Peak Detection in Comprehensive Two-Dimensional Chromatography

Indu Latha

Contents

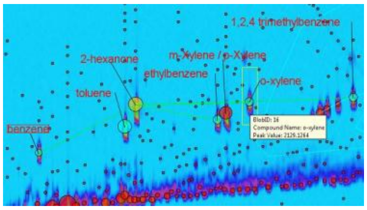
Introduction

- Gas Chromatography (GC) is a technique for separating and analyzing compounds from a mixture by injecting them into a gas stream passed through a glass tube or column coated with material that interacts with the compounds.
- Comprehensive two-dimensional chromatography (GCxGC) is a technology to separate compounds indistinguishable by one-dimensional GC by directing them onto a second column that separates the sample according to a different set of properties.
- Each compound in a mixture has a characteristic **retention time** which is the time it takes pass through a column. A detector measures the compounds as they exit the column as a function of the retention time.

Introduction

- GCxGC data processing extracts relevant information from the chromatogram to identify the compounds.
- The data processing follows a sequence of steps :
  - Eliminate background noise present in the chromatogram.
  - Detect peaks, each of which corresponds to a single chemical compound.
  - Distinguish peaks for compounds that have similar retention time.
  - Identify and quantify compounds.
  - Report information.

Introduction



BlobID: 16
Compound Name: o-xylene
Peak Value: 2129.1264

Fig 1: Peaks detected by GC Image software are highlighted by circles. A peak that is selected is highlighted by a box

Peak Detection

- Peak detection is the process of aggregating cluster of pixels that form a peak. A peak is a two-dimensional spot or elution curve that relates a compounds concentration with its retention time.

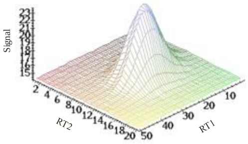


Fig 2: Single two-dimensional peak induced by a compound

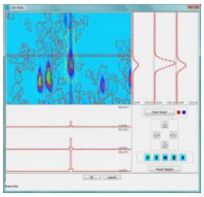


Fig 3: Column and row slices of the highlighted peak

Peak Detection

- Two techniques for peak detection in GCxGC:-
  - Extension of the traditional one-dimensional peak detection algorithm – The two-step 1D approach
  - The inverted Watershed algorithm or Drain algorithm

Motivation

- A recent paper by G.Vivó-Truyols & H.G.Janssen analyzed the effects of retention time shifts in the second column separation on the performance of both the peak detection techniques.
- It was concluded that the skew in successive secondary separations reduced the performance of the inverted watershed algorithm.
- For the two-step 1D approach, geometric alignment was used while merging the 1D peaks but no measure was taken to correct the skew in the inverted watershed algorithm.
- This talk is aimed at disputing the conclusions of G.Vivó-Truyols & H.G.Janssen.

Two-Step One-Dimensional Peak Detection Algorithm

- Extends the approach used in one-dimensional gas chromatography.
- The 1D chromatogram is a single column of the data matrix.
- Individual peaks are detected for each chromatogram.
  - The maximum peak height should exceed a threshold value (Th_0) above the baseline.
  - Peak region in a second-dimension chromatogram is defined by peak start and peak end.

Two-Step One-Dimensional Peak Detection Algorithm

- A peak-merging algorithm is applied that determines if consecutive 1D peaks belong to a single compound.
  - The first 1D peak in the secondary chromatogram forms the first 2D cluster.
  - All 1D peaks found in the next second-dimension chromatogram are considered as candidates for merging.
  - Several criteria are applied to accept the merger of the 1D peaks.
  - G.Vivó-Truyols & H.G.Janssen also correct for skew while merging the 1D peaks.

Two-Step One-Dimensional Peak Detection Algorithm

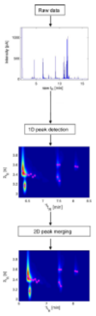


Fig 4. Flow chart representing two-step 1D peak detection algorithm

Inverted Watershed Algorithm

- The watershed transformation has been widely used in image segmentation.
- An inversion of the watershed algorithm , referred to as the Drain algorithm, was adapted for peak detection by Reichenbach et al.
- The inverted watershed algorithm proceeds by starting a peak at the highest elevation and iteratively adding smaller pixels bordering the peak until there are no more smaller positive-valued pixels in the data matrix.

