Signal Detection for Data Sets with a Signal-to-Noise Ratio of 1 or Less with the Use of a Moving Product Filter

H. G. SCHULZE, L. S. GREEK, C. J. BARBOSA, M. W. BLADES, and R. F. B. TURNER*

Biotechnology Laboratory, The University of British Columbia, 237-6174 University Boulevard, Vancouver, B. C., Canada, V6T 1Z3 (H.G.S., L.S.G., C.J.B., R.F.B.T.); Department of Electrical Engineering, The University of British Columbia, 434-2356 Main Mall, Vancouver, B. C., Canada, V6T 1Z4 (L.S.G., R.F.B.T.); and Department of Chemistry, The University of British Columbia, E257-2036 Main Mall, Vancouver, B. C., Canada, V6T 1Z1 (C.J.B., M.W.B.)

We report on a method to reduce background noise and amplify signals in data sets with low signal-to-noise ratios (SNRs). This method consists of taking a data set with mean 0 and normalized with respect to absolute value, adding 1 to all values to adjust the mean to 1, and then applying a moving product (MP) to the transformed data set (similar to the application of a moving average or 0-order Savitzky–Golay filtering). A data point in the presence of a signal raises the probability of that data point having a value >1, while the absence of a signal increases the probability of that data point having a value <1. If the autocorrelation lag of the signal is larger than the autocorrelation lag of the associated noise, the use of an MP with window comparable to that of the signal width (i.e., 2–3 times the signal standard deviation) will tend to reduce the values of data points where no signal is present and similarly amplify data points where signal is present. Signal amplification, often to a considerable degree, is gained at the cost of signal distortion. We have used this method on simulated data sets with SNRs of 1, 0.5, and 0.33, and obtained signal-to-background noise ratio (SBNR) enhancements in excess of 100 times. We have also applied this procedure to low SNR measured Raman spectra, and we discuss our findings and their implications. This method is expected to be useful in the detection of weak signals buried in strong background noise.

Index Headings: **Noise reduction; Signal amplification; Moving product; Low signal-to-noise ratios.**

INTRODUCTION

Noise is ubiquitous in scientific measurements. We shall consider noise to mean here "instantaneously irreproducible" (although statistical characterization is possible) following the usage by Barclay et al.[1] Where signal levels are large, the presence of noise is generally not a problem. However, weak signals can often be obscured by noise, thus rendering signal detection and/or identification difficult. Therefore, the ability to enhance the signal relative to the noise, especially the background noise, can be of considerable importance in signal detection. Consequently, various methods have been developed to reduce the influence of noise in data sets. These operate in the time domain, (e.g., Savitzky–Golay filters)[2] and in the frequency domain (e.g., Fourier transform filters),[3] employ discrete wavelet transforms,[1] or use maximum entropy processing,[4] to name but a few approaches.

We are reporting here on a method that employs a moving product (MP) to selectively amplify the signal and suppress the noise content of a data set employing synthetic data and real data (consisting of Raman scattering measurements). These results are compared to those for Savitzky–Golay filtering and ensemble averaging.

THEORY

Measured data can often be modeled as

$$\begin{aligned} D_i &= \sum B_{ki}(f_i + b_i) + \sigma_i \\ &= B{*}(f_i + b_i) + \sigma_i \end{aligned} \tag{1}$$

where the D_i are the measured data points, the f_i represent the ideal underlying spectrum and the b_i those of the background convolved ($*$) with the instrumental response or blurring function B (of bandwidth k), and the σ_i represents the standard deviation of the noise. Since we will not consider deconvolution here, we can simplify Eq. 1 by restating it as

$$D_i = f_i + b_i + \sigma_i. \tag{2}$$

A basic problem of signal processing is how to detect, recover, identify, and quantify the underlying spectrum f_i from the imperfect, incomplete, and contaminated measured data. An inspection of Eq. 2 suggests a number of approaches: (1) removal of the background b_i; (2) removal or reduction of the noise σ_i; and (3) amplification of the signal f_i. Let us assume that the background can be removed from the data set without undue difficulty by the subtraction of a blank or the fitting of a polynomial to the data set and subtracting the fitted values from the data set.[5,6] Equation 2 now reduces to the basic additive model

