Streamlining Group Type Analysis with Standard GCxGC Templates through Computer Vision-Assisted Alignment

Daniel Geschwender, Qingping Tao, Chase Heble

GC Image, LLC (Lincoln NE, USA)

Outline

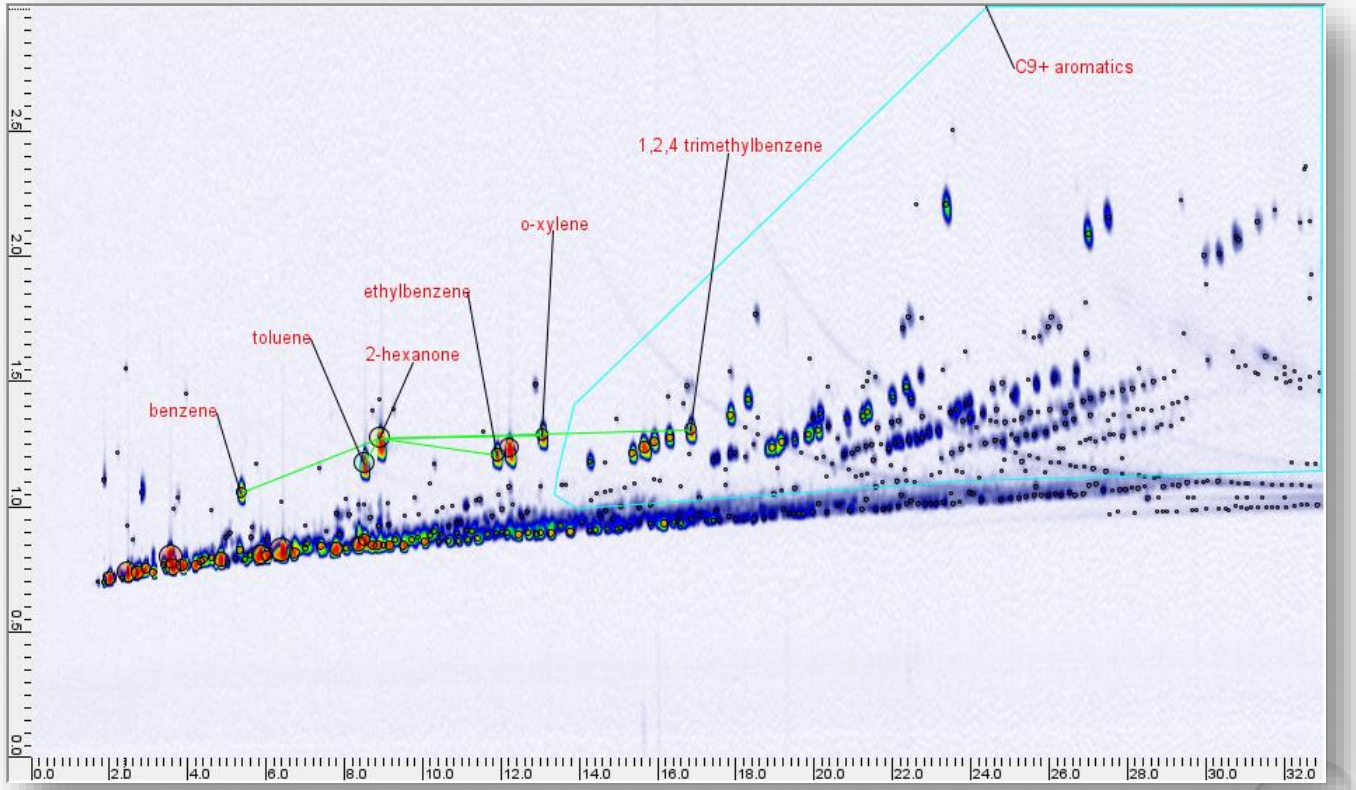
- Introduction
  - Template-based Methods
- A Workflow for Retention Time Alignment
  - Visual peak matching with local chromatographic patterns
  - Transformations for Alignment
  - Verification with group chromatographic patterns
- Conclusions

Introduction

- Hydrocarbon type analysis is an important type of characterization of petroleum products
  - To optimize production operations and quality control
  - To meet government regulations and environment standards
- Traditional methods for hydrocarbon type analysis are mostly based on one dimensional gas chromatography (GC)
  - There are known limitation of overlapping hydrocarbon types especially for C9 and heavier compounds
- Comprehensive two-dimensional gas chromatography (GCxGC) offers much greater separation capacity than traditional one dimensional GC.

