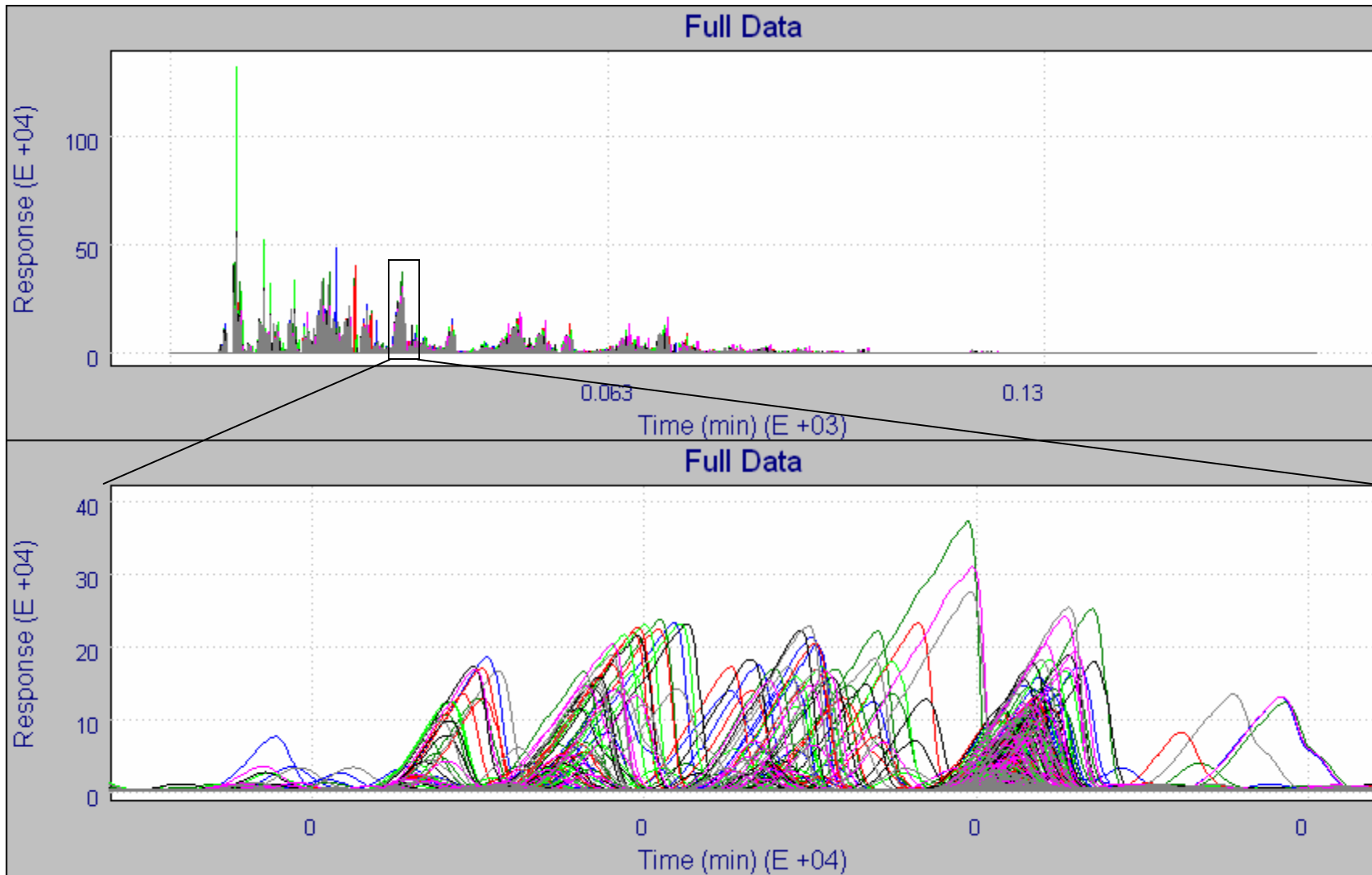


Alkylate Profiles, Aligned



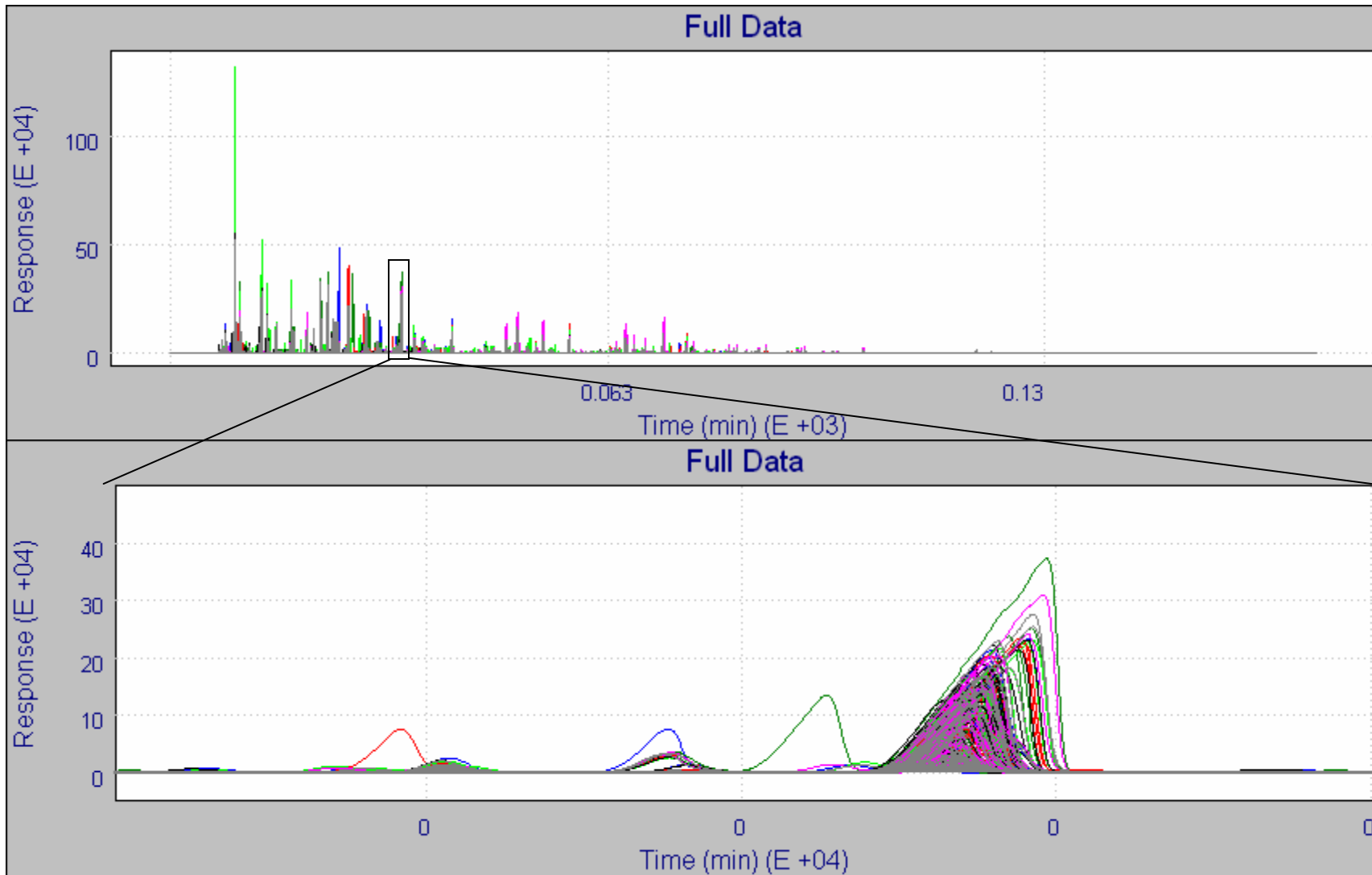


Gasoline Profiles



