Image Background Removal in Comprehensive Two-Dimensional Gas Chromatography

Stephen E. Reichenbach,[1] Mingtian Ni,[1] Dongmin Zhang,[1] and Edward B. Ledford, Jr.[2]

[1] Computer Science and Engineering Department, University of Nebraska – Lincoln, Lincoln NE 68588-0115 USA; e-mail: [reich|mni|dzhang]@cse.unl.edu
[2] Zoex Corporation, 2611 W M St, Suite D, Lincoln NE 68522 USA; e-mail address: ed@zoex.com

Key Words: image background removal, two-dimensional gas chromatography, imaging system calibration and correction, digital image processing

Summary

This paper describes a new technique for removing the background level from digital images produced in comprehensive two-dimensional gas chromatography (GCxGC). The approach estimates the background level across the chromatographic image based on a few structural and statistical properties of GCxGC data. Then, the background level is subtracted from the image, producing a chromatogram in which the peaks rise above a zero-mean background.

The structural and statistical properties of data obtained by a well-controlled GCxGC separation are:

1) There are *dead-bands* (regions devoid of chemical signal) in the secondary chromatograms.
2) The mean background level varies slowly with respect to characteristic peak widths.
3) Background noise has the statistical properties of random white noise.

An algorithm based on these properties is demonstrated experimentally to be effective at determining and removing the background level from GCxGC images. The algorithm also provides several parametric controls that permit adjustment of background detection. The background-removal algorithm is incorporated into an interactive program with graphical interface for rapid and accurate detection of GCxGC peaks.

1 Introduction

In gas chromatography (GC), the signal peaks, which correspond to chemical constituents in the sample, rise above a background level in the output. Under controlled conditions, the background level consists primarily of the sum of two slowly varying components: a steady-state standing-current offset (characteristic of standard GC detectors) and temperature-induced column-bleed.

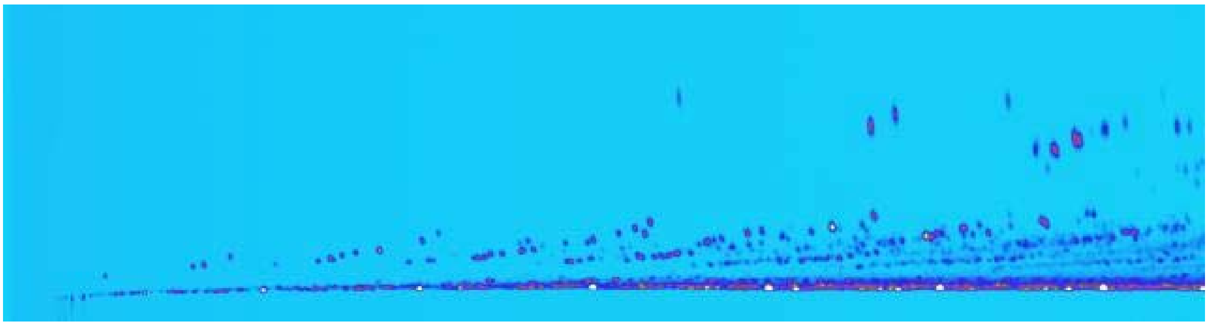

Figure 1: The left half (columns 1-1500) of an image from two-dimensional gas chromatography (GCxGC)

Accurate quantification of the chemical-related peaks requires subtraction of the background level from the signal.

This paper describes a new technique for removing the background level from images produced by comprehensive two-dimensional gas chromatography (GCxGC). Comprehensive two-dimensional gas chromatography separates chemical species with two capillary columns interfaced by a two-stage thermal desorption [1,2]. As shown in **Figure 1**, the GCxGC output can be displayed as an image, with pixels arranged so that the abscissa (X-axis, left-to-right) is the elapsed time for the first column separation and the ordinate (Y-axis, bottom-to-top) is the elapsed time for the secondary column separation. Each pixel value indicates the rate at which molecules are detected at a specific time. Each resolved chemical substance in a sample produces a small *blob* or cluster of pixels with values that are larger than the background values. In Figure 1, the smaller values of the background are colorized light blue and the larger values of the blob are colorized with dark blue, magenta, and white to show increased values.

Figure 2 illustrates a perspective plot of a typical GCxGC blob (peak) from Figure 1, corresponding to naphthalene. The background level in the vicinity of the naphthalene peak is about 14 pico-amps and the peak maximum rises to more than 23 pico-amps. In order to accurately quantify the volume of the blob, it is necessary to remove the background level from all pixels in the blob so that, in effect, the peak rises above a zero-mean background.

