Investigation of Selected Baseline Removal Techniques as Candidates for Automated Implementation

Article *in* Applied Spectroscopy · June 2005
DOI: 10.1366/0003702053945985 · Source: PubMed

6 authors, including:

Andrew Jirasek
University of British Columbia - Okanagan
97 PUBLICATIONS **1,723** CITATIONS

SEE PROFILE

Investigation of Selected Baseline Removal Techniques as Candidates for Automated Implementation

GEORG SCHULZE, ANDREW JIRASEK,* MARCIA M. L. YU, ARNEL LIM, ROBIN F. B. TURNER,† and MICHAEL W. BLADES†

Michael Smith Laboratories, The University of British Columbia, 301-2185 East Mall, Vancouver, BC, Canada, V6T 1Z4 (G.S., A.J., R.F.B.T.); Department of Chemistry, The University of British Columbia, E257-2036 Main Mall, Vancouver, BC, Canada, V6T 1Z1 (A.J., M.M.L.Y., M.W.B.); Department of Physics and Astronomy, The University of British Columbia, 325-6224 Agricultural Road, Vancouver, BC, Canada, V6T 1Z1 (A.L.); and Department of Electrical and Computer Engineering, The University of British Columbia, 434-2356 Main Mall, Vancouver, BC, Canada, V6T 1Z4 (R.F.B.T.)

Observed spectra normally contain spurious features along with those of interest and it is common practice to employ one of several available algorithms to remove the unwanted components. Low frequency spurious components are often referred to as 'baseline', 'background', and/or 'background noise'. Here we examine a cross-section of non-instrumental methods designed to remove background features from spectra; the particular methods considered here represent approaches with different theoretical underpinnings. We compare and evaluate their relative performance based on synthetic data sets designed to exemplify vibrational spectroscopic signals in realistic contexts and thereby assess their suitability for computer automation. Each method is presented in a modular format with a concise review of the underlying theory, along with a comparison and discussion of their strengths, weaknesses, and amenability to automation, in order to facilitate the selection of methods best suited to particular applications.

Index Headings: **Automated baseline determination; Baseline correction; Baseline removal; Noise median; First derivative; Peak picking; Peak stripping; Artificial neural networks; Local regression; Spectral shifting; Maximum entropy; Fourier transforms; Wavelet transforms.**

INTRODUCTION

Acquired spectra often contain undesirable elements such as noise and background features in addition to the desired signal itself. Hence, it may be desirable, or even necessary, to separate the signal from both background and noise. The specific reasons for baseline removal are manifold and application dependent. These may range from relatively simple requirements such as presentation, to more exacting requirements such as quantification or preparation for further numerical processing, and the stringency required of any baseline-correction method (BCM) may vary accordingly. Moreover, the accuracy obtainable, modes of failure, and computation time required vary between, and sometimes within, different BCMs and an understanding of these trade-offs may facilitate the selection of a BCM suitable to the problem at hand.

Ruckstuhl et al.[1] observed that the available literature on BCMs is widely dispersed amongst many fields of research and comment that useful review articles are relatively scarce. This highlights at once the widespread interest in, and utility of, BCMs and underscores the difficulties in undertaking a systematic study of the available methods. Not only do they have to be retrieved from the literature of different specialty areas, but they may also have to be reformulated to make valid comparisons that reveal their relative merits. This article attempts to address these issues and to provide guidance to aid selection of the most appropriate method for a given application.

Early reports of baseline correction that appear in the literature were predominantly based on hardware implementations or modifications.[2–5] In 1969, some of the first computer-based methods designed for the automated removal of baselines were published.[6,7] Activity in the development and use of BCMs accelerated through the 1970s and 1980s, partly facilitated by the increasing availability of distributed computing equipment.[8–48] The first comparisons, reviews, and evaluations of BCMs also appear during this period.[49–53] Since then, a few additional review and/or evaluation articles have appeared,[54,55] as well as a great number of new methodological developments. Some of these methods were specifically designed for automation, while others clearly offer such potential. It is likely that increases in small computer performance, improved instrumentation yielding vast amounts of data to be processed, and the increasing costs of manual methods of baseline removal will provide continued driving forces for automation. Therefore, along with an evaluation of methods under a variety of conditions, we also specifically address issues relating to automation.

Measured spectra can be considered to consist of the

Received 30 September 2004; accepted 19 January 2005.

* Present address: Department of Physics and Astronomy, University of Victoria, Victoria, BC, Canada, V8W 3P6.

† Authors to whom correspondence should be sent. E-mail: blades@chem.ubc.ca; turner@msl.ubc.ca.

0003-7028/05/5905-0545$2.00/0

signals of interest plus a background, both convolved with an instrumental blurring function, and with noise added to the result. Intuitively, the background is taken to comprise low frequency spectral components, while noise is considered to have high frequency components. The signals of interest usually have frequency components between those of background and noise. We will adopt this somewhat simplified view to recast a number of different BCMs to facilitate comparisons between them. We acknowledge that, in reality, the situation is more complex since different sources of noise, with widely different frequency components, are typically present.[56] On occasion, one may even have an interest in studying the noise itself or in the background itself. However, these additional dimensions would unnecessarily complicate the comparative study presented here and will be neglected.

Implicitly, the real analytical problem is essentially a signal recovery problem, and baseline correction is an intermediate step towards this goal. Nevertheless, a variety of approaches have been designed specifically for the baseline removal step. These methods differ, in general, by the extent or type of explicit knowledge or assumptions about the spectral components such as background, noise, blurring function, etc., required, and/or in the order with which they are dealt. Clearly, to include every baseline removal method developed to date would be intractable (although we offer an extensive bibliographic reference list as an aid to interested readers). In this communication, we focus on techniques that are designed to be applied to spectral data *a posteriori* (i.e., after spectra have been acquired using whatever instrumentation is appropriate). Additional techniques rely, at least in part, on the instrumentation itself[2–5,57–64] and we will not discuss them further here. Of the non-instrumental methods, we have chosen BCMs that represent techniques that (a) require no *a priori* information, or (b) do require some *a priori* information (e.g., regarding the noise, baseline, blurring function, signal, etc.). These represent a cross-section of methods both conceptually and in terms of the underlying assumptions they make about and the treatment of the data.

We begin by introducing the problem of baseline correction in general theoretical terms, followed by the materials and methods used in the evaluation of the BCMs. Following this introduction, each BCM to be evaluated is discussed separately in some detail. Each BCM discussion 'module' includes some of its history, theory of operation, implementation, evaluation on a wide range of synthetic spectra, advantages and disadvantages, modes of failure encountered, and suitability for automation. A general summary of results is then presented to permit researchers to select the most suitable BCM, given their objectives and specific constraints, from those techniques presented explicitly here (or, by extrapolation, from similar methods). This 'modular' structure was chosen in the hope that, by having each method 'self-contained', this article can also serve as a useful reference for non-experts in these methodologies.

THEORY

The Baseline-Estimation Problem. A given vector $\mathbf{y} = \{y_1, y_2, \ldots, y_i\}$ of i observed frequency domain spectral intensities can be modeled as the sum of an ideal spectrum **s** and a background **b**, convolved with a blurring function **p**, with noise **n** added to the result:

$$\mathbf{y} = (\mathbf{s} + \mathbf{b}) * \mathbf{p} + \mathbf{n} \quad (1)$$

with * denoting convolution. The noise is often taken to be Gaussian,[1] but is perhaps more correctly described as Poissonian.[65] Either way, the two distributions become highly similar under some conditions, and one could consider their means as contributing to the baseline **b** while their variances constitute the noise **n** of Eq. 1. The issue of noise and its characterization is also addressed elsewhere.[66]

The problem for the data analyst is now to recover **s** from Eq. 1, hence to determine

$$\mathbf{s} = (\mathbf{y} - \mathbf{n}) * \mathbf{p}^{-1} - \mathbf{b} \quad (2)$$

with $\mathbf{p}^{-1}$ being the inverse of the blurring function. It is clear that a solution to Eq. 2 requires knowledge of $\mathbf{p}^{-1}$, **b**, and **n**. Such knowledge is often incomplete or totally lacking, and the approaches taken to the recovery of **s** reflect then, in general, various ways of dealing with these deficits. We shall, therefore, base our discussion and evaluation of BCMs on a classification of their treatment, whether implicit or explicit, of the incomplete information. The main purpose of this classification is to impose some order on the collection of methods to facilitate discussion. It has as added benefit an indication of the relative extent of information required for baseline correction, but does not reflect on the computational complexity of the various methods.

The missing or incomplete information can often be determined either explicitly, such as the characterization of the blurring function **p**, implicitly by way of assumptions and/or estimates, or ignored altogether. For example, noise is often assumed to be independent and identically distributed with a zero-mean normal distribution.[1,67] The standard deviation σ_n of the noise distribution is then estimated from a signal-free subset of the observed data **y**. Likewise, the background **b** is assumed to consist of broad, slowly varying, correlated features,[1,61,64,68] and an estimate of its intensity can be made from background-containing but signal-free subsets of **y**.

MATERIALS AND METHODS

Numerical Implementation of Baseline Correction Methods. The BCMs investigated here were implemented in HiQ script (National Instruments Corporation, Austin, TX) running under Windows 98 (Microsoft, Redmond, WA) on a 1 GHz AMD Athlon platform (AMD, Sunnyvale, CA). Manual baseline removals were done using Grams software (Galactic Software, Salem, NH).

Synthetic Spectra. Standardized spectra consisting of 1001 channels were created to simulate, primarily, spectra obtained from the vibrational spectroscopies (Raman, infrared (IR), etc). However, results presented here may be applied to other data types with relatively narrow peaks (e.g., nuclear magnetic resonance (NMR)) as well, but potentially not to those with broad peaks (e.g., near-infrared (NIR), ultraviolet (UV)). A cumulative normal function (sigmoid) was used as baseline. The slope of the linear part of the baseline was set to one of four values:

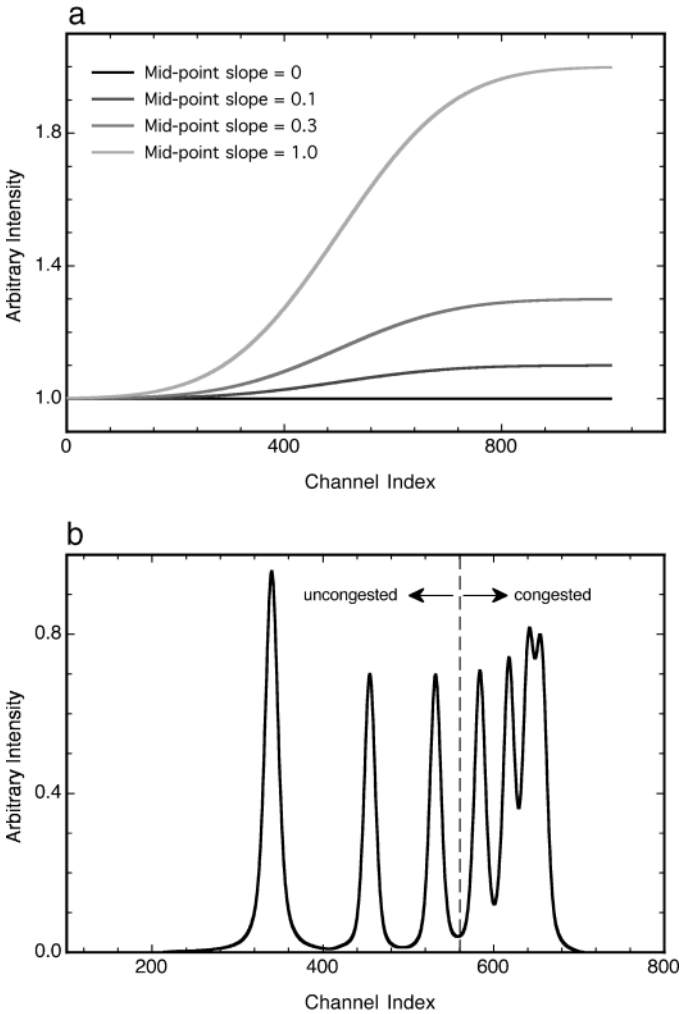


FIG. 1. Standardized data. (**a**) Baselines of midpoint slope = 0 (black), 0.1 (dark gray), 0.3 (gray), and 1.0 (light gray). (**b**) Synthetic spectrum consisting of seven Lorentzian peaks. The spectrum has been divided into uncongested and congested regions.

0.0, 0.1, 0.3, or 1.0. In the latter case, the baseline amplitude (baseline maximum − baseline minimum) reached a maximum value of 1.0.

Seven Lorentzian peaks were added to every baseline with peak centers at 340, 455, 532, 584, 618, 641, and 656 channels, respectively. Hence, one part of the spectrum was congested but the other part was not. Lorentzians had a full-width at half maximum (FWHM) of 5.7 channels, except for the first peak (at position 340), which had a width of 10 channels. They were then convolved with a 5-channel standard deviation normal distribution.

All Lorentzian signals had the same height within a spectrum, but the heights were varied between spectra through 0.0, 0.05, 0.1, 1.0, and 10.0, respectively, hence effectively varying the signal-to-baseline ratio (SBR), i.e., signal strength (signal maximum − signal minimum) relative to baseline amplitude (baseline maximum − baseline minimum). Therefore, for every baseline with a given slope, five spectra with Lorentzians of different height were generated.

Noise was added to each of the spectra above to give signal-to-noise ratios (SNR) of 2, 3, 5, 10, and 100. Since there were four baseline slopes, each with five intensities of peaks added, and each of these with five different levels of noise added, a total of 100 synthetic spectra were generated. Both SBR and SNR, when used in the context of a spectrum, are based on the most intense signal in the spectrum; when no signal was present, no noise was added. The different synthetic baselines used are shown in Fig. 1a, and a sample convolved synthetic signal, before superimposing on these baselines, is shown in Fig. 1b.

