Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Chromatography A, 985 (2003) 47–56
JOURNAL OF CHROMATOGRAPHY A
www.elsevier.com/locate/chroma

Review

Image background removal in comprehensive two-dimensional gas chromatography

Stephen E. Reichenbach[a,*], Mingtian Ni[a], Dongmin Zhang[a], Edward B. Ledford Jr.[b]

[a]*Computer Science and Engineering Department, University of Nebraska, Lincoln, Lincoln, NE 68588-0115, USA*
[b]*Zoex Corporation, 2611 W M Street, Suite D, Lincoln, NE 68522, USA*

Abstract

This paper describes a new technique for removing the background level from digital images produced in comprehensive two-dimensional gas chromatography (GC×GC). Background removal is an important first step in the larger problem of quantitative analysis. The approach estimates the background level across the chromatographic image based on structural and statistical properties of GC×GC data. Then, the background level is subtracted from the image, producing a chromatogram in which the peaks rise above a near-zero mean background. After the background level is removed, further analysis is required to determine the quantitative relationship between the peaks and chemicals in the sample. The algorithm is demonstrated experimentally to be effective at determining and removing the background level from GC×GC images. The algorithm has several parametric controls and is incorporated into an interactive program with graphical interface for rapid and accurate detection of GC×GC peaks.

Keywords: Gas chromatography, comprehensive two-dimensional; Image background removal

Contents

*Corresponding author. Tel.: +1-402-472-5007; fax: +1-402-472-7767.
E-mail addresses: reich@cse.unl.edu (S.E. Reichenbach), http://cse.unl.edu/~reich (S.E. Reichenbach).

0021-9673/02/$ – see front matter
PII: S0021-9673(02)01498-X

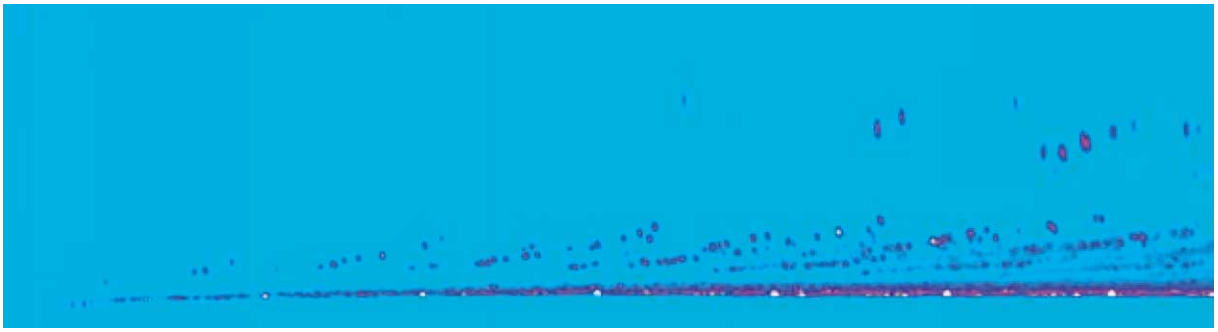

Fig. 1. The left half (columns 1–1500) of an image from two-dimensional gas chromatography (GC×GC).

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 many GC detectors) and temperature-induced column-bleed. 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 (GC×GC). Comprehensive two-dimensional gas chromatography separates chemical species with two capillary columns interfaced by two-stage thermal desorption [1,2]. As shown in Fig. 1, the GC×GC 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 signal intensity 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 Fig. 1, the image is colorized with the smaller values of the background appearing light blue and the larger values of the blob peaks appearing dark blue, magenta, and white.

Fig. 2 illustrates a perspective plot of a typical GC×GC blob peak from Fig. 1, corresponding to

