September 2020

Bayesian Blocks for Alignment of Chromatographic Traces

Anonymous

Bayesian Blocks for Alignment of Chromatographic Traces

ABSTRACT

This disclosure describes using a Bayesian Blocks partitioning technique with a Bayesian fitness measure for aligning chromatographic traces for chromatography and mass spectrometry applications.

KEYWORDS

- Bayesian Blocks
- Alignment
- Chromatographic Traces
- Chromatography
- Mass Spectrometry
- Mass Spectrometer
- LCMS
- GCMS

BACKGROUND

A mass spectrometer is an analytical instrument used to measure the mass-to-charge ratios (m/z) of ions of a sample-under-analysis, or an analyte. Typically, the analyte 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 for ionization, and the resulting ions are subject to transport, confinement, and separation by the components of the mass spectrometer for analysis. The analysis can include generating a mass spectrum depicting a plot of intensity (relative abundance) as a function of the m/z. The mass spectrum is useful for the identification, quantification, and structural elucidation of the analyte, for example, peptides, proteins, and related molecules. The analysis can also include generating a chromatogram depicting retention time and the response (e.g., concentration) created by the components of the analyte exiting the chromatography system.

In liquid chromatography-mass spectrometry (LCMS) or gas chromatography-mass spectrometry (GCMS), alignment of chromatographic traces is an important step for data analysis. In some scenarios, alignment is performed by identifying comparable peaks in a candidate and a reference chromatogram, calculating a retention time offset between the identified comparable peaks, and then using a linear interpolation to align peaks in the candidate chromatogram with corresponding peaks in the reference chromatogram.

DESCRIPTION

As described herein, a Bayesian Blocks partitioning technique is used for alignment of chromatographic traces. 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. Figure 1 below shows a Bayesian Blocks fit to an example chromatographic trace. The Bayesian Blocks fit includes all the information of the original data set, but is more tractable. For example, the area, the centroid, and the moments of the peaks in the step function are the best Bayesian fits to the area, the centroid, and the moments of every valid peak in the original data set.

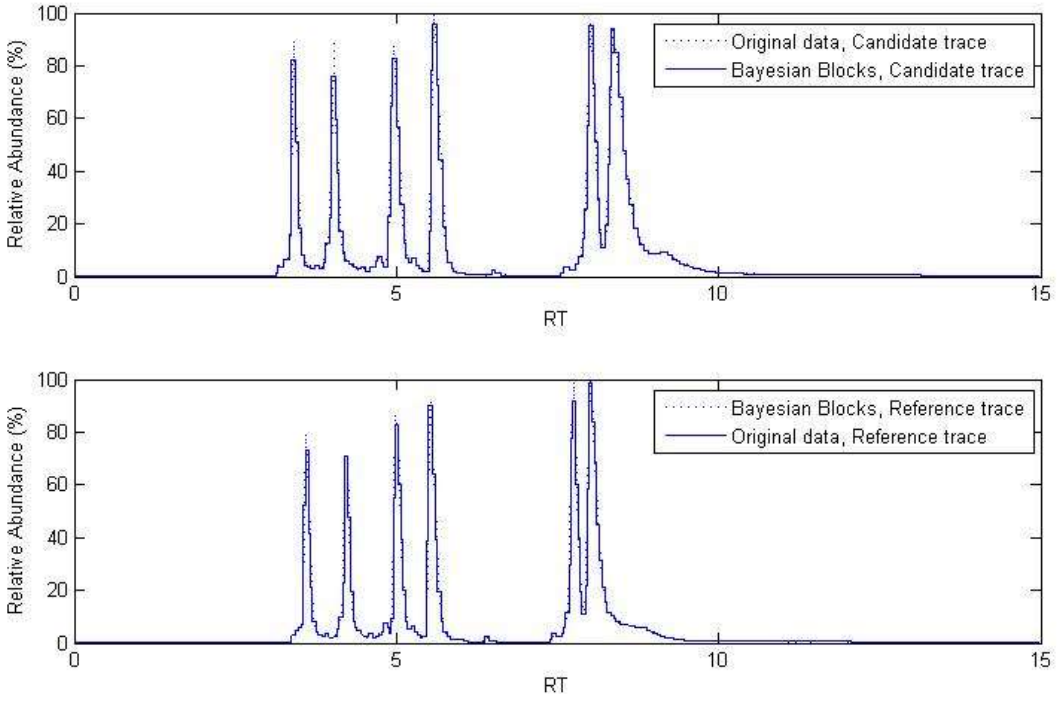

Figure 1

A Bayesian fitness measure is also used to detect chromatographic peaks in a candidate chromatogram and a reference chromatogram with which the candidate chromatogram is to be aligned. The most likely offset in retention time between the peaks is determined and used to perform the alignment.