$$D_i = f_i + \sigma_i. \tag{3}$$

It is now assumed (for the sake of the argument only) that the noise has a Gaussian probability distribution with 0 mean, that the signal has a single peak with full width at half-maximum (FWHM) of n data points, that the peak is situated at point k in the spectrum and is 0 at all other points, that the data set is normalized with regard to its maximum absolute value, that the constant $c = 1$ is added to all data points, and that no data point exactly equals 0 (by adjusting c somewhat if necessary). In the absence of more information, the expected value of any single specimen from a population is equal to the mean of that population.[7] Let the mean of the signal be x. Let the maximum number of data points in the data set be max. Then the expected values are $E(\sigma_i) = 0$ and $E(f_i) = x$. Consequently for those data points containing the signal,

Received 18 August 1997; accepted 11 December 1997.

* Author to whom correspondence should be sent.

0003-7028/98/5204-0621$2.00/0

$$E(D_i) = E(f_i) + E(\sigma_i) + 1, \qquad (k-n <= i <= k+n)$$
$$= x + 0 + 1$$
$$> 1, \tag{4}$$

and for those containing noise only

$$E(D_i) = E(f_i) + E(\sigma_i) + 1, \qquad (i < k-n \text{ or } i > k+n)$$
$$= 0 + 0 + 1$$
$$= 1. \tag{5}$$

In particular, the product of $2n + 1$ points containing the signal would be

$$\prod_{i=k-n}^{k+n} E(D_i) = (x + 0 + 1)^{2n+1}$$
$$\gg 1, \tag{6}$$

and those containing noise only

$$\prod E(D_i) = (0+0+1)^a, \qquad (i < k-n \text{ or } i > k+n)$$
$$= 1 \tag{7}$$

where a is any arbitrary integer such that $a <= \max - 2n + 1$ (maximum minus signal width).

However, since the noise is symmetrically distributed around the mean (i.e., approximately equal numbers of data points lie on either side of the mean), the product of any number of noise data points would tend to 0. In order to see this, assume (as an approximation) that every data point larger than 1 can be paired with a data point an equal amount smaller than 1. Let, for any such pair of data points, the distance from the data point above or below the mean to the mean be d. Then the product of this pair can be represented as

$$(1 - d)(1 + d) = 1 - d^2$$
$$< 1. \tag{8}$$

Consequently, the product of such pairs of products would tend to 0. The preceding analysis suggests that assigning

$$D_k = \prod_{i=k-n}^{k+n} (D_i) \text{ where } n/2 <= k <= max - n/2 \tag{9}$$

would amplify the data points containing the signal while reducing the data points containing only noise. This procedure, performed on a preprocessed data set (normalized, mean adjusted to 1), replaces the center data point within a moving window with the product of all the data points within that window, a process somewhat similar to using a moving average smoothing algorithm.[3]

EXPERIMENTAL PROCEDURES

Synthetic data sets were created by adding Gaussian noise (0, 1) to three synthetic signal sets. Each signal set consisted of 2000 points that contained four Gaussian peaks with equal maxima. The peaks were centered at points 1000, 1500, 1800, and 1950 and had standard deviations of 50, 25, 12, and 6 points, respectively. The peak maxima in the three sets were 1 (signal set 1), 0.5 (signal set 2), and 0.33 (signal set 3), respectively, to create signal-to-noise ratios (SNRs) of 1, 0.5, and 0.33. In order to obtain greater statistical insight, 10 different sequences of Gaussian noise (0, 1) were added to signal set 2 to create 10 synthetic data sets, all with SNR = 0.5. Figure 1 shows the synthetic signal, a noise sequence, and signal plus noise (SNR = 0.5).

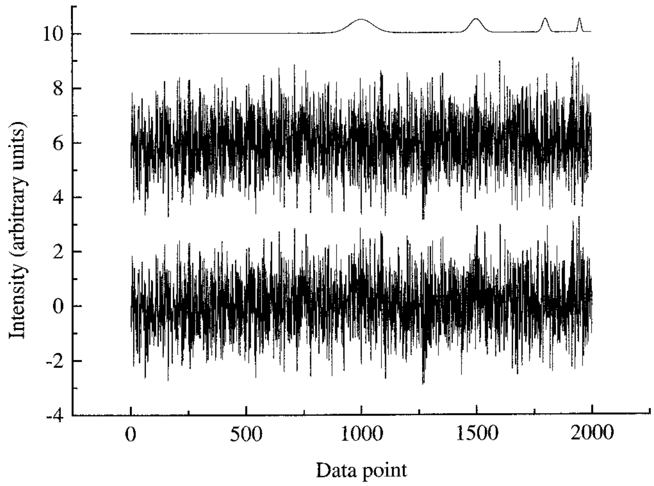

FIG. 1. An example of a synthetic data set showing the signal (top trace), noise, and signal plus noise (bottom trace, SNR = 0.5), vertically offset for ease of viewing.