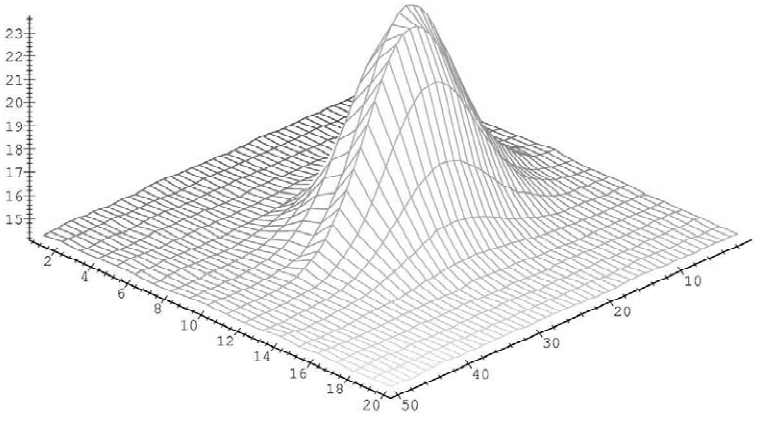

Fig. 2. Perspective plot of the GC×GC data containing the isolated blob of pixels in the upper middle of Fig. 1, corresponding to Naphthalene.

naphthalene. The background level in the vicinity of the naphthalene peak is about 14 pA and the peak maximum rises to more than 23 pA. 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 the peak rises above a near-zero mean background. Background removal is an important first step in the larger problem of quantitative analysis. After the background level is removed, further analysis is required to determine the quantitative relationship between peaks in the signal and chemicals in the sample (e.g., dealing with chemically dependent detector responses).

The approach developed in this paper estimates the local background level across the entire chromatogram based on structural and statistical properties of GC×GC data. Then, the local background level is subtracted across the entire image, producing a chromatogram in which the peaks rise above a near-zero mean background. This algorithm is demonstrated experimentally to be effective at determining and removing the background level of GC×GC 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 GC×GC peaks.

2. Experimental instrumentation

The image in Fig. 1 was acquired with an Agilent Model 6890 gas chromatograph with a flame ionization detection (FID) system. The GC×GC separation was implemented with a thermal modulator (prototype Model KT2002) fabricated by Zoex (Lincoln, NE, USA), employing dual pairs of pulsed hot and cold gas jets. The cold jet temperature was approximately −189 °C, cooled by heat exchange with liquid nitrogen contained in a holding dewar mounted atop the GC system. The hot jet temperature was at all times approximately 100 °C above the programmed oven temperature, maintained by a cartridge heater under the control of the AUX 2 channel of the 6890.

For the image in Fig. 1, the first dimension column was 4.0 m long×0.1 mm I.D. and coated with a 3.5 μm thick film of methyl silicone stationary phase (Quadrex, New Haven, CT, USA). Modulation was performed on the head of the secondary column, which was 0.2 m×0.1 mm I.D. and coated with 0.1 μm thick film of carbowax (Quadrex). The primary and secondary columns were connected by means of a borosilicate glass press-fit connector (Zoex). The GC oven was programmed from 35 to 225 °C over 100 min. Diesel fuel obtained from a local fueling station was injected neat [3]. Injection volume was 1.0 μl using split injection mode, a split ratio of 100, and an injector temperature 280 °C.

The image in Fig. 1 was acquired with a 2-s thermal modulation, at a rate of 200 samples/s, collecting 400 pixels for each of 3000 image columns over a period of 100 min. (Only the first 1500 image columns are 4 shown in Fig. 1.)

The images presented in Section 5 were acquired with a Zoex production model KT2002 and a secondary column 0.4 m long. Those images were acquired with a 3-s thermal modulation, at a rate of 200 samples/s, yielding a 600 pixels in each of 1200 columns over 60 min. Other images, not presented in this paper, were acquired with similar configurations and with similar results.

3. GC×GC image structure and statistics

Under well-controlled conditions, GC×GC 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 GC×GC image in Fig. 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 Fig. 3, which is the right half of the data partially pictured in Fig. 1. About half way across the sub-image in Fig. 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—steady-state standing-current baseline in GC detectors and temperature-induced secondary column bleed—vary on a much larger time-scale than the peak widths. The dead-bands in GC×GC images facilitate determination of the background level. This is a significant advantage for estimating the background level in GC×GC 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 GC×GC 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. Fig. 4 plots the average of the five smallest pixels in each half-column stride of the image in Figs. 1 and 3 (under the assumption that there are at least five dead-band pixels in each half-column stride).