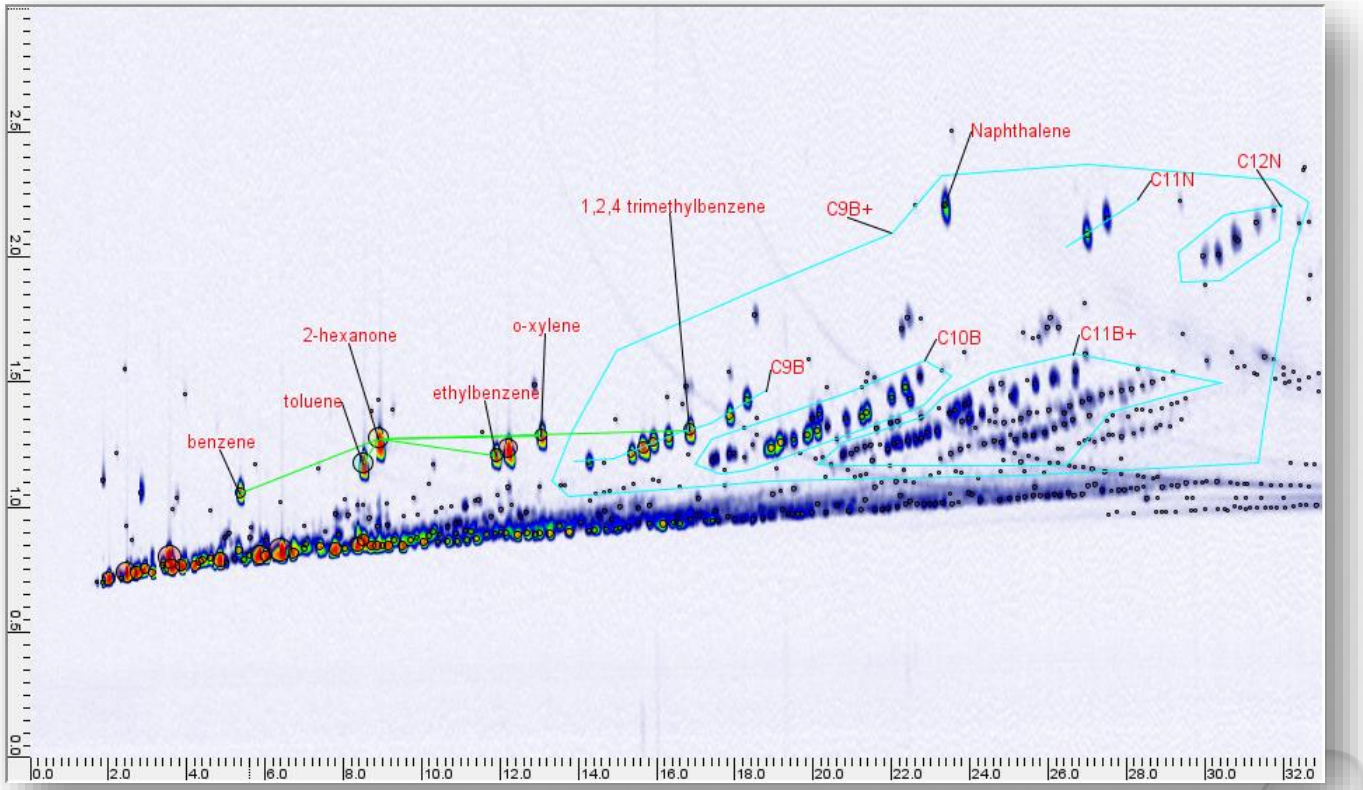
Hydrocarbon Type Analysis – Example 1

- A simple GCxGC analysis similar to ASTM D5580 - "Standard Test Method for Determination of Benzene, Toluene, Ethylbenzene, p/m-Xylene, o-Xylene, C9 and Heavier Aromatics, and Total Aromatics in Finished Gasoline by Gas Chromatography"
  - Identify peaks of benzene, toluene, ethylbenzene, p/m-xylene, o-xylene, …
  - Determine the RT region of C9 and heavier aromatics
    - "...Nonaromatic hydrocarbons having a boiling point greater than *n*-dodecane may cause interferences with the determination of the C_9 and heavier aromatics..."
  - Quantify responses with internal standard calibration



Hydrocarbon Type Analysis – Example 1 (Cont'd)

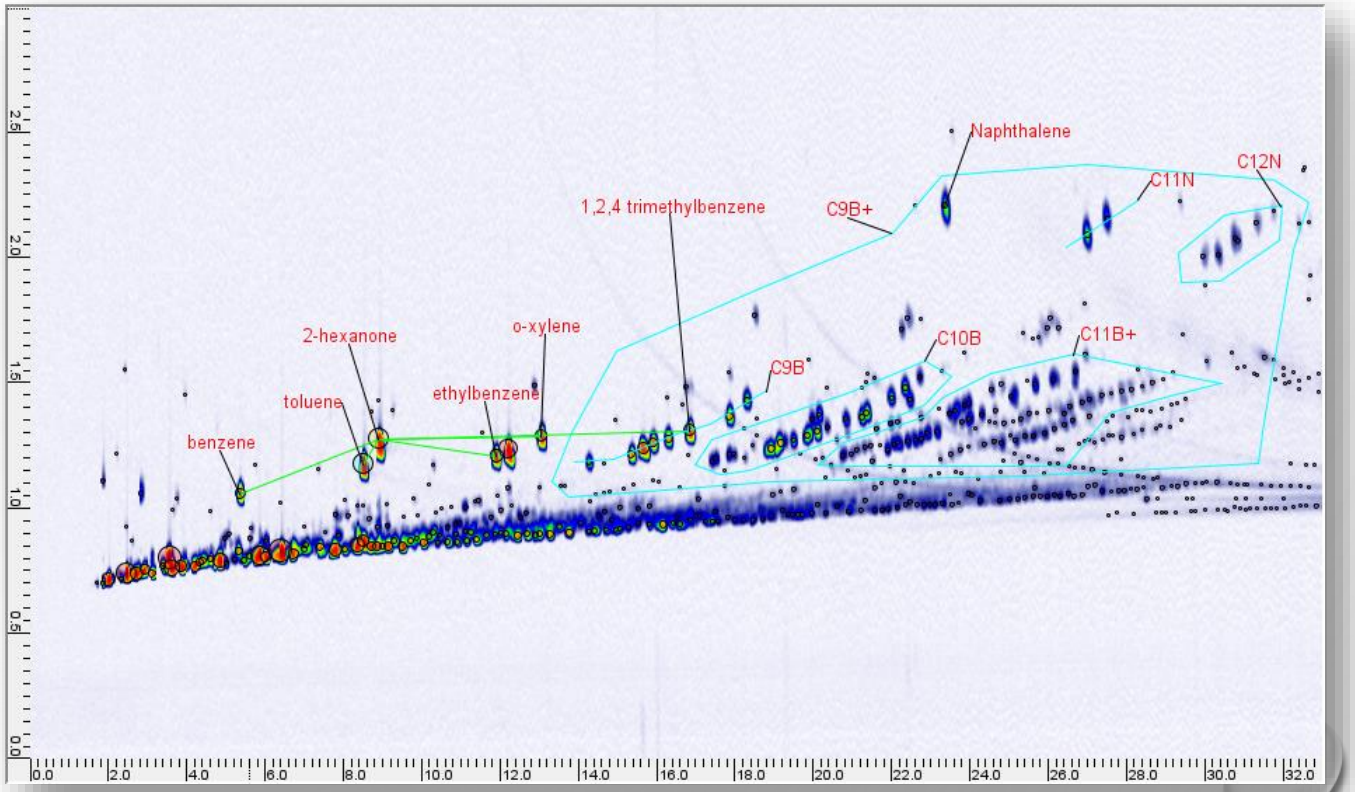
- More detailed type analysis can be achieved
- Automated analysis requires identifying peaks and chromatographic regions for compound groups



GC IMG

Automate Type Analysis with Template

- A template records:
  - Peaks including retention times, spectra, etc.
  - Chromatographic regions (as polygons) for compound groups
  - Chemical logic expressions for peak matching constraints & quality assurance (QA) assessment
  - Other metadata, e.g., descriptive annotations and additional chemical properties



Standard Method – ASTM D8396

- ASTM D8396-22 – "Standard Test Method for Group Types Quantification of Hydrocarbons in Hydrocarbon Liquids with a Boiling Point between 36°C and 343°C by Flow Modulated GCxGC–FID"
  - Identify 42 reference peaks
  - Apply a template with regions for paraffins, naphthenes, 1-R aromatics, 2-R aromatics, and 3-R aromatics to a chromatogram
  - With response factor calibration and quantification