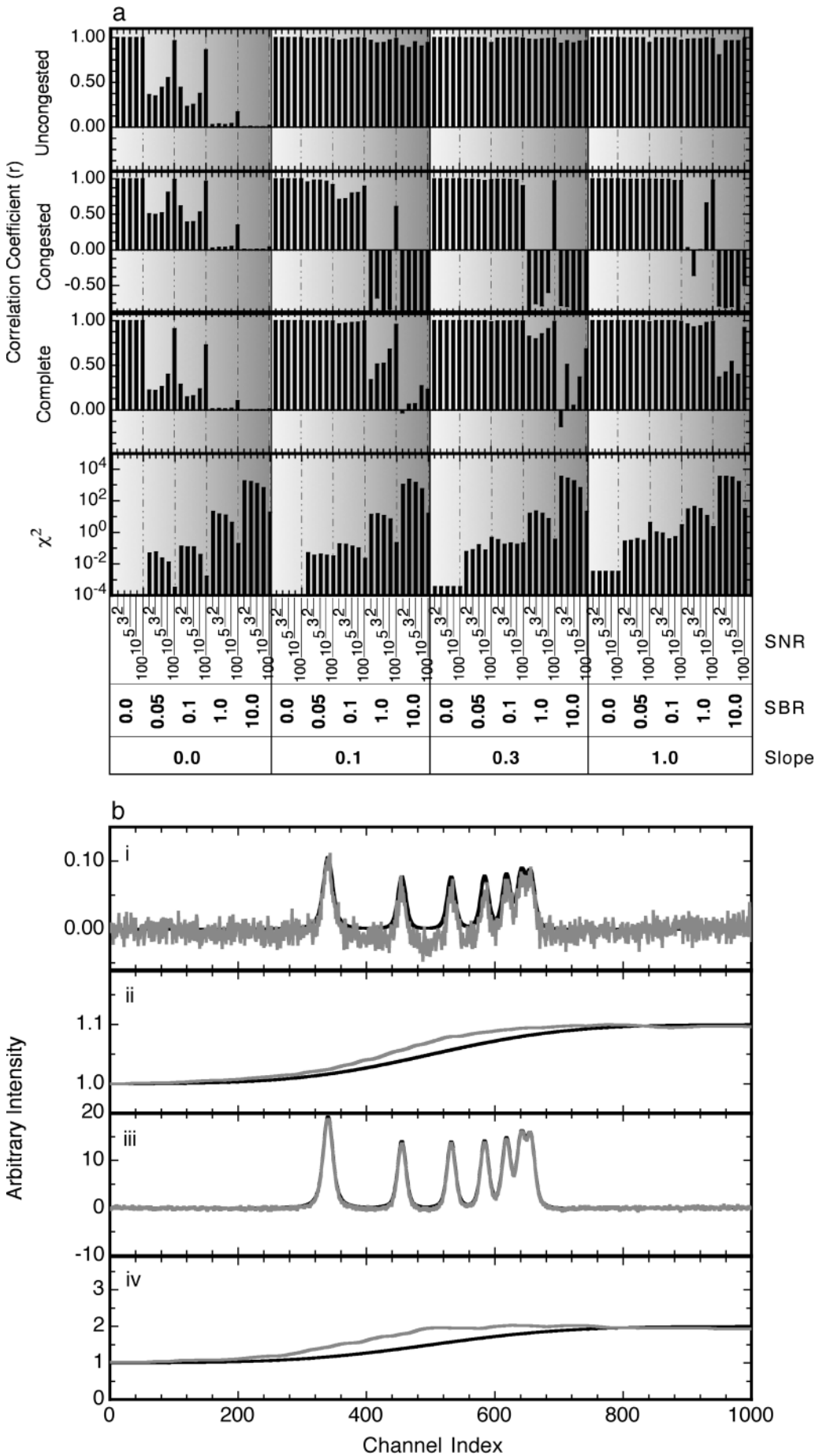
Figures-of-Merit (FOM). Due to the number of methods tested (11) and the number of spectra subjected to each method (100), reporting multiple FOM for each trial, as well as the representation of the results of baseline estimation, became problematic. Hence, ''bulk'' FOM were chosen that best represent the trends in the baseline estimation capabilities of each method. To this end, the correlation coefficient was calculated between estimated and given baselines for both the congested and uncongested regions of each spectrum, as well as for the total spectrum. Furthermore, the χ^2 value between the baseline estimate and the actual baseline was calculated for the entire spectrum. This gives an indication of offsets between estimated and actual baselines. Furthermore, for each method representative examples of baseline estimates as well as baseline-removed spectra are presented. This allows for a qualitative visual inspection of the types of distortions imposed by the given method on the baseline-removed spectra.

We have chosen to use both the correlation coefficient and χ^2 as FOM for the following reason. Although the

TABLE I. Classification of baseline removal methods based on the type of required information.[a]

Class	Method
Methods requiring no explicit knowledge of **p**, **b**, or **n**	Noise median method (NMM)
	First derivative method (FDM)
Methods requiring estimates of **b**	Artificial neural networks (ANN)
	Threshold-based classification (TBC)
	Signal removal methods (SRM)
	Composite (linear-sine-cosine) baseline method (CBM)
	Spectral shift methods (SSM)
Methods requiring estimates of **b** and **n**	Manual methods (MM)
Methods requiring use of **p**, **b**, and **n**	Maximum entropy method (MEM)
Methods requiring information about frequency	Fourier transform method (FTM)
	Wavelet transform method (WTM)

[a] The methods are referenced in the order of their appearance in the section titled Baseline Correction Methods and Their Results.


a
Correlation Coefficient (r)
Uncongested
Congested
Complete
χ^2
SNR
SBR
Slope
0.0
0.1
0.3
1.0
b
Arbitrary Intensity
Channel Index
i
ii
iii
iv

chi squared (χ^2) statistic is often used as a FOM, it represents a bulk value and hence the nature of the error cannot be inferred. A result with a large χ^2 value may nevertheless have a very high correlation coefficient, indicating that the estimated and actual baselines are separated by a constant offset. In contrast, a smaller χ^2 value could be associated with a small correlation coefficient indicative of pronounced distortions in the estimated baseline despite average values similar to that of the actual baseline.

For every method discussed a figure containing bar graphs for these FOM (correlation coefficients and χ^2 values) is given. Each bar graph reflects differences between theoretical and estimated baselines on individual spectra with spectral attributes as shown on the abscissa. The bar graph panels, from top to bottom, show the correlation coefficients for uncongested, congested, and total ('complete') spectral regions, and the χ^2 values for the total spectral region, respectively. The abscissa shows the attributes for each of the 100 individual spectra with SNR nested within SBR nested within slope. For example, the rightmost bar in the top panel of Fig. 2a shows the correlation coefficient for the uncongested (see Fig. 1) spectral region for a spectrum with SNR = 100, SBR = 10, and slope = 1.0.

To gain insight into the overall ability of each method to estimate the baseline, regardless of SBR or SNR, a mean correlation coefficient and χ^2 over all SNR and SBR (i.e., 100 spectra), as well as the standard deviation of this mean, was calculated for each method over the entire spectrum. Finally, the mean time and the standard deviation of this mean, required by each method to perform the baseline estimation for a given spectrum, were calculated.

BASELINE CORRECTION METHODS AND THEIR RESULTS

We have selected a number of methods belonging to each of the categories in Table I. As mentioned before, these categories reflect the amount and type of prior, assumed, or additional knowledge available. Since it is not practical to include all methods, some have been omitted. Omission does not reflect on the value of these methods and their particular approaches or features. Methods not described here include those based on principal component analysis or singular value decomposition[69–72] and Bayesian methods.[65,73–77] Also not included were methods primarily aimed at mixture resolution and calibration.[12,14,26,28,32,35,37,43,44,47,67,75,77–82] The latter methods often require information about the signal **s** and involve the resolution of background and signal components through modeling.

METHODS REQUIRING NO EXPLICIT KNOWLEDGE OF p, b, OR n.

Noise Median Method (NMM). ***Background.*** Introduced by Friedrichs[83] in 1995 to remove baseline distortion in nuclear magnetic resonance (NMR) spectra, this method estimates the baseline as the median value in a moving window. If the median is based on extrema in the window, and if the window is suitably large, the median is not unduly affected by signal peaks. This method could be said to have antecedents in minimum search and interpolation methods based on the identification of local and global minima in different successive segments of a spectrum and in methods based on the means in different successive segments of a spectrum.[6,13,19,20,22,23,25,29,36] Lewis and Chatwin[84] also used segment minima to estimate the position of the baseline. In later work a median filter has been used by Keselbrener et al.[85]

Theory. The data are modeled here implicitly in the form

$$\mathbf{y} = \mathbf{s}*\mathbf{p} + \mathbf{b}*\mathbf{p} + \mathbf{n} \qquad (3)$$

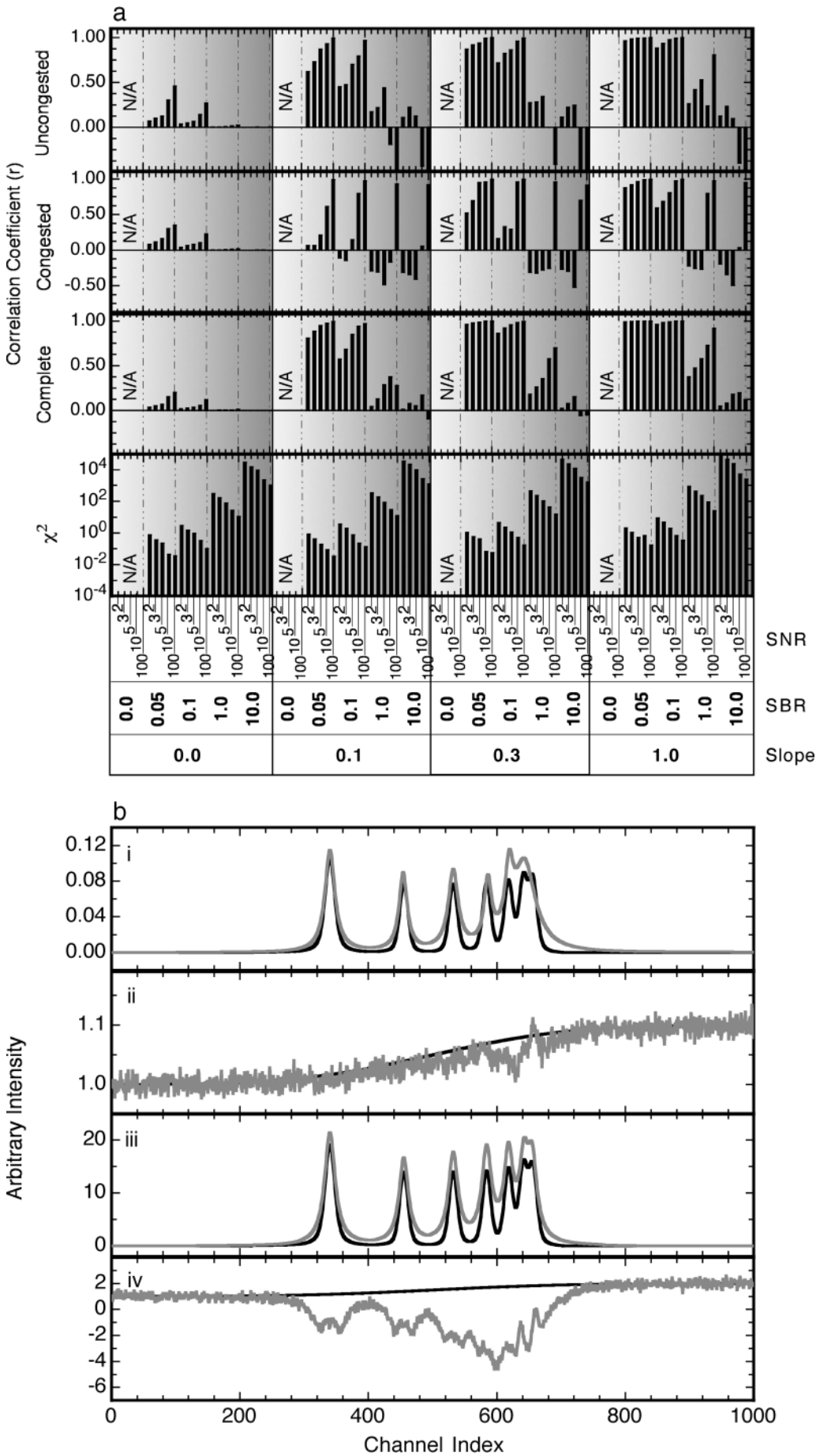
with the objective to estimate $\mathbf{b}*\mathbf{p} \approx \mathbf{b}$. Since the baseline itself is not modeled (e.g., with some specific function), but locally estimated, it is considered a 'model-free' method.[83]

In this method, the baseline is estimated as the median value of all those points within a window of given size. When the points are rank ordered, the median value is the middle value of the set, implying that there are as many points greater than the median as there are smaller than the median. It is generally unimportant by *how much* the values are greater (or smaller), as long as an equal number are greater and smaller than the median, respectively. For a suitably large window, the presence of a peak simply means that some of the numbers above the median are much greater (in value or intensity) than they would be in the presence of only noise. Under these conditions, there would be neither more nor fewer points above the median. The median remains unaffected by the presence of a signal if the signal is relatively narrow. For wider signals, the effects of the signal can be attenuated by considering the median of all the extrema, rather than the median of all the points, in the window. Hence, in the presence of a signal, the baseline is defined by the requirement of equal representation of local extrema above and below the baseline. If the window is too narrow relative to the signal, this interpretation no longer holds, leading to errors in baseline determination, and a wider window has to be used (see also the Modes of Failure subsection below).

Within a window of suitable size, which defines a local neighborhood, the baseline is approximated by the median value of the local extrema. A given element of **y** is an extremum if both of its neighbors are either strictly

←

FIG. 2. Baseline removal using the noise median method. (**a**) Correlation coefficient between theoretical and estimated baselines for the uncongested, the congested, and the total spectral regions of given spectra. Also shown is the χ^2 between theoretical and estimated baselines for the total spectral region. (**b**) (***i***) Baseline removed (gray trace) and initial input spectrum (less noise and true baseline, black trace) for SNR = 10, SBR = 0.1, and baseline slope = 0.1; (***ii***) estimated baseline (gray trace) and input baseline (black trace), SNR, SBR, and baseline slope as for (i); (***iii***, ***iv***) as for (i) and (ii) with SNR = 100, SBR = 10, and baseline slope = 1.0.


a
Correlation Coefficient (r)
Uncongested
Congested
Complete
χ²
N/A
SNR
SBR
Slope
0.0
0.05
0.1
1.0
10.0
0.0
0.1
0.3
1.0
b
Arbitrary Intensity
i
ii
iii
iv
Channel Index

larger or strictly smaller than itself. The estimated baseline obtained by moving the window along the spectrum can then be smoothed by convolving it with a Gaussian function to remove sharp discontinuities where these exist.

Implementation. We used a fixed window size of 400 points (approximately 3 times that of the congested spectral region peak). A standard deviation of 5 was used for the Gaussian smoothing function applied to the estimated baseline to reduce sharp discontinuities in case any were present.[83]

Results and Discussion. The quantitative results are shown in Fig. 2a and comparative examples of baseline correction on two spectra (SNR 10, SBR 0.1, slope 0.1; SNR 100, SBR 10, slope 1.0) are shown in Fig. 2b. The results, consistent with the literature (see below), indicated that the method tends to have difficulties with congested spectral regions. Further difficulties are observed in high SBR spectra (i.e., spectra with pronounced signal). Nevertheless, it produced relatively good results and was also comparatively fast as shown in Fig. 13.

This method requires that a suitable window size be determined. A window size of at least twice the FWHM is recommended for uncongested spectral regions and a larger window is recommended for congested and low SNR regions.

Modes of Failure. This method depends on the presence of noise and will fail if very little or no noise is present. Furthermore, if the window is too small relative to a peak or a congested region and is moved along until it reaches this peak or congested region, most of the extrema in the window would derive from the peak or congested peaks and the baseline estimate will become biased. If too wide a window is used, changes in the baseline cannot be tracked accurately. Ideally, one should use as narrow a window as is consistent with adequate processing of broad peaks and/or congested spectral regions. Friedrichs describes a method of implementing an elastic window that is automatically adjusted to incorporate a suitable number of extrema.[83] Also, although we did not encounter such problems, it is possible that the choice and width of the smoothing function could fail to remove sharp discontinuities in the estimated baselines.

Suitability of Method for Automation. Given that few parameters have to be determined and that generally good results can be obtained for a range of SNR, SBR, and slope, this method is suitable for automation purposes. Incorporating the suggestions for an elastic window[83] may further enhance its suitability for automation.

First-Derivative Method (FDM). ***Background.*** The first derivative represents the slopes of a spectrum. Because the slopes of a background are generally lower than those of signal peaks, taking the first derivative of a spectrum will discriminate against slowly varying background components irrespective of their absolute values. The use of derivatives for the identification of peaks and the rejection of baselines has been recognized in early efforts.[4,5,27,39,49,86] Although we follow here the implementation of a derivative method by Mosier-Boss et al.[61] proposed for the rejection of fluorescence in Raman spectroscopy, others have also employed the use of derivatives for baseline correction.[67,87–92]

Theory. After smoothing, a measured spectrum consisting of m Lorentzian Raman peaks on a broad background can be approximated by

$$\mathbf{y} = \mathbf{s}*\mathbf{p} + \mathbf{b}*\mathbf{p} \tag{4}$$

where differentiation of Eq. 4 yields

$$\mathbf{y}' = (\mathbf{s}*\mathbf{p})' + (\mathbf{b}*\mathbf{p})' \tag{5}$$

with $\mathbf{y}' = d\mathbf{y}/dx$ (x = spectral variable). Often, depending on the blurring function, Eq. 5 can be approximated as

$$\mathbf{y}' = \mathbf{s}' + \mathbf{b}' \tag{6}$$

Since $\mathbf{b}' \approx \{0, 0, \ldots, 0\}$, we have $\mathbf{s}' \approx \mathbf{y}'$ and $\mathbf{s}$ is numerically reconstructed from $\mathbf{s}'$ using

$$s_n = \sum_{i=1}^{m} \left\{ \frac{a_i}{1 + \left[\frac{(x - x_0)^2}{w_i} \right]} \right\} \quad \text{and} \tag{7}$$

$$s'_n = \sum_{i=1}^{m} \left(\frac{2}{w_i} \right) \left(\frac{a_i(x_0 - x)}{\left\{ 1 + \left[\frac{(x - x_0)^2}{w_i} \right] \right\}^2} \right) \tag{8}$$

where a_i is the maximum amplitude of peak i, x_0 is the center position of peak i, and w_i is the width (variance) of peak i. The values of the parameters in Eq. 8 are determined with a suitable nonlinear optimization procedure and substituted in Eq. 7 to calculate the signal $\mathbf{s}$.