The approach developed in this paper estimates the local background level across the entire chromatogram based on structural and statistical properties of GCxGC data. Then, the local background level is subtracted across the entire image, producing a chromatogram in which the peaks rise above a zero-mean background. This algorithm is demonstrated experimentally to be effective at determining and removing the background level of GCxGC data. The algorithm also provides several parametric controls that permit adjustment of background detection. The background-removal algorithm is incorporated into an interactive program with graphical interface for rapid and accurate detection of GCxGC peaks.

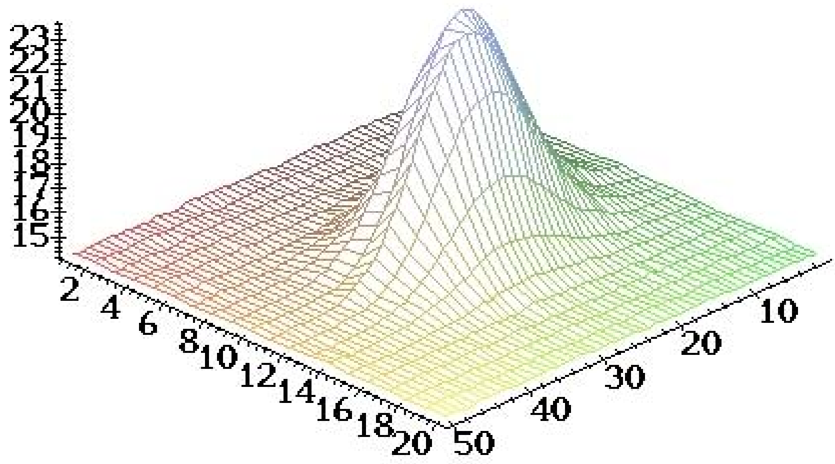

Figure 2: Perspective plot of the GCxGC data containing the isolated blob of pixels in the upper middle of Figure 1, corresponding to Naphthalene.

2 Experimental Instrumentation

The image in Figure 2 was acquired with an Agilent Model 6890 gas chromatograph with flame ionization detector (FID). The GCxGC separation was implemented with a thermal modulator (prototype Model KT2002) fabricated by Zoex Corporation (Lincoln, NE, USA), employing dual pairs of pulsed hot and cold gas jets. The cold jet temperature was approximately –189ºC, cooled by heat exchange with liquid nitrogen contained in a holding dewar mounted atop the GC. The hot jet temperature was at all times approximately 100ºC above the programmed oven temperature, maintained by a cartridge heater under the control of the AUX 2 channel of the 6890.

The first dimension column was 4.0 meters long, with 0.1 mm inside diameter and coated with a 3.5 micron thick film of methyl silicone stationary phase (Quadrex Corporation, New Haven, CT). Modulation was performed on the head of the secondary column, which was 0.8 meters long, with a 0.1 mm inside diameter and coated with 0.1 micron thick film of carbowax (Quadrex). The primary and secondary columns were connected by means of a borosilicate glass press-fit connector (Zoex Corporation). The GC oven was programmed from 35ºC to 225ºC over 100 minutes. Diesel fuel obtained from a local fueling station was injected neat [3]. Injection volume was 1.0 micro-liter using split injection mode, a split ratio of 100, and an injector temperature 280ºC.

The image in Figure 1 was acquired at a rate of 200 samples/second, collecting 400 pixels for each of 3000 image columns over a period of 100 minutes. (Only the first 1500 image columns are shown in Figure 1.) Other images were acquired with similar configurations and with similar results.

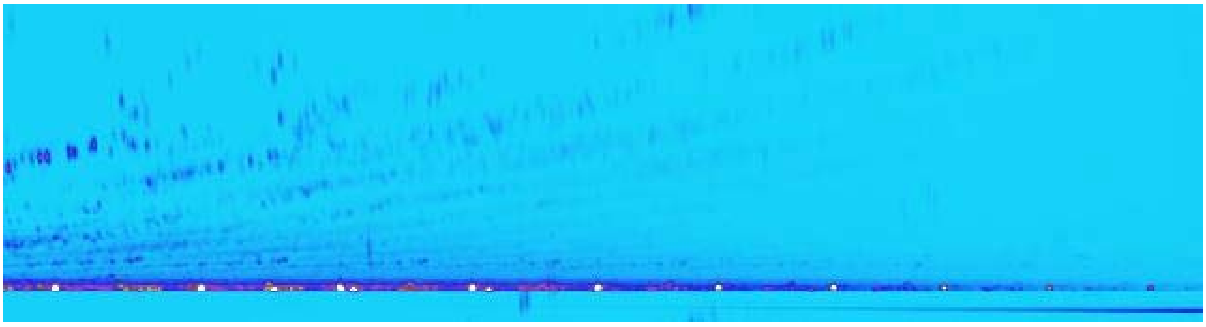

Figure 3: The right half (columns 1501-3000) of the image partially pictured in Figure 1.

