# Comparative Analysis of Peak Detection Methods for Comprehensive Two-Dimensional Chromatography

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

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

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

## Objective: Evaluate the performance two-dimensional peak detection algorithms

– Peak detection aggregates data points of analyte peaks based on retention times and intensities.

– Two most common two-dimensional (2D) peak detection algorithms: Two-Step algorithm and Watershed algorithm.

– Vivó-Truyols and Janssen [*J. Chromatography A*, **1217**:1375-1385, 2010] showed that undesirable shifting of second-column retention times can degrade the performance of 2D peak detection algorithms. They accounted for shift in the Two-Step algorithm but not with the Watershed algorithm.

– This research conducted experiments to compare performance of these 2D peak detection algorithms with shift correction for both algorithms.

## Results: Watershed algorithm outperforms Two-Step algorithm for 2D peak detection

– Watershed algorithm is consistently more accurate for 2D peak detection with various levels of noise, peak widths, and retention-time shifts.

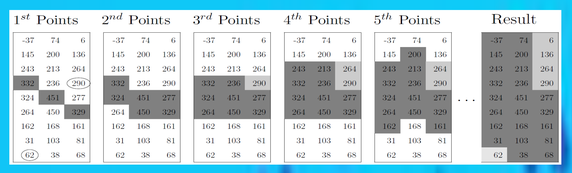
## 2D Peak Detection Algorithms

– **Two-Step** algorithm: One-dimensional (1D) peak detection on each secondary chromatogram followed by merging detected 1D peaks.

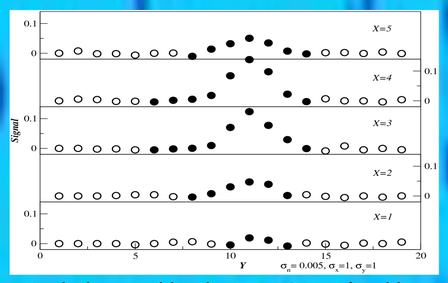
– **Watershed** algorithm: Peak detection on 2D neighborhoods in both retention-time dimensions simultaneously.

## Two-Step Algorithm

1. Perform 1D peak detection on each secondary chromatogram.
2. Merge detected 1D peaks subject to overlap and unimodality constraints.
   – Overlap constraint parameterizes allowable shifts of the second-column retention times for merging1D peaks.
   – Unimodality constraint ensures that each 2D peak has a single apex.




Progressive operations of the Two-Step algorithm: Each column of data is a secondary chromatogram. Points included in the main peak are shown in dark gray and other points are shown in light gray.



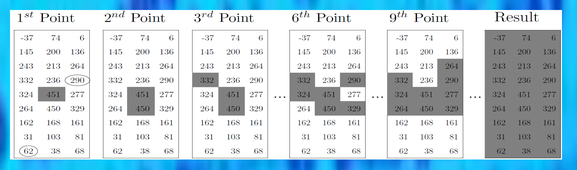

Peak detected by the Two-Step algorithm shown with filled circles.

## Watershed Algorithm

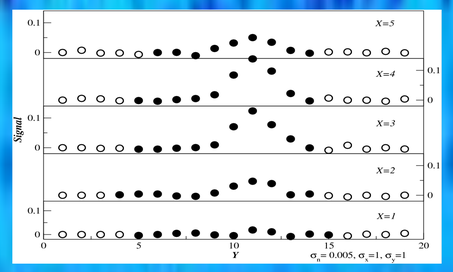
Traditional Watershed algorithm is inverted to find peaks rather than basins; Drain algorithm [Reichenbach *et al.*, *Chemo. Intell. Lab. Sys.*,**71**:107-120, 2004].

1. Data points that have the largest value in their neighborhood indicate a new peak.
2. Other data points belong to the same peak as the largest of their neighbors.

Retention-time shifts can be corrected for by shifting the data before peak detection or by adjusting the neighborhood.




Progressive operations of the Watershed algorithm: Data points are labeled in intensity order in the 2D chromatogram.




Peak detected by watershed algorithm shown with filled circles.

# Simulation of Two-Dimensional Chromatograms to Compare Peak Detection Algorithms

Simulation allows controlled experimentation with varying levels of noise, peak widths, and retention-time shifts.

1. Two-dimensional, standard Gaussian peak model, centered at ($\mu_x$,$\mu_y$), with unit-integral sampling, parameterized by:
   - First-column peak width, $\sigma_x$.
   - Second-column peak width, $\sigma_y$.
2. Second-dimension retention-time shift, parameterized by:
   - Skew, $s$.
3. Zero-mean, Gaussian-distributed noise, $G$, parameterized by:
   - Standard deviation, $\sigma_n$.

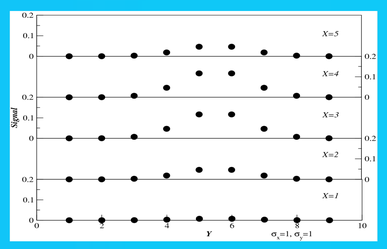
$$g[i,j]=\frac{1}{2\pi\sigma_x\sigma_y}\int_{i-\frac{1}{2}}^{i+\frac{1}{2}}\int_{j-\frac{1}{2}}^{j+\frac{1}{2}}\exp\left(\frac{(x-\mu_x)^2}{2\sigma_x{}^2}+\frac{\left(y-\mu_y+(i-\mu_x)s\right)^2}{2\sigma_y{}^2}\right)dxdy+\sigma_n G_{i,j}$$

# Retention-Time Shift Correction

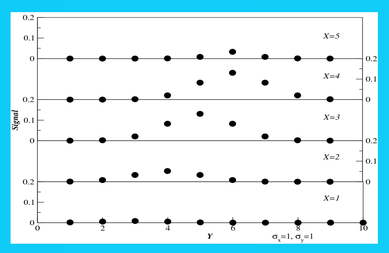
Retention-time shift correction is implemented for both algorithms as preprocessing for peak detection.