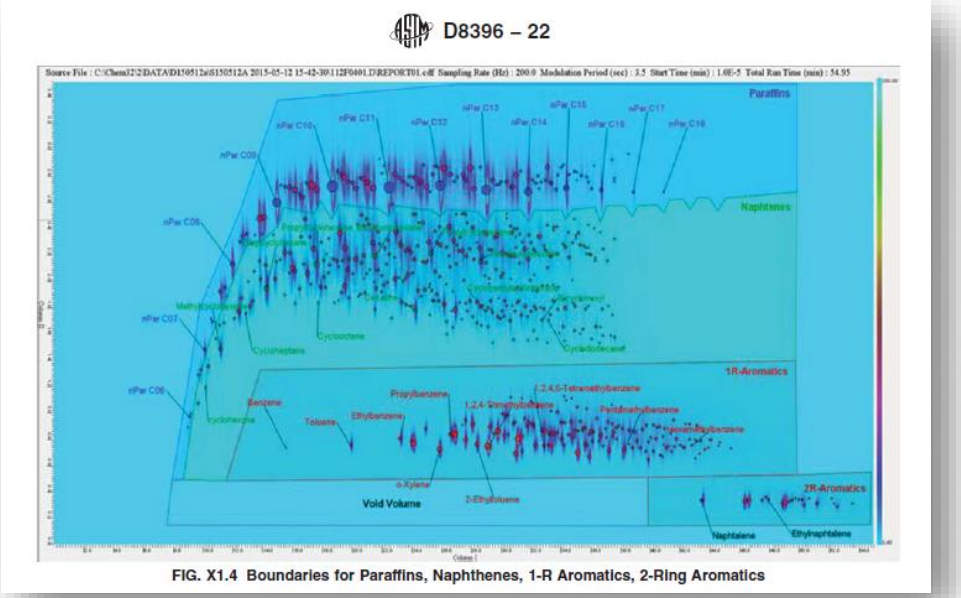
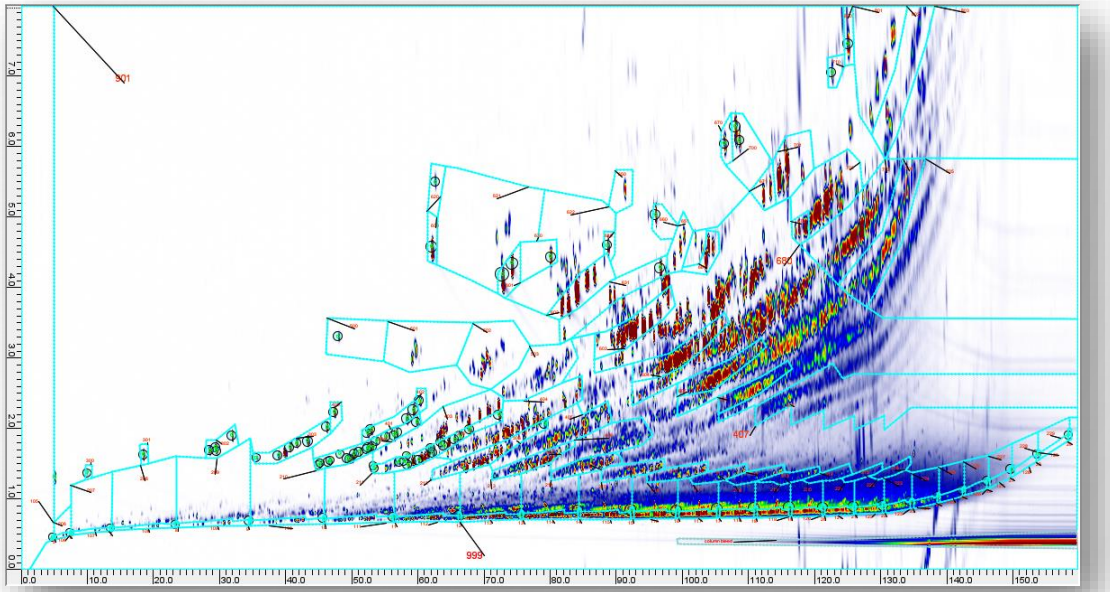


FIG. X1.4 Boundaries for Paraffins, Naphthenes, 1-R Aromatics, 2-Ring Aromatics

Targeted Analysis – UOP 990

- UOP 990-11 – "Organic Analysis of Distillate by Comprehensive Two-Dimensional Gas Chromatography with Flame Ionization Detection"
  - Identify a list of landmark peaks
  - Align a template of regions for n-Alkanes index, homologous series, and other specific molecular types to a chromatogram
  - With response factor calibration and quantification
- The analysis requires
  - Matching more than 90 peaks
  - Creating or aligning hundreds of RT regions
  - Very time consuming if done by hand

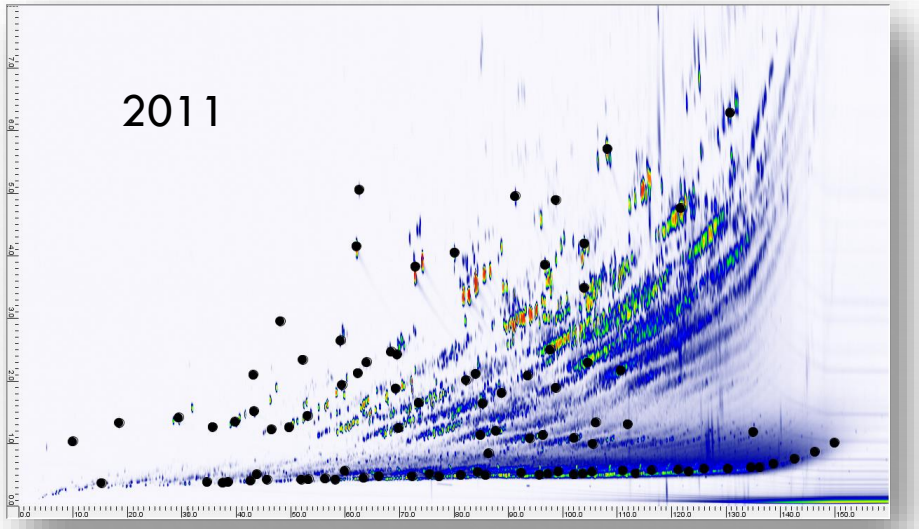


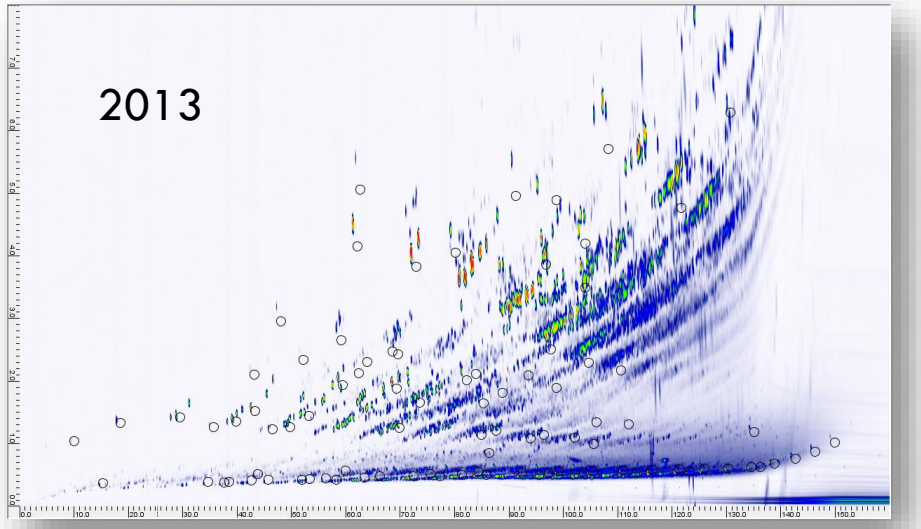
Automated Type Analysis - Challenges

- One of the main difficulties when analyzing GCxGC data is retention time variations due to instrumental conditions.
  - Run-to-run system variations (e.g. pressure and temperature fluctuations), and column aging
  - Retention times may vary between chromatograms, even when acquired on the same system.

