Comparative Analysis of Peak Detection Techniques for Comprehensive Two-Dimensional Gas Chromatography

Indu Latha[1], Stephen E. Reichenbach[1], Qingping Tao[2]

[1]Computer Science & Engineering Dept., University of Nebraska – Lincoln; [2]GC Image, LLC

Email: ilatha@cse.unl.edu, reich@cse.unl.edu , qtao@gcimage.com

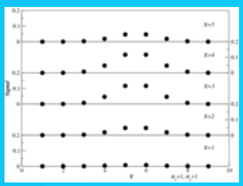
Objective: Evaluate the effect of retention-time shifts on 2D peak detection algorithms

– Peak detection aggregates data points of analyte peaks based on retention times and intensities.
– Undesirable second-column retention-time shifts can degrade the performance of two-dimensional (2D) peak detection algorithms.
– This research conducted experiments to compare performance of two popular 2D peak detection algorithms with shift correction.

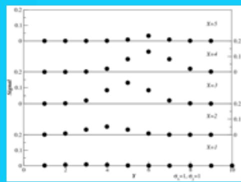
Peak Detection Algorithms

– Two-step algorithm: One-dimensional (1D) peak detection on each secondary chromatogram followed by merging detected 1D peaks.
– Watershed algorithm: Peak detection on 2D neighborhoods in both retention-time dimensions simultaneously.

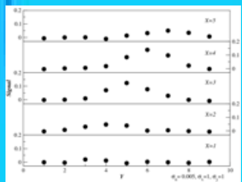
Simulation of Two-Dimensional Chromatograms to Compare Peak Detection Algorithms



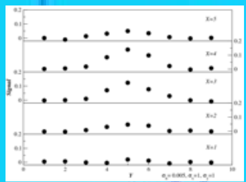
Slices of a sampled, simulated 2D peak displaying each secondary 1D peak



Shifted slices incorporating a skew in the 2D peak



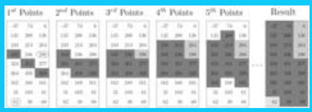
Slices of a skewed 2D peak with random Gaussian noise



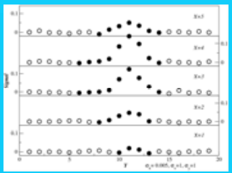
Slices of skewed 2D peak with noise after skew correction

Two-Step Algorithm

– Identify and correct retention-time shift for skewed peaks using cross correlation.
– Perform 1D peak detection on each secondary chromatogram.
– Identify largest peak apex in the 2D chromatogram.
– Peak merging with the overlap and unimodality constraints.



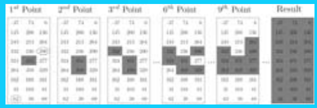
Each column is a secondary chromatogram. Points included in the main peak are shown in dark gray and other points are shown in light gray.



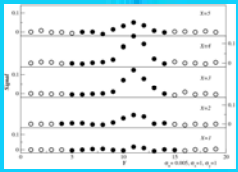
Peak detected by two-step algorithm.

Watershed Algorithm

– Identify and correct retention-time shift for skewed peaks using cross correlation.
– Identify and label largest data point in the 2D peak.
– Compare each data point with its neighbors.
  – Give a new label if data point is large than its neighbors.
  – Otherwise, give the same label as its largest neighbor.



Data points are labeled in intensity order in the 2D chromatogram.



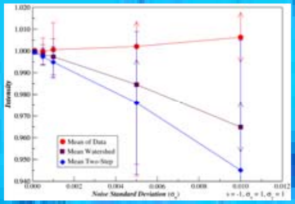
Peak detected by watershed algorithm.

Experimental Results

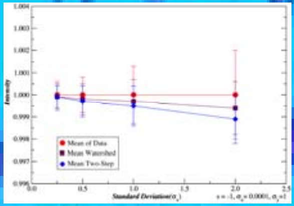
– Four parameters are varied:
  – Noise standard deviation, σ_n, from 0.0001 to 0.01.
  – Peak-width standard deviation along the x-dimension, σ_x, from 0.25 to 2.00.
  – Peak-width standard deviation along the y-dimension, σ_y, from 1 to 8.
  – Skew, s, from -8 to -1.
– Each experiment is conducted 1000 times.
– Compare intensity mean and standard deviation of 2D peaks for the two-step and watershed algorithms.
– Results indicate that the watershed algorithm has better accuracy than the two-step approach when skew correction is applied for both methods.
– Statistical significance tests suggest that the results of watershed algorithm are significant and would be observed in repeated experiments.