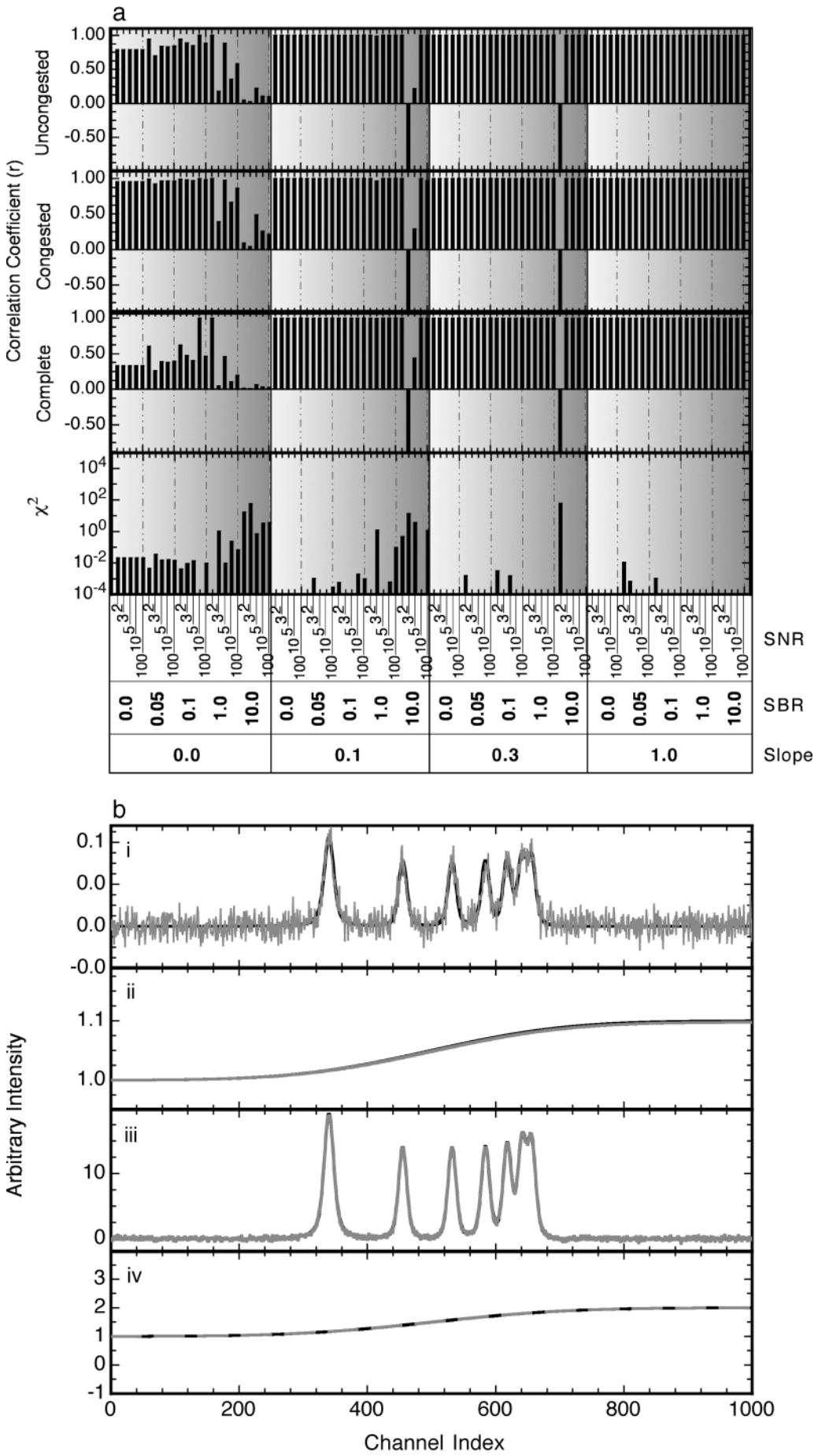
Implementation. We smoothed the spectra with an 11-point second-order Savitzky–Golay filter[93] and used Levenberg–Marquardt[94] optimization to obtain the parameters for the peak derivatives. Peak positions for fitting were taken as the positions of zero crossings on the first derivative of any given spectrum in the neighborhood of derivative peaks exceeding a threshold value of five times the standard deviation of the derivative spectrum.

Results and Discussion. The quantitative results are shown in Fig. 3a and comparative examples of baseline correction on two spectra (SNR 10, SBR 0.1, slope 0.1; SNR 100, SBR 10, slope 1.0) are shown in Fig. 3b. The results (see also below) illustrate the difficulties encountered by the FDM with spectra where no peaks were present, spectra with high noise levels, and spectra with congested regions. Computation times were comparatively long as shown in Fig. 13.

The benefit of the fitting procedure is that it removes noise while estimating the convolved signal. It is there-

←

FIG. 3. Baseline removal using the first-derivative method. (**a**) Correlation coefficient between theoretical and estimated baselines for the uncongested, the congested, and the total spectral regions of given spectra. Also shown is the χ^2 between theoretical and estimated baselines for the total spectral region. (**b**) (***i***) Baseline removed (gray trace) and initial input spectrum (less noise and true baseline, black trace) for SNR = 10, SBR = 0.1, and baseline slope = 0.1; (***ii***) estimated baseline (gray trace) and input baseline (black trace), SNR, SBR, and baseline slope as for (i); (***iii***, ***iv***) as for (i) and (ii) with SNR = 100, SBR = 10, and baseline slope = 1.0. N/A = not applicable; this method relies on peak picking, and spectra without peaks (SBR = 0) could not be processed.


a
Correlation Coefficient (r)
Uncongested
Congested
Complete
χ²
SNR
SBR
Slope
0.0
0.1
0.3
1.0
0.05
10.0
b
Arbitrary Intensity
i
ii
iii
iv
Channel Index

fore important to keep this point in mind when examining the results reported for this method since the estimated baseline is derived by subtracting the estimated convolved signal from the synthetic spectrum. The result is that, unlike most other methods, the noise is forced into the baseline estimate and this negatively affects correlation coefficients and χ^2 values (a comparison of Figs. 2b and 3b will reveal this allocation of noise).

Modes of Failure. If the selection of parameters for the smoothing step required before taking the first derivative results in a first derivative that is too noisy, peak picking and peak fitting may become more difficult or fail. The threshold setting determines how many peaks are selected from the smoothed differentiated spectrum. If the threshold is low, many spurious noise-induced peaks are generated. If the threshold is high, some peaks, especially overlapping ones, are missed. Since the computation time depends primarily on the fitting procedure, the threshold level set for peak detection strongly influences the overall computation time: the higher the threshold, the fewer peaks are detected and the shorter the fitting process. The accuracy of the method also depends on peak selection since missing some peaks would force them into the baseline and hence lead to a poor baseline estimate. Furthermore, unless the baseline is simply a constant offset, its derivative, although small, will not be zero everywhere. Whenever peaks are riding on non-zero segments of the baseline derivative, less effective and/or difficult peak fitting occurs and in some cases the algorithm may fail to converge.

Suitability of Method for Automation. Due to the need for, and difficulty of, peak picking, the method may have only limited appeal for automated procedures unless substantially modified to incorporate more accurate peak picking procedures. Some of the methods discussed below incorporate automated peak picking procedures[95,96] and they could possibly be adapted for this purpose. Computation time due to the fitting procedure may be an additional, but lesser, concern.

METHODS REQUIRING EXPLICIT ESTIMATES OF b.

Artificial Neural Networks (ANNs). ***Background.*** Artificial neural networks are highly simplified models of naturally occurring nervous system assemblies. Background information about ANNs can be obtained from the reviews by Lippmann[97] and Hush and Horne.[98] To the best of our knowledge, they have not been used before for baseline removal.

Theory. ANNs or multilayer perceptrons consist of an input, hidden, and output layer.[97,98] Each layer consists of a number of computational units or nodes. For a fully connected network, all nodes in one layer are connected to all nodes in the next layer by weighted connections. At a given hidden or output node, all the inputs to that node arriving along their respective weighted connections are summed. The function value of the sum is then determined. The function used at a given node is referred to as the activation or transfer function and it is frequently a nonlinear function. The logistic function is commonly used as the transfer function in hidden and output layer nodes. Input layer nodes customarily use linear transfer functions. For example, an input datum at an input node is multiplied with the strength of the connection weight between that input node and a given hidden node. The product that arrives at that given hidden node is then summed along with all the other products arriving from the other input nodes, respectively. In addition, a single further input is provided to all the nodes in the hidden and output layers, over weighted connections, from an input node that supplies a constant value (usually +1). The product of this fixed value with the respective connection weight is summed at the hidden or output nodes along with all the other products that arrive as input. In this way, a bias or offset can be provided to affect the sum of inputs and the subsequent function value determined by a node. This can improve the performance of a network.

Consider now a conventional single hidden layer neural network, as described above, processing a data matrix $\mathbf{Y}$ of n samples, each consisting of l spectral channels. An additional channel with a fixed value of +1 is appended for each sample to represent the bias node; thus, $\mathbf{Y}$ is an $n \times (l + 1)$ matrix. Let the network have l input nodes plus bias node, p hidden nodes plus bias node, and m output nodes. Hence, the weights between input and hidden layers can be represented by $\mathbf{W}_1$, an $(l + 1) \times p$ matrix, and the weights between hidden and output layers by $\mathbf{W}_2$, a $(p + 1) \times m$ matrix.

The ANN can now be stated in matrix notation as the following sequence of operations:

$$\mathbf{YW}_1 = \mathbf{U} \tag{9}$$

$$f(\mathbf{U}) = \mathbf{V}' \tag{10}$$

$$\mathbf{VW}_2 = \mathbf{X} \tag{11}$$

$$g(\mathbf{X}) = \mathbf{Z} \tag{12}$$

or in more compact form:

$$g\{[f(\mathbf{YW}_1) + \text{bias column}](\mathbf{W}_2)\} = \mathbf{Z} \tag{13}$$

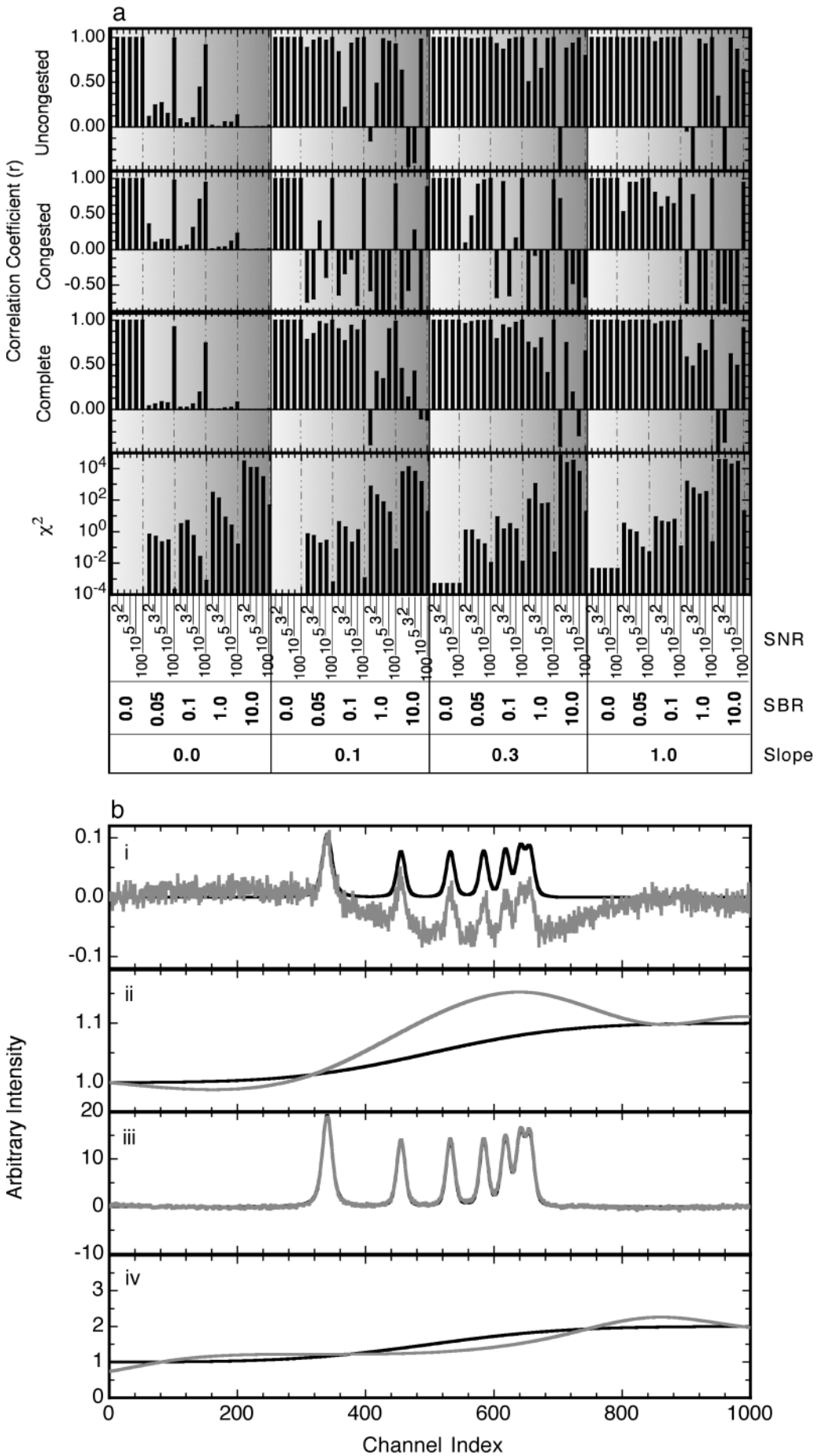
where $\mathbf{V}$ is $\mathbf{V}'$ with bias added, $\mathbf{X}$ is the $n \times m$ matrix of outputs from the output layer, $\mathbf{Z}$ is the $n \times m$ matrix of the final network output, and $f(x)$ and $g(x)$ are (nonlinear) transfer functions for which their respective inverses $f^{-1}(x)$ and $g^{-1}(x)$ exist. Equation 13 can be conceived as a linear first part

$$\mathbf{U} = \mathbf{YW}_1 \tag{14}$$

and a nonlinear second part

←

FIG. 4. Baseline removal using artificial neural networks. (**a**) Correlation coefficient between theoretical and estimated baselines for the uncongested, the congested, and the total spectral regions of given spectra. Also shown is the χ^2 between theoretical and estimated baselines for the total spectral region. (**b**) (***i***) Baseline removed (gray trace) and initial input spectrum (less noise and true baseline, black trace) for SNR = 10, SBR = 0.1, and baseline slope = 0.1; (***ii***) estimated baseline (gray trace) and input baseline (black trace), SNR, SBR, and baseline slope as for (i); (***iii, iv***) as for (i) and (ii) with SNR = 100, SBR = 10, and baseline slope = 1.0.


a
Correlation Coefficient (r)
Uncongested
Congested
Complete
χ^2
SNR
SBR
Slope
0.0
0.05
0.1
1.0
10.0
0.0
0.1
0.3
1.0
b
i
ii
iii
iv
Arbitrary Intensity
Channel Index

$$\mathbf{U} = f^{-1}[g^{-1}(\mathbf{Z})\mathbf{W}_2^+] = f^{-1}(\mathbf{X}\mathbf{W}_2^+ - \text{bias column}) \quad (15)$$

where $\mathbf{W}_2^+$ is the pseudo-inverse of $\mathbf{W}_2$. The network error is the Euclidean norm of the difference between the computed output $\mathbf{Z}$ and the desired or target output $\mathbf{D}$:

$$E = \|\mathbf{D} - \mathbf{Z}\|_2 \quad (16)$$

where $\mathbf{D}$ also has dimensions $n \times m$. Normally ANNs are trained by initializing the weight matrices to small random values and using an iterative optimization procedure to minimize the error function. The network could also be trained by initializing and optimizing $\mathbf{U}$ (and solving for the weights after each iteration) instead.

