Retention-Time Based Peak Clustering in Comprehensive Two-Dimensional Gas Chromatography

Qingping Tao, GC Image LLC, Lincoln, NE, USA

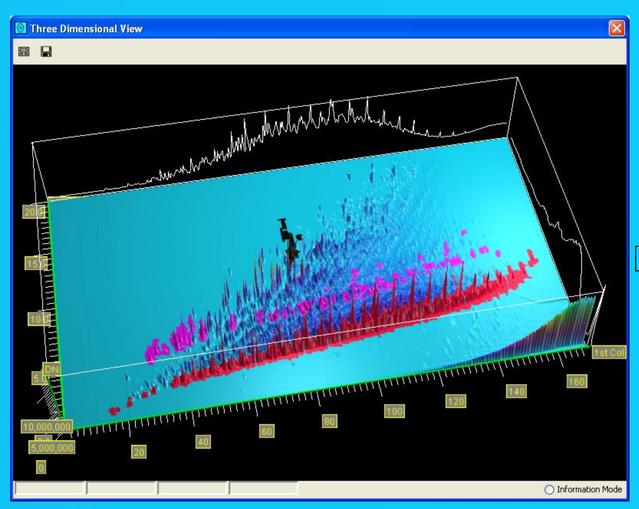
Stephen E. Reichenbach, Shilpa L. Deshpande

University of Nebraska, Lincoln, NE, USA

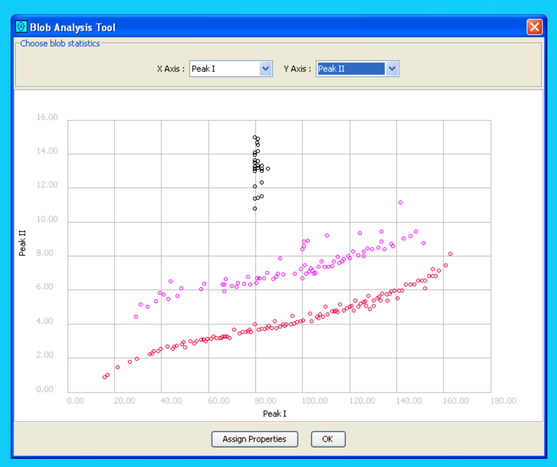
Contact: GC Image LLC, PO Box 57403, Lincoln NE 68505-7403; Phone: +1.402.310.4503;
Email: info@gcimage.com, qtao@gcimage.com, reich@cse.unl.edu

- Goal: Sample identification and chemical group analysis
  - –GCxGC analyses utilize the inherent relationship between chemical structure and peak position on the retention-time plane.
  - –A template that captures the pattern of peaks in the retention-time plane can be used for registering GCxGC analyses in order to identify chemicals in a sample or to identify similarities and differences between samples.
- Method: Automate construction of templates with retention-time based peak clustering.
  - –Find natural clusters in GCxGC data with the Complete Linkage with PCA algorithm.
  - –Adjust clustering with the Primary Column Weight.
  - –Identify markers for chemical groups.

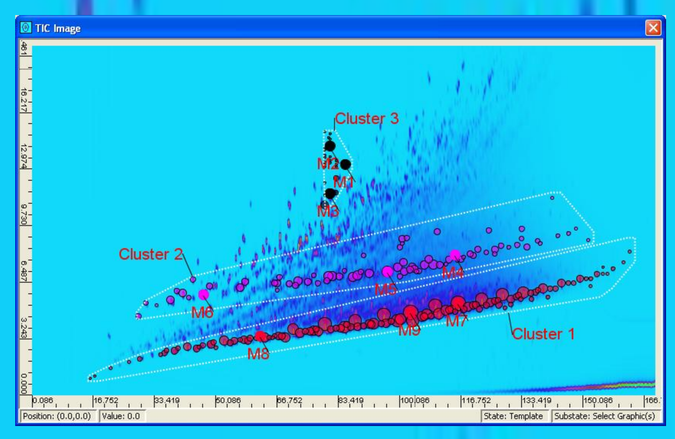
Automated Template Construction with Clustering