3 GCxGC Image Structure and Statistics

Under well-controlled conditions, GCxGC images have structural and statistical properties that facilitate accurate determination of the background level:

1) There are *dead-bands* in the secondary chromatograms — regions where there are no sample chemical constituents and the pixel values are determined by the background level and noise.
2) The mean background level varies slowly with respect to characteristic peak widths.
3) The background noise (low-level variation about the mean background level) has the statistical properties of random white noise.

These three properties are examined in the following subsections.

3.1 Dead-bands

Under well-controlled conditions, retention-time windows within successive modulation periods do not overlap or wraparound. This ensures the presence of dead-bands at the beginning and end of each secondary column separation. The dead-bands at the bottom and top of each column of pixels are clearly visible in the structure of the GCxGC image in Figure 1.

If the operating conditions of the instrument are not carefully controlled, retention-time windows may overlap or wraparound. An example of this problem is illustrated in **Figure 3**, which is the right half of the data partially pictured in Figure 1. About half way across the sub-image in Figure 3, the secondary column separations wrap-around, overlapping the following secondary column separation. Also, in the lower portion of the image, there is a horizontal streak due to bleed from the primary column, which shortens the leading dead-band.

3.2 Mean Background Level

Under well-controlled operating conditions, both of the major components of the background level — the steady-state standing-current baseline in standard GC detectors and temperature-induced secondary column bleed — vary on a much larger time-scale than the peak widths. The dead-bands in GCxGC images facilitate determination of the background level. This is a significant advantage