If additional information about the weights or $\mathbf{U}$ is available, the network parameters can be directly calculated and/or the iteration process abbreviated. This can be done, for example, by considering the linear part of the network to constitute a set of linear filters that permit only 'essential' information to reach the output. This information is sampled at a frequency corresponding to the number of hidden nodes chosen for the network and an effective receptive field that corresponds to the filter window (i.e., non-zero weights). The 'receptive' field of a node is the number of nodes to which it is connected on the input side. Hence, instead of connecting a given hidden or output layer node to all the nodes in the previous layer, it can be given a restricted receptive field by connecting it only to a few of the input nodes. If the input data are linearly modeled as

$$\mathbf{y} = \mathbf{s}*\mathbf{p} + \mathbf{b}*\mathbf{p} + \mathbf{n} \quad (17)$$

then a finite impulse response filter can be constructed such that the filtered output at point i, u_i, is defined by:

$$u_i = \sum_{k=-M}^{M} (y_k)(h_{i-k}) \quad (18)$$

where h_k is the filter weight function, the window size is $2M + 1$, and $i \geq M$.

Restated in matrix notation:

$$\mathbf{U} = \mathbf{Y}\mathbf{H}^{\mathrm{T}} \quad (19)$$

where $\mathbf{H}^{\mathrm{T}}$ is the transpose of the matrix of filter weights (compare to Eq. 14). Hence, we consider the weights $\mathbf{W}_1$ and $\mathbf{W}_2$ as filters of their respective inputs. This implies that we view $\mathbf{Y}$ and $\mathbf{X}$ in Eqs. 14 and 15, respectively, as a set of 'measurements' to be filtered. Note that the use of recurrent connections (where the output from a given node is connected back to itself as well as to nodes in the next layer) implements infinite impulse response filters.

Savitzky–Golay filters have become widely used for spectroscopic applications with the zero-order variant being commonly known as a moving average filter. In this case, for example, the filter coefficients or network weights are all equal to 1/(filter window size). In the ANN implementation, weights outside of the receptive fields (filter windows) are set to zero and those within a field are set to the appropriate filter coefficients. When $\mathbf{W}_1$ is assigned on an *a priori* basis as shown here, $\mathbf{W}_2$ can be solved for analytically or both can be further refined by iteration to minimize the output error function.

Implementation. We trained a network with 1001 input nodes plus 1 bias node, 50 hidden nodes plus 1 bias node, and 1001 output nodes. Every hidden node was connected to 300 input nodes (hence having a receptive field of 300) such that all input nodes were connected to at least one hidden node. The input weights in every receptive field were taken as the Savitzky–Golay first-order filter coefficients of a correspondingly sized filter window. Hidden nodes used a logistic transfer function and output nodes used a linear transfer function. The network was trained in a single pass with a randomly selected set of 240 spectra comprising 80% of the total number of 300 spectra. These spectra were given as input at the input layer nodes with the corresponding baseline given as target at the output nodes. The network was tested with the remaining 60 spectra.

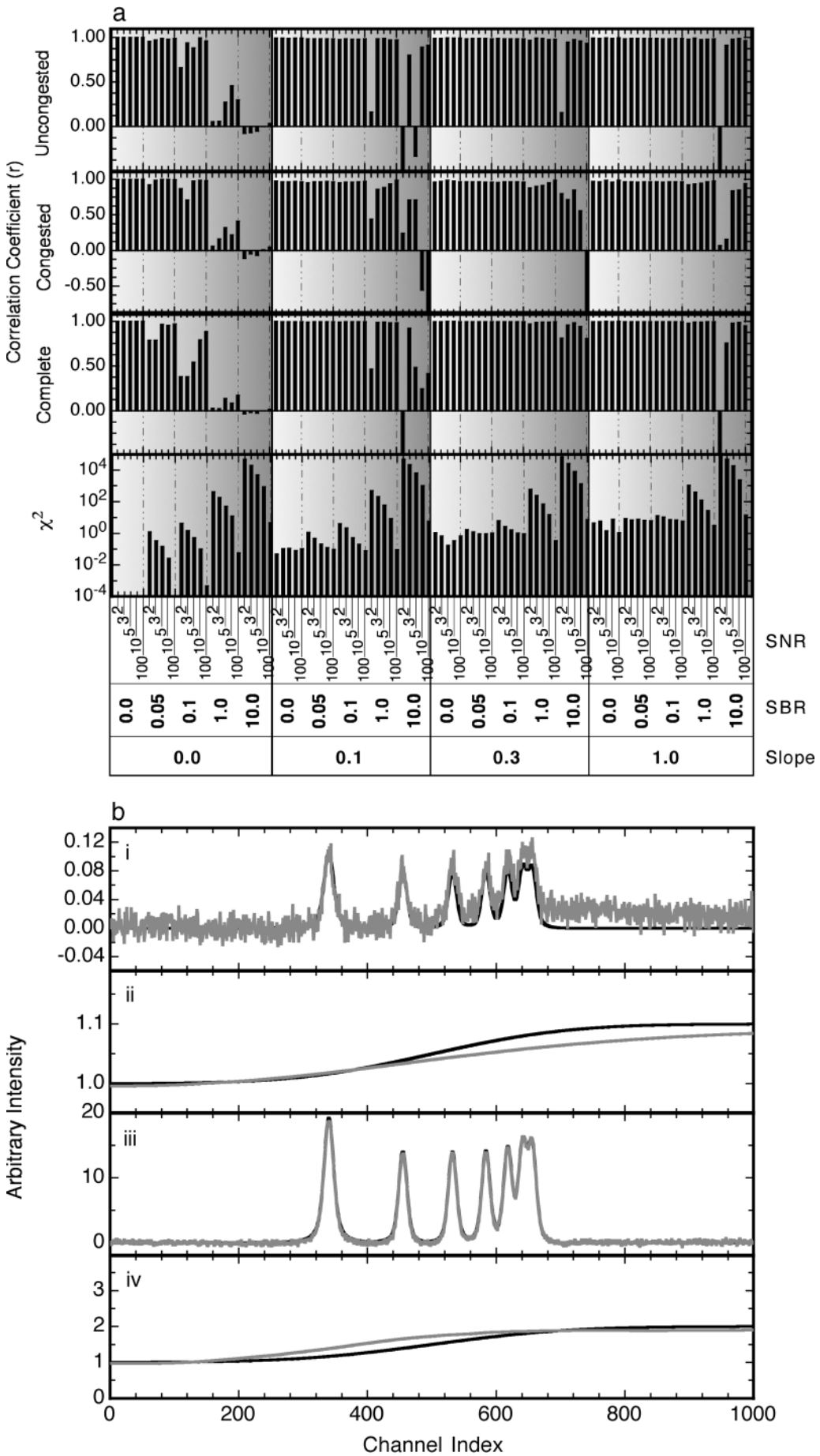
Results and Discussion. The root mean square error for the network on the 60 testing spectra was 0.001, and when tested on all the 300 spectra the error was 0.042. The other results reported here were obtained on all 300 spectra. The quantitative results are shown in Fig. 4a and comparative examples of baseline correction on two spectra (SNR 10, SBR 0.1, slope 0.1; SNR 100, SBR 10, slope 1.0) are shown in Fig. 4b. The results (see also below) showed little difficulty with congested spectra. Computation times were very short as shown in Fig. 13, but training times were longer (seconds).

Although the results for this method were good, it should be noted that despite having information about the general shape of the baseline, the combination of low SNR, high SBR, and mild or no slope combined to sufficiently obscure the baseline to make its estimation difficult. Furthermore, the neural network may have to be retrained for different baseline types. It is not currently known whether a single network could be trained to perform baseline estimation on spectra with varying shapes of baseline. If the initial network weights are randomly assigned, instead of using some *a priori* assignment such as the filter coefficients used here, iterative optimization of the network weights will be necessary. Training a network in this manner (e.g., using gradient descent, conjugate gradient, or other method) may take considerable time depending on the difficulty of the problem and the number of spectra in the training set. However, once trained, baseline estimation for individual spectra would proceed quite rapidly.

Modes of Failure. This method may produce poor results if the optimum number of hidden nodes is not used: too few hidden nodes results in poor performance on all

←

FIG. 5. Baseline removal using threshold-based classification. (**a**) Correlation coefficient between theoretical and estimated baselines for the uncongested, the congested, and the total spectral regions of given spectra. Also shown is the χ^2 between theoretical and estimated baselines for the total spectral region. (**b**) (***i***) Baseline removed (gray trace) and initial input spectrum (less noise and true baseline, black trace) for SNR = 10, SBR = 0.1, and baseline slope = 0.1; (***ii***) estimated baseline (gray trace) and input baseline (black trace), SNR, SBR, and baseline slope as for (i); (***iii***, ***iv***) as for (i) and (ii) with SNR = 100, SBR = 10, and baseline slope = 1.0.


a
Correlation Coefficient (r)
Uncongested
Congested
Complete
χ^2
SNR
SBR
Slope
0.0
0.05
0.1
1.0
10.0
0.0
0.1
0.3
1.0
b
Arbitrary Intensity
i
ii
iii
iv
Channel Index

spectra and too many hidden nodes in good performance on the training data but poor performance on the testing data (i.e., poor generalization). The size of the receptive field (filter window) will also influence the results: too small a receptive field produces poor results while too large a receptive field leads to increases in training and processing time and in some cases also a reduced performance. The network performance also depends on the filter weights chosen for the connections in each receptive field and these weights are governed by the selected filter parameters (e.g., using the coefficients of a first order Savitzky–Golay filter as weights). Finally, the selection of training data may lead to poor results if it provides insufficient knowledge about the baseline(s) during the training step (see also below).

Suitability of Method for Automation. If the shape of the baseline is not known in advance, the network will have to be trained on different baseline types. This requirement may limit its general appeal for automated procedures and reduce its performance. On the other hand, precisely this adaptive quality and the fast processing times provide valuable potential for automation under well-characterized conditions. These well-characterized conditions essentially mean that the shape (though not the relative intensities) of the baseline has to be known (e.g., manually or experimentally determined) since it is required for training.

Threshold-Based Classification (TBC). ***Background.*** This method, published by Dietrich et al. in 1991,[95] is an automated implementation of a previous suggestion by Dietrich and Gerhards made in 1981.[21] TBC involves automatic peak recognition using the squared derivative of the spectrum, polynomial fitting to the remaining spectral regions to obtain an estimate of the baseline, and subsequent removal of the fit from the original spectrum to produce a baseline-corrected result. Some earlier[8,13,16,17,19,20,25] and later works[30,36,38,41,84,99,100] employ one or more similar concepts. The feature-based classification method by Rouh et al.,[96] based on pattern classification principles,[101] is also in many respects conceptually related to TBC.

Theory. The TBC uses a smoothing step to reduce the effects of noise on the derivative of the spectrum. This is followed by a differentiation step to facilitate peak location. After smoothing, the measured spectrum is approximated by Eq. 4 and subsequent derivation by Eq. 6. The power spectrum $\mathbf{y}'^2$ of $\mathbf{y}'$ is obtained by squaring every element of $\mathbf{y}'$. The effect of this operation is to convert all values associated with signal peaks to positive values and to generally enhance them relative to the derivative of the baseline. The mean $\bar{x}$ and standard deviation σ of $\mathbf{y}'^2$ are then determined and a threshold of $\bar{x} + 3\sigma$ is established to discriminate between elements of $\mathbf{y}'^2$. All elements are then classified as provisionally belonging to either $\mathbf{b}'$ or $\mathbf{s}'$ (and by extension to **b** or **s**) depending on whether their values are less or greater than the threshold, respectively. This classification step is iterated by obtaining a new mean and standard deviation, and thus a new threshold, from those elements assigned to **b**. Iteration is terminated when the elements assigned to **b** and **s** no longer change. The classification of elements of $\mathbf{y}'^2$ as belonging to either **b** or **s** is completed by moving all those elements of **s** where both of their neighbors belong to **b**, to **b** as well (and *vice versa*). Since the elements in **b**, obtained from $\mathbf{y}'^2$, identify those in **y** that belong to the baseline, an approximation of the true baseline can now be obtained by replacing all elements assigned to **b** with their corresponding elements in **y** (the measured data). A fifth-order polynomial or natural cubic spline is then fitted to **b** to estimate the baseline and the fitted values are subtracted from **y** to yield **s**.

Implementation. We smoothed the data with an 11-channel filter window using a moving average filter (i.e., Savitzky–Golay zero-order filter). A natural cubic spline was then fitted through 10 points chosen from the set of points classified as belonging to the baseline. These points were evenly spaced with regard to the elements of the baseline set, but not necessarily evenly spaced with regard to spectral (domain) position.

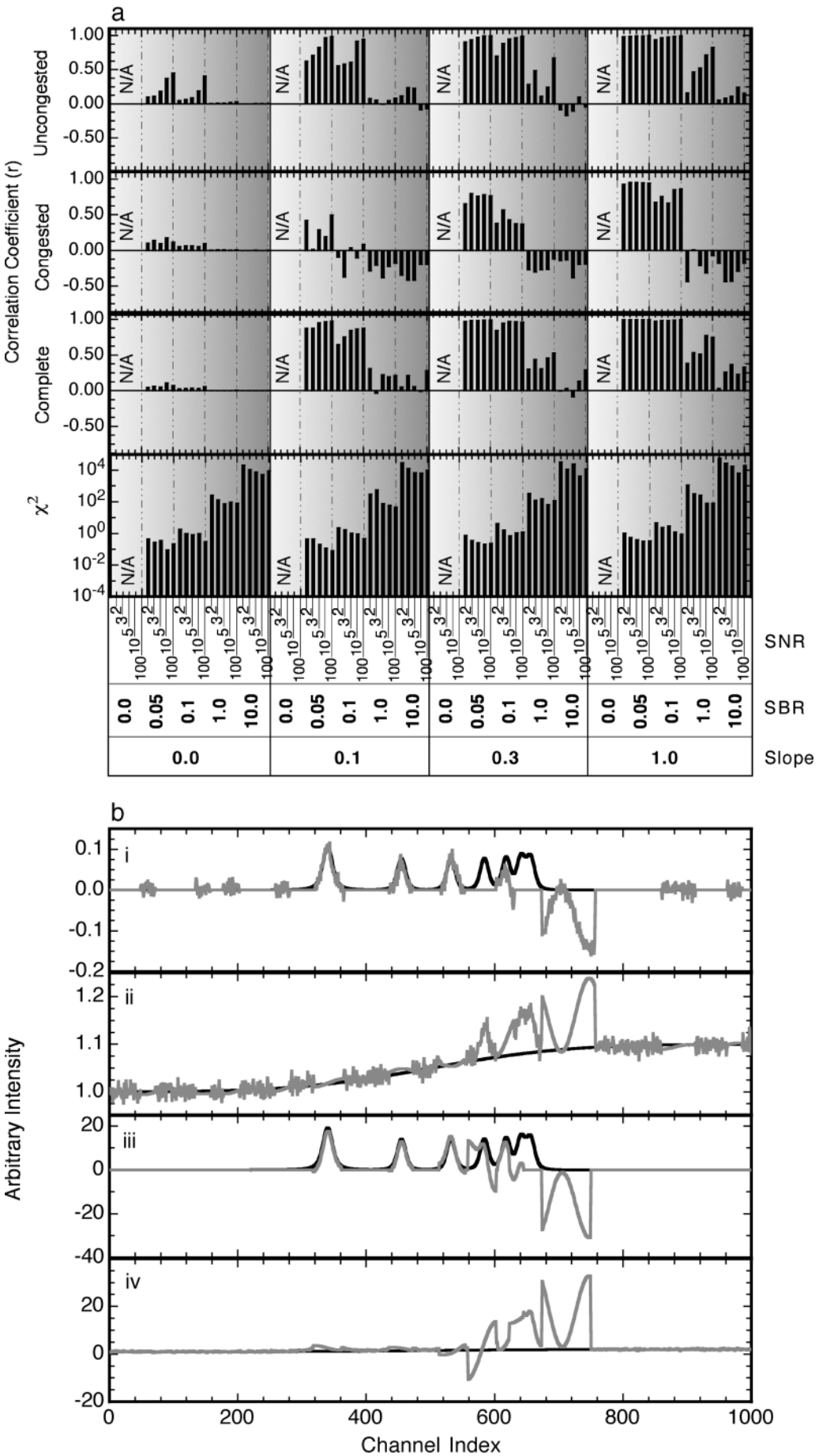
Results and Discussion. The quantitative results are shown in Fig. 5a and comparative examples of baseline correction on two spectra (SNR 10, SBR 0.1, slope 0.1; SNR 100, SBR 10, slope 1.0) are shown in Fig. 5b. The results showed difficulties with noisy spectra, congested spectral regions, and with baseline fitting (see below). Computation times were comparatively short as shown in Fig. 13.