Colorized Diesel Chromatogram with Characteristic MS Grouping



Peak Retention Pattern of Groups



Template Constructed with Clustering and Markers

Cluster 1 (**red**) : Ordinal(57)<5&Ordinal(71)<5&Ordinal(43)<5, Cluster 2 (**purple**) : Ordinal(105)<=2|Ordinal(106)<=2, Cluster 3 (**black**) : Ordinal(156)=1

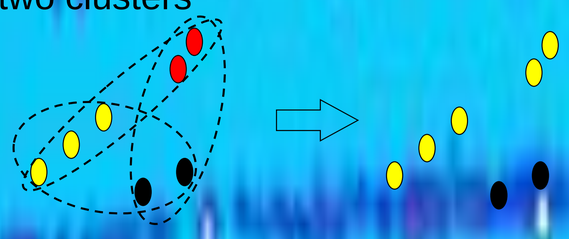
Complete Linkage with Principle Component Analysis

- The objective is to group chemical peaks into “clusters” based on retention times
- Natural clusters of GCxGC data commonly appear as striated bands .
- Complete linkage with PCA is a hierarchical clustering algorithm that uses the “retention-time area”, calculated by the PCA of two clusters as the proximity measure.
- The algorithm computes the covariance matrix of the peaks in PC space. The area is equal to the square root of the determinant of the covariance matrix
- The algorithm finds natural clusters in GCxGC data.

Algorithm:

- Initially, each cluster contains one peak with Gaussian variance.

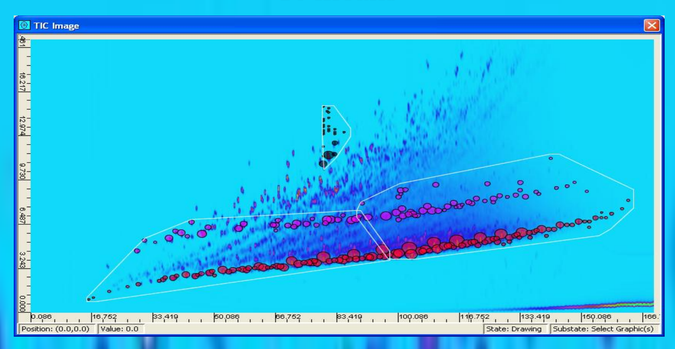
- Choose the pair of the clusters which, when merged, has the least area.
- Merge these two clusters



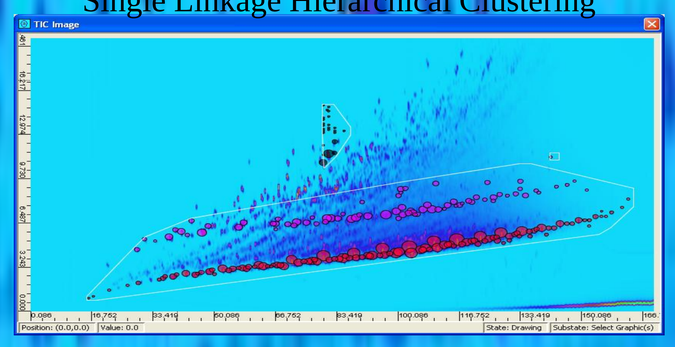
- Repeat Steps (2) and (3) until the desired number of clusters are found.

Comparison on Clustering Algorithms on GCxGC

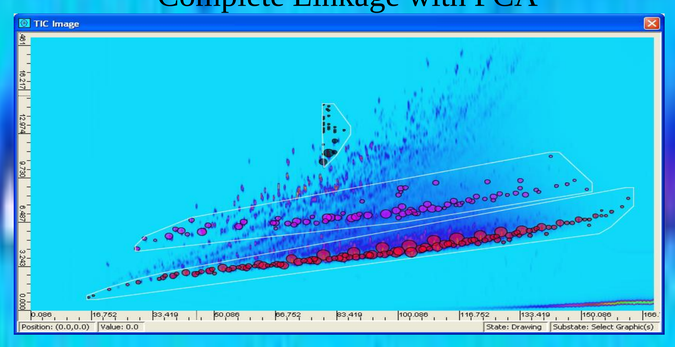
K-means



Single Linkage Hierarchical Clustering



Complete Linkage with PCA



GC Image
Software for Comprehensive Two-Dimensional Gas Chromatography

University of Nebraska Lincoln