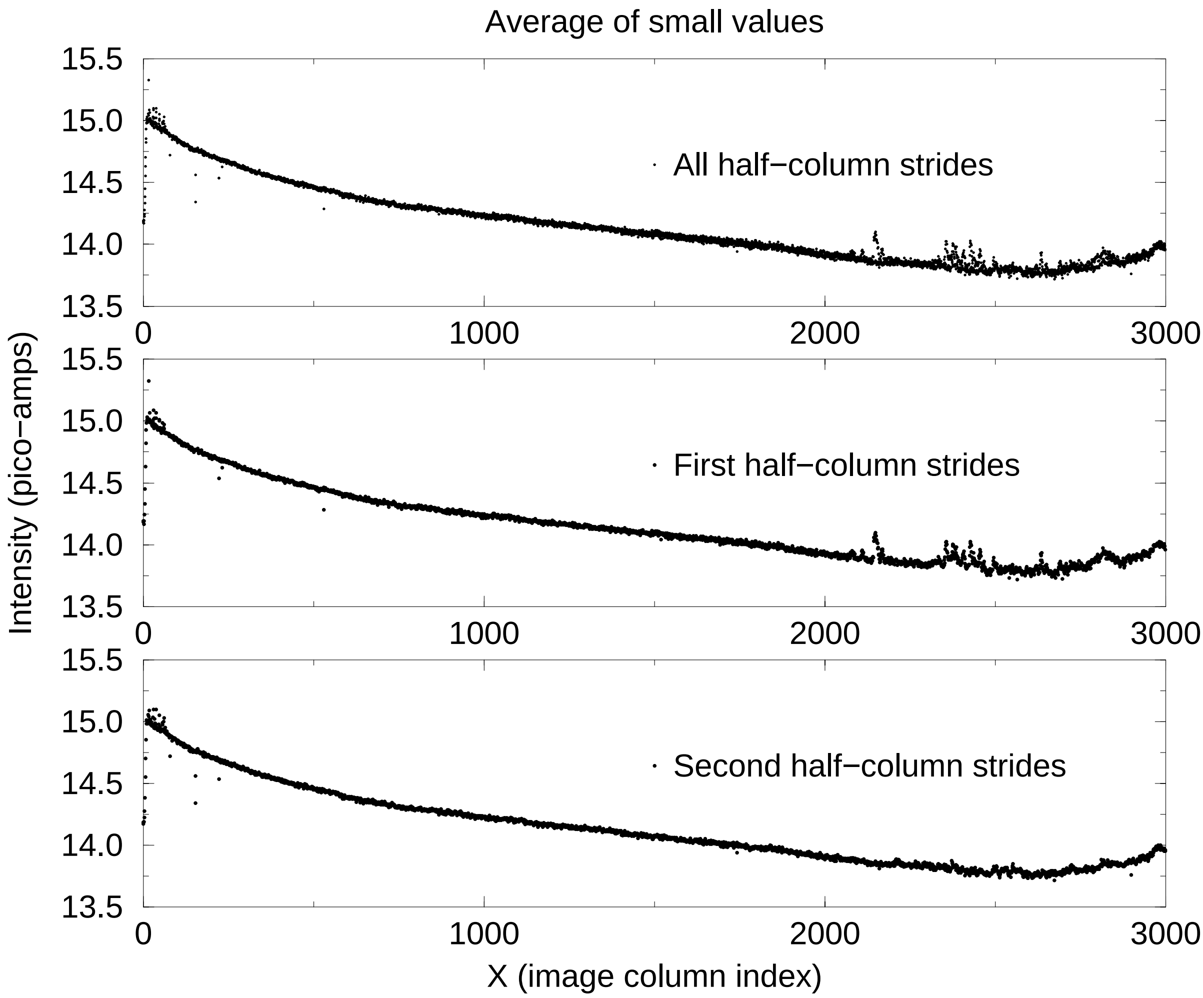

Figure 4: Smallest values (averaged by half-column) indicating background level.

for estimating the background level in GCxGC images as compared with traditional one-dimensional gas-chromatograms in which there are commonly few dead-bands.

To illustrate tracking of the background level in a GCxGC image, each column of pixels is divided into two halves or *strides*, under the assumption that there is a dead-band in each half column. Then, the smallest values in each half-column stride are extracted and averaged. This provides a good indication of the lower limits of the background pixel values as perturbed by noise). **Figure 4** plots the average of the five smallest pixels in each half-column stride of the image in Figures 1 and 3 (under the assumption that there are at least five dead-band pixels in each half-column stride).

The most important observation from Figure 4 and similar plots from other images is that the background level varies slowly with respect to peak width (which is typically only a few columns wide). Although the averages are perturbed by noise, it is clear that the smallest background values in this image rise quickly to about 15 pico-amps, then gradually decrease to about 13.75 pico-amps near the end of the image, before rising back to about 14 pico-amps. Some other observations are:

- The first and second half-column strides (from the image top and bottom, respectively) have virtually identical averages, again indicating that the background level varies slowly.
- There are occasional outliers that are not accurate indicators of the background level.
- The background level can be difficult to discern in strides where there are no true dead-bands (as in the first half-column strides near the end of the image).

The averages presented in Figure 4 are from the smallest pixel values in each stride and therefore are biased estimates of the mean background level. If there are more than five dead-band pixels in the stride, then it is likely that the five smallest values average to less than the mean background level. Therefore, the mean background level in each stride is probably greater than the average of the five smallest values. However, as described later in this paper, if the statistical properties of the noise are inferred, then it is possible to eliminate the bias and to accurately estimate the range of background noise and the mean background level.

3.3 Noise

To evaluate the statistics of the background noise, it is necessary to obtain a relatively unbiased sampling of background pixels. The pixels neighboring the smallest-valued pixels can provide an unbiased sampling of the background if a) pixels adjacent to the smallest pixels are also in dead-bands and b) the neighboring background pixels are uncorrelated. To test the first condition, **Figure 5A** plots the value differences between the five smallest pixels and their nearest neighbors. The differences are small and fairly constant (averaging 0.0345 pico-amps in this image). The values of what we will call the *adjacent pixels* closely track the values of the smallest pixels, indicating that most of the neighboring pixels are also in the dead-band. The second condition is tested in **Figure 5B**, which plots the autocorrelation in each stride between the small-valued pixels (index t) and the adjacent pixels (indexes t-1 and t+1). The autocorrelation coefficient of neighboring pixels is near zero (about 0.120 for this image), indicating that the adjacent pixels are relatively uncorrelated. These tests indicate that most adjacent pixels are in dead-bands and are relatively uncorrelated and therefore can provide relatively unbiased estimates of the background noise statistics.

In addition to near-zero autocorrelation, the background noise exhibits other attractive statistical properties including high signal-independence and time-invariance and a Gaussian normal distribution. **Figure 6A** illustrates the variations in each stride of the adjacent pixels about their mean. The noise distributions are nearly the same across half-column strides and are relatively independent of both the background level and time. **Figure 6B** illustrates the probability density