Modes of Failure. In this method, the baseline is estimated by fitting through spectral points classified as belonging to the baseline (polynomial fit) or a selection of these points (spline fit). For spectra with low SNRs, polynomial fitting produced better results since the fit minimizes deviations from the polynomial and the results are, relatively speaking, less influenced by noise, but when spline fitting is used, the spline is forced through the selected points and the fit is therefore much distorted by noise. For spectra with higher SNRs, spline fitting produced better results since a spline fit could track the baseline more accurately than a polynomial. In either case, the fitting procedure produced artifacts in some of the spectra. Additionally, poor parameter selection for the smoothing step prior to differentiation could adversely affect locating peaks.

Suitability of Method for Automation. Since few parameters need to be determined, this method is suitable for automation. Provided that little or moderate amounts of noise are present in spectra and that a suitable baseline-fitting approach could be found or that the fitting approaches mentioned here are suitable for the problem

←

FIG. 6. Baseline removal using signal removal (moving average). (**a**) Correlation coefficient between theoretical and estimated baselines for the uncongested, the congested, and the total spectral regions of given spectra. Also shown is the χ^2 between theoretical and estimated baselines for the total spectral region. (**b**) (***i***) Baseline removed (gray trace) and initial input spectrum (less noise and true baseline, black trace) for SNR = 10, SBR = 0.1, and baseline slope = 0.1; (***ii***) estimated baseline (gray trace) and input baseline (black trace), SNR, SBR, and baseline slope as for (i); (***iii***, ***iv***) as for (i) and (ii) with SNR = 100, SBR = 10, and baseline slope = 1.0.


a
Correlation Coefficient (r)
Uncongested
Congested
Complete
χ²
N/A
SNR
SBR
Slope
0.0
0.1
0.3
1.0
0.05
10.0
b
Arbitrary Intensity
i
ii
iii
iv
Channel Index

at hand (see also the general discussion below), the results may be satisfactory.

Signal Removal Methods (SRM). ***Background.*** Signal removal methods require the estimation of the baseline with some procedure such as polynomial fitting to the entire spectrum. Points more than a given distance above this baseline estimate are stripped from the spectrum by replacing them with the values of the baseline estimate. After stripping, a new baseline estimate is generated and the procedure is often iterated until new baseline estimates change little or not at all between iterations. Because peaks are iteratively stripped from the spectra, they do not have to be explicitly located in advance.

One of the first reported SRMs was that by Ralston and Wilcox in 1969 using a moving window smoothing procedure and 10 iterations.[7] Variants using polynomials, splines, moving averages or other smoothing filters, and wavelets followed.[9,102–104] Other approaches, to a lesser or greater extent related to signal removal or peak stripping, have also been advanced.[15,24,31,33,43,47,54,65,78,105]

Theory. We have further developed and refined SRMs in-house with the aim to effect automated baseline correction. No smoothing step is applied and the data are implicitly modeled as in Eq. 3. A crude estimate of the baseline, $\hat{\mathbf{b}}$, is obtained from $\mathbf{y}$ using either a low-order polynomial fit, a spline fit, or a large window filter (e.g., moving average or Savitzky–Golay zero-order filter), depending on the variant employed. A high threshold is then established based on the standard deviation σ of $\mathbf{y}$ (e.g., 5σ), and the value of points above the threshold value from the estimated baseline, if any, are reduced by some pre-specified amount (e.g., 25% of its value) to yield a residual spectrum $\mathbf{r}$. The value of the threshold is then reduced by some small amount (e.g., 1%) and the process iterated by obtaining a new $\hat{\mathbf{b}}$ from $\mathbf{r}$ and reducing those values of $\mathbf{r}$ exceeding the corresponding value of $\hat{\mathbf{b}}$ plus current threshold until a stopping criterion is reached. The stopping criterion can be based on either the signal-to-noise ratio (SNR) of the residual spectrum $\mathbf{r}$ (here defined as maximum of $\mathbf{r} - \hat{\mathbf{b}}$/standard deviation of $\mathbf{r} - \hat{\mathbf{b}}$) or on a predetermined cut-off value for the threshold. Where splines or moving averages are used to determine $\hat{\mathbf{b}}$, the splines are fitted through segments with initial lengths about 5–10% of the spectrum or with the window size set at about 5–10% of the spectrum, respectively. On successive iterations, the segment or the window will be given a random component that can be biased to gradually reduce or increase its size. The effect of this approach is to remove the convolved signals from the original data $\mathbf{y}$, leaving $\mathbf{b}*\mathbf{p} + \mathbf{n}$ from which $\mathbf{b}*\mathbf{p}$ is obtained with the last polynomial fit, spline fit, or moving average operation on $\mathbf{r}$. If, instead of the final estimate $\hat{\mathbf{b}}$ based on the final $\mathbf{r}$, the final $\mathbf{r}$ itself is subtracted from the initial spectrum $\mathbf{y}$, noise is also removed from the recovered spectrum. Hence, initially $\mathbf{r} = \mathbf{s}*\mathbf{p} + \mathbf{b}*\mathbf{p} + \mathbf{n}$ and finally $\mathbf{r} \approx \mathbf{b}*\mathbf{p} + \mathbf{n}$, and thus $\hat{\mathbf{b}} \approx \mathbf{b}*\mathbf{p}$.

Implementation. We used an initial window size of 51 channels and subsequently increased that by 5% per iteration. In every channel, we removed 25% of that portion of the peak extending above the current threshold. The threshold was initially set to 5 times the standard deviation of the original spectrum. The threshold was adjusted to 90% of its previous value after each iteration. We used stopping criteria of either SNR = 3 or a threshold level less than 0.05, whichever was attained first. For the spline variant we used a natural cubic spline and for the polynomial variant we used a fourth-order polynomial.

Results and Discussion. The quantitative results are shown in Fig. 6a and comparative examples of baseline correction on two spectra (SNR 10, SBR 0.1, slope 0.1; SNR 100, SBR 10, slope 1.0) are shown in Fig. 6b for the moving average variant. The results showed fairly good consistency under most conditions, but with some difficulties in noisy spectra and congested spectral regions. Computation times were comparatively moderately short as shown in Fig. 13.

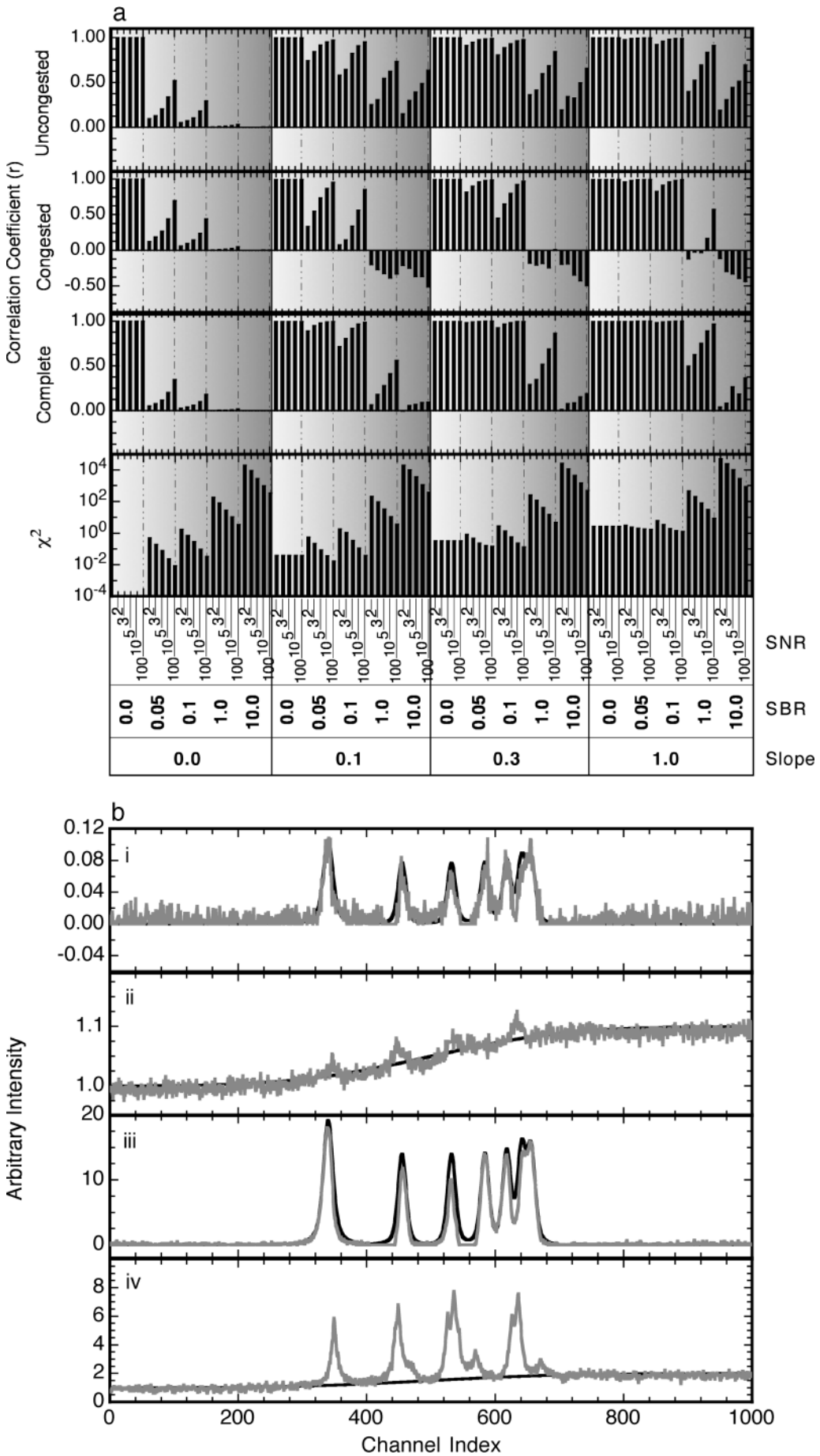
Modes of Failure. This general approach requires the selection of initial window sizes, rates of change of these window sizes, and amounts of peaks in excess of current thresholds to remove from the spectra. Choosing too small an initial window size will result in more of the peaks being considered part of the baseline, while too large an initial window size will result in prominent parts of the baseline being considered as belonging to a peak and hence removed. A choice of initial window size of about three times that of the average FWHM of spectral peaks proved satisfactory. If a larger window is used, it is advisable to reduce the window size on subsequent iterations. Reducing peaks by larger amounts will speed up computation time but may reduce the accuracy with which the baseline is estimated. For low SNR spectra, outliers due to noise are removed from the residual spectrum $\mathbf{r}$ along with peaks; hence, the baseline is generally underestimated but retains its shape. This problem can be addressed by adjusting the final baseline-corrected spectrum by a constant amount to eliminate the offset.

Suitability of Method for Automation. This general approach requires relatively few parameters to be determined, is modestly fast, consistently good under most conditions, and therefore suitable for automation. Residual baseline artifacts could potentially be removed in a second pass of the baseline-corrected spectrum.

Composite (Linear-Sine-Cosine) Baseline Method (CBM). ***Background.*** In their 1977 evaluation of the effects of polynomial and spline baseline estimating methods on the limits of quantification in chromatography, McNulty and MacFie[106] also evaluated a composite meth-

←

FIG. 7. Baseline removal using composite baseline estimation. (**a**) Correlation coefficient between theoretical and estimated baselines for the uncongested, the congested, and the total spectral regions of given spectra. Also shown is the χ^2 between theoretical and estimated baselines for the total spectral region. (**b**) (***i***) Baseline removed (gray trace) and initial input spectrum (less noise and true baseline, black trace) for SNR = 10, SBR = 0.1, and baseline slope = 0.1; (***ii***) estimated baseline (gray trace) and input baseline (black trace), SNR, SBR, and baseline slope as for (i); (***iii***, ***iv***) as for (i) and (ii) with SNR = 100, SBR = 10, and baseline slope = 1.0. N/A = not applicable; this method relies on detecting peak edges, and spectra without peaks (SBR = 0) could not be processed.


a
Correlation Coefficient (r)
Uncongested
Congested
Complete
χ^2
SNR
SBR
Slope
0.0
0.1
0.3
1.0
b
Arbitrary Intensity
i
ii
iii
iv
Channel Index

od influenced by Synovec and Yeung[107] that appeared in 1985, estimating the baseline under signal peaks as a straight line plus sinusoids to accommodate nonlinearities in the baseline.

This method requires the location of peaks in a spectrum, as well as their leading and trailing edges. Regions near the leading and trailing edges are then used to estimate the baselines locally with a regression procedure and they are then interpolated across the feet of the peaks. In earlier work the baselines were generally approximated by straight lines[10] and in later similar work polynomials and splines were used to approximate nonlinear baseline regions.[13,22] In addition, linear estimates were refined to provide greater robustness.[1] The minimum search and interpolation methods are somewhat related to the CBM as is the robust local regression estimation method.[1]

Theory. The measured data are implicitly modeled as in Eq. 3 and the baseline is estimated for regions under signal peaks. A small number of data points, δ, preceding the leading edge of a given peak, *l*, and following the trailing edge, *t*, are used to perform a least squares regression and the baseline value underneath a given peak is estimated according to:

$$b_i = mi + c + \lambda_1 \sin\left[\pi \frac{i - (l - \delta)}{(t + \delta) - (l - \delta)}\right] - \lambda_2 \cos\left[\pi \frac{i - (l - \delta)}{(t + \delta) - (l - \delta)}\right] \quad (20)$$

where *i* is the index of the local region of the baseline, *m* and *c* are slope and intercept respectively, and λ_1 and λ_2 are optimization parameters. Baselines estimated in this manner are generally smooth (i.e., devoid of noise), while baseline regions not under signal peaks may contain noise.

Implementation. We used 21 channels for δ. We determined the leading and trailing edges of peaks by smoothing a given spectrum with a moving average using a window five times the size of δ and subtracting the same spectrum when smoothed with a moving average using a window of the same size as δ. In the resulting difference spectrum, peak leading and trailing edges were defined by points more than three times above the difference spectrum standard deviation.

Results and Discussion. The quantitative results are shown in Fig. 7a and comparative examples of baseline correction on two spectra (SNR 10, SBR 0.1, slope 0.1; SNR 100, SBR 10, slope 1.0) are shown in Fig. 7b. The results showed difficulties with noisy spectra and spurious peak detection and difficulties with congested spectral regions. Baseline estimation was comparatively moderately slow as shown in Fig. 13.

