

Technical Disclosure Commons

Defensive Publications Series

September 2020

Using Bayesian Blocks to Compare Chromatogram Traces

Anonymous

Follow this and additional works at: https://www.tdcommons.org/dpubs_series

Recommended Citation

Anonymous, "Using Bayesian Blocks to Compare Chromatogram Traces", Technical Disclosure Commons, (September 11, 2020)

https://www.tdcommons.org/dpubs_series/3596



This work is licensed under a [Creative Commons Attribution 4.0 License](https://creativecommons.org/licenses/by/4.0/).

This Article is brought to you for free and open access by Technical Disclosure Commons. It has been accepted for inclusion in Defensive Publications Series by an authorized administrator of Technical Disclosure Commons.

Using Bayesian Blocks to Compare Chromatogram Traces

ABSTRACT

This disclosure describes using Bayesian Blocks to compare and determine similarities between chromatogram traces for chromatography and mass spectrometry applications.

KEYWORDS

- Bayesian Blocks
- Chromatography
- Liquid Chromatography
- Mass Spectrometry
- Mass Spectrometer
- LCMS
- Chromatographic Traces

BACKGROUND

Analytical instruments, such as mass spectrometers, are used to measure the mass-to-charge ratios (m/z) of ions. Typically, a sample is separated into components via a chromatographic instrument (e.g., via liquid chromatography, gas chromatography, or capillary electrophoresis), the separated components are introduced into an ion source of the mass spectrometer to be ionized, and the resulting ions are subject to transport, confinement, and separation by the components of the mass spectrometer for analysis.

Chromatography traces are generated from data analysis of the liquid chromatography (LC) process. The traces indicate the relationship between retention time (e.g., on the x-axis) and an abundance or concentration of an analyte exiting the LC system (e.g., on the y-axis). In liquid chromatography and mass spectrometry (LCMS) data analysis, chromatography traces are often compared with each other to determine their similarities. For example, correlation may be performed.

DESCRIPTION

As disclosed herein, a Bayesian Blocks algorithm is applied with a Bayesian probability measure to compare the traces. For example, a measure of similarity between the traces can be calculated which, in turn, can be related to the probability that the two traces are identical.

Bayesian Blocks is a partitioning scheme that represents a time series as the best Bayesian fit of a step function to an original data set. This fit contains the information in the original set but is significantly more tractable. Figure 1 below shows Bayesian Blocks fit to a chromatogram trace. The step function

excludes features that are indistinguishable from noise while preserving the area, centroids, and higher-order moments of the original data.

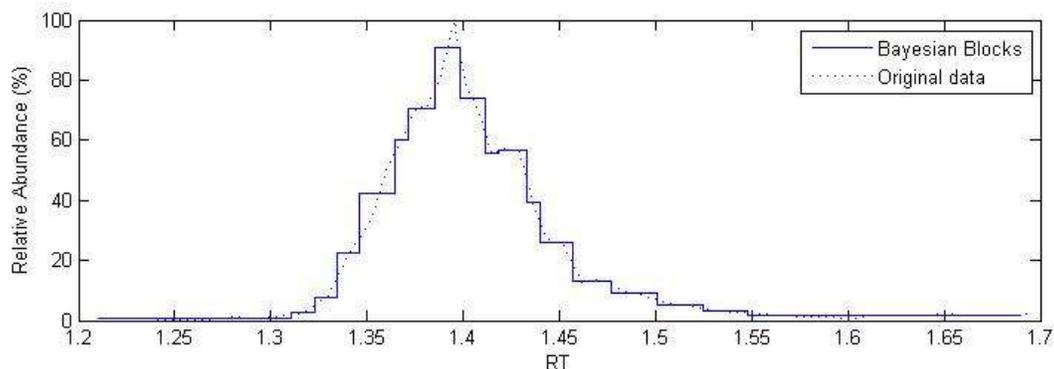


Figure 1

In more detail, Bayesian Blocks is applied to traces that are to be compared with each other. In Figure 2 below, the fit is shown regarding two traces. Each fit retains every valid peak in the original data, along with its area, centroid, and moments.