Real data sets were obtained by measuring three Raman spectra of 1×10^{-3} M acetylcholine in physiological saline (0.9% NaCl) from 500 to 2000 cm^{-1}. Raman scattering was excited with the 488 nm line from an Ar^+ laser operating at 200 mW and measured with a 1 m focal length spectrometer (JASCO Model NR-1100, Tokyo, Japan) scanning at 120 cm^{-1}/min and with slits set at 500 μm. The spectra were digitized at 1 cm^{-1} intervals, resulting in 1501 points per spectrum.

The MPs for the synthetic and real data sets were calculated with a custom-written computer software program (QuickBasic 4.0, Microsoft, Redmond, WA). Spectral baselines were removed by polynomial fitting (as described earlier), except where stated otherwise. The size of the moving window could be controlled by the user. The program determined the absolute value maximum of the data set and normalized the data set with respect to the maximum. Where necessary, "normalization" was effected by division by a number slightly larger than the maximum (e.g., maximum + 0.1) to prevent any value from being exactly 0, since this would cause the product in any window containing the 0 value to be 0 also. A constant (=1) was then added to all data points in the set. Starting at the beginning of the data set, a window of preselected size was advanced by one point at a time to cover the entire data set. The product of all the data points in the window was calculated and the data point in the center of the window was replaced by this product before the window was advanced and the next product calculated on the basis of the original data values.

RESULTS

The Effect of Window Size. Increasing the window size caused the progressive loss of narrow peaks, as can be seen from Fig. 2 (SNR = 1). Even with the smallest

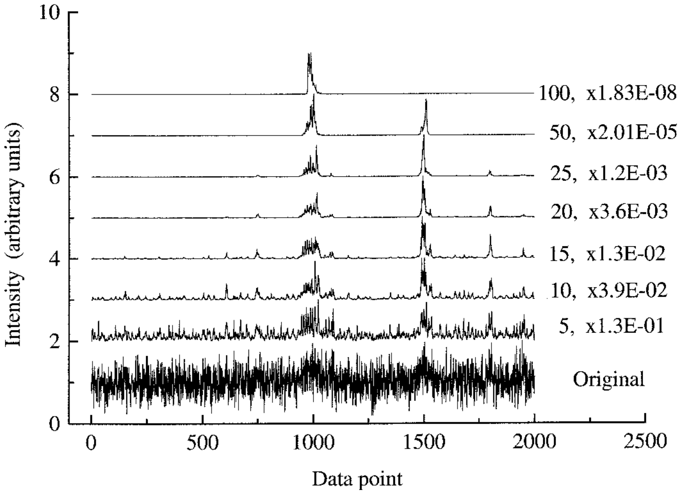

FIG. 2. The signal amplification and noise suppression for a data set with SNR = 1, showing the effect of window size on the recovery of the narrower peaks (see Fig. 1 for signal trace). The window size and scaling factor are shown on the right-hand side of each trace. The traces are vertically offset for ease of viewing.

window used (five points) the narrowest peak could not be adequately discriminated from noise, a situation that worsened with decreasing SNR. In contrast, the widest peak became progressively more amplified with increasing window size, reaching more than 54×10^6, while the largest noise peak was less than 25. Progressively narrower signal peaks were amplified correspondingly less. Most of the noise was suppressed to levels less than 1. These results were in general agreement with those expected on the basis of the preceding theoretical analysis.

With the use of different data sets ($n = 10$) with signal-to-background noise ratio (SBNR) = 0.5 (for these data sets SBNR $\approx$ SNR), the results, shown in Table I, were obtained. It is evident from Table I that enormous amplifications of the broader signal peaks can be obtained. Given that the ratio of the broadest peak maximum to the largest noise peak (not noise standard deviation) is in excess of 36 000 when the original set had an SBNR of 0.5, the implication is that the method improved the SBNR by a factor of at least 72 000. In all 10 samples, the maximum of the broadest peak was located within two standard deviations of the original position. For the peak with a 50-point standard deviation and centered at point 1000, the implication is that all the amplified maxima were observed between 900 and 1100. The use of larger windows produced greater signal amplification, but also greater variance in the position of the maximum of the amplified peak.