RT Variation – Example 1

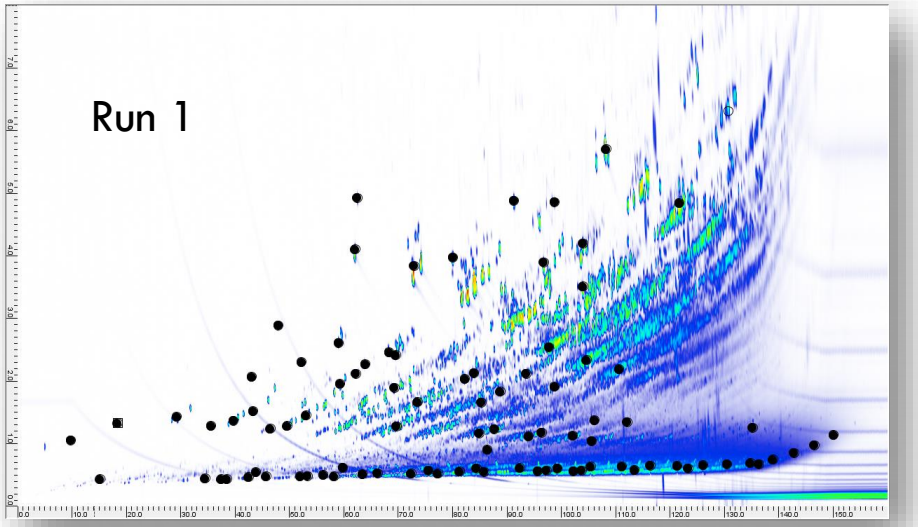
- A single distillate sample was run four different times on the same system over a period of about two and a half years (D. Rempe et al., Anal. Chem. 2016).
  - "Each of these runs were far apart in time, so the chromatograms have moderate misalignments from column differences, such as aging and replacement."

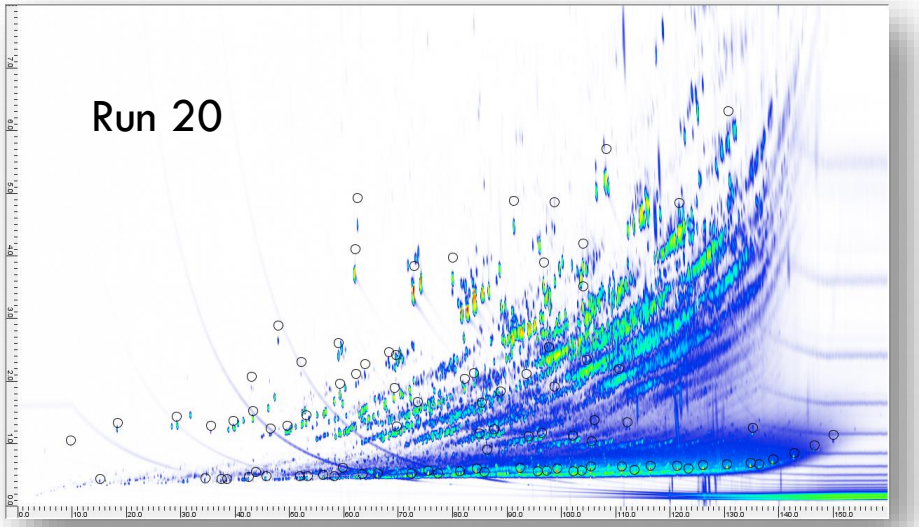




RT Variation – Example 2

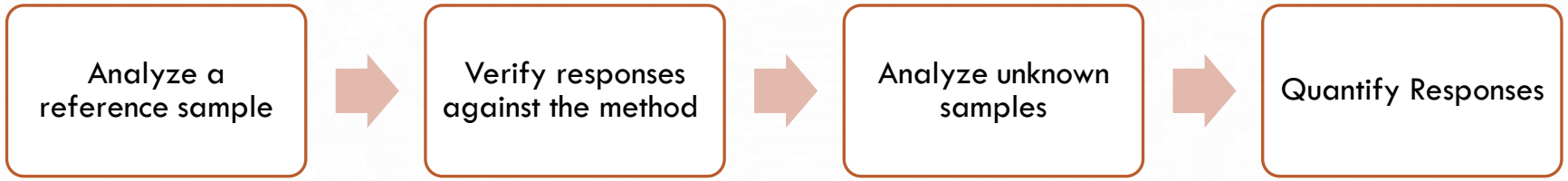
- A single distillate sample was run consecutively multiple times on the same system over a period of three days (D. Rempe et al., Anal. Chem. 2016).
  - "Misalignment between two replicate chromatograms acquired one after another with the same sample on the same system can be considered the level of random retention-times noise inherent to the system itself."





Workflow for Retention Time Alignment

- It is necessary to perform chromatographic alignment by mapping the retention times of one chromatogram to the times of another chromatogram, in order to perform routine analysis.
- The data processing procedure defined in the standard methods:

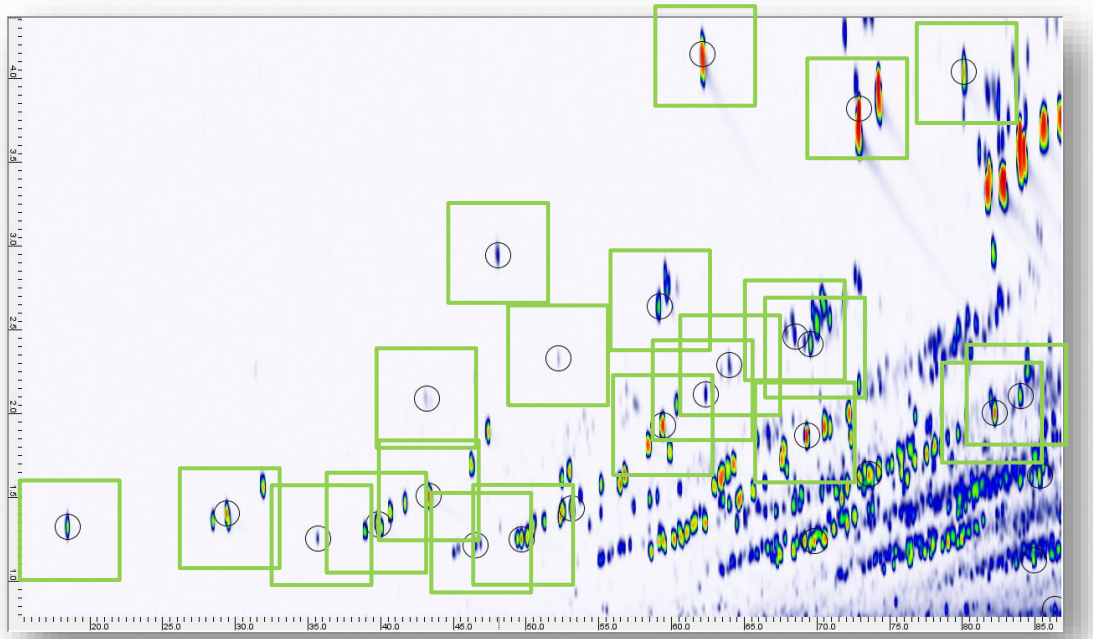


- The RT alignment procedure required at each analyzing step:

- The data processing software is better able to tolerate imperfection in data, and assist or reduce manual intervention.