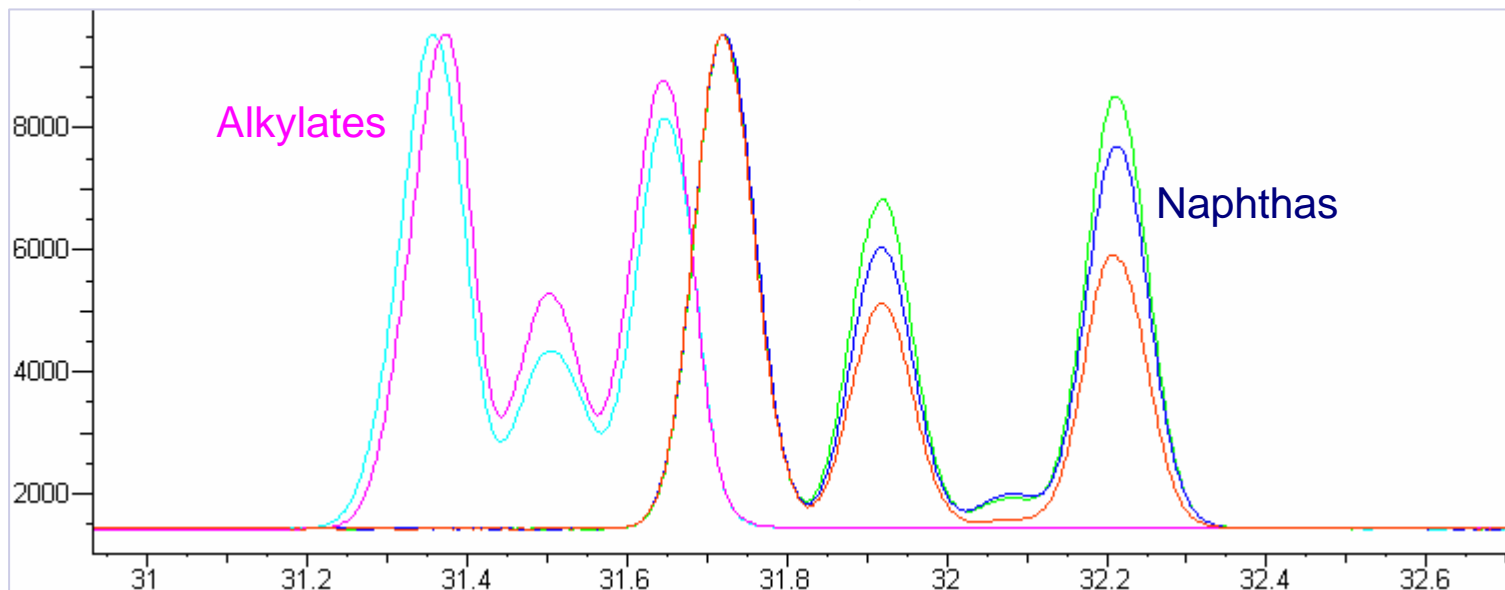


Gasoline Profiles, Aligned



Database of Chromatograms

- Database made for each material:
Alkylates, Naphthas, Gasolines
- Combined into a master KIA database for all fuels
 - Software: KnowItAll (Bio-Rad)
- Problem: different materials don't align