TABLE I. The effects of applying a moving product (150 points wide) to synthetic data sets (SNR = 0.5, n = 10) on peak maximum position (mean ± SEM), maximum peak height (mean ± SEM), ratio of peak maximum to largest peak (mean ± SEM), and number of peak maxima (out of 10) within 2 original standard deviations of the original position.[a]

Original peak width	50 Points	25 Points	12 Points
Position	995 ± 7	1505 ± 11	1824 ± 10
Maximum	$4.3 \times 10^6 \pm 4.0 \times 10^6$	$19{,}628 \pm 15{,}313$	49.9 ± 27.6
Ratio	35,607	163	0.42
Number	10	8	3

[a] Note: SEM = standard error of the mean.

TABLE II. The effects of applying a moving product (150 points wide) to synthetic data sets (SBNR = 0.5, n = 3) on line shape as evidenced by fitting a Gaussian curve to the amplified signal peaks. The table shows fitted peak center position (mean ± SEM), peak height (mean ± SEM), peak FWHM (mean ± SEM), and χ^2.[a]

n	Peak center	Peak height	Peak FWHM	χ^2
1	994.2	792	25.02	11,486
2	994.3	810	24.49	11,474
3	990.4	284,247	35.53	135,743,966
Original	1000	1	118.26	···

[a] Note: SEM = standard error of the mean.

The Effect on Line Shape. Signal amplification occurs at the cost of line shape distortion. The MP method severely distorted the original Gaussian signals with somewhat displaced peak positions, as shown in Table II.

The Effect of SNR and Cross-Multiplication of Data Sets. Spurious peaks were more problematic with reduced SBNR (hence SNR) data. Occasionally these were as large as the two most highly amplified peaks of the signal. However, due to the random nature of the noise, these spurious peaks occurred in random positions in the synthetic data sets. This behavior made possible the point-for-point multiplication of independent data sets, thereby augmenting real peaks because the large values of a peak in one set were multiplied by the large values of a peak in the next data set. Spurious peaks were reduced since the large values of a spurious peak were multiplied by the very small values of suppressed noise in the next set. The results of such cross-multiplication of five independent synthetic data sets (SNR = 0.5) are shown in Fig. 3.

Savitzky–Golay Filtering. The Savitzky–Golay filter is a popular windowed time-domain filter used to render visible spectral lines in noisy spectrometric data.[3,8,9] The same five independent synthetic data sets used above were ensemble-averaged and then processed with a 0-order Savitzky–Golay filter. The result is also shown in Fig. 3 for

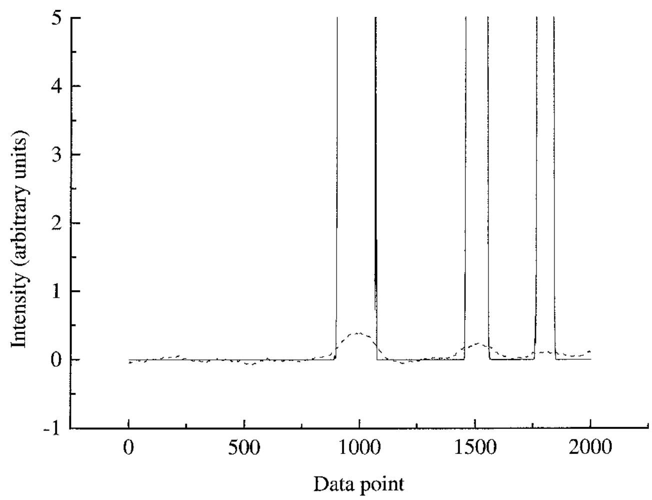

FIG. 3. The signal amplification and noise suppression for five independent synthetic data sets with SNR = 0.5 treated with a moving product (150-point window) and then cross-multiplied point by point. The vertical scale is expanded to show the suppression of noise. The maximum value of the peak centered at 1000 points was 1×10^{22}. The same five data sets ensemble averaged and then processed with a 0-order Savitzky–Golay filter (150-point window, dashed trace) are shown for comparison.