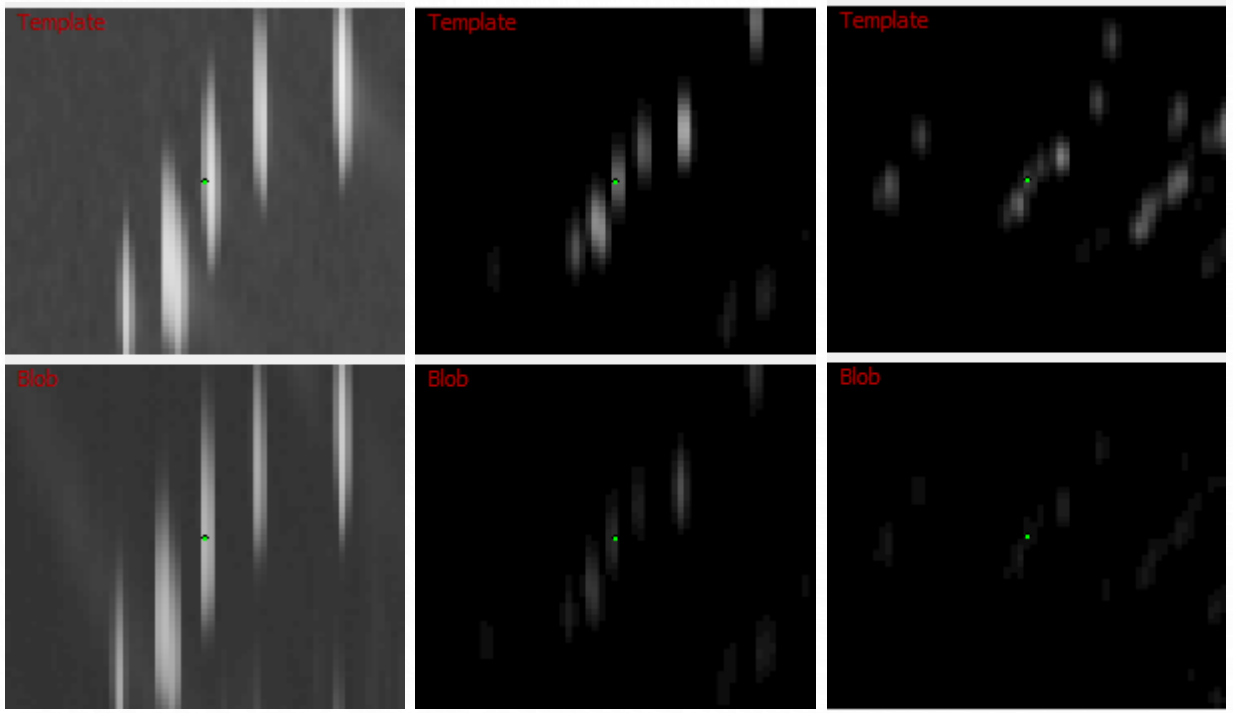
Visually Match Peaks with Chromatographic Pattern

- How to match reference peaks reliably?
  - Manual matching performed by the analyst
  - Auto matching with computer vision chromatographic pattern
- For some peak groups, elution order is consistent and provide easy identification.
- For some peaks, local chromatographic patterns are distinct and easily identifiable.
- Local chromatographic patterns may be extracted and used during matching.



GC IMG

Enhanced Correlation Coefficient for Visual Match



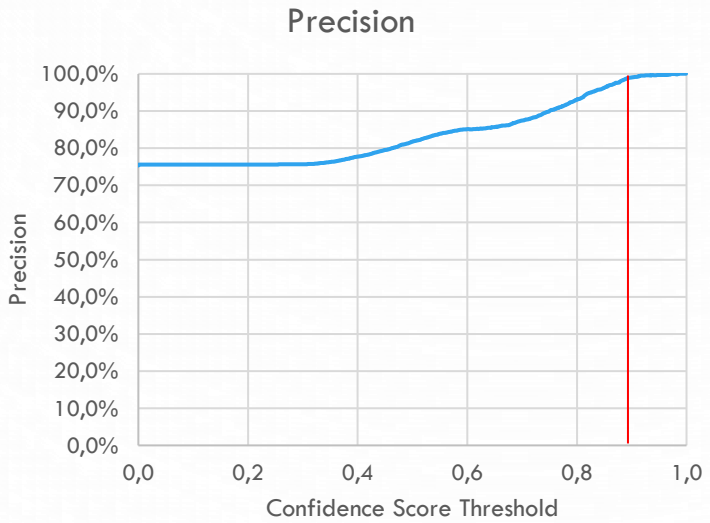
- **Enhanced Correlation Coefficient (ECC)** is a measure of image similarity from computer vision (E.Z. Psarakis et al., ICCV 2005)
- It is invariant to photometric distortions (lighting and contrast)
  - Helps with handling intensity variations
- We compute at multiple different image scales and log/linear/exponential mappings

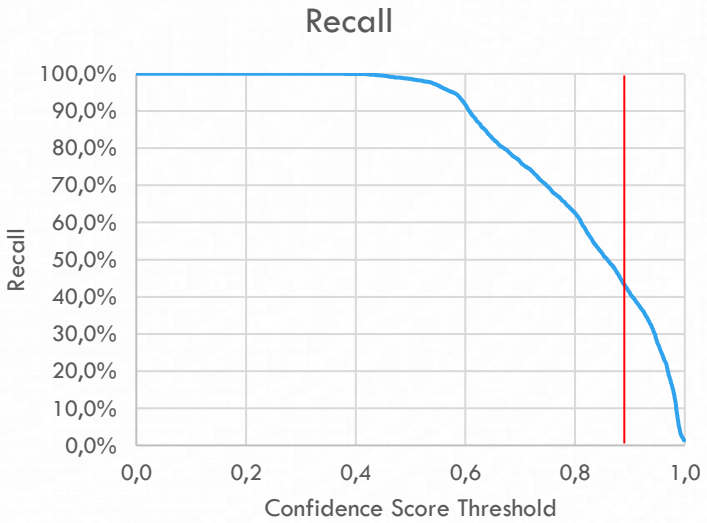
Visual Match Evaluation

- Evaluate visual matching using ECC on 166 chromatograms with 7618 individual peaks considered
- The correct match is ranked as the top scoring hit 82.2% of the time
- A confidence score is computed for estimating the likelihood of the top match being correct

