Journal of Chromatography A, 1369 (2014) 92–97

Contents lists available at ScienceDirect

Journal of Chromatography A

journal homepage: www.elsevier.com/locate/chroma

Diagnostic application of the exponentially modified Gaussian model for peak quality and quantitation in high-throughput liquid chromatography–tandem mass spectrometry

Adam P.R. Zabell[a,*], Tyler Foxworthy[a,1], Kimberly Napoli Eaton[b], Randall K. Julian[a]

[a] Indigo BioSystems, 7820 Innovation Boulevard, Suite 250, Indianapolis, IN 46278, United States
[b] Andor Labs, 4134 South Alston Avenue, Suite 107, Durham, NC 27713, United States

ARTICLE INFO

Article history:
Received 9 June 2014
Received in revised form 2 October 2014
Accepted 5 October 2014
Available online 13 October 2014

Keywords:
Exponentially modified Gaussian
Column degradation
High throughput chromatography
Mass spectrometry

ABSTRACT

Typical area calculation for a chromatographic peak assumes the observed signal strength at every measurement is an exactly accurate count of the signal. We compared that approach to one using the exponentially modified Gaussian (EMG) in an automated, clinical production setting. Peak areas in a 47 analyte high throughput clinical production liquid chromatography–tandem mass spectrometry assay were compared across four months of production data to determine trends over the lifespan of a chromatographic column. The EMG parameters were superior to traditional quality control methods for monitoring data reproducibility, accuracy and precision. Because the EMG calculations are performed for every peak in the system, a constant monitor of system health is integrated into the operational workflow. Parameter trends confirmed the need for column replacement, and indicated the opportunity for a reduced schedule of preventive and routine maintenance.

1. Introduction

Clinical diagnostic test procedures rely on accurate quantification of biologically important analytes in complex mixtures such as blood and urine. Calculating the chromatographic peak area in the absence of noise is a straightforward numerical integration process of peak detection, peak boundary assignment, and baseline estimation [1]. The move to apply liquid chromatography–tandem mass spectrometry (LC–MS/MS) in this setting has revolutionized accuracy, precision and selectivity across a wide range of target molecules [2–6] and allowed for the use of smaller sample volumes while achieving lower detection limits. Biological analytes have benefited from atmospheric pressure ionization techniques, but the drive to report at the detection limit necessarily produces results that are significantly more noise-prone than systems investigated by the older electron impact methods.

Noise combined with low concentration or weak signal strength presents special challenges when calculating peak area. Simple first and second derivative procedures fail to reliably find peak start and end times in a noisy context, and valley-to-valley peak assignment will significantly skew the baseline [7,8]. Complex mixture chromatograms will over-represent both chemical and ionization noise in simplistic area and baseline estimators [9]. Finally, because the system-associated chemical and ionization noise is heteroscedastic, the true peak start, peak apex, and peak end may be significantly different from the observed signal [10].

Early in the development of chromatographic numerical analysis, models of concentration profiles were developed and their integration was shown to be superior to integration of the raw data [11]. Numeric integration based on the raw data ignores the fact that a datapoint contains noise. Smoothing the raw data can improve the overall signal to noise ratio, but creates distortion in the observation and does not of itself remove the noise in the measurement [12,13]. In contrast, fitting a model to the data acknowledges the collected signal contains noise and represents a sampling of the concentration within a confidence interval. One early model, the exponentially modified Gaussian (EMG) [14] has remained in the forefront of chromatographic peak fitting due to its handling of asymmetry, easily derived statistical moments, and robust handling of overlapped peaks. In addition to extensive application in GC and LC methods [15–20], the EMG has been shown to improve quantitation in areas as diverse as solid phase extraction [21], two dimensional GC [22], and metabolomics [23].

We present here an evaluation of the EMG applied in an automatic fashion for the needs of high throughput production chromatography, where data reproducibility and instrument

* Corresponding author. Tel.: +1 317 493 2400.
E-mail address: azabell@indigobio.com (A.P.R. Zabell).
[1] Present address: KSM Consulting, 800 East 96th Street, Indianapolis, IN 46240, United States.

http://dx.doi.org/10.1016/j.chroma.2014.10.005
0021-9673/

stability are equally important factors for result reporting [24,25]. First, we examine how these methods for peak area calculation differ in their reported values. Those differences are then examined in the context of standard curve calculation to confirm statistical similarity. Finally, we apply the EMG parameters as an assessment for column suitability and peak quality.

2. Theory

The general assumption for column chromatography is a normal (Gaussian) distribution of solute in the elution profile:

$$f(x) = \frac{A}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad (1)$$

where the intensity at every time point is subject to the standard deviation (σ), the mean (μ), and the area (A). This method is sufficient for well behaved and baseline resolved peaks, but more complex peak profiles will create poor alignment between observation and model.