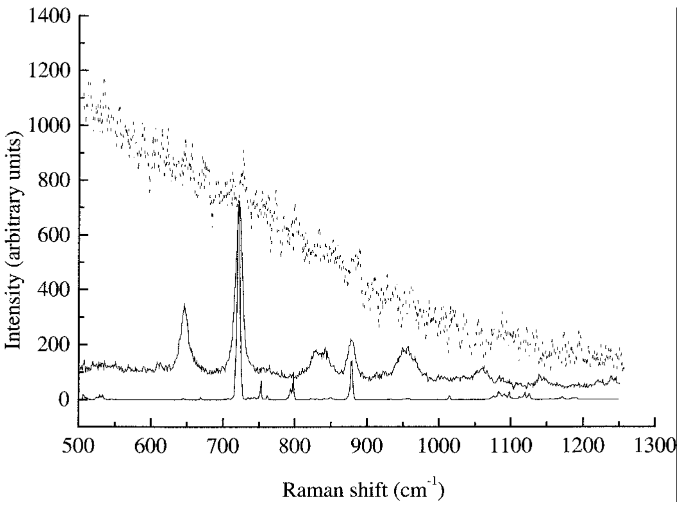

FIG. 4. The signal amplification and noise suppression for three independent Raman spectra of acetylcholine (1×10^{-3}M) treated with a moving product (12-point window) and then cross-multiplied point by point (bottom trace). The same three spectra ensemble averaged (dotted trace, top) and a higher SNR spectrum of acetylcholine (middle trace).

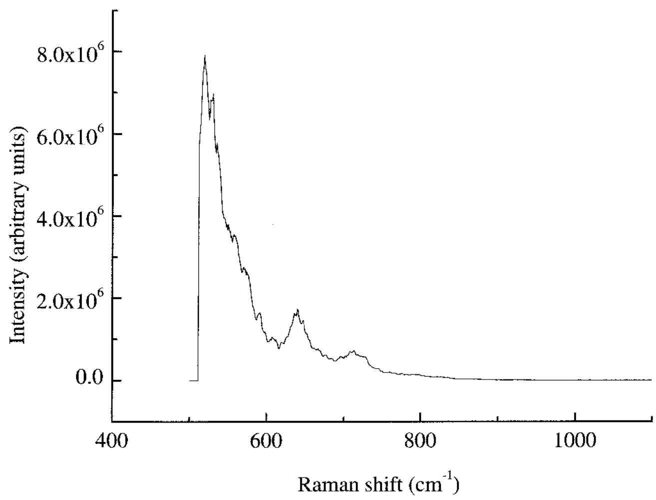

FIG. 5. A Raman spectrum (acetylcholine, 1×10^{-3}M) without baseline correction, treated with a moving product (12-point window).

comparison. It can be seen from Fig. 3 that the MP method produced narrower and much pronounced peaks while suppressing background noise to a far greater extent than the Savitzky–Golay filter.

Real Data. Although good results were obtained when applying the MP to synthetic data sets, real-world data often provide complexities rendering processing more difficult. The MP method, including cross-multiplication, was applied to three Raman spectra of acetylcholine (1×10^{-3} M each). Figure 4 shows the results from the MP method and, for comparison, the ensemble-averaged spectrum of the three individual Raman spectra as well as a high SNR acetylcholine spectrum.

Figure 4 shows that the Raman C–N symmetric stretch vibrational mode at 722 cm^{-1} can clearly be identified in the spectrum processed with the MP and cross-multiplication method. Importantly, however, the O–C=O in-plane deformation near 645 cm^{-1} is not detectable. One possible reason for this was that the third-order polynomial used to remove the background was not capable of producing a flat background without distorting the original spectrum. It is rather to be expected that baseline correction may introduce artifacts into a spectrum. Consequently, the MP method was applied to one of the three low SNR Raman spectra without baseline correction. Although the presence of the uncorrected baseline in the original resulted in an even more pronounced baseline after manipulation, it is evident from Fig. 5 that both acetylcholine peaks at 645 and 722 cm^{-1} could be clearly identified, the former being more pronounced. This result demonstrates the need for accurate baseline correction in order to obtain the maximum benefit from the MP method.