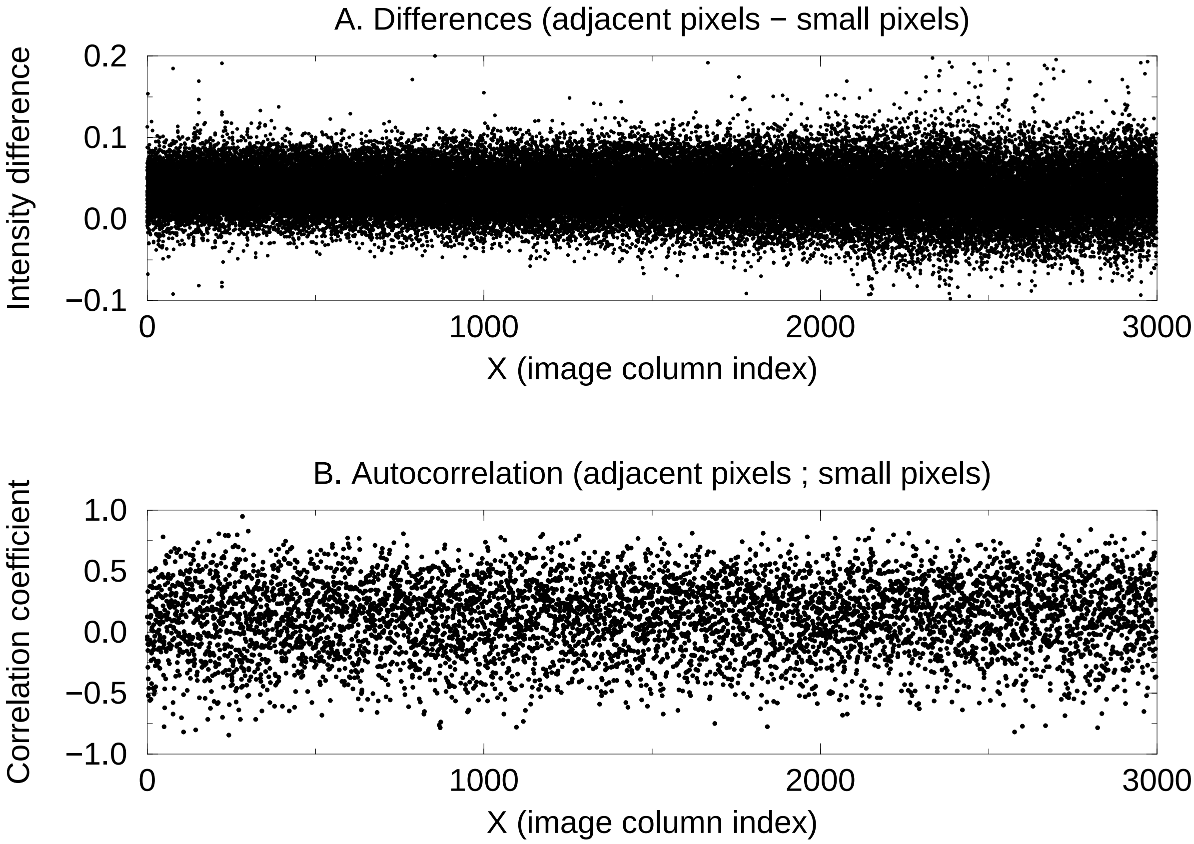

Figure 1: A. Differences between the adjacent pixels and the small pixels. B. Autocorrelations between adjacent pixels and small pixels.

function in all strides of the variations about the local mean. The distribution function of the noise in the adjacent pixels closely matches the Gaussian normal distribution.

The estimated standard deviation of the noise in the adjacent pixels in this image is about 0.244 pico-amps. The average difference between the adjacent pixels and the smallest pixels is relatively constant (Figure 5A) and about 1.41 times the estimated standard deviation of the noise (Figure 6B). This is consistent with the expectation that the smallest values in each stride fall on the low side of the expected range of background values.

4 Background Estimation and Removal

As seen in the previous section, the structure of GCxGC images provides regions of pixels from which the statistical properties of the background can be analyzed. Based on these structural and statistical properties, the background level is estimated and removed in steps:

1) Locate the pixels with the smallest values in each stride.
2) Locate the adjacent pixels (pixels before and after the smallest-valued pixels in each stride), and then use mean and median computations to estimate the local mean and standard deviation for each stride.