The benefit of estimating the baseline only under peaks is that noise in the baseline-only regions of the spectrum are assigned to the baseline, thus effecting a partial noise removal from the baseline-corrected spectrum. It is therefore important to keep this point in mind when examining the results reported for this method since the estimated baseline, unlike most other methods, will contain noise. A more consistent approach may use the same procedure or a modification to estimate the baseline also in regions where signal peaks are absent. This would result in the spectrum-wise separation of baseline, signal, and noise.

Modes of Failure. As with other methods requiring peak picking, the threshold setting determines how many peaks are selected and how accurately the baseline is estimated. For more details in this regard, see the Modes of Failure subsection for the first-derivative method above. This method also requires peak edges to be determined, thus further complicating the process. In particular, for congested spectra, the leading and trailing edges of a given peak may overlap with those of other peaks. In such cases, regression parameters may be invalid or difficult to determine, rendering proper baseline estimation difficult. Extensive interpolation, such as required under peaks in congested regions, can also lead to incorrect baseline estimation.

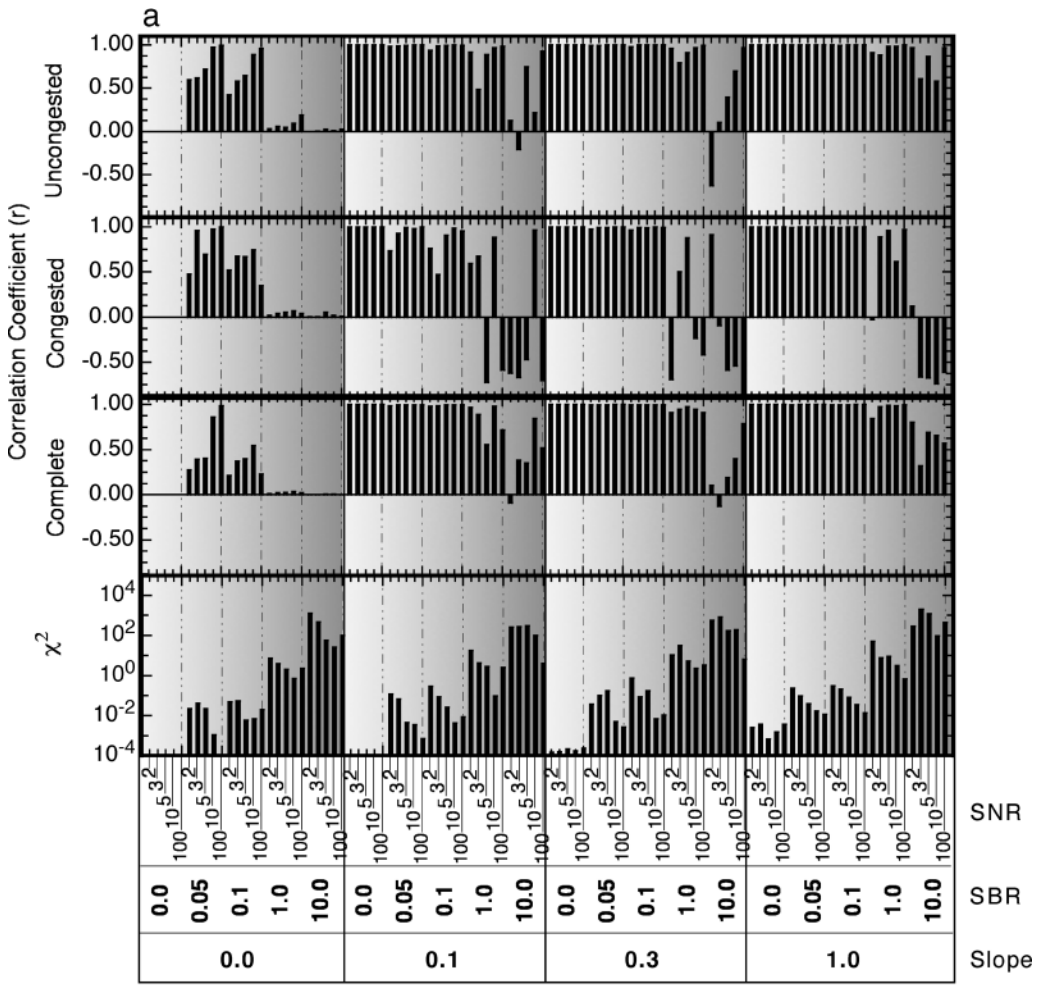
Suitability of Method for Automation. The number of parameters to be determined and especially the difficulties in establishing them render this approach in general not very suitable for automation.

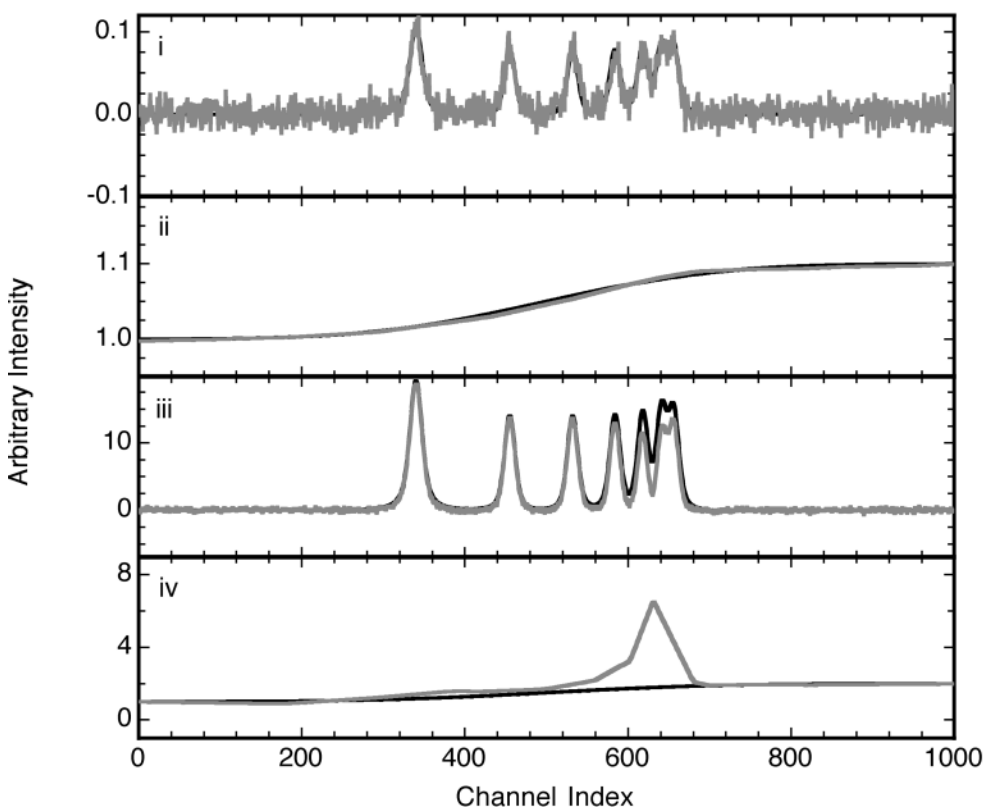
Spectral Shift Methods (SSM). ***Background.*** In spectral shift methods, one spectrum is subtracted from another highly similar spectrum of the same sample. The highly similar spectrum can be obtained, depending on the spectroscopic method, by changing the excitation frequency or changing the grating position in the spectrometer by some small amount. A highly similar spectrum can also be digitally generated by shifting the original spectrum along the abscissa by some small amount (e.g., a few wavenumbers for Raman spectroscopy). The smaller this shift becomes, the more the resulting difference spectrum between the original and the shifted spectrum approximates a first derivative. When the spectrum is shifted by only one channel, a first derivative by the simple difference method results. SSMs are therefore related to FDMs. Spectral shift approaches have been developed to discriminate against fluorescence backgrounds in Raman spectroscopy.[61–64,108,109]

Theory. If the baseline or background in a spectrum has broad, slowly varying features relative to the signal peaks, displacing the spectrum by some small amount would affect the signal peaks more significantly than the background. In other words, the background intensity in any given channel is similar to the background intensity in a nearby channel, but the same is not true for a signal peak. Hence, subtracting a displaced or shifted spectrum from the non-shifted spectrum will succeed in removing

←

FIG. 8. Baseline removal using spectral shifts. (**a**) Correlation coefficient between theoretical and estimated baselines for the uncongested, the congested, and the total spectral regions of given spectra. Also shown is the χ^2 between theoretical and estimated baselines for the total spectral region. (**b**) (***i***) Baseline removed (gray trace) and initial input spectrum (less noise and true baseline, black trace) for SNR = 10, SBR = 0.1, and baseline slope = 0.1; (***ii***) estimated baseline (gray trace) and input baseline (black trace), SNR, SBR, and baseline slope as for (i); (***iii, iv***) as for (i) and (ii) with SNR = 100, SBR = 10, and baseline slope = 1.0.


a
Correlation Coefficient (r)
Uncongested
Congested
Complete
χ²
SNR
SBR
Slope
0.0
0.1
0.3
1.0
0.0
0.05
0.1
1.0
10.0


Arbitrary Intensity
i
ii
iii
iv
Channel Index

most of the background effectively but will either remove too much or too little of the signal peaks, giving rise to a derivative-like difference spectrum.

The measured spectra are modeled as in Eq. 1, where the values of signal plus baseline can be calculated according to

$$y_{1x} = b_{1x} + \sum_{i=1}^{m} \left\{ \frac{a_i}{1 + \left[\frac{(x - x_0)^2}{w_i} \right]} \right\} + n_{1x} \quad \text{and} \tag{21}$$

$$y_{2x} = b_{1x+\delta} + \sum_{i=1}^{m} \left(\frac{a_i}{1 + \left\{ \frac{[x - (x_0 + \delta)]^2}{w_i} \right\}} \right) + n_{1x+\delta} \tag{22}$$

for the original and shifted observations, respectively. Parameters are as defined for the FDM and n denotes noise. If the amount of shift, δ, covers only a small number of data points, it is assumed that $b_{1x+\delta} - b_{1x} \approx 0$. Hence, subtracting the original from the shifted data (ignoring the additive noise) yields a spectrum with derivative-like features:

$$\begin{aligned} y_{2x} - y_{1x} &= s_{2x} - s_{1x} \\ &= \sum_{i-1}^{m} \left(\frac{a_i}{1 + \left\{ \frac{[x - (x_0 + \delta)]^2}{w_i} \right\}} \right\} \\ &\quad - \sum_{i=1}^{m} \left\{ \frac{a_i}{1 + \left[\frac{(x - x_0)^2}{w_i} \right]} \right\} \end{aligned} \tag{23}$$

Analogous to the FDM, the values of the parameters in Eq. 23 are determined with a suitable nonlinear optimization procedure and substituted in Eq. 7 to calculate the signal vector **s**.

Implementation. We have modified this method by using instead of a single shifted spectrum two shifted spectra, each shifted by the same amount, but in opposite directions from the original. These were subtracted from the original, the negative values in the resulting spectra were set to zero, and the two difference spectra were then averaged. Due to the complementary nature of the opposing shifts, and the elimination of negative values, derivative-like features were eliminated from the resultant spectra. Therefore, no fitting process is required and there is no need to determine spectral parameters from Eq. 23 and to calculate the signal **s** from these parameters, hence greatly reducing computation time. We used a spectral shift or offset of 21 channels.

Results and Discussion. The quantitative results are shown in Fig. 8a and comparative examples of baseline correction on two spectra (SNR 10, SBR 0.1, slope 0.1; SNR 100, SBR 10, slope 1.0) are shown in Fig. 8b. The results showed that distortions in peak shape and relative intensities occur and that baseline estimation is more difficult in congested spectral regions. Computation times were very short as shown in Fig. 13.

Although the implementation by Mosier-Boss et al.[61] requires several parameters to be determined, the opposing offset variant implemented here has only one parameter to be determined, namely, the size of the offset or shift. An additional benefit of the opposing shift implementation is that spectra do not appear in derivative representation.

Modes of Failure. If the offset is too large, then regions of the baseline with strong curvature are retained in the baseline-corrected spectra. If the offset is too small, then some portions of peaks are rejected along with the baseline. This seems to produce an enhancement of spectral resolution, but carries with it a loss of peak intensity and a distortion of peak shapes. The enhancement of sharp peaks was also observed in another implementation of the SSM.[109]

Suitability of Method for Automation. Since only one parameter has to be determined and computation times are very short for the opposing shift variant of SSMs, it is appealing for automation purposes. However, the spectral distortions induced constitute a drawback and modification may be needed to overcome this problem.

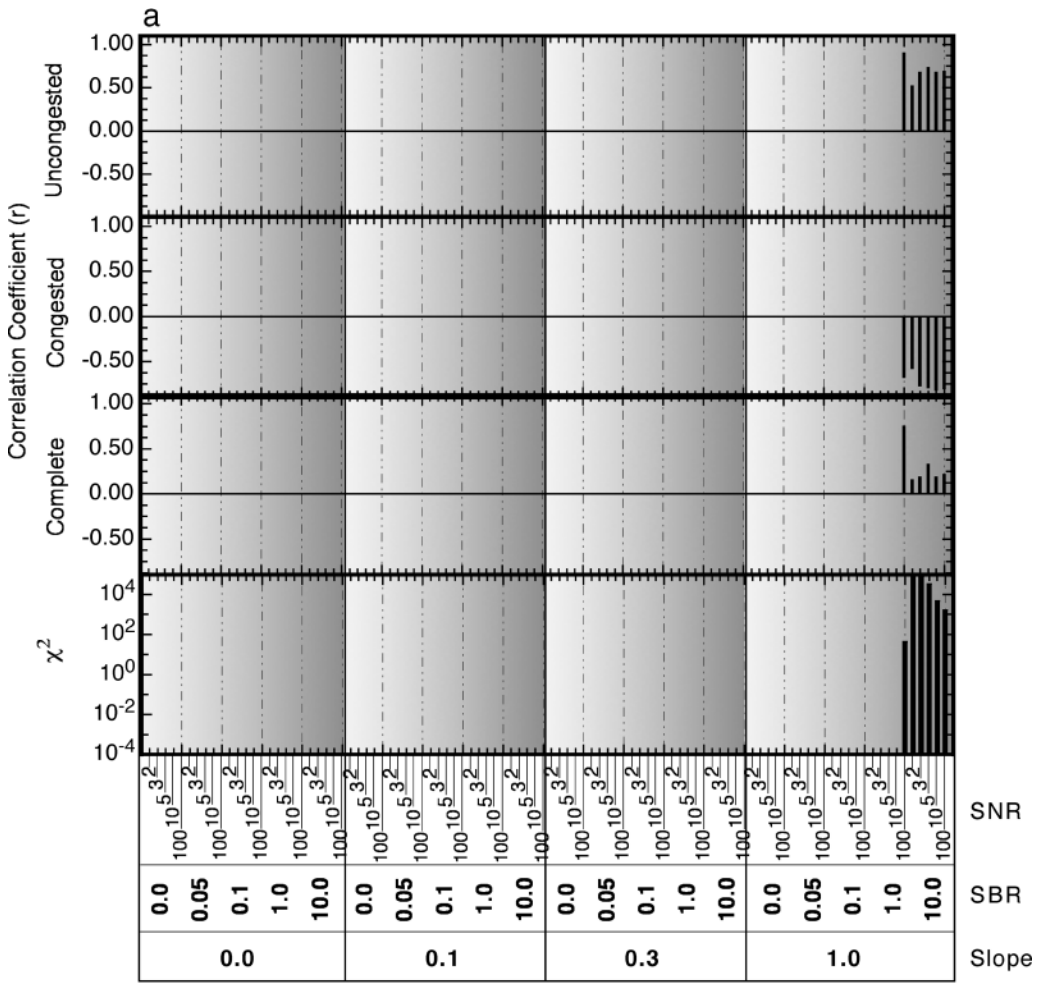
METHODS REQUIRING EXPLICIT USE OF b AND n.