DISCUSSION

We have demonstrated a novel method to selectively amplify signals and suppress background noise on synthetic and real data with unity and subunity SNRs. The method consists of the application of a moving product to data first transformed to a mean of 0, normalized with regard to absolute value, and then transformed to a mean of 1. Broad signals (with an autocorrelation lag much larger than that of the noise in which they were embedded) could be amplified by several orders of magnitude with the use of a moving window of approximately the FWHM of the signal peak while simultaneously suppressing the background noise to near zero. This enhancement of the signals was obtained at the cost of distortion of the line shapes, relative peak heights, and displacement of peak maxima. The cross-multiplication of a very few additional MP-processed spectra was shown to further eliminate noise while augmenting signal(s). The method was tested with synthetic data with the use of Gaussian peaks (it has also been found to work with Lorentzian peaks) and measured Raman spectra.

The degree of amplification and its effects on the signal were dependent on the size of the MP window. For example, an inspection of Fig. 2 reveals that for window sizes 50 and 100, the broader peak at 1000 was most strongly amplified. For window sizes from 10 to 25, the narrower peak positioned at 1500 was more strongly amplified. It can also be seen from this figure that, for window sizes from 5 to 20, the still narrower peak situated at 1800 is evident while the narrowest peak at 1950 is perhaps visible only when using a window five points in size. Therefore, care should be exercised when selecting the size of the window to be employed. The degree of amplification as a function of window size for the synthetic peak situated at 1500 (FWHM approximately 60) is shown in Fig. 6. Figure 6 shows that an optimum window size exists; moreover, the optimum window size is somewhat dependent on the SNR of the spectrum. This observation probably reflects the fact that more of the signal wings have an effect on the moving product as the SNR increases.

Another issue that should receive careful consideration is the treatment of the baseline. As shown in Figs. 4 and 5, baseline removal may result in the concomitant removal of important data or the introduction of artifacts. This problem is especially true where signals are weak and completely buried in the background noise. The selection of the baseline correction method would therefore have an

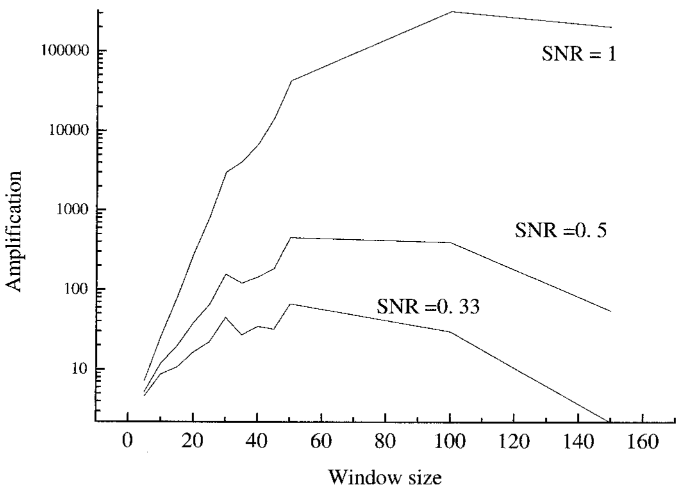

FIG. 6. The amplification of the synthetic signal situated at 1500 (Fig. 1) as a function of the moving product window size.

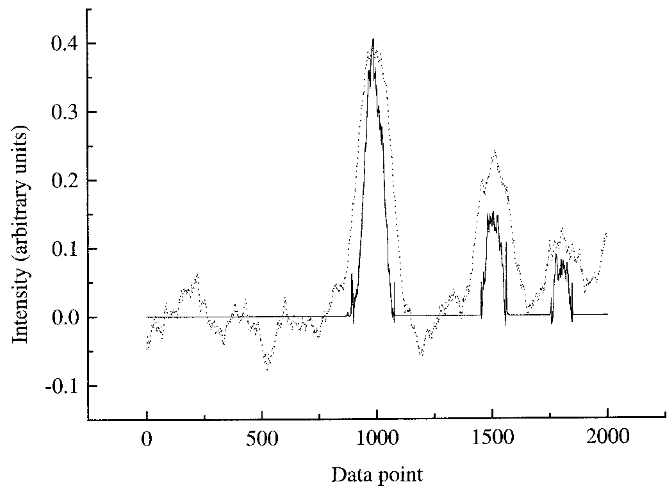

FIG. 7. The signal amplification and noise suppression for five independent synthetic data sets with SNR = 0.5 treated with a moving product (150-point window) and then cross-multiplied point by point. The 150th root of all points >0.1 was then taken and, finally, 1 was subtracted from all data values. The same five data sets ensemble averaged and then processed with a 0-order Savitzky–Golay filter (150-point window, dotted trace) are shown for comparison.

important impact on the success of the application of the MP method to real-world data. Furthermore, an understanding of systematic instrumental error, which contributes to the baseline, will facilitate its removal from data sets and is therefore recommended.