With liquid chromatography systems, the most common issues involve peak tailing and can be modeled with the distribution function of the exponentially modified Gaussian (EMG):

$$f(x) = \frac{A\sigma}{\tau} e^{\left(\frac{1}{2}\left(\frac{\sigma}{\tau}\right)^2 - \frac{x-\mu}{\tau}\right)} \times \sqrt{\frac{\pi}{2}} \operatorname{erfc}\left(\frac{\sigma}{\tau\sqrt{2}} - \frac{x-\mu}{\sigma\sqrt{2}}\right) \quad (2)$$

where erfc is the complementary error function:

$$\operatorname{erfc}(n) = 1 - \frac{2}{\sqrt{\pi}} \int_0^n e^{-t^2} dt \quad (3)$$

and a factor for skewness (τ) is added. The peak center and area are still defined as μ and A, respectively, but the because the EMG function is a convolution of a Gaussian process with an Exponential process, it does not simplify to the Gaussian function as τ approaches zero. Instead, σ is a width term which broadly approximates the normal distribution. As the skew decreases ($\tau < 1$) the first term in Eq. (2) gets very large while the second term gets very small, and the product of the two terms is driven to the same order of magnitude as the result from Eq. (1). For ease of discussion, σ is taken to be the normal curve contribution to peak width and τ is taken to be the exponential curve contribution.

3. Materials and methods

3.1. Data source

A production LC–MS/MS test procedure monitoring forty-seven analytes was performed using electrospray on a tandem quadrupole system (Agilent 6410B; Agilent Technologies, Santa Clara, CA, USA) with multiple reaction monitoring in positive ion mode and a reversed phase analytical column (Poroshell 120 EC-C18, 3.0 mm × 50 mm, 2.7 μm; Agilent) for analyte separation. Each batch run was controlled with a 9-point calibration curve and duplicate urine-based quality control samples for negative, low and high concentration values. Forty-one internal standards were used to correct for variations in analyte recovery. The system was optimized to provide the strongest signal to noise for a given signal strength, has been in clinical use since August 2010, and follows industry guidelines for validation and external quality assessment (see Supplemental Table 1). Data was collected across the lifetime of a single column on a single instrument and included data from three weeks before to three weeks after the column was replaced. The column was in use from 20 August 2013 to 14 November 2013. Unless otherwise specified, data from only the calibration standard and quality control samples in each batch were used for evaluation.

Supplementary Table S1 related to this article can be found, in the online version, at http://dx.doi.org/10.1016/j.chroma.2014.10.005.

3.2. Naive peak integration

In order to provide a vendor-agnostic comparison, the peak area calculation based from raw data used a simple threshold evaluation. We defined the simple peak as all contiguous points greater than 5% of the peak maximum. This includes more actual area than the classic peak asymmetry methods that advocate measurements at 10% of peak maximum [9,13] and is a reflection on the quality of

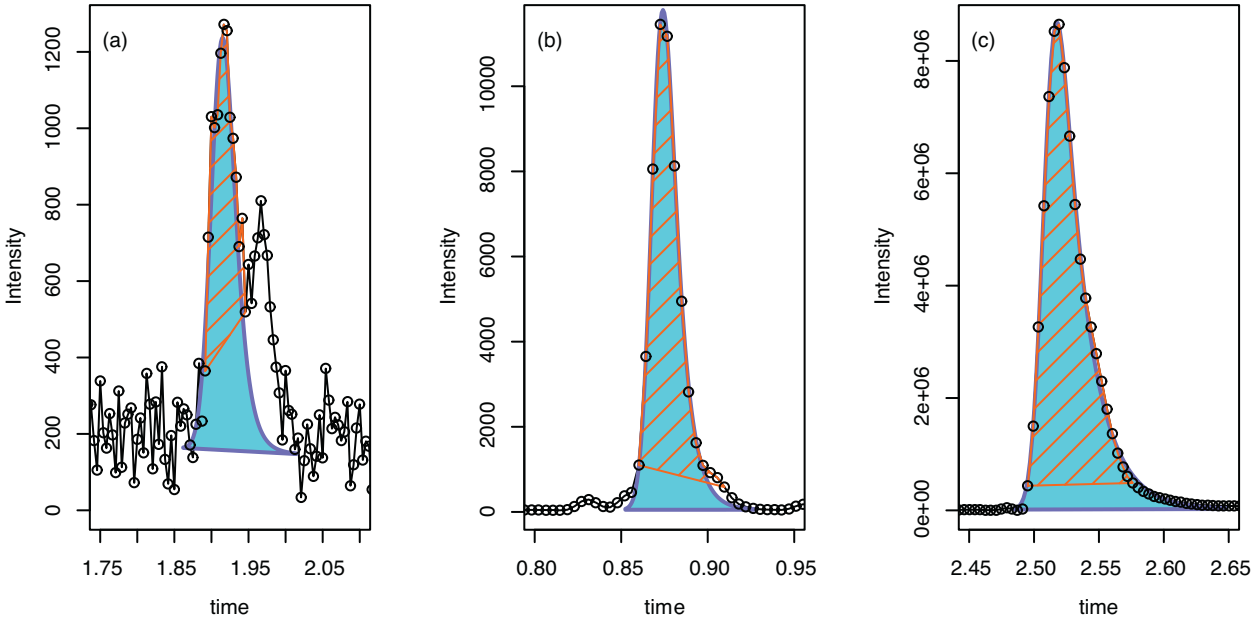

Fig. 1. EMG (blue area) and simple integration (red hashed area), demonstrating artifacts introduced by simple integration. Artificial data (a) showing high noise and coeluting peak with a manual "best fit" of numeric integration as compared with experimental data showing a trivial (b) and more substantial (c) tailing component. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)