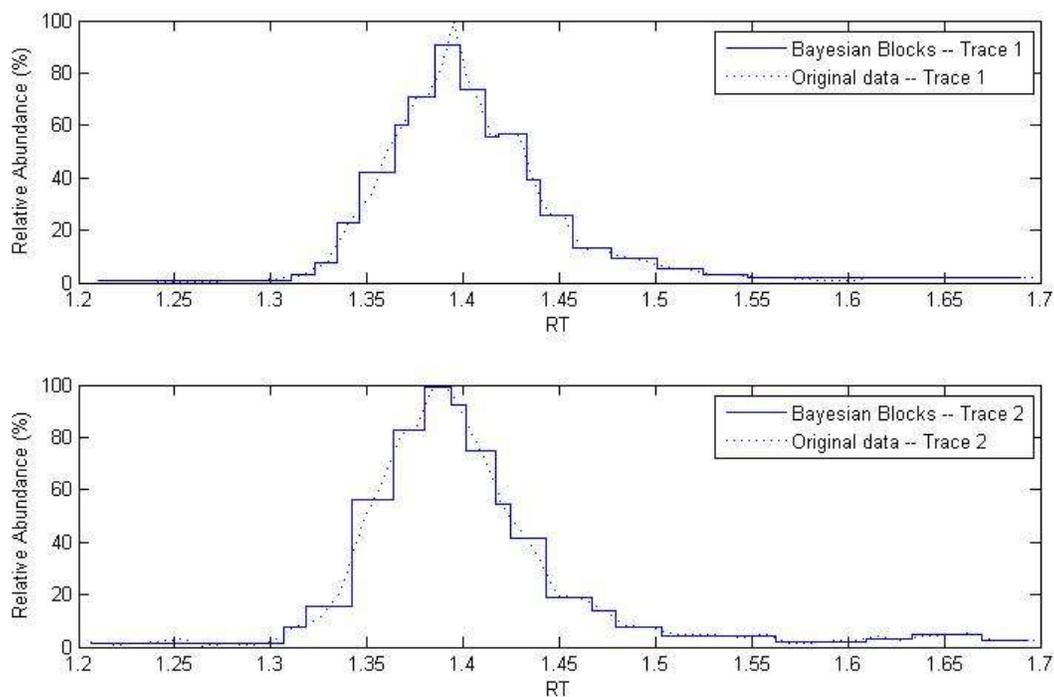


Figure 2

Next, normalization of each fit by area is performed and compared with each other, as shown in Figure 3 below. In Figure 3, the two fits are overlaid upon each other to visually depict the differences.