Given the similarity between a moving average smoothing method and the MP method, the computational complexity of the latter is comparable to that of the former. It appears therefore that the MP method introduces no additional computational burden.

By introducing a threshold value and taking the window sized root of the values larger than the threshold, one can reduce the distortion in peak shapes, peak maxima, and relative peak heights while retaining the degree of noise suppression achieved initially. The result of such a procedure, applied to the data shown in Fig. 3, is shown in Fig. 7. After the data were processed with the MP, every data point larger than 0.1 was replaced by its nth root (n = window size) followed by the subtraction of 1. The same data processed with a 0-order Savitzky–Golay filter are shown for comparison. The comparison shows that the MP method produces greater noise suppression and less peak broadening than the Savitzky–Golay filter. Ensemble averaging increases the SNR by the square root of the number of added spectra.[9] Therefore, averaging five independent spectra, each with an SNR of 0.5, would improve the SNR to 1.12. The SNR improvement that is obtained with Savitzky–Golay filtering is approximately equal to the number of points used in the smooth,[9] implying that a 150-point smooth would provide a further enhancement of 12.24, giving a total improvement from 0.5 to 13.72. In contrast, with application of an MP, cross-multiplication, and a threshold to reject noise, as in the example shown in Fig. 6, the SNR is enhanced from 0.5 to 10 125 with the use of the first 700 points of the processed spectrum to determine the noise standard deviation.

The careful choice of window size and the application of a threshold as demonstrated above may mitigate greatly against line shape and relative peak height distortions. The choice of threshold value and window size would also influence the artifacts on the peak sides, as evident from Fig. 7. Specifically, setting the threshold higher would reduce the negative spectral artifacts, and a value of 1 is suggested. Furthermore, the recovery of an acceptable representation of the underlying signal may be possible with the judicious application of MP windows of varying size and an effort to combine them in a predetermined weighted manner.

Although signal recovery may be possible as outlined above, the considerable degree of background noise suppression and signal amplification that can be obtained with the MP method is likely to make it primarily useful for signal detection applications. Signal detection is achieved primarily by suppressing the background noise as opposed to improving the SNR. Where very low SNRs are involved, the discrimination of a signal from its background is difficult, and the MP method could be employed as demonstrated in this report. In particular, the method may be useful to extend the ultimate detection limit of a given instrument or experimental setup.

ACKNOWLEDGMENTS

Financial support for this work was provided by the Natural Sciences and Engineering Research Council of Canada and the British Columbia Health Research Foundation.

1. V. J. Barclay, R. F. Bonner, and I. P. Hamilton, Anal. Chem. **69,** 78 (1997).
2. A. Savitzky and M. J. E. Golay, Anal. Chem. **36,** 1627 (1964).
3. W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C* (Cambridge University Press, Cambridge, 1994) 2nd. ed., p. 537.
4. L. S. Greek, H. G. Schulze, M. W. Blades, A. V. Bree, B. B. Gorzalka, and R. F. B. Turner, Appl. Spectrosc. **49,** 425 (1996).
5. A. P. Shreve, N. J. Cherepy, and R. A. Mathies, Appl. Spectrosc. **46,** 7070 (1992).
6. H. G. Schulze, M. W. Blades, A. V. Bree, B. B. Gorzalka, L. S. Greek, and R. F. B. Turner, Appl. Spectrosc. **48,** 50 (1994).
7. E. Kreysig, *Advanced Engineering Mathematics* (Wiley International, New York, 1972), 3rd ed., p. 699.
8. M. U. A. Bromba and H. Ziegler, Anal. Chem. **53,** 1583 (1981).
9. N. Demas and S. E. Demas, *Interfacing and Scientific Computing on Personal Computers* (Allyn and Bacon, Boston, 1990), p. 188.