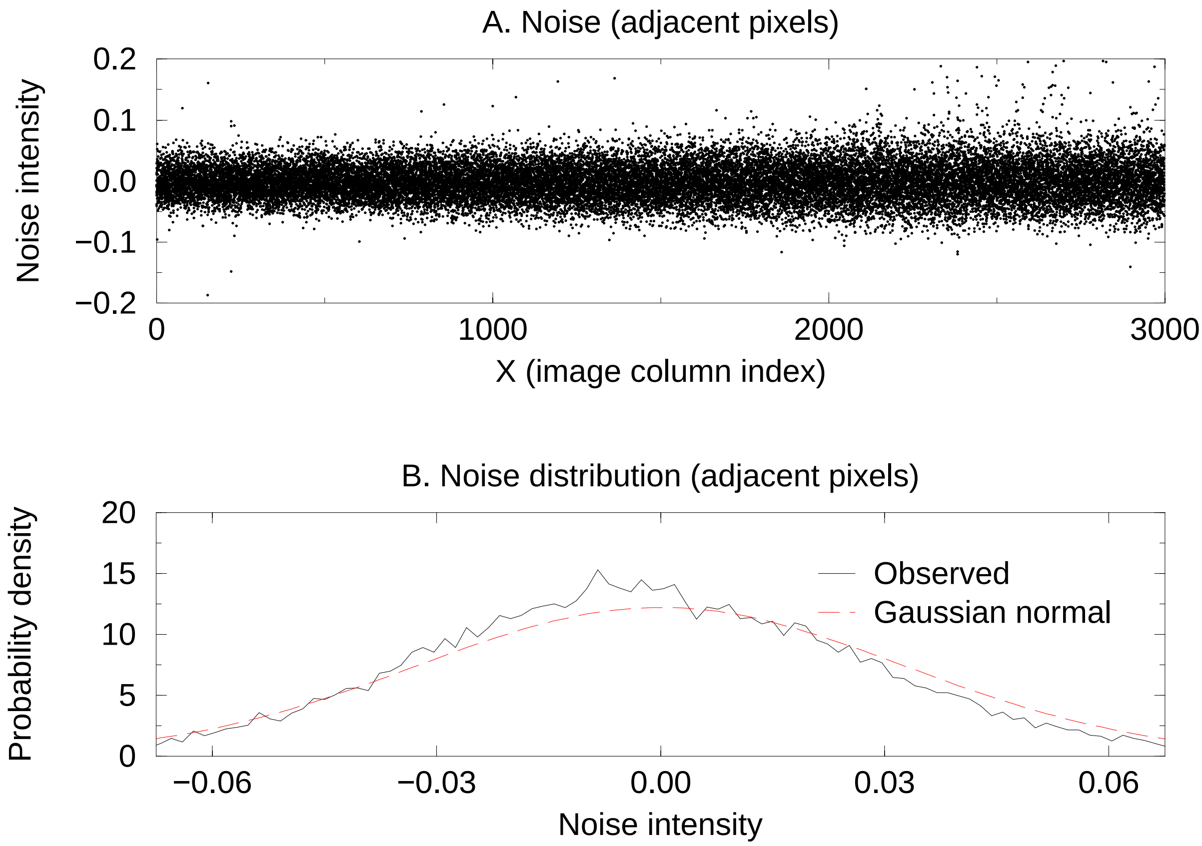

Figure 6: A. Noise in the adjacent pixels (value – stride mean). B. Probability density of noise in the adjacent pixels.

3) Locate all qualifying pixels with values in the estimated effective range of the background values in each stride and allowable gradient (explained below), then use mean and median computations to estimate the background level in each stride.
4) Interpolate the estimated background level in each stride to yield per-pixel background levels, and then subtract the per-pixel background level from each pixel.

Steps 1 and 2 follow the analysis of the statistical properties presented in the previous section with mean and median filtering to eliminate outliers and smooth estimates. Step 3 computes the background level from all of the pixels with values in the estimated effective range of the background values and with allowable gradient. (The gradient at pixel t is computed as half of the difference between the values at t+1 and t-1.) The limit on the gradient helps reject pixels at sharp intensity transitions such as on blob edges. Step 3 could be implemented by simply taking the mean computed in Step 2, particularly as the background values have a Gaussian normal distribution. Using all pixels in Step 3, rather than the mean from Step 2, requires additional computation, but the estimate is based on a larger sampling of pixels. Step 4 removes the estimated local background level from each pixel in the GCxGC image.