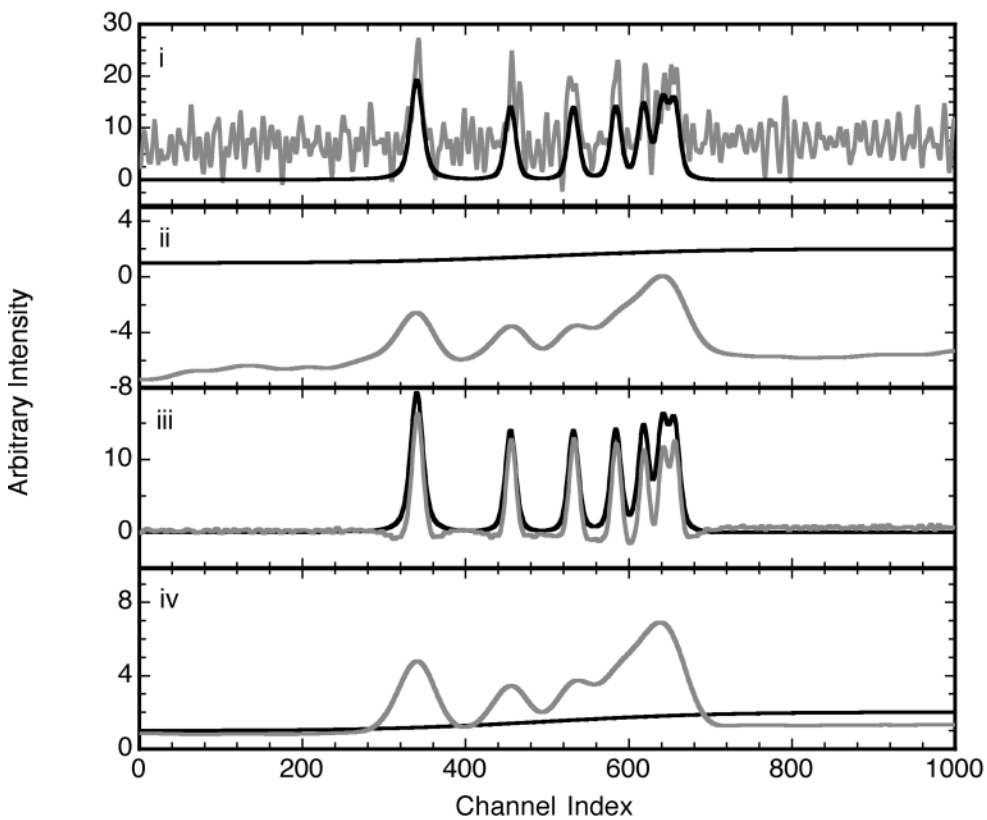
Manual Methods (MM). ***Background.*** Manual methods consist of selecting manually or 'by eye' a number of points from a spectrum deemed to lie on the baseline. The baseline is then interpolated between these points using linear, polynomial, or spline interpolations. The manual removal of baselines may be the most commonly used baseline correction method and it is often used as a standard.[66] Indeed, many automated methods such as NMM, TBC, and CBM, and methods related to them such as minimum search and interpolation methods could be seen as attempts to simulate these MMs. Several MM implementations, evaluations, and comparisons with new methods can be found in the literature.[38,40,48,55,66,103,110–112]

Theory. In the global variant of this approach, a number of more-or-less evenly spaced points judged to be on the true baseline are identified. Line segments, a low-order polynomial, or splines are then fitted through these points to estimate the baseline. In the local variant, points judged to be on the baseline at the start and end of peaks are identified and a linear interpolation is made to estimate the baseline below spectral peaks. In both variants, the data are modeled implicitly as in Eq. 3 and the selection of points judged to be on the true baseline involve some explicit accounting of baseline and noise characteristics even though these are not mathematically formulated. It is the latter factor that may vary from person to person and within the same person based on experi-

←

FIG. 9. Baseline removal using manual methods. (**a**) Correlation coefficient between theoretical and estimated baselines for the uncongested, the congested, and the total spectral regions of given spectra. Also shown is the χ^2 between theoretical and estimated baselines for the total spectral region. (**b**) (***i***) Baseline removed (gray trace) and initial input spectrum (less noise and true baseline, black trace) for SNR = 10, SBR = 0.1, and baseline slope = 0.1; (***ii***) estimated baseline (gray trace) and input baseline (black trace), SNR, SBR, and baseline slope as for (i); (***iii, iv***) as for (i) and (ii) with SNR = 100, SBR = 10, and baseline slope = 1.0.


a
Correlation Coefficient (r)
Uncongested
Congested
Complete
χ2
SNR
SBR
Slope
0.0
0.1
0.3
1.0
0.05
0.1
1.0
10.0


Arbitrary Intensity
i
ii
iii
iv
Channel Index

ence. Hence, automated numerical methods have been pursued in an attempt to minimize the need for judgment calls and to improve reproducibility. Therefore we include here some results based on MMs to provide an indication of their comparative precision and accuracy as well as to provide a benchmark for other methods. A more detailed analysis of MMs is provided in a recent research article.[66]

Implementation. The 100 spectra were randomly assigned to 10 groups of 10 spectra each. Volunteers familiar with spectroscopic concepts were recruited to remove the baselines from one or more of these 10 sets of spectra so that, in total, triplicates of each set were available for analysis. The shapes of the individual baselines, number of simulated peaks, and the peak positions were not revealed to the volunteers prior to their spectral subtractions.

Results and Discussion. The results were based on the means of the triplicate baseline estimates. The quantitative results are shown in Fig. 9a and comparative examples of baseline correction on two spectra (SNR 10, SBR 0.1, slope 0.1; SNR 100, SBR 10, slope 1.0) are shown in Fig. 9b. Computation times were moderately long as shown in Fig. 13.

This method produced results that were, generally speaking, very accurate and precise, confirming its general utility and eminent position as a standard of baseline removal. The method is characterized in greater detail in a recent research article in which its advantages and deficiencies are highlighted.[66]

Modes of Failure. These are reported in detail elsewhere.[66]

Suitability of Method for Automation. Not applicable.

METHODS REQUIRING EXPLICIT USE OF p, b, AND n.

Maximum Entropy Method (MEM). ***Background.*** The maximum entropy method is a way to reconstruct a spectrum in such a manner that it is as smooth as possible while its convolution with the instrumental blurring function conforms as closely as possible to the original spectrum. The smoothness of the reconstruction, governed by its entropy, and the closeness of its convolution with the blurring function to the original spectrum, governed by the χ^2 value, are traded off against each other with a regularization parameter λ to find the optimum reconstruction. The MEM variant discussed here implements separate reconstructions of background and signal peaks.[68] The background reconstruction is convolved with a blurring function several times wider than the blurring function used for the signal peaks to ensure that its features are broad and slowly varying. The sum of the background and signal reconstructions is then compared to the original spectrum for fidelity. Possibly the first use of this method for baseline removal was that by Durman and Wood in 1988.[113] It should also be noted that Bayesian methods are in some respects related to MEMs.[65,73–77]

Theory. Durman and Wood[113] model the data as in Eq. 1 and Phillips and Hamilton[68] as in a modified version of Eq. 3:

$$\mathbf{y} = \mathbf{s}*\mathbf{p} + \mathbf{b}*\mathbf{q} + \mathbf{n} \tag{24}$$

where **q** is a broad blurring function with a width of 5 times or more that of the blurring function **p**. The signal **s** and baseline **b** are then separately reconstructed in such a way that their entropy is maximized (negative entropy minimized) and that, after convolving with the blurring function(s), the added signal and baseline $\hat{\mathbf{y}}$ conform to the measured data **y** within acceptable limits. Phillips and Hamilton[68] applied their method to absorption spectra, in which case all values of **s** are negative. MEM involves a trade-off between the smoothness constraints imposed by the (negative) entropy function

$$S = \left[\sum_{i=1}^{n} s_i \ln\left(\frac{s_i}{\varepsilon}\right) + \sum_{i=1}^{n} b_i \ln\left(\frac{b_i}{\varepsilon}\right)\right] \tag{25}$$

where ε is an arbitrarily small constant, and the demand that the reconstructed spectrum $\hat{\mathbf{y}}$ agree with the observed data within

$$\chi^2 = \sum_{i=1}^{k} \frac{(y_i - \hat{y}_i)^2}{\sigma^2} \leq k \tag{26}$$

where k is the number of points in the spectrum and σ is based on **n**. Computation then proceeds by minimization of a weighted sum of Eqs. 25 and 26 ($\chi^2 + \lambda S$) where the weight (Lagrange multiplier or regularization parameter) $0 < \lambda < \infty$ also has to be optimized. Note that several formulations of entropy functions exist.[114,115]

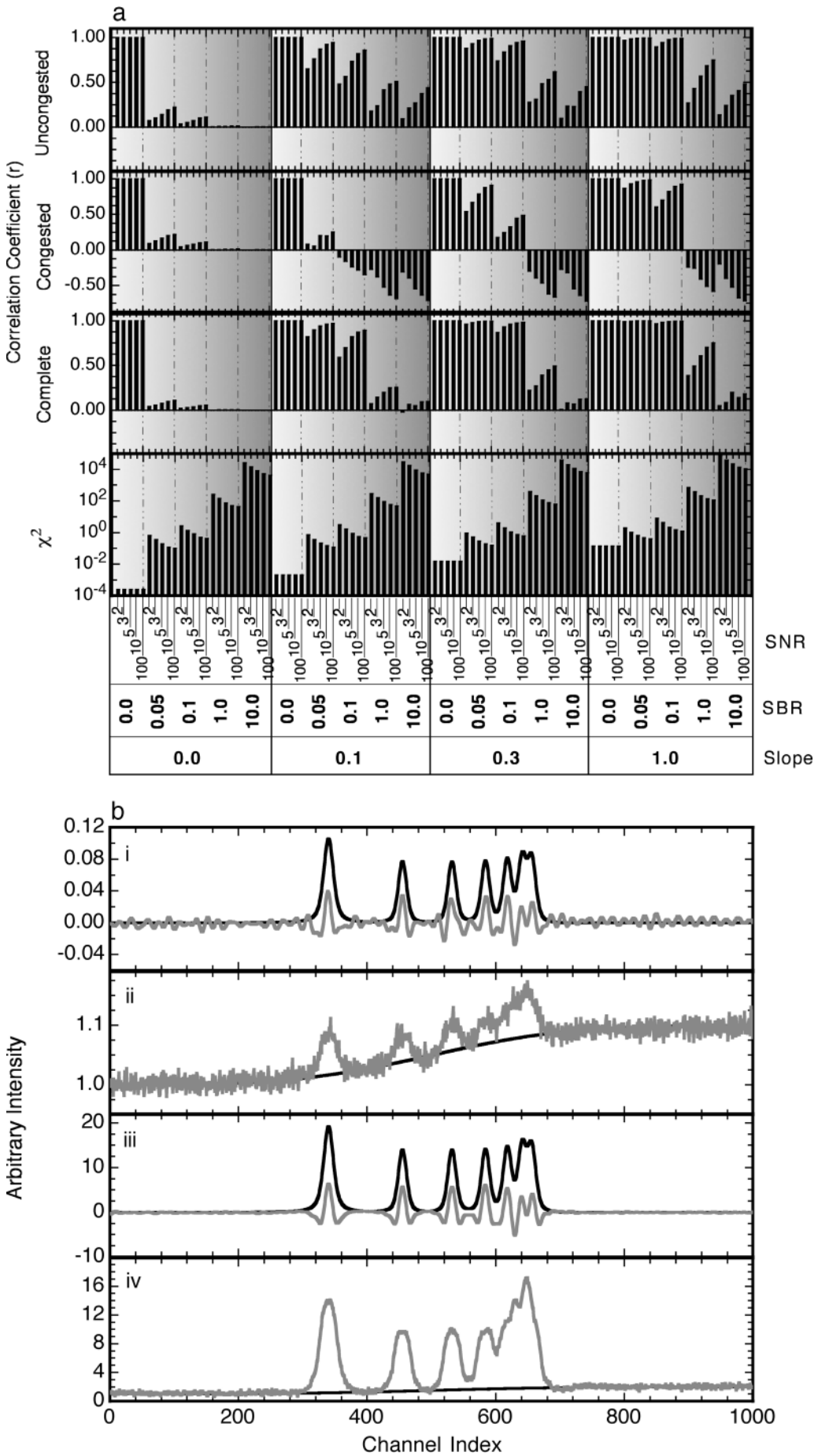
Implementation. We implemented this method with a three-point entropy window analogous to Greek et al.,[116] 15-point and 75-point Gaussian blurring functions for the signal and background components of the spectra, respectively, and used a λ (background) of 1000 and a λ (signal) of 0.01. The Levenberg–Marquardt[94] method was used for optimization.

Results and Discussion. The quantitative results are shown in Fig. 10a and comparative examples of baseline correction on two spectra (SNR 5, SBR 10, slope 1.0; SNR 100, SBR 10, slope 1.0) are shown in Fig. 10b. The results on those spectra processed indicated that baseline estimation was difficult in all regions, especially in more congested spectral regions. Computation times were very long as shown in Fig. 13 and precluded effective optimization.

This method requires that the instrumental blurring function be determined or known. It is a very slow meth-

←

FIG. 10. Baseline removal using maximum entropy regularization. (**a**) Correlation coefficient between theoretical and estimated baselines for the uncongested, the congested, and the total spectral regions of given spectra. Also shown is the χ^2 between theoretical and estimated baselines for the total spectral region. (**b**) (***i***) Baseline removed (gray trace) and initial input spectrum (less noise and true baseline, black trace) for SNR = 5, SBR = 10, and baseline slope = 1.0; (***ii***) estimated baseline (gray trace) and input baseline (black trace), SNR, SBR, and baseline slope as for (i); (***iii***, ***iv***) as for (i) and (ii) with SNR = 100, SBR = 10, and baseline slope = 1.0. Note: due to the computational demands of this method, only six spectra were processed.


a
Correlation Coefficient (r)
Uncongested
Congested
Complete
1.00
0.50
0.00
-0.50
χ^2
10^4
10^2
10^0
10^{-2}
10^{-4}
SNR
SBR
Slope
0.0
0.05
0.1
1.0
10.0
0.0
0.1
0.3
1.0
b
Arbitrary Intensity
i
ii
iii
iv
0.12
0.08
0.04
0.00
-0.04
1.1
1.0
20
10
0
-10
16
12
8
4
0
0
200
400
600
800
1000
Channel Index

od due to its computational load and it made optimization of the various parameters impractical. It is therefore likely that better results could be achieved as computation systems improve. An important consideration is that the potential for deconvolution and smoothing is offered by this method along with baseline removal.[116]

Modes of Failure. Congested spectra hinder optimization and convergence failure may occur. Since the same general approach is used to estimate both the signal and baseline components of the data, the baselines under peaks are generally overestimated. This difficulty can be countered by using a broader blurring function for the background, but at the cost of losing background estimation accuracy. The selection of the background blurring function size relative to the peak blurring function size (a factor of 5 was used here) will therefore affect the outcome of this method. Selecting too large a stopping tolerance will lead to early termination of the procedure and poor results while too low a stopping tolerance will increase processing times and possibly lead to convergence failure. The choice of the optimization method (e.g., Levenberg–Marquardt, conjugate gradient descent, etc.) and the selection or determination of regularization parameters (Lagrange multipliers) may also affect the accuracy of the results and the time required to achieve convergence.

Suitability of Method for Automation. Due to its computational load and the difficulty in determining appropriate parameters the MEM is not very attractive, as it currently stands, for automation. However, this assessment may have to be revised in the future as computation equipment becomes more powerful.

METHODS REQUIRING FREQUENCY INFORMATION.

Fourier Transform Method (FTM). ***Background.*** The Fourier transform is used to decompose a spectrum into a sum of sine and cosine waves of different frequencies. A plot of amplitude as a function of frequency will reveal which frequencies are contributing most to the spectrum. Hence, the low (background), middle (signal), and high frequencies (noise) of the spectrum can be observed and separated by retaining only those frequency components contributing to the signal and discarding the others. One of the first applications of Fourier methods for baseline correction was that of Atakan et al. in 1980.[18] Other variants and implementations have appeared[34,42,45,46,117–119] and we are following here that by Mosier-Boss et al.[61]

Theory. Fourier transform methods assume that a given spectrum (or dataset) can be decomposed into a series of frequency varying sine and cosine functions. Fourier transformation of the initial spectrum yields the frequency components of that spectrum. Fourier transformed spectral data can then be used to discriminate between baseline (low frequency), signal (mid frequency), and noise (high frequency) components within the given dataset. These frequency components can be filtered by application of a low-pass, band-pass, or high-pass filter to the Fourier domain spectral data to eliminate unwanted spectral components. For example, application of a low-pass filter suppresses spectral noise (high frequency components) while a high-pass filter suppresses spectral baseline (low frequency components). A band-pass filter can be utilized in similar fashion.