Results for peak detection algorithms with various noise, σ_n, and peak widths, σ_x and σ_y.

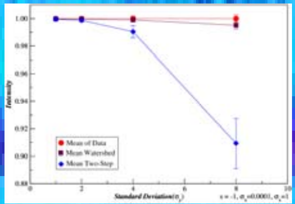
Skew s	Peak σ_x	Peak σ_y	Noise σ_n	Array Size	Signal Mean	Signal Stdv	WS Mean	WS Stdv	WS Error	WS Failed	2-Step Mean	2-Step Stdv	2-Step Error	2-Step Failed	Signif. $(1-p)$
-1	1.00	1	0.0001	11×21	1.0000	0.0013	0.9997	0.0010	-0.0003	0	0.9995	0.0009	-0.0005	0	1.0000
-1	1.00	1	0.0005	11×21	0.9999	0.0061	0.9984	0.0045	-0.0015	0	0.9974	0.0040	-0.0024	0	1.0000
-1	1.00	1	0.0010	11×21	1.0006	0.0124	0.9974	0.0082	-0.0032	0	0.9949	0.0073	-0.0057	0	1.0000
-1	1.00	1	0.0050	11×21	1.0021	0.0597	0.9846	0.0368	-0.0174	0	0.9761	0.0328	-0.0260	0	1.0000
-1	1.00	1	0.0100	11×21	1.0063	0.1227	0.9649	0.0706	-0.0415	0	0.9451	0.0598	-0.0612	0	1.0000
-1	0.25	1	0.0001	5×15	1.0000	0.0006	0.9999	0.0006	-0.0001	0	0.9999	0.0005	-0.0001	0	0.0000
-1	0.50	1	0.0001	7×17	1.0000	0.0008	0.9998	0.0007	-0.0001	0	0.9997	0.0007	-0.0003	0	0.9986
-1	1.00	1	0.0001	11×21	1.0000	0.0013	0.9997	0.0010	-0.0003	0	0.9995	0.0009	-0.0005	0	1.0000
-1	2.00	1	0.0001	20×30	1.0000	0.0020	0.9994	0.0012	-0.0006	0	0.9989	0.0011	-0.0011	0	1.0000
-1	1.00	1	0.0001	11×21	1.0000	0.0013	0.9997	0.0010	-0.0003	0	0.9995	0.0009	-0.0005	0	1.0000
-1	1.00	2	0.0001	11×30	1.0000	0.0015	0.9996	0.0012	-0.0003	0	0.9987	0.0013	-0.0013	0	1.0000
-1	1.00	4	0.0001	11×48	1.0000	0.0019	0.9990	0.0015	-0.0010	0	0.9904	0.0045	-0.0096	0	1.0000
-1	1.00	8	0.0001	11×84	1.0000	0.0028	0.9951	0.0028	-0.0048	39	0.9093	0.0182	-0.0906	341	1.0000



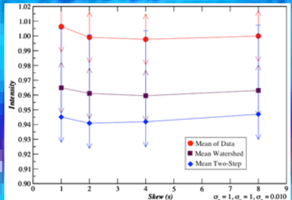
Performance of peak detection algorithms as a function of noise standard deviation, σ_n.



Performance of peak detection algorithms as a function of first-column peak width, σ_x.



Performance of peak detection algorithms as a function of second-column peak width, σ_y.



Performance of peak detection algorithms as a function of skew, s.

University of Nebraska Lincoln Department of Computer Science & Engineering

GC IMAGE Software for Multidimensional Chromatography

Water, Energy and Agriculture Initiative WEAI

University of Nebraska Lincoln Nebraska Center for Energy Sciences Research