Inverted Watershed Algorithm

- Two-dimensional peaks in GCxGC are detected by the following steps:
  - A pixel with the largest value is chosen and labeled.
  - The neighboring pixels are inserted into a priority queue with a priority level corresponding to the value of the pixel.
  - The pixel with the highest priority level is extracted from the priority queue.
  - The pixel is compared with all its neighbors in a 3x3 matrix.
    - If the neighboring pixels are of equal or larger value, give the same label to the extracted pixel.
    - If the pixel is the largest in its neighborhood, give it a different label.
  - Iterate until the priority queue is empty.

Inverted Watershed Algorithm

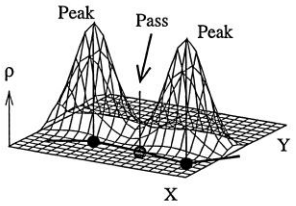


Fig 5. 2D peaks in a chromatogram can be detected by labeling each successive pixel in the priority queue.

Simulation

- Simulate a 2D chromatographic peak by generating a 2D Gaussian with parametric peak width.
- To simulate retention time shifts in the secondary chromatogram, a skew is introduced in the peak that shifts each second-dimension peak.

$$\gamma(i) = (i - \mu_X) \times s$$

where, $\gamma(i)$ is the skew in each secondary chromatogram
i is the position of each data point along the x-dimension
μ_X is the mean along the x-axis
s is the skew parameter that controls the shift

Simulation

- A random Gaussian noise is added to the skewed peak at each point to generate the chromatographic peak

$$f(i,j) = A \times exp[-(\frac{((i+\gamma)-\mu_x)^2}{2\sigma_x^2} + \frac{(j-\mu_y)^2}{2\sigma_y^2})]$$

where, $f(i,j)$ is the chromatographic peak at position (i, j),
i and j are the position of the pixel along x and y-dimension,
$\gamma(i)$ is the skew,
μ_x and μ_y represent the mean along the x and y-dimension,
σ_x and σ_y represents the spread of the peak along (x, y),
σ_N is the standard deviation of noise,
$G_N(i, j)$ is the random Gaussian noise at (i, j),
and $A = \frac{1}{2\pi\sigma_X\sigma_Y}$ denotes the maximum amplitude of the peak

Simulation

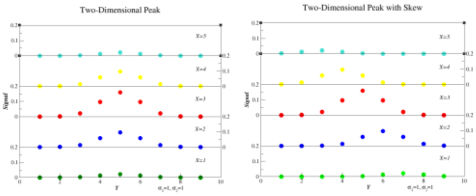


Fig 6. Slices of a 2D peak displaying each secondary 1D peak

Fig 7. Slices of a 2D peak after skew simulating the retention time shifts

Simulation

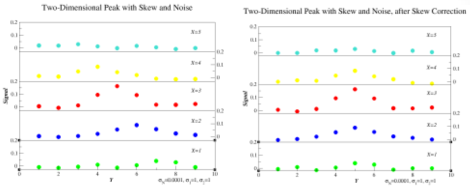
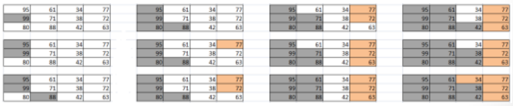


Fig 8. Slices of a skewed 2D peak with random Gaussian noise.

Fig 9. Skewed 2D peak with noise, after skew correction.

Experiment

- The inverted watershed algorithm is applied to the 2D signal after skew correction shown in Fig 9.
- The largest pixel in the signal is identified and labeled.
- Each successive pixel is given the label of its largest neighbor in a 3x3 neighborhood.