Correct Match in	% of Cases*
Top 1	82.2%
Top 5	92.5%
Top 10	94.7%

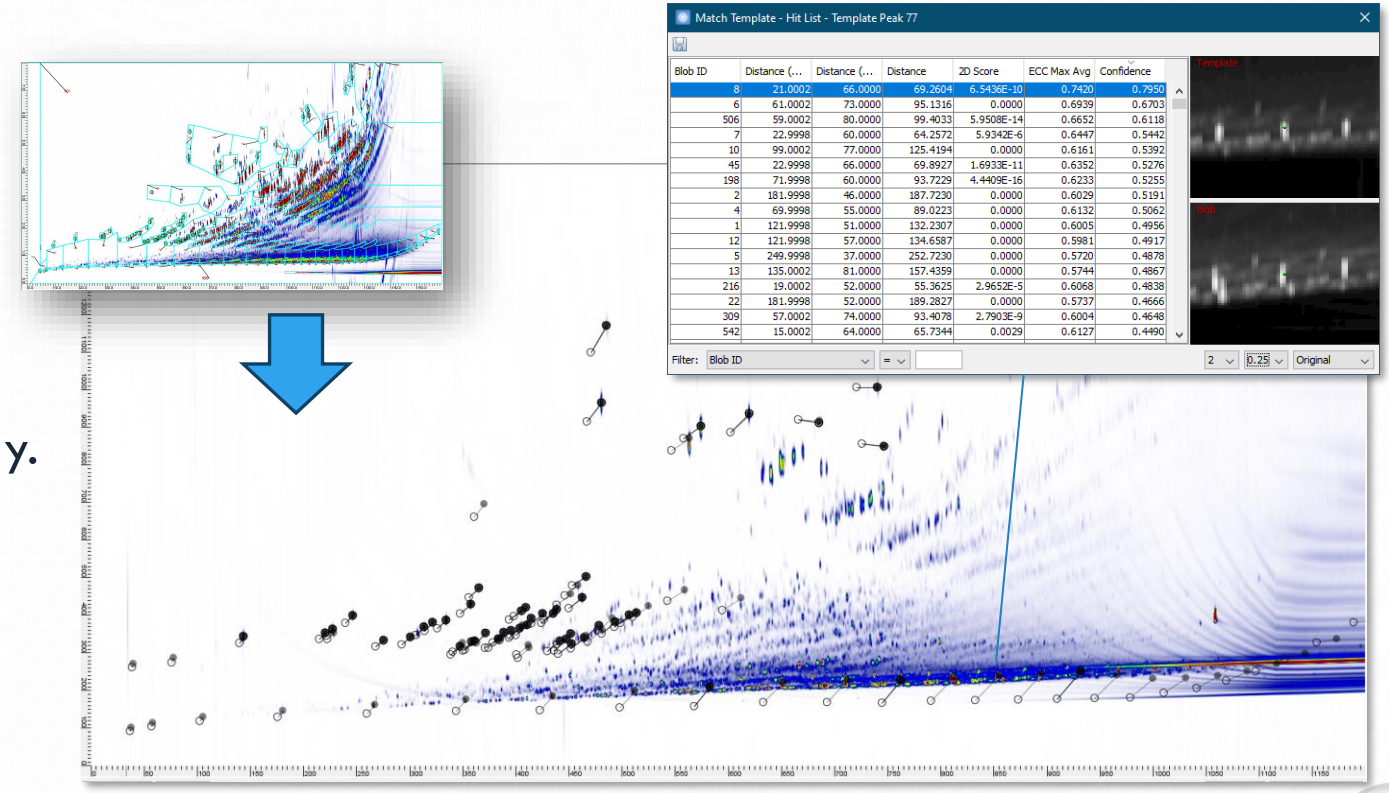
* Out of cases with a valid match





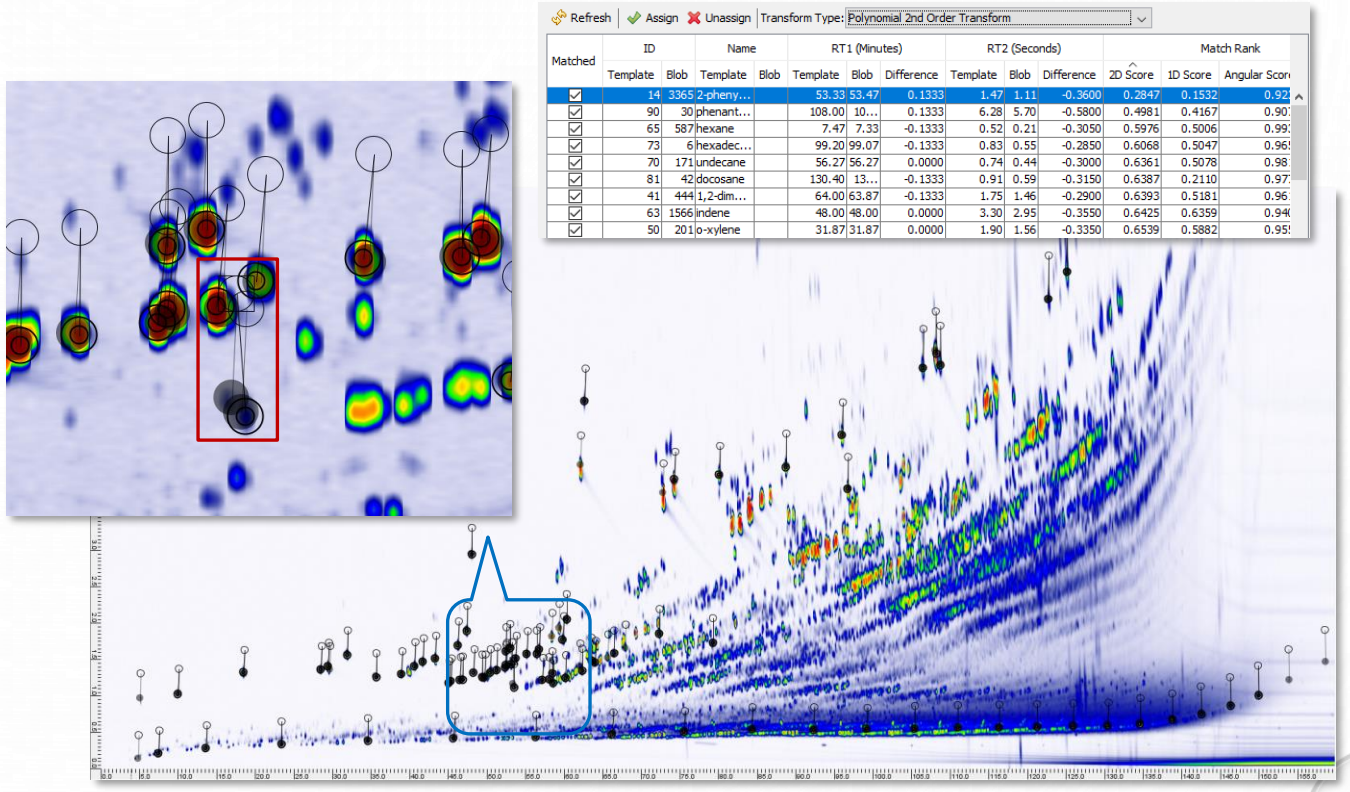
Match Example – Inter Lab

- The reference template and the chromatogram were acquired by two different labs.
- Well separated reference peaks with distinct pattern are still matched correctly.
- Some mismatches and unmatched peaks need to be manually corrected
- Ranking candidate matches by visual match score speeds the process