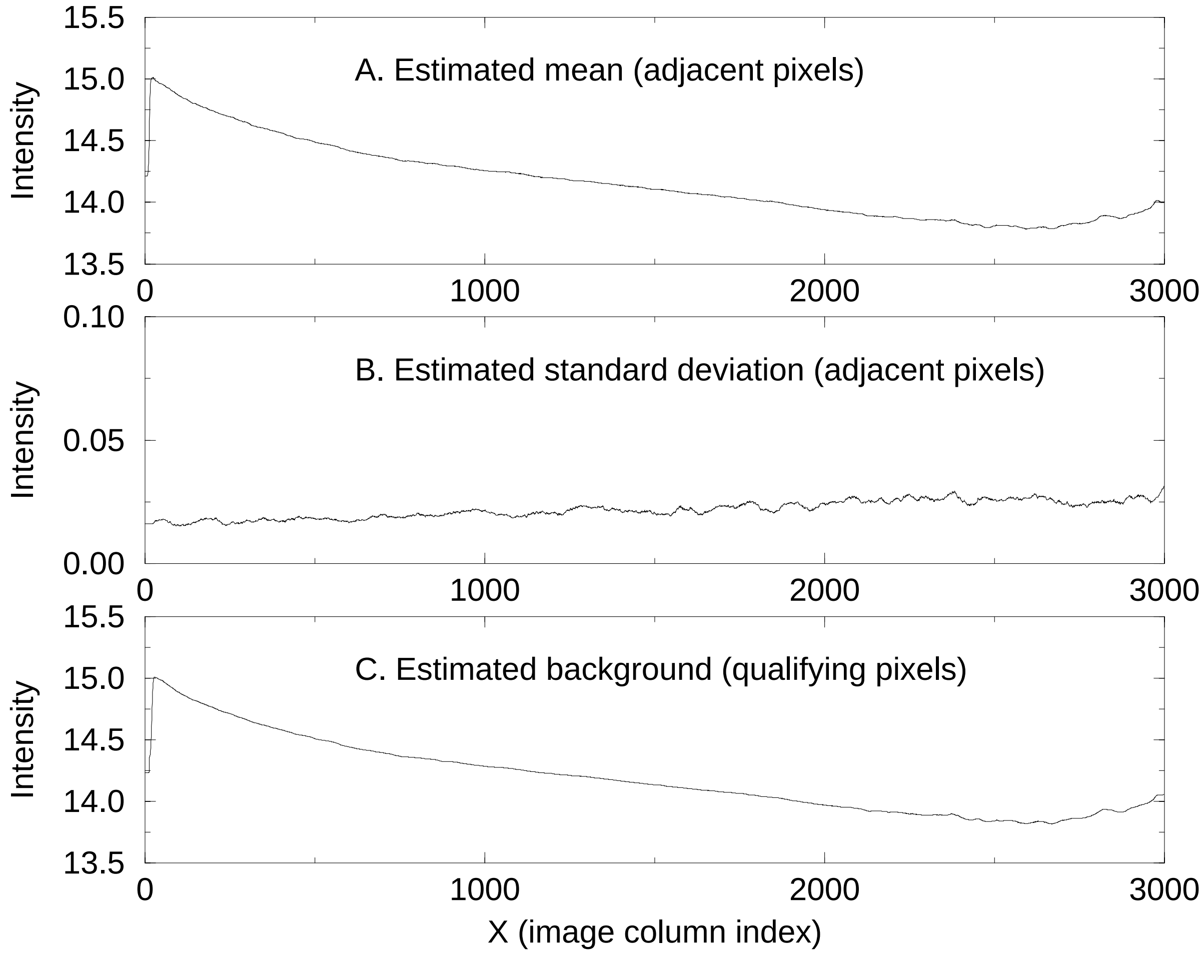

Figure 2: A. Initial estimate of mean background level (Step 2). B. Initial estimate of background noise standard deviation (Step 2). C. Final estimate of background level (Step 3).

Figure 7A illustrates the estimated mean of background values in each stride (Step 2). Mean and median filters provide a smoother function as compared to Figure 4. **Figure 7B** illustrates the estimated standard deviation of background values in each stride (Step 2). Mean and median filters provide a smooth estimate. **Figure 7C** illustrates the estimated background level in each stride, computed with mean and median filtering of all the pixels in the estimated effective range of the stride (Step 3). The estimated background level is relatively smooth. **Figure 8** illustrates the qualifying pixels used in computing the background level in Step 3.

This approach is demonstrated experimentally to be effective at determining and removing the background level from GCxGC images. **Table 1** reports statistics, after background removal, of several rectangular regions without apparent peaks. The background is corrected to be nearly zero-mean. (As in Figure 6B, the distribution is nearly Gaussian.) In these regions, the residual level after background removal is on the order of 0.005 pico-amps, which is about one-fifth of one