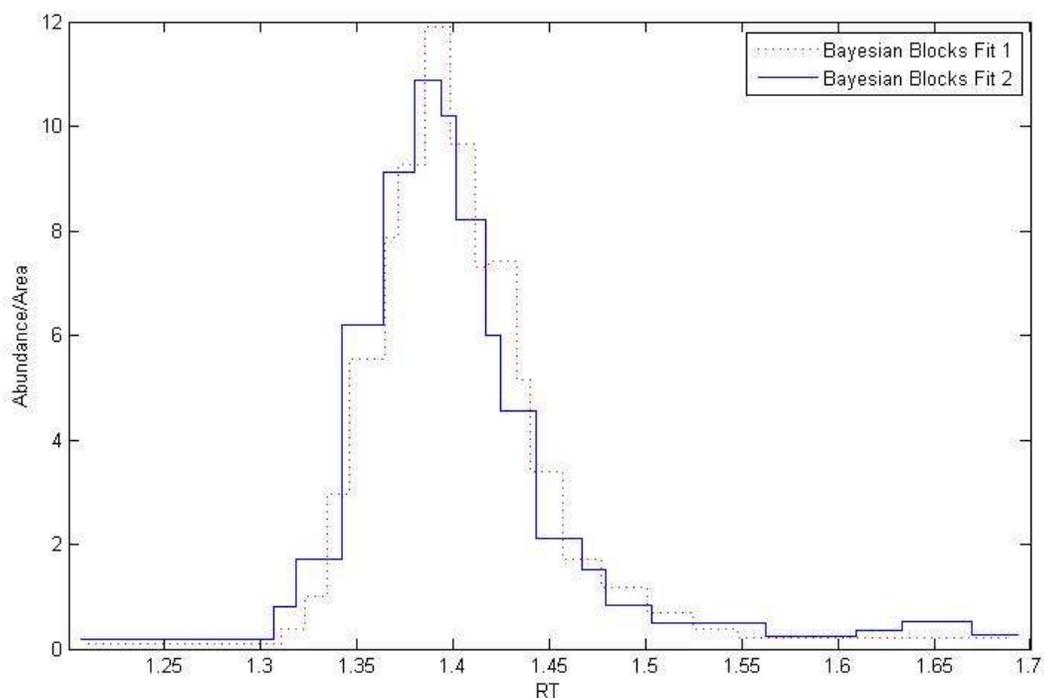


Figure 3

The Bayesian probability $P(\text{identical} \mid \text{fits})$ is then calculated to determine whether the fits represent or are derived from identical or similar data sets.

For example, $P(\text{identical} \mid \text{fits}) = P(\text{fits} \mid \text{identical}) \cdot P(\text{identical}) / P(\text{fits})$ is calculated. $P(\text{identical} \mid \text{fits})$ and $P(\text{partitionings})$ are then marginalized to obtain $P(\text{identical} \mid \text{fits}) = P(\text{fits} \mid \text{identical})$.

If the uncertainty in the observations has a Gaussian distribution, then $P(\text{fits} \mid \text{identical})$ will be $P(\text{fits} \mid \text{identical}) = \exp(-\sum(\delta Y_i / \sigma)^2 / 2)$ where Y_i is the difference between the two fits at retention time i and σ is the uncertainty in the measurements. $\sum(\delta Y_i)$ is a function of the area of the difference between the two fits, as shown in the shaded regions in Figure 4.