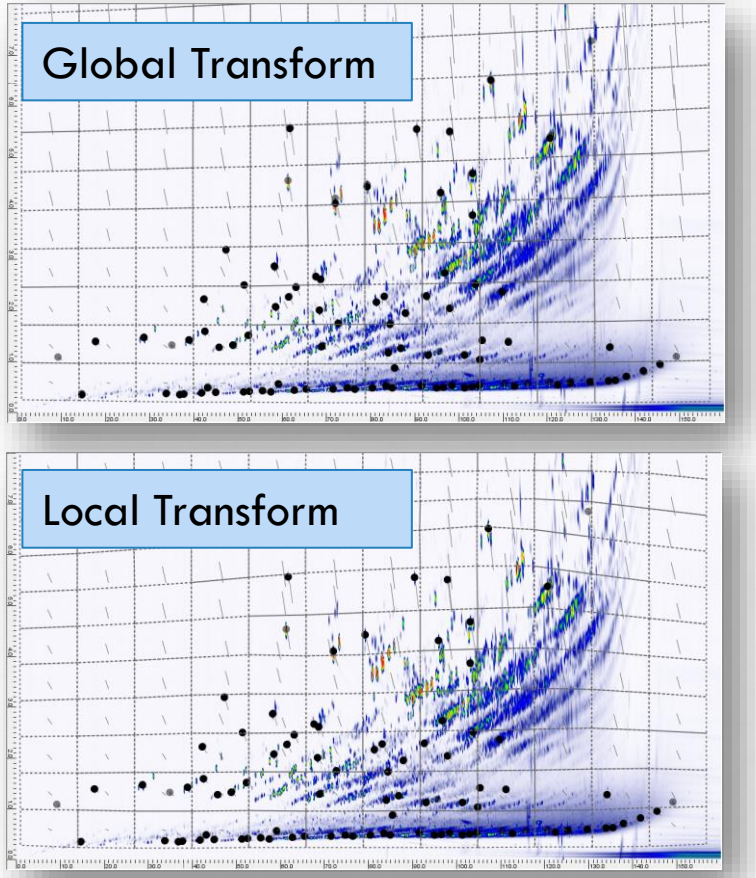
Match Example – Intra Lab

- The reference template and the chromatogram were acquired by the same lab.
- Most of reference peaks are matched correctly.
- Match scores can be used to rank and screen unreliable matches
  - 2D Score – a normalized distance score between the transformed peak and the matched peak.



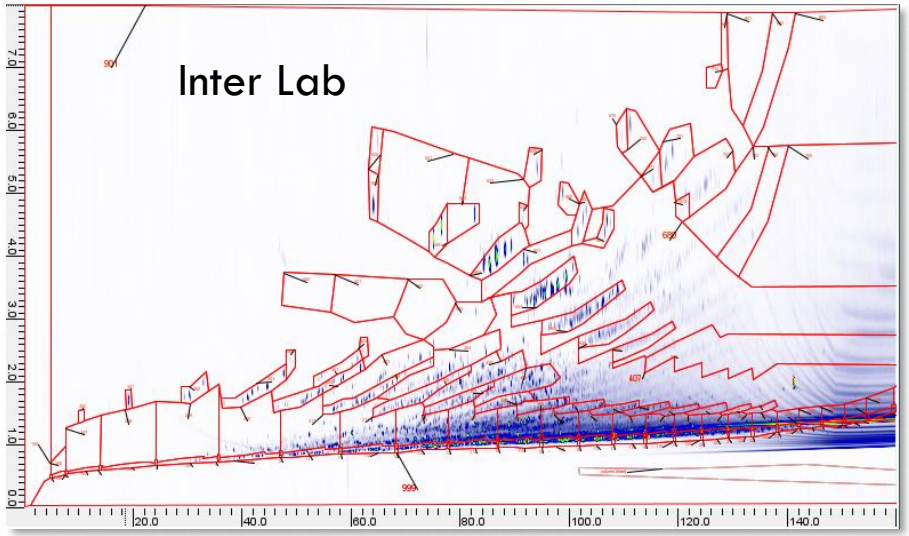
Transform and Apply Template

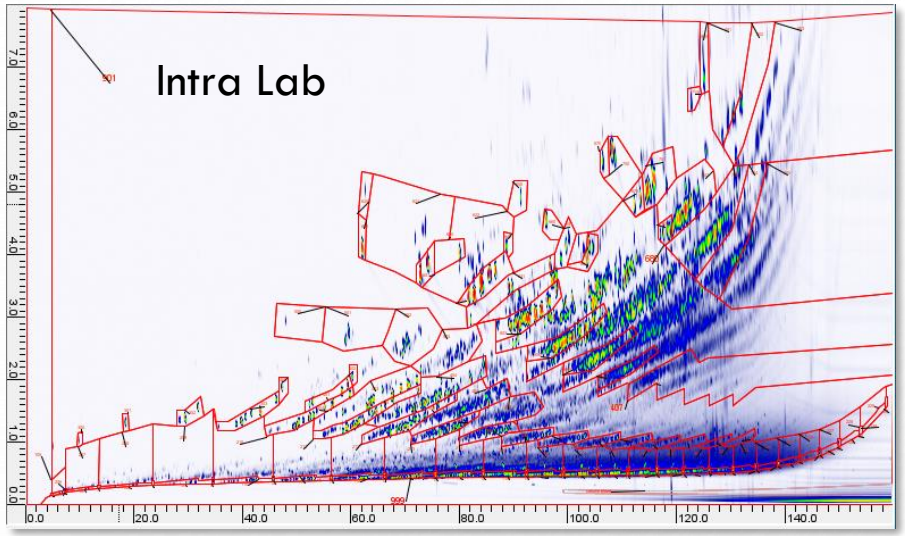
- Transformation indicated by correctly matched peaks can be useful to guide the matching of other peaks and groups.
- Global Transformation
  - A single function for the entire chromatogram, e.g. affine or 2nd order polynomial.
  - Global functions may be able to capture systemic properties and structure that underlie retention-time differences.
- Local Transformation
  - A combination of many functions for different regions of the chromatogram.
  - Local functions may be able to capture retention-time variations that are not related to systemic properties and structure.