In more detail, Bayesian Blocks fits to the traces of the candidate and reference chromatograms are generated, as shown in Figure 2 below.

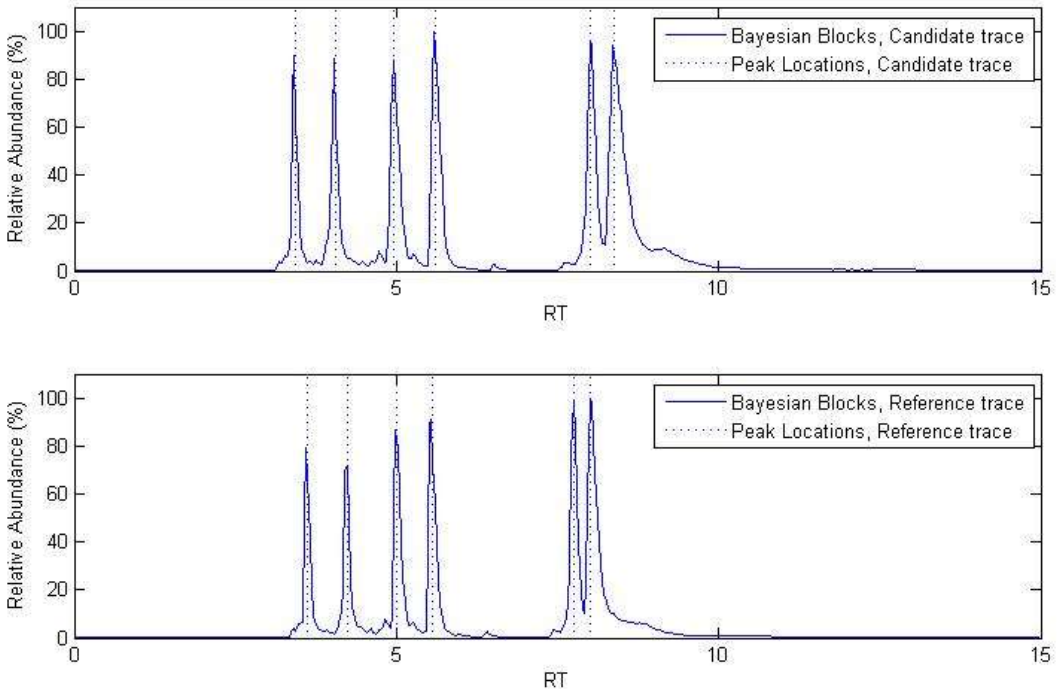

Figure 2

Next, local maxima in these fits are determined to identify the peaks in Figure 2. Because the Bayesian Blocks fits preserves valid information in the original data set, any valid peak in the original trace appears as a peak in the fit, and its most probably centroid location will be identical to the centroid of the fit.

A Bayesian fitness measure is then defined for equivalent peaks, as depicted in the right-hand side of Figure 3 below.