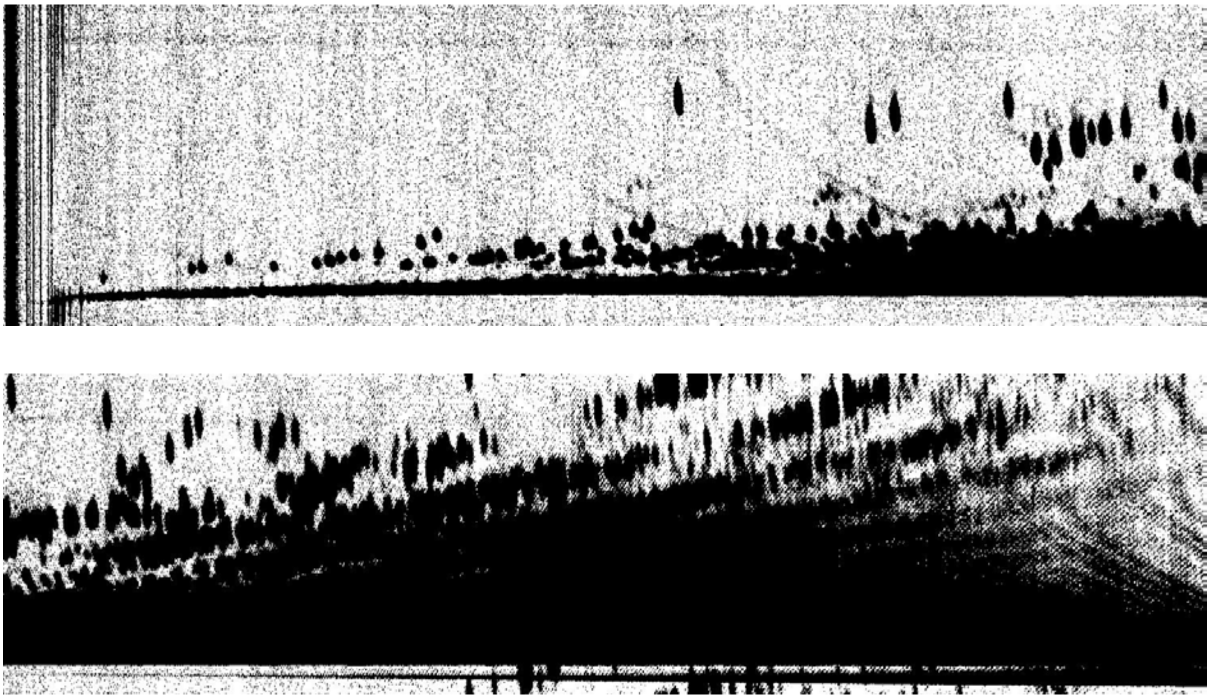

Figure 3: Qualifying pixels for background estimation (Step 3), top image is left half, bottom image is right half.

standard deviation of the background noise. This small offset may be due to small non-background signal or to small, systematic error in background removal.

The effect of residual background offset on accuracy in quantifying chemical constituents depends on the quantity of the constituent. For the peak shown in Figure 2, the volume is more than 700 pico-amps in an area of 300 pixels. An offset error of 0.005 pico-amps over an area of 300 pixels would cause an error in the volume measure of about 1.5 pico-amps. For this peak, the relative error is 0.2%. For the largest peak in this image, which has a volume greater than 21,000 pico-amps, the relative error would be about 0.01%. The bias error would be reduced by calibration to internal standards.

The algorithm provides several parametric controls, including:

Size	Position (X,Y)	Mean	Standard Deviation
420x30	(0,1640)	0.00396	0.03340
680x220	(80,55)	0.00691	0.02716
1400x30	(170,370)	0.00526	0.02699
500x15	(1570,385)	–0.00085	0.03218

Table 1: Statistics of several regions without apparent peaks.

- The size of the stride (*e.g.*, the default is half of the number of pixels in each secondary separation).
- The number of small pixels examined in each stride (*e.g.*, the default is five pixels in each stride).
- The mean and median filters used to estimate the low-end and magnitude of the effective range of background values and the mean background level.
- The size of the effective range of background values in terms of the estimated standard deviation of the background noise (*e.g.*, the default figures a range of four standard deviations from low-end that is two standard deviations below the mean, encompassing more than 95% of a Gaussian normal distribution).
- The interpolation function used to compute the per-pixel background level (*e.g.*, the default is a piecewise-cubic spline).

The background-removal algorithm is included in the GC Image software system, which provides a suite of tools with graphical user interface (GUI) for examining, processing, and analyzing GCxGC images. GC Image is being developed in Computer Science and Engineering Department of the University of Nebraska – Lincoln (http://cse.unl.edu/~gcimage/usersguide.html) and is available for license through Zoex Corporation (http://www.zoex.com).

5 Related Operations

Removing the background level is only the first step in analyzing GCxGC images. After the background is removed, the image must be segmented into background and individual blobs. Removing the background is necessary for accurate quantification of blob volume, but the algorithm developed in this paper provides other information useful in GCxGC analysis. The estimate of the noise variance determined during background removal is useful in algorithms for detecting the presence of small blobs, for accurately detecting the extent (or area) of blobs, and for separating closely adjacent blobs. Moreover, accurate extents, detection of small blobs, and separation of adjacent blobs contribute to the automated matching of observed blobs to catalogued GCxGC blob templates.

References

[1] E.B. Ledford, Jr., C.A. Billesbach, *Journal of High Resolution Chromatography*, **2000**, 23(3):202-204.

[2] J. Beens, M. Adahchour, R.J.J. Vreuls, K. van Altena and U.A.T. Brinkman, *Journal of Chromatography A*, **2001**, 919(1):127-132.

[3] G.S. Frysinger, R.B. Gaines, E.B. Ledford, Jr., *Journal of High Resolution Chromatography*, **1999**, 22(4):195-200.