Transform and Apply Template - Results

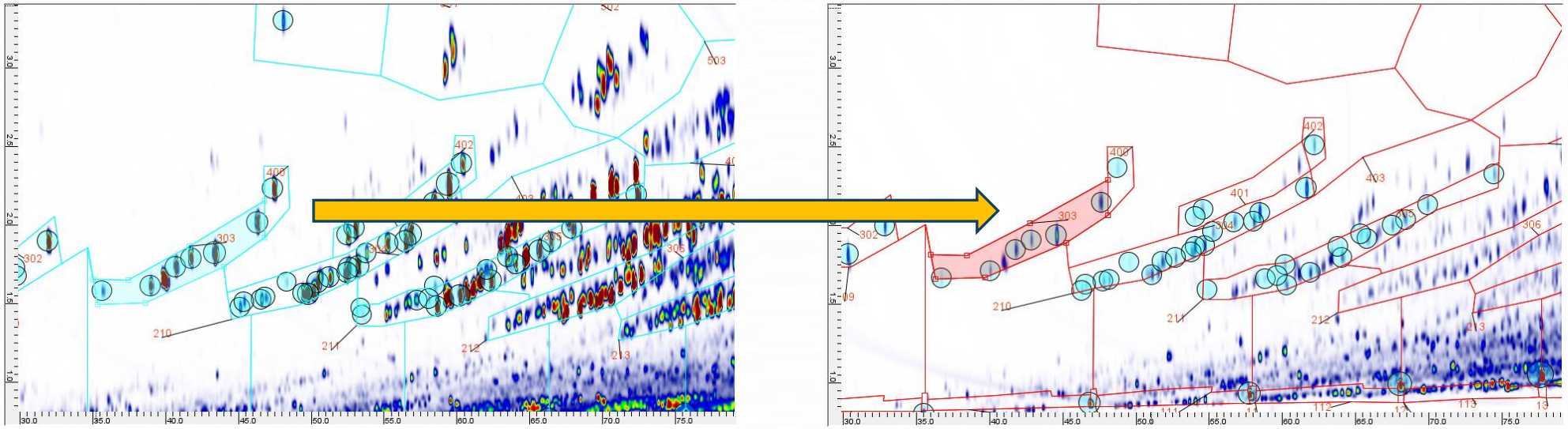
- The template is transformed with a combination of global and local transforms.





Verify Groups with Chromatographic Patterns

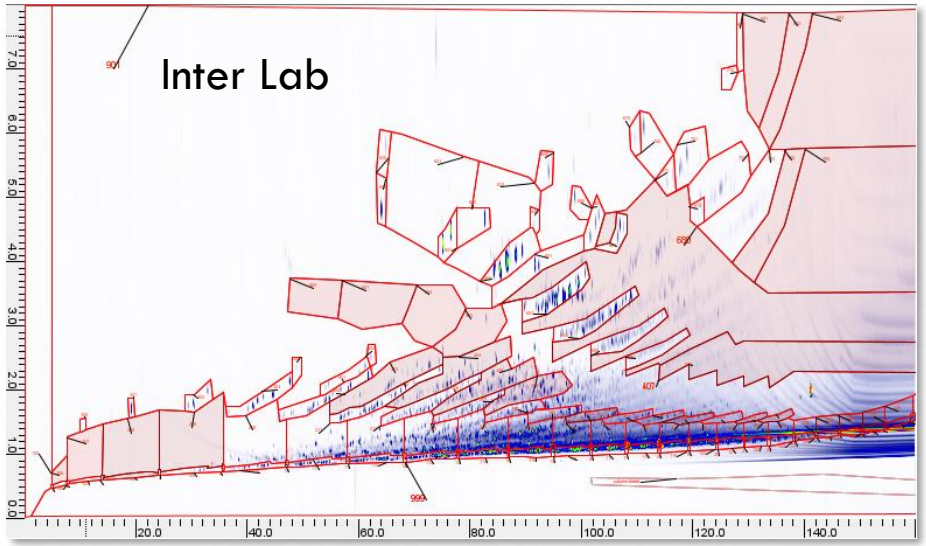
- Similar to peak matching, chromatographic patterns may be extracted and used to verify group boundaries.

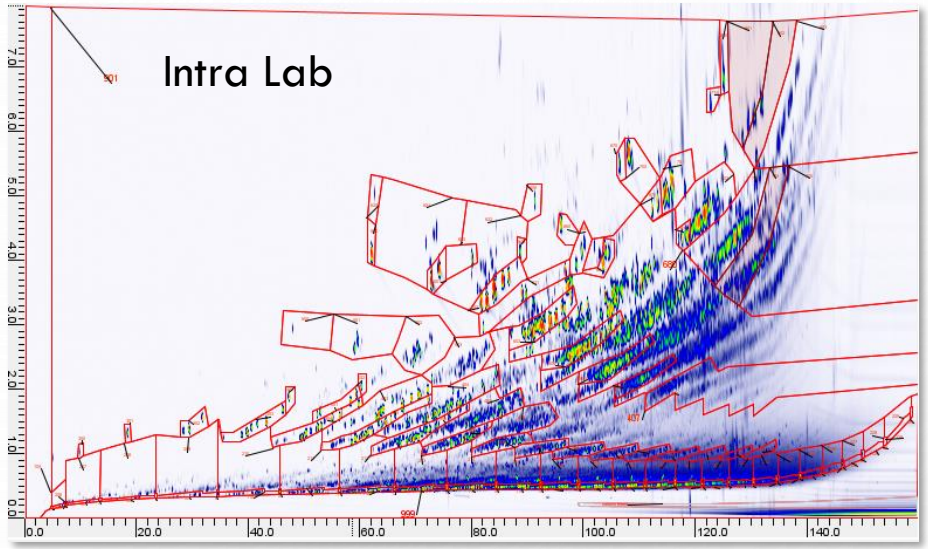


Comparing corresponding groups between the reference and a new sample

Verify Groups with Chromatographic Patterns - Results

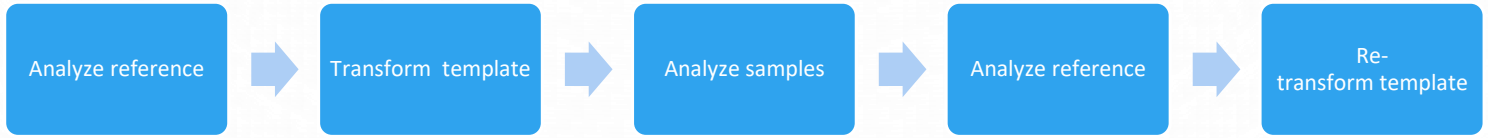
- A normalized correlation score is calculated between standard and reference samples.
- Groups with low correlation scores are highlighted.





Conclusions

- With chromatographic pattern matching and RT transformations,
  - A standard template can be transformed to a reference chromatogram acquired by a lab.
  - The standard template can be re-transformed to a newly acquired reference chromatogram by the same lab easily.
- Matching scores can guide an analyst to verify matching and transformation results.
- A routine RT alignment workflow is possible by combining these tools.



THANK YOU

Questions?