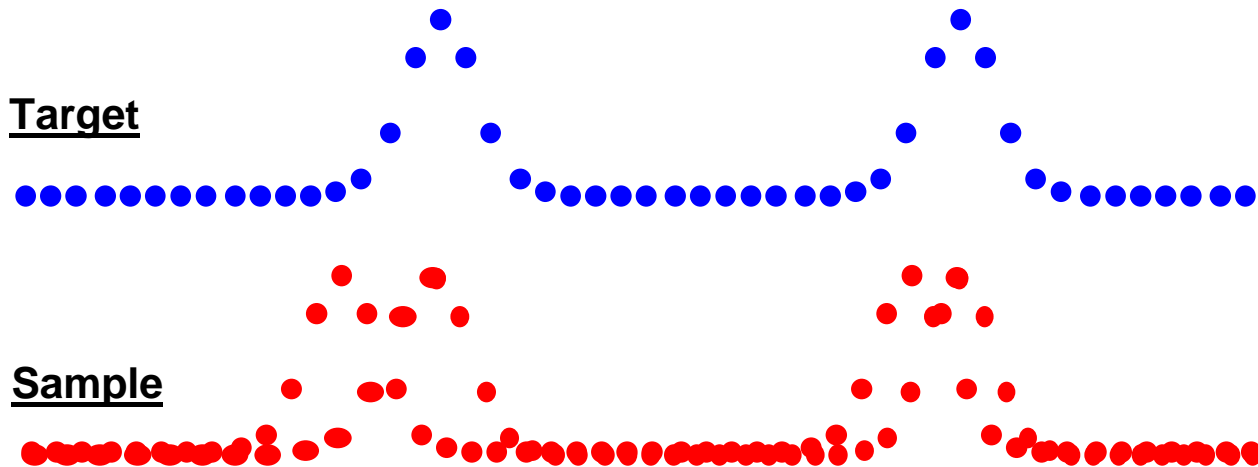
Alignment

- Algorithm – MBA (marker based alignment)
 - Target selection
 - One per material
 - Parameters
 - Marker retention values for 9 n-paraffins
 - Peak width – 51
 - Software: Pirouette (Infometrix)

Malmquist, G. *Chemometrics for Characterization, Classification and Prediction in Chromatography*, Ph D Dissertation, Uppsala, Sweden, 1993, p. 51.

Marker Based Alignment

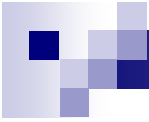
- Peg markers to same position, then interpolate point positions between markers
- Can re-use marker positions to align samples lacking marker peaks