1. Estimate retention-time skew using cross correlation.
2. Shift simulated data to correct retention-time skew.

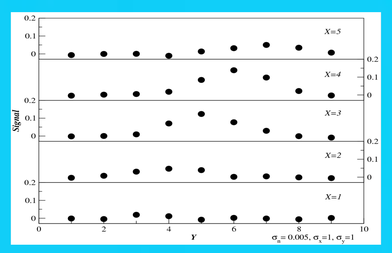
Slices of a sampled, simulated 2D peak displaying each secondary 1D peak.



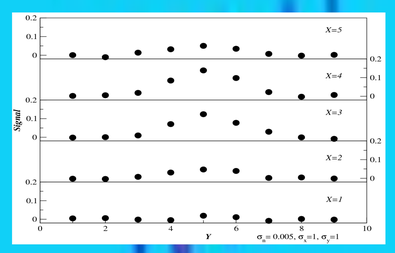

Shifted slices incorporating a skew in the 2D peak.




Slices of a skewed 2D peak with random Gaussian noise.




Slices of skewed 2D peak with noise after skew correction.

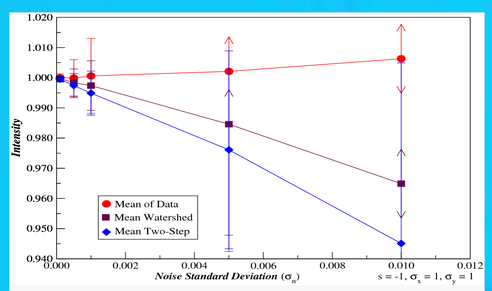



# Experimental Results

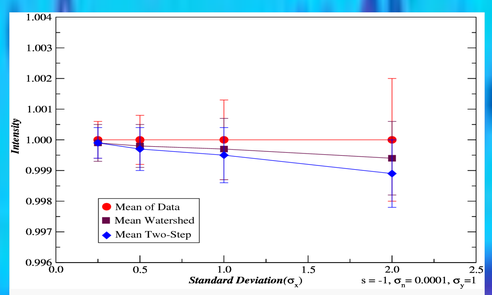
- Four parameters are varied:
  - Noise standard deviation, $\sigma_n$, from 0.0001 to 0.01.
  - First-dimension peak-width standard deviation, $\sigma_x$, from 0.25 to 2.00.
  - Second-dimension peak-width standard deviation, $\sigma_y$, from 1 to 8.
  - Skew, $s$, from -8 to -1.
- Each experiment is conducted 1000 times.
- Compare intensity mean and standard deviation of 2D peaks detected by the Two-Step and Watershed algorithms with the actual peak signal.
- The Watershed algorithm has better accuracy than the Two-Step algorithm when retention-time shift correction is used with both methods.
- Statistical significance indicates that the superiority of the Watershed algorithm is strongly supported and almost certainly would be observed in repeated experiments.

| Skew $s$ | Peak $\sigma_x$ | Peak $\sigma_y$ | Noise $\sigma_n$ | 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 |
| -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 |
| -2 | 1.00 | 1 | 0.0100 | 11×31 | 0.9991 | 0.1503 | 0.9611 | 0.0671 | -0.0380 | 0 | 0.9409 | 0.0598 | -0.0581 | 0 | 1.0000 |
| -4 | 1.00 | 1 | 0.0100 | 11×51 | 0.9977 | 0.1904 | 0.9595 | 0.0676 | -0.0382 | 0 | 0.9419 | 0.0616 | -0.0558 | 0 | 1.0000 |
| -8 | 1.00 | 1 | 0.0100 | 11×91 | 0.9990 | 0.2589 | 0.9631 | 0.0696 | -0.0359 | 0 | 0.9470 | 0.0603 | -0.0520 | 1 | 1.0000 |

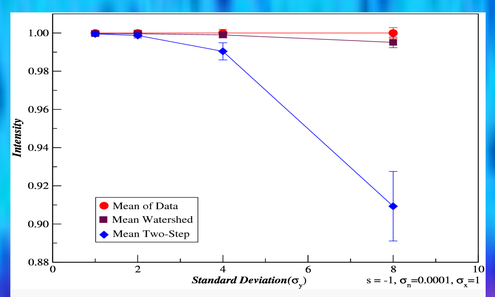
Results for 2D peak detection algorithms with various levels of noise ($\sigma_n$), peak widths ($\sigma_x$ and $\sigma_y$), and retention-time skew ($s$). More comprehensive results are available by request.



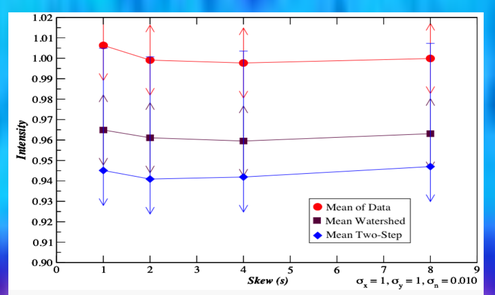

Performance of peak detection algorithms as a function of noise standard deviation, $\sigma_n$.




Performance of peak detection algorithms as a function of first-column peak width, $\sigma_x$.




Performance of peak detection algorithms as a function of second-column peak width, $\sigma_y$.




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