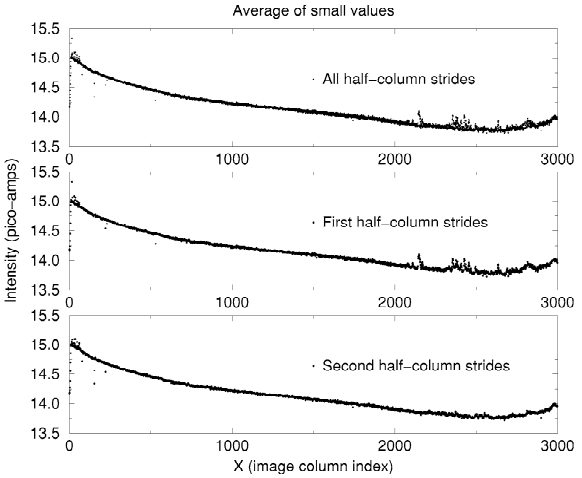

Fig. 4. Smallest values (averaged by half-column) indicating background level.

The most important observation from Fig. 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

Fig. 3. The right half (columns 1501–3000) of the image partially pictured in Fig. 1.

image rise quickly to about 15 pA, then gradually decrease to about 13.75 pA near the end of the image, before rising back to about 14 pA. Some other observations are:

(i) 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.

(ii) There are occasional outliers that are not accurate indicators of the background level.

(iii) The background level can be difficult to discern in strides where there are not sizeable dead-bands (as in the first half-column strides near the end of the image).

The averages presented in Fig. 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, Fig. 5A plots the value differences between the five smallest pixels and their nearest neighbors. The differences are small and fairly constant (averaging 0.0345 pA 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 Fig. 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

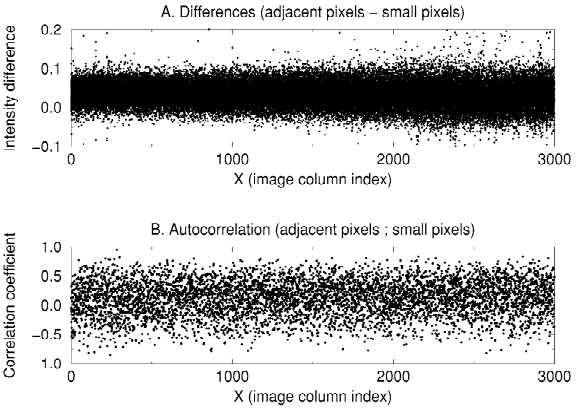

Fig. 5. Analysis of pixels adjacent to the smallest pixels. (A) Differences between the adjacent pixels and the small pixels. (B) Autocorrelations between adjacent pixels and small pixels.

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. Fig. 6A illustrates the variations in each stride of the adjacent pixels about their mean. The noise dis-

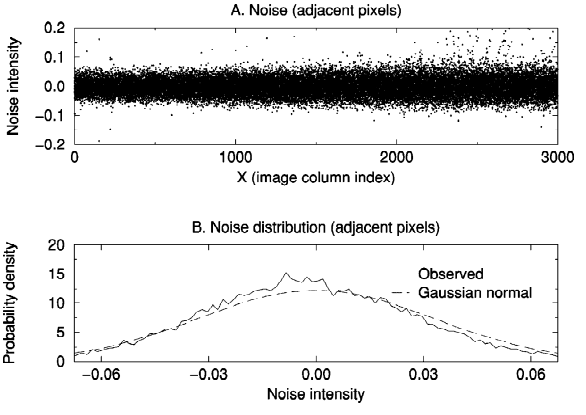

Fig. 6. Analysis of noise. (A) Noise in the adjacent pixels (value minus mean). (B) Probability density of noise in the adjacent pixels.

tributions are nearly the same across half-column strides and are relatively independent of both the background level and time. Fig. 6B illustrates the probability density function in all strides of the variations of adjacent pixels about their 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 pA. The average difference between the adjacent pixels and the smallest pixels is relatively constant (Fig. 5A) and about 1.41 times the estimated standard deviation of the noise (Fig. 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 GC×GC 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 background mean and standard deviation for each stride.

(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 GC× GC image.

Fig. 7A illustrates the estimated mean of background values in each stride (Step 2). Mean and median filters provide a smoother function as compared to Fig. 4. Fig. 7B illustrates the estimated standard deviation of background values in each stride (Step 2). Mean and median filters provide a smooth estimate. Fig. 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. Fig. 8 illustrates the qualifying pixels used in computing the background level in Step 3.