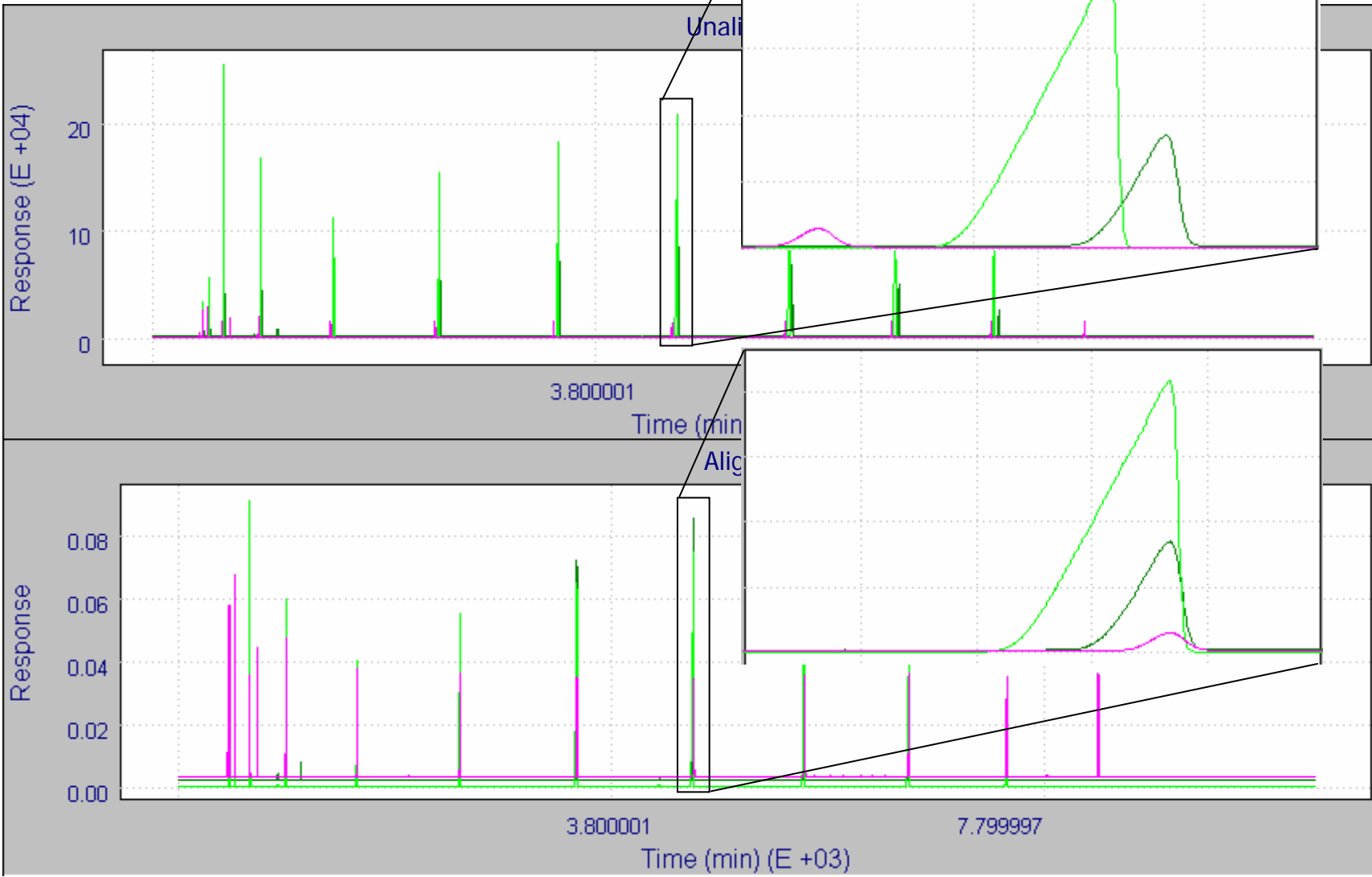


Alignment Procedure – 2 Stage

- For each target (of 3 materials)
 - identify immediate prior QC sample (same batch)
- Choose one material type as global target
 - Gasoline
- Align each QC sample to global QC sample (using MBA)
- Align each target to global QC sample using alignment positions from previous step
- (Re)Align sample profiles to aligned targets (using COW)