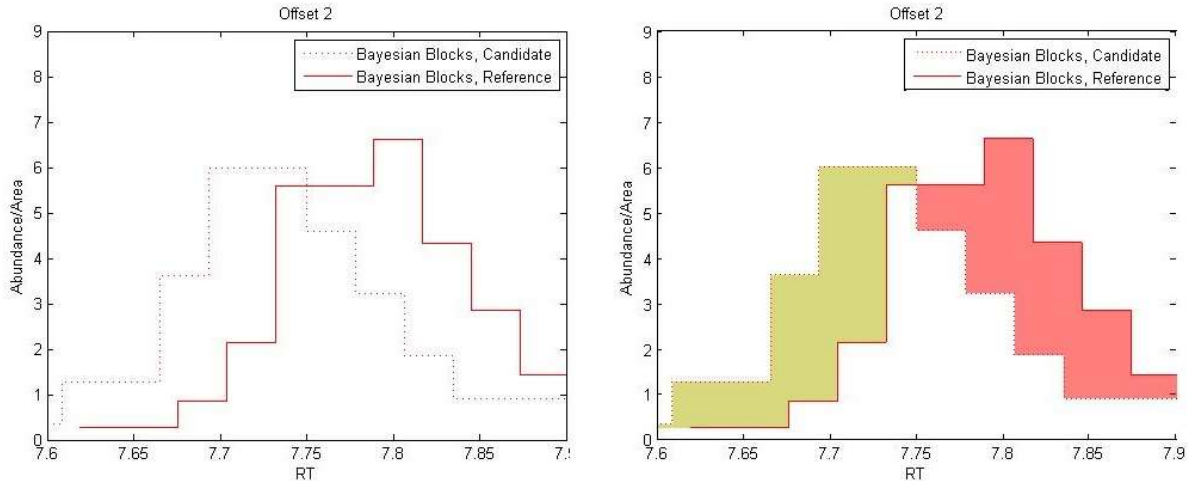

Figure 3

P(identical| observed fits) = P(observed fits | identical)*P(identical)/P(observed fits) is then calculated. P(identical) and P(observed fits) are marginalized to obtain P(identical| observed fits) = P(observed fits | identical). P(observed fits | identical) is proportional to exp(- sum((deltaYi / sigma)^2 / 2)), which is proportional to exp(- (Area / sigma)^2/2), where Area is the area of the difference between the traces, as shown in the left-hand of Figure 3. The area of the difference between the traces is used as a fitness measure.

Next, the fitness measure is calculated for different retention time offsets of peaks in the candidate chromatogram, as depicted in Figure 4, and the retention time offset that maximizes the fitness is determined, which is the most probable R offset between the peaks.

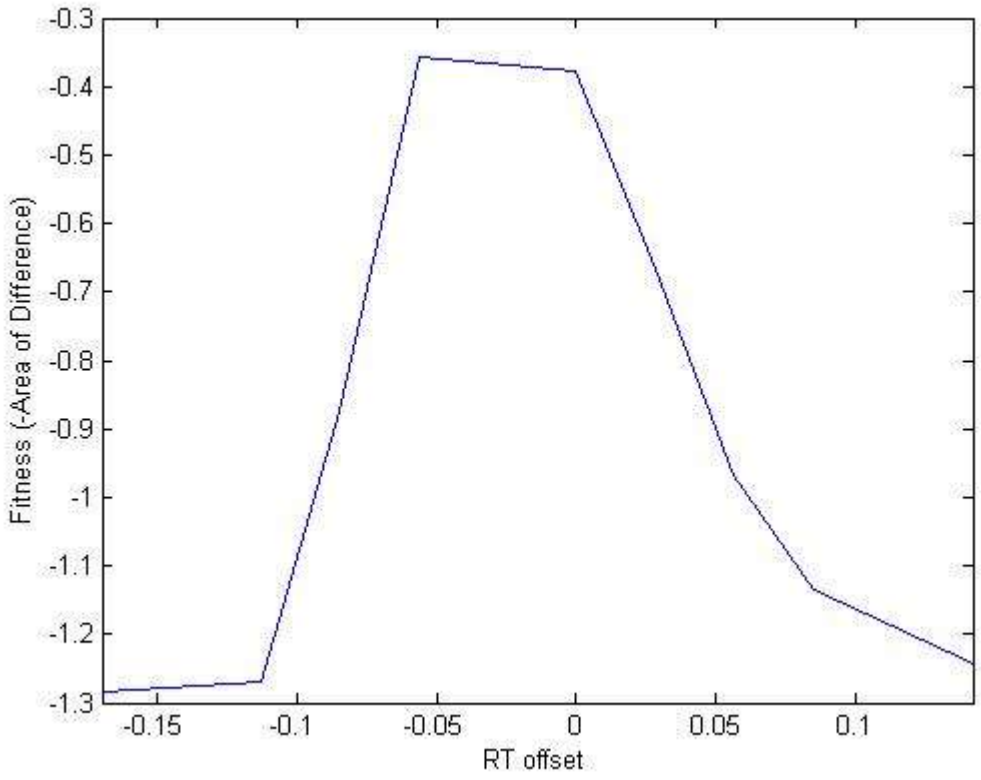

Figure 4

Points in the candidate chromatogram are shifted by a spline fit or by a linear interpolation of the segments between peaks to result in shifting peaks in the candidate chromatogram by the offsets calculated above. This brings the candidate chromatogram into alignment with the corresponding peaks in the reference chromatogram, as depicted in Figure 5.