95	61	34	77
99	71	38	72
80	88	42	63

95	61	34	77
99	71	38	72
80	88	42	63

95	61	34	77
99	71	38	72
80	88	42	63

95	61	34	77
99	71	38	72
80	88	42	63

95	61	34	77
99	71	38	72
80	88	42	63

95	61	34	77
99	71	38	72
80	88	42	63

95	61	34	77
99	71	38	72
80	88	42	63

95	61	34	77
99	71	38	72
80	88	42	63

95	61	34	77
99	71	38	72
80	88	42	63

95	61	34	77
99	71	38	72
80	88	42	63

95	61	34	77
99	71	38	72
80	88	42	63

95	61	34	77
99	71	38	72
80	88	42	63

Fig 10. Pixels assigned in the order of their intensity in a peak by the watershed algorithm

Experiment

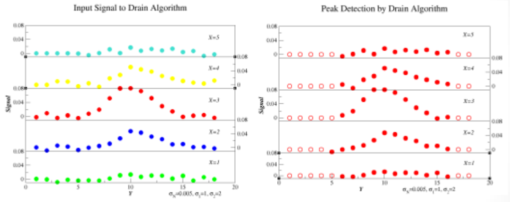
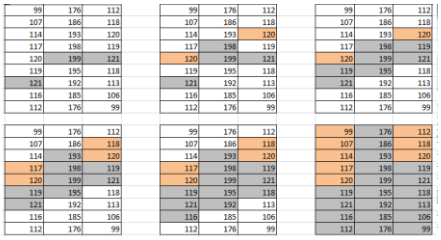


Fig 11: The drain algorithm applied to the 2D input signal for peak detection

Experiment

- The two-step one-dimensional algorithm is also applied to the 2D signal with noise and skew. Skew correction is performed for each 1D peak prior to merging them to form the 2D detected peak.
- In the first step, 1D peak detection is done for each 1D slice of the 2D signal.
- In the second step, the skew is corrected and all the peaks are merged to form the 2D detected peak.

Experiment



99	176	112
107	186	118
114	193	120
117	198	119
120	199	121
119	195	118
121	192	113
116	185	106
112	176	99

99	176	112
107	186	118
114	193	120
117	198	119
120	199	121
119	195	118
121	192	113
116	185	106
112	176	99

99	176	112
107	186	118
114	193	120
117	198	119
120	199	121
119	195	118
121	192	113
116	185	106
112	176	99

99	176	112
107	186	118
114	193	120
117	198	119
120	199	121
119	195	118
121	192	113
116	185	106
112	176	99

99	176	112
107	186	118
114	193	120
117	198	119
120	199	121
119	195	118
121	192	113
116	185	106
112	176	99

99	176	112
107	186	118
114	193	120
117	198	119
120	199	121
119	195	118
121	192	113
116	185	106
112	176	99

Fig 12. Pixels assigned in the order of their intensity for each second-dimension 1D chromatogram

Experiment

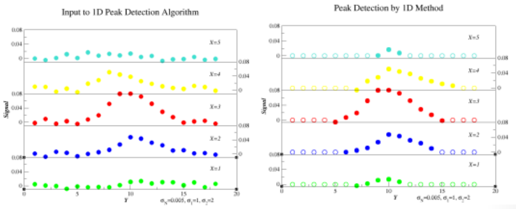


Fig 13. 1D peak detection applied to each 1D slice of the signal followed by peak-merging

Results

- It can be noticed that both algorithms detect the peak fairly accurately for the signal.
- A set of experiments were performed by varying:
  - The noise introduced in the signal by steadily increasing the Noise standard deviation, σ_N from 0.0001 to 0.01.
  - Increasing the peak width along the x-dimension by increasing the standard deviation, σ_x from 0.25 to 4.
  - The standard deviation, σ_Y is increased from 1 to 8 thereby increasing the peak width along the y-dimension.
- For each value assigned to σ_N, σ_x and σ_Y the experiment was conducted a 1000 times to observe the performance.