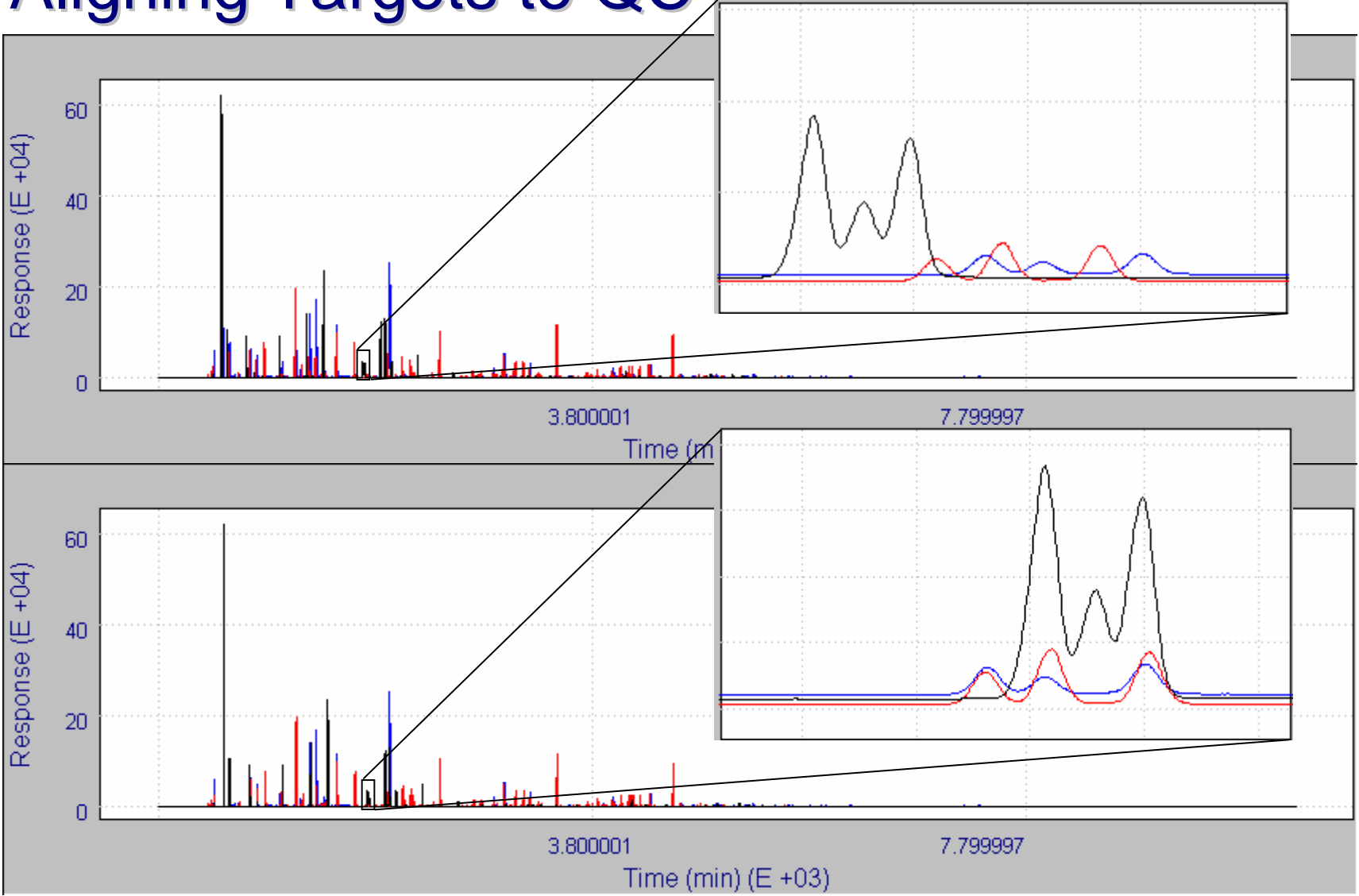


Aligning QC Samples



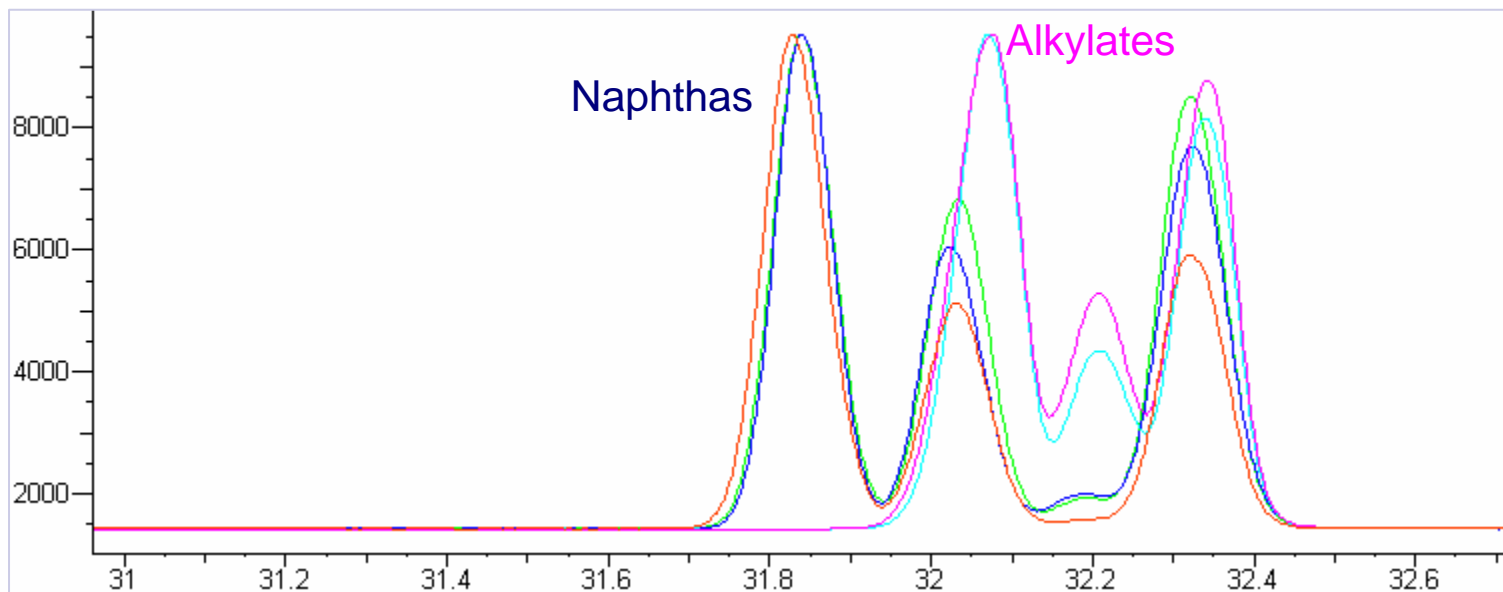


Aligning Targets to QC

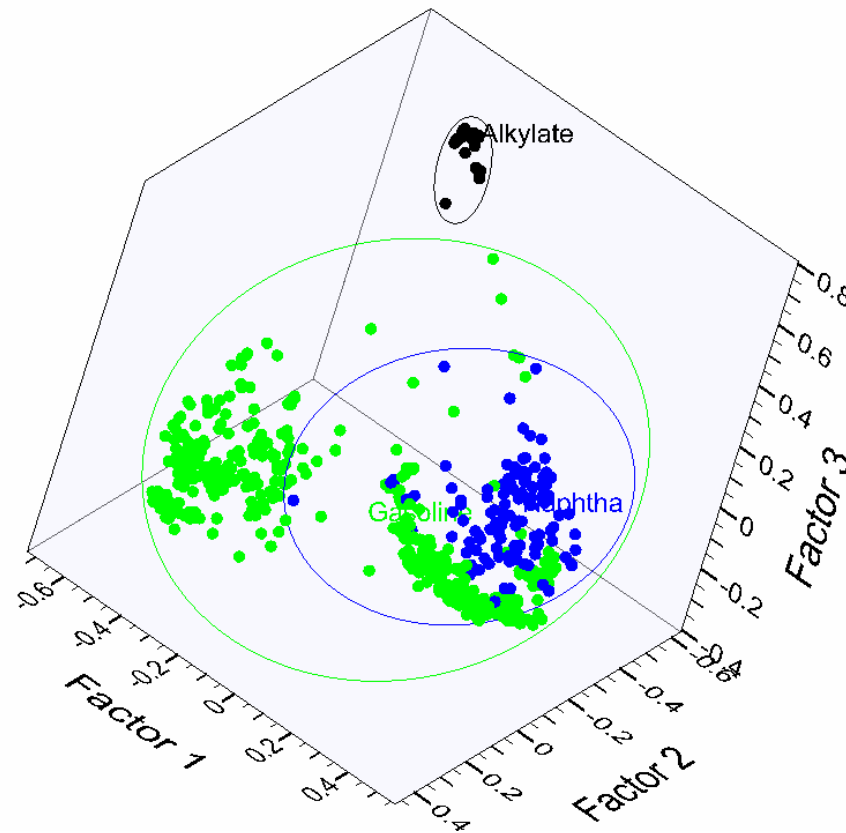


Database of Chromatograms

- KnowItAll database made for each material: Alkylates, Naphthas, Gasolines
- Combined into a master KIA database for all fuels
- Problem solved: align targets before aligning profiles



PCA Analysis of Aligned Fuel Profiles





Conclusions

- Correlation optimized warping works well for chromatograms collected on different instruments and columns
- Profiles of samples with few peaks in common can be aligned using marker positions from other samples (with sufficient peaks in common)
- Realignment of targets and references not necessary
- Process can be used to create and maintain corporate-wide databases
- Applicable to historical and future analyses



Acknowledgment

- Carl Rechsteiner
Chevron ETC
Richmond, CA