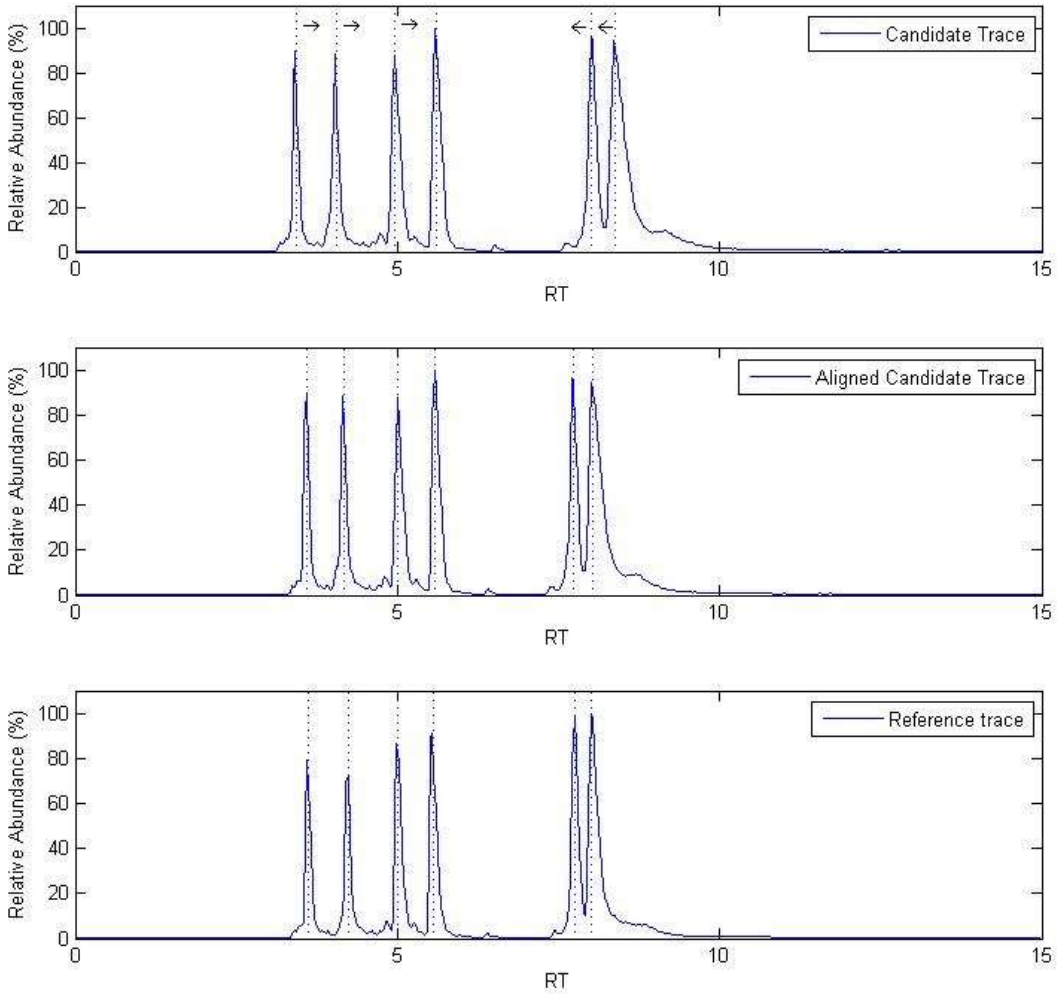

Figure 5

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

CONCLUSION

Thus, using a Bayesian Blocks partitioning technique with a Bayesian fitness measure for aligning chromatographic traces for chromatography and mass spectrometry applications is described. Per the techniques of the disclosure, a candidate chromatogram is brought into alignment with corresponding peaks in a reference chromatogram.