Results

σ_X	σ_Y	σ_N	Mean Data	Variance	Mean Watershed	Var WS	Mean 1D	Var 1D	Error(WS)	Error(1D)	Size	Watershed Negative Cases	1D Negative Cases
1	1	0.0000	0.9903	0.0000	0.9903	0.0000	0.9903	0.0000	0.0000	0.0000	9x5	0	0
1	1	0.0001	0.9904	0.0000	0.9904	0.0000	0.9903	0.0000	0.0000	0.0001	9x5	0	0
1	1	0.0005	0.9902	0.0000	0.9901	0.0000	0.9898	0.0000	0.0001	0.0004	9x5	0	0
1	1	0.0010	0.9905	0.0000	0.9899	0.0000	0.9892	0.0000	0.0006	0.0013	9x5	0	0
1	1	0.0050	0.9894	0.0009	0.9840	0.0008	0.9745	0.0008	0.0054	0.0149	9x5	0	0
1	1	0.0100	0.9876	0.0040	0.9710	0.0036	0.9491	0.0032	0.0166	0.0385	9x5	0	0

σ_X	σ_Y	σ_N	Mean Data	Variance	Mean Watershed	Var WS	Mean 1D	var 1D	Error(WS)	Error(1D)	Size	Watershed Negative Cases	1D Negative Cases
0.25	1	0.0001	1.5968	0.0000	1.5968	0.0000	1.5966	0.0000	0.0000	0.0002	9x3	0	0
0.5	1	0.0001	1.0143	0.0000	1.0143	0.0000	1.0142	0.0000	0.0000	0.0001	9x5	0	0
1	1	0.0001	0.9989	0.0000	0.9988	0.0000	0.9988	0.0000	0.0001	0.0001	9x9	0	0
2	1	0.0001	0.9578	0.0000	0.9576	0.0000	0.9579	0.0000	0.0002	-0.0001	9x18	0	0
4	1	0.0001	0.7270	0.0000	0.7261	0.0000	*	*	0.0009	*	9x36	0	1000

σ_X	σ_Y	σ_N	Mean Data	Variance	Mean Watershed	Var WS	Mean 1D	var 1D	Error(WS)	Error(1D)	Size	Watershed Negative Cases	1D Negative Cases
1	1	0.0001	0.9989	0.0000	0.9988	0.0000	0.9988	0.0000	0.0001	0.0001	9x9	0	0
1	2	0.0001	0.9999	0.0000	0.9996	0.0000	0.9987	0.0000	0.0003	0.0012	18x9	0	0
1	4	0.0001	1.0000	0.0000	0.9990	0.0000	0.9903	0.0001	0.0010	0.0097	36x9	0	0
1	8	0.0001	1.0001	0.0000	0.9952	0.0000	0.9173	0.0002	0.0049	0.0828	72x9	57	361

Fig 14. The table depicts the results produced by both the peak detection methods as the standard deviation of noise, standard deviation along the x and y dimensions are increased.

* - All the test cases for the two-step 1D peak detection failed

Results

- The mean and variance of the peak intensity detected by both the techniques are plotted against the changing function, here noise standard deviation and standard deviation along the y-dimension.

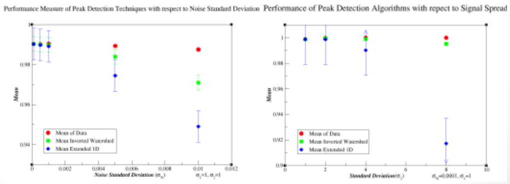


Fig 15. The performance of the peak detection algorithms wrt increasing noise and peak spread

Results

- The amount of noise added to the signal is increased gradually and the mean is computed for both the algorithms.
- It can be seen that the inverted watershed (drain) algorithm performs better than the extended 1D approach.
- Similarly, for the set of experiments conducted by varying the standard deviation of the signal along both x and y axis, it can be observed that the performance of the two-step one-dimensional approach reduces with increase in the peak width even for low random Gaussian noise.

Results

- The table in Fig 14. shows the number of test cases that failed for both the inverted watershed and two-step 1D approach.
- All the pixels within σ_X and σ_Y of the largest pixel should be included in the peak for accurate peak detection. The resulting peaks that do not meet this threshold are not considered while computing the mean.
- The number of failing cases in 1000 runs is higher for the two-step 1D approach when compared to the watershed algorithm. This is because the two-step 1D approach detects 1D peaks for each secondary chromatogram separately.

Conclusion

- From the experiments, it is evident that the inverted watershed algorithm for peak detection performs better than the two-step 1D approach when skew correction is applied to both the methods.
- However, the inverted watershed algorithm also does not ensure successful peak detection always.
- With better noise suppression techniques we may be able to generate better results from both the algorithms.

Thank you!!