This approach is demonstrated experimentally to

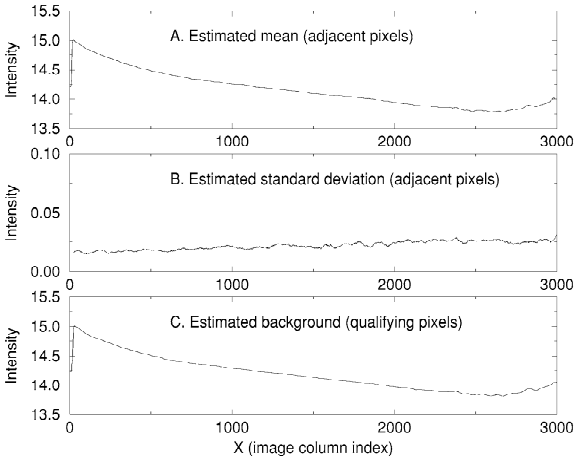

Fig. 7. Estimating the background level. (A) Initial estimate of background level (Step 2). (B) Initial estimate of background noise standard deviation (Step 2). (C) Final estimate of background level (Step 3).

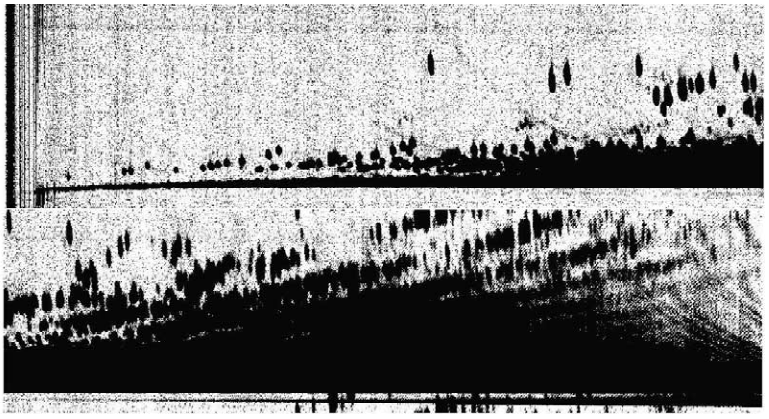

Fig. 8. Qualifying pixels (in white) for background estimation (Step 3), top image is left half, bottom image is right half.

be effective at determining and removing the background level from GC×GC 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 Fig. 6B, the distributions are nearly Gaussian.) In these regions, the residual level after background removal is on the order of 0.005 pA, which is about one-fifth of one 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 Fig. 2, the volume is more than 700 pA in an area of 300 pixels. An offset error of 0.005 pA over an area of 300 pixels would cause an error in the volume measure of about 1.5 pA. For this peak, the relative error would be about 0.2%. For the largest peak in this image, which has a volume greater than 21 000 pA, the relative error would be about 0.01%. The bias error would be reduced by calibration against internal standards.

Table 1
Statistics of several regions without apparent peaks

Size	Position (*X*,*Y*)	Mean	Standard deviation
420×30	(0,1640)	0.00396	0.03340
680×220	(80,55)	0.00691	0.02716
1400×30	(170,370)	0.00526	0.02699
500×15	(1570,385)	0.00085	0.03218

The algorithm provides several parametric controls that can be used to adjust background detection, including:

(i) The size of the stride (e.g., the default is half of the number of pixels in each secondary separation).

(ii) The number of small pixels examined in each stride (e.g., the default is five pixels in each stride).

(iii) 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.

(iv) 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).

(v) The interpolation function used to compute the per-pixel background level (e.g., the default is a piecewise-cubic spline).

5. Experiments with blank images

In blank GC×GC images, run under set conditions but without sample input, the background level is not obscured by sample-related peaks. Figs. 9 and 10 illustrate two blank images before and after the background was removed by the algorithm

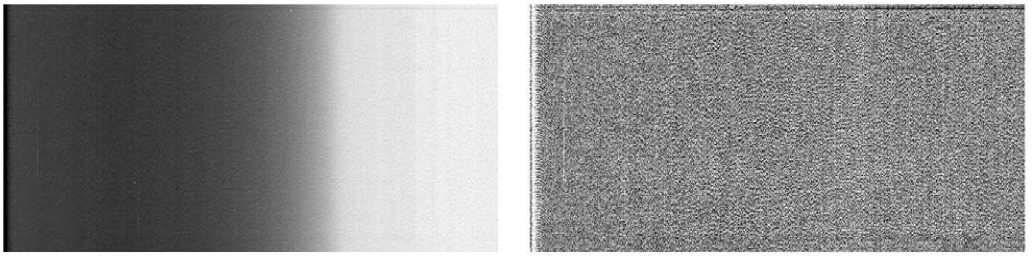

Fig. 9. Blank image 1, before and after background removal.