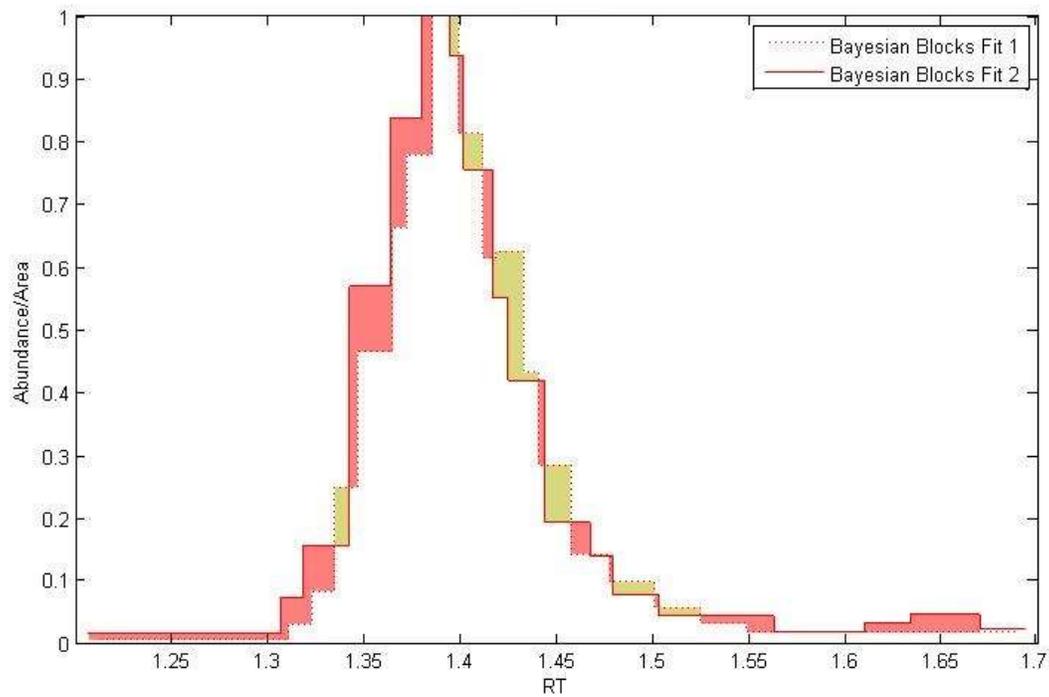


Figure 4

This allows the use of the area of the difference as a similarity measure. This measure is proportional to the log probability that the two data sets are identical. Thus, Similarity Measure = - Area of the difference between the two fits. Therefore, the similarity measure is calculated for different retention times to determine the best similar and retention time shift, as shown in Figure 5 below.

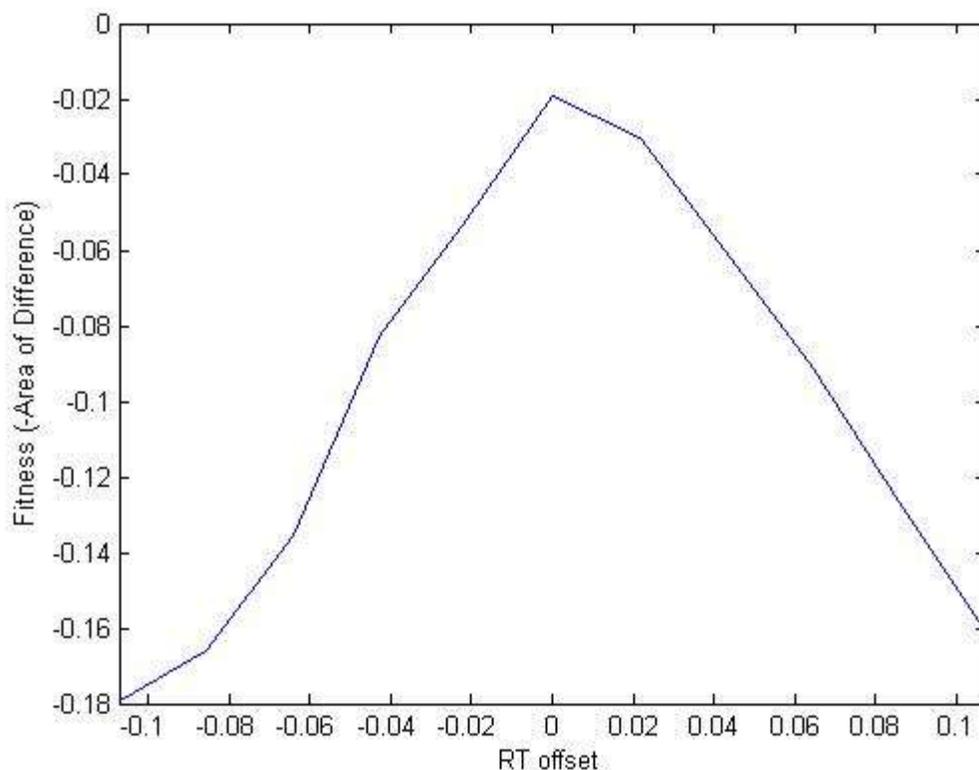


Figure 5

Based on the results, insight into different chromatographic traces is provided for the experimenter. Additionally, the determined information described herein can be used to adjust operation of the LC or mass spectrometry system.

These techniques can be performed on a chromatography or mass spectrometry system, or using a computing system for post-acquisition analysis.

CONCLUSION

This disclosure describes using Bayesian Blocks to compare and determine similarities between chromatogram traces for chromatography and mass spectrometry applications. Per the techniques of the disclosure, similarities between chromatographic traces can be performed without relying on correlation techniques.

REFERENCES

1. Scargle, Jeffrey D. "Studies in Astronomical Time Series Analysis. V. Bayesian Blocks, a New Method to Analyze Structure in Photon Counting Data." *The Astrophysical Journal* 504.1 (1998): 405–418.