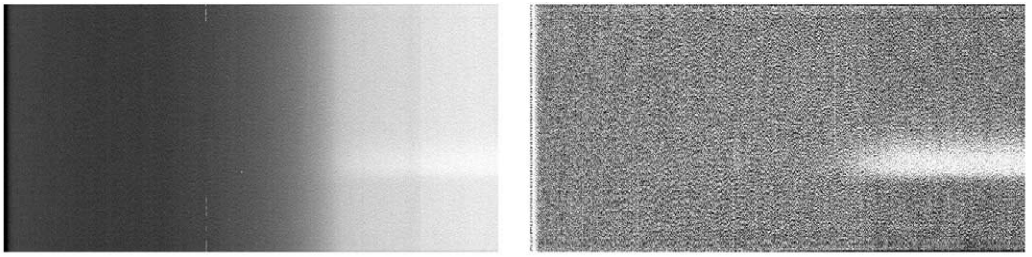

Fig. 10. Blank image 2, before and after background removal.

described in this paper. The images before background removal are displayed with a grayscale from 0.5 pA less than the mean to 0.5 pA greater than the mean. The background level varies across the image left-to-right from well below the mean to well above the mean. The images after background removal are displayed with a grayscale from −0:05 pA to +0:05 pA—one-tenth of the range used to display the images before background removal. In the images after background removal, even with a very small grayscale range, the image appears homogeneous and closely resembles zero-mean white noise.

The background removal algorithm is effectively removes the background, but it does not remove some artifacts observed in these images. Both images have a few pixels with spike noise and the second blank image has an area of elevated values on the right side of the image. The background removal algorithm is not designed to remove these artifacts. Also, the background removal algorithm does not accurately remove the background in the first few columns. As the acquisition begins, the baseline rises very quickly in the first few columns. The baseline in these first few columns violates one of the assumptions underlying the algorithm—that the baseline is slowly varying. Therefore, the background removal algorithm fails to track the background in these first few columns. Typically, this is not a problem because analytes do not elute in the first few columns of the image.

Figs. 11 and 12 illustrate the column and row averages after background removal. In both images, in both columns and rows, the pixel average is near zero. The column averages illustrate two artifacts not corrected by the algorithm. As described above, the

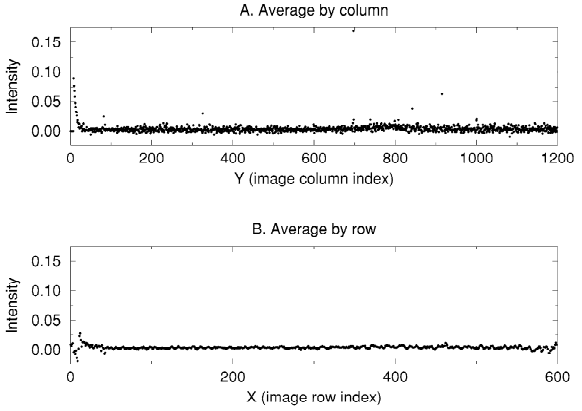

Fig. 11. Blank image 1 column and row averages after background removal.

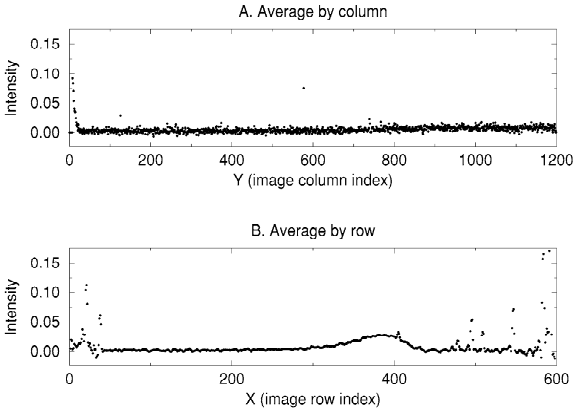

Fig. 12. Blank image 2 column and row averages after background removal.

algorithm does not accurately track the background in the first few columns where the baseline rises sharply during startup. (Because of this artifact, the first few columns are not included in the row averages.) Also, a few columns have a few pixels with spike noise that elevates the average. The row averages in Fig. 12 show the presence of the region of elevated pixels. Except for these artifacts, the background removal algorithm yields images that are homogeneous with near-zero mean.

Fig. 13 illustrates a GC–GC image of diesel fuel separated under the same conditions as the two blank images. The pattern of a rising background left-to-right is evident in the image before background removal (displayed with grayscale range of 1.0 pA), even though the presence of sample chemicals obscures the background in much of the right side of the image. After background removal, the background appears to be homogeneous with near-zero mean.

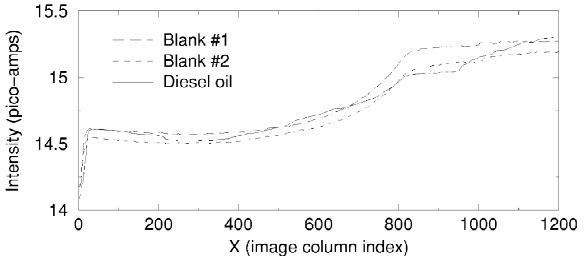

Fig. 14. Background detected and removed in background images 1 and 2 and diesel fuel image.

Fig. 14 illustrates the background that is estimated and removed by the algorithm. Even though the conditions are nominally the same, the background levels of the two blank images vary slightly. The background level detected in the diesel fuel image closely tracks the background levels observed in the blank image. In most of the image, the background level detected in the diesel fuel image is between the levels of the two blank images.

6. Related operations

Removing the background is necessary for accurate quantification of blob volume, but it is only the first step in analyzing GC×GC images. After the background is removed, the image must be segmented into background and individual blobs. Then, the blobs must be quantified, accounting for a variety of factors, including detector response. It also is desirable to identify the chemical associated with each peak.

The background removal algorithm developed in this paper provides information that is useful in subsequent GC×GC analyses. The estimate of the

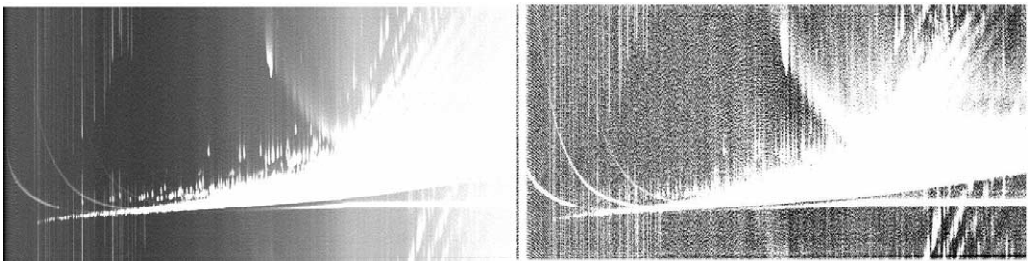

Fig. 13. Diesel fuel image, before and after background removal.

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 GC×GC blob templates.

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 GC×GC 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) in collaboration with Zoex Corporation (http://www.zoex.com).

7. Conclusion

This paper describes a new technique for removing the background level from digital images produced in comprehensive two-dimensional gas chromatography (GC×GC). The approach estimates the background level across the chromatographic image based on a few structural and statistical properties of GC–GC data. Then, the background level is subtracted from the image, producing a chromatogram in which the peaks rise above a zero-mean background. Background removal is an important first step in the larger problem of quantitative analysis. After the background level is removed, further analysis is required to determine the quantitative relationship between the peaks and chemicals in the sample (e.g., dealing with chemically dependent detector responses).

The structural and statistical properties of data obtained by a well-controlled GC×GC 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 GC×GC images. The algorithm 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 GC×GC peaks.

References

[1] E.B. Ledford Jr., C.A. Billesbach, J. High Resolut. Chromatogr. 23 (2000) 202.
[2] J. Beens, M. Adahchour, R.J.J. Vreuls, K. van Altena, U.A.Th. Brinkman, J. Chromatogr. A 919 (2001) 127.
[3] G.S. Frysinger, R.B. Gaines, E.B. Ledford Jr., J. High Resolut. Chromatogr. 22 (1999) 195.