The Fourier transform $Y(f)$ of a spectrum $y(t)$ is given by

$$Y(f) = \int_{-\infty}^{\infty} y(t)e^{-i2\pi ft}\,\mathrm{d}t \tag{27}$$

and the inverse transform is written as

$$y(t) = \int_{-\infty}^{\infty} Y(f)e^{i2\pi ft}\,\mathrm{d}t \tag{28}$$

The Fourier transformed spectral data can be filtered by application of a suitable Fourier domain filter,

$$Y'(f) = c(x)Y(f) \tag{29}$$

where $c(x)$ are the bandpass filter coefficients and $Y'(f)$ is the filtered Fourier domain signal. Recovery of the filtered spectrum proceeds via inverse transformation (Eq. 28) of $Y'(f)$. It is noted that an added benefit of Fourier transform methods is the potential for spectral deconvolution.

Implementation. We followed the approach by Mosier-Boss et al.[61] and applied a band pass filter to remove the baseline and noise components of the spectrum. Filter coefficients were determined by:

$$c(x) = \frac{x}{N}, \qquad 0 < x \le lN \tag{30}$$

$$c(x) = \frac{k}{l}\left(\frac{x - lN}{N}\right), \qquad lN < x < hN \tag{31}$$

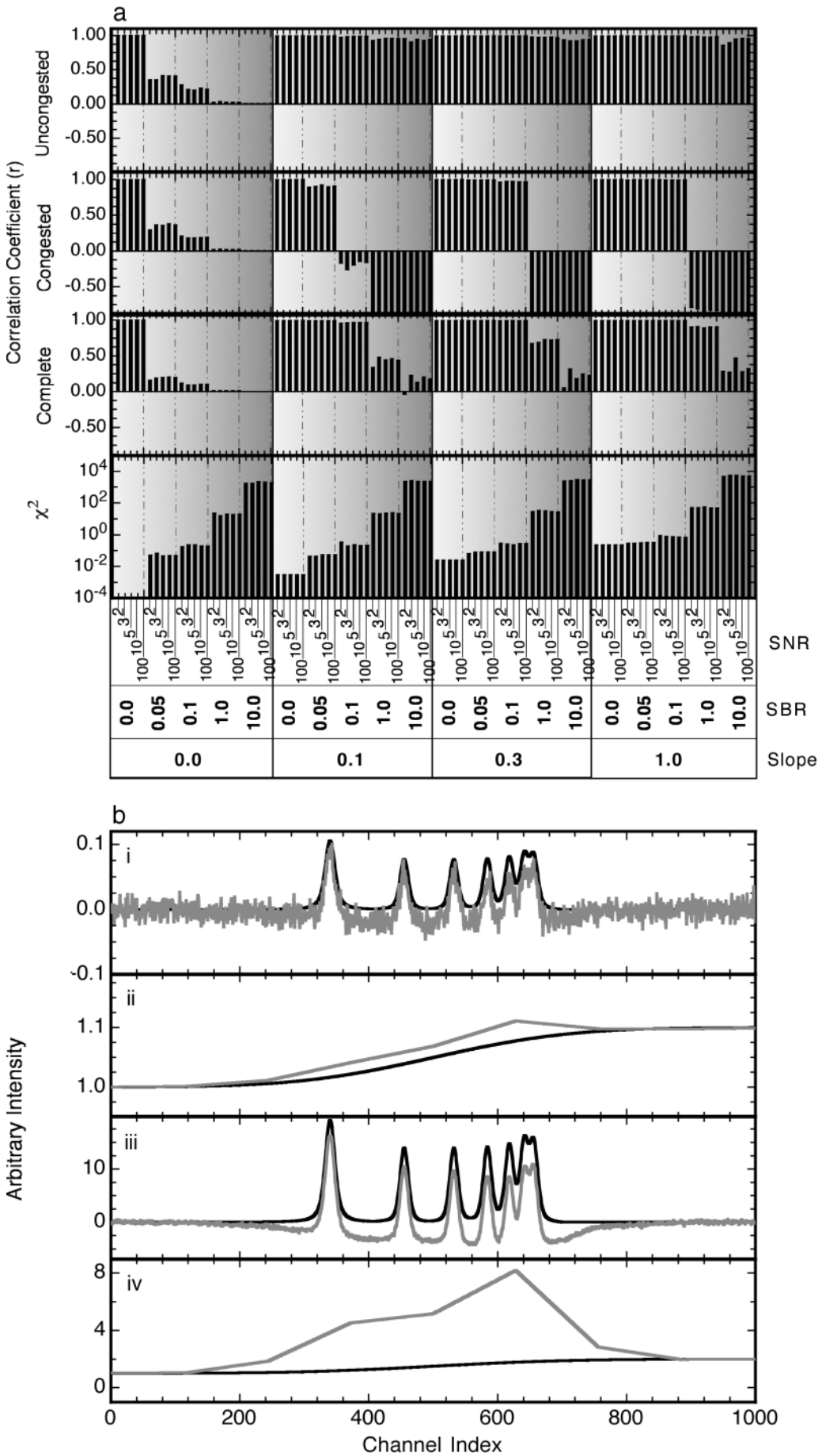
$$c(x) = \frac{h}{1 - h}\left(\frac{x - hN}{N}\right), \qquad hN < x < N \tag{32}$$

where $c(x)$ are the band-pass coefficients varying from 0 to 1, N is the number of points in the original spectrum, k is a filter coefficient, and l and h are constants determining the low and high frequency regions of the filter. In this implementation, $k = 0.69$, $l = 0.01$, and $h = 0.07$.

Results and Discussion. The quantitative results are shown in Fig. 11a and comparative examples of baseline correction on two spectra (SNR 10, SBR 0.1, slope 0.1; SNR 100, SBR 10, slope 1.0) are shown in Fig. 11b. The results indicated that baseline estimation was difficult. Computation times were short as shown in Fig. 13.

←

FIG. 11. Baseline removal using Fourier transforms. (**a**) Correlation coefficient between theoretical and estimated baselines for the uncongested, the congested, and the total spectral regions of given spectra. Also shown is the χ^2 between theoretical and estimated baselines for the total spectral region. (**b**) (***i***) Baseline removed (gray trace) and initial input spectrum (less noise and true baseline, black trace) for SNR = 10, SBR = 0.1, and baseline slope = 0.1; (***ii***) estimated baseline (gray trace) and input baseline (black trace), SNR, SBR, and baseline slope as for (i); (***iii, iv***) as for (i) and (ii) with SNR = 100, SBR = 10, and baseline slope = 1.0.


a
Correlation Coefficient (r)
Uncongested
Congested
Complete
χ²
SNR
SBR
Slope
0.0
0.05
0.1
1.0
10.0
0.0
0.1
0.3
1.0
b
Arbitrary Intensity
i
ii
iii
iv
Channel Index

Although information such as the sine and cosine basis functions is required, in practice these are inherently specified by the FTM. However, low and high frequency cutoffs also have to be determined as well as filter parameters that govern the filter coefficients obtained. Some of these variables such as low and high frequency cutoffs may change from spectrum to spectrum, and means to determine them must be devised for automation.

Modes of Failure. The selection of filter parameters (Eqs. 30 through 32) that will effectively discriminate between baseline and signal can be difficult.[61] A further drawback of the method is that it is sensitive to noise in the input data.[61] Improper selection of filter parameters can lead to considerable distortions in the estimated baseline: if the low frequency cutoff is too low, some residual amount of baseline will remain in the baseline-corrected spectra; if the low frequency cutoff is too high, some of the signal peak frequency components will be removed along with those of the baseline and the signal peaks will be reduced. Likewise, the amount of noise in the spectrum will depend on the high frequency cutoff selected. In addition, a 'tuning constant' k used in Eq. 31 governs the transition between low and high frequency regions and has a more complex effect on the results: lower values lead to reduced peak heights and reduced baseline artifacts in the corrected spectra while higher values produce more appropriate peak intensities and also more pronounced artifacts in the corrected spectra. Although we have not done so, the reader is reminded that Eq. 31 can be altered to effect a different type of transition between frequency regions, hence producing different results.

Suitability of Method for Automation. The number of parameters to optimize and the difficulty in doing so make this method generally not practical for automated procedures, except possibly under well-characterized (e.g., low noise) conditions. As noted above, this method offers the potential for simultaneous deconvolution and noise removal that could be utilized as part of another automated procedure.

Wavelet Transform Method (WTM). ***Background.*** When representing a given function as a sum of sinusoids such as in the FTM, quite a large number of sines and cosines of increasing frequency may have to be used to cancel each other out in featureless regions of the spectrum. Instead, a given function could be represented as a sum of wavelets. Wavelets are localized in space and also in frequency.[114] For example, the second derivative of the Gaussian probability density function is a wavelet termed the 'Mexican hat'. The spatial localization of wavelets is a unique feature not found in their sinusoidal counterparts (sine, cosine). This makes wavelets more suited to representing a large class of signals. For example, spectra can be represented with far fewer wavelets than sinusoids. Wavelets also allow for a representation of spectra at different resolution scales[115] and this is of interest here since it offers the potential to discriminate between high frequency noise, low frequency baseline, and mid-frequency signals at these different levels of resolution.

Wavelets have been used to estimate baselines in capillary electrophoresis data by iterative peak stripping and wavelet application.[104] This method is analogous to the SRMs, but employs wavelet transforms instead of moving averages, polynomials, or splines to estimate the baseline. The effects of wavelet shrinking schemes on estimating baselines in synthetic and real data at different resolution scales have also been investigated.[120,121]

Theory. Wavelet decomposition proceeds by the recursive application of a matrix of wavelet filter coefficients to a spectrum modeled as in Eq. 3. The spectrum must have 2^n resolution elements (zero padding of data may be necessary). The wavelet filter coefficients constitute quadrature mirror filters, i.e., the first filter passing smooth data and the second passing rough or detail data:

$$\begin{pmatrix} c_0 & c_1 & c_2 & c_3 & & & & & & & \\ c_3 & -c_1 & c_2 & -c_0 & & & & & & & \\ & & c_0 & c_1 & c_2 & c_3 & & & & & \\ & & c_3 & -c_1 & c_2 & -c_0 & & & & & \\ \vdots & & & & & & & & & & \\ & & & & \cdots & & c_0 & c_1 & c_2 & c_3 & \\ & & & & & & c_3 & -c_1 & c_2 & -c_0 & \\ c_2 & c_3 & & & & & & & c_0 & c_1 & \\ c_1 & -c_0 & & & & & & \cdots & c_3 & -c_2 & \end{pmatrix} \begin{pmatrix} y_0 \\ y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} s_1 \\ d_1 \\ s_2 \\ d_2 \\ \vdots \\ s_{n/2} \\ d_{n/2} \end{pmatrix} \quad (33)$$

In Eq. 33 the wavelet smooth-pass filter coefficients are shown in row 1 and that of the rough-pass filter in row 2 of the wavelet filter coefficient matrix (c matrix). The requirements of these filter coefficients and how to derive them can be found elsewhere.[114] The spectral data are denoted by the vector **y** and the smooth and detail wavelet coefficients by the s_i and d_i coefficients, respectively.

A spectrum vector, **y**, processed in this way gives rise to a vector of interleaved wavelet coefficients that alternately specify the smooth and detail (i.e., non-smooth) components of the original spectrum as shown in Eq. 33. If these wavelet coefficients are sorted into smooth and detail sections, the smooth coefficients would amount to one-half the size of the original spectrum vector. Recur-

←

FIG. 12. Baseline removal using wavelet transforms. (**a**) Correlation coefficient between theoretical and estimated baselines for the uncongested, the congested, and the total spectral regions of given spectra. Also shown is the χ^2 between theoretical and estimated baselines for the total spectral region. (**b**) (***i***) Baseline removed (gray trace) and initial input spectrum (less noise and true baseline, black trace) for SNR = 10, SBR = 0.1, and baseline slope = 0.1; (***ii***) estimated baseline (gray trace) and input baseline (black trace), SNR, SBR, and baseline slope as for (i); (***iii, iv***) as for (i) and (ii) with SNR = 100, SBR = 10, and baseline slope = 1.0.

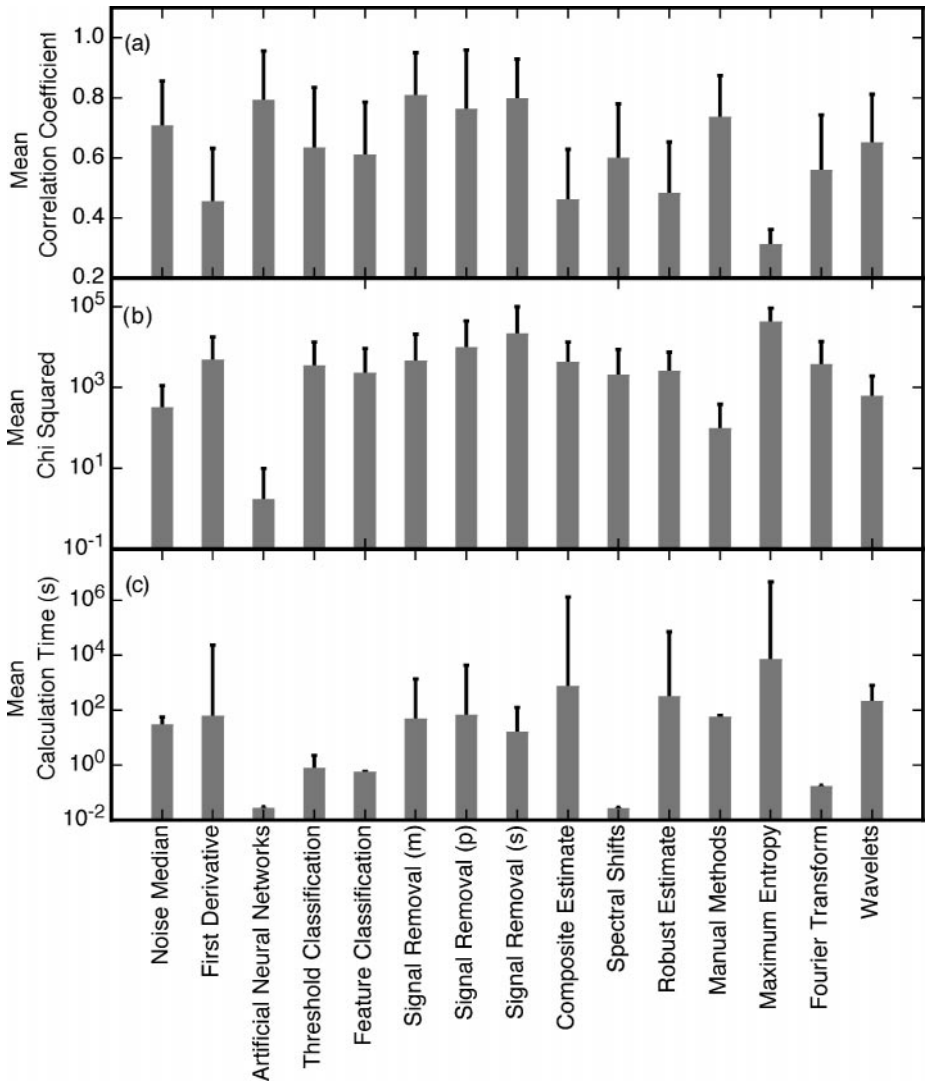


FIG. 13. Global comparisons for baseline estimation methods based on all spectra. (**Top**) Mean correlation coefficient and variance. (**Middle**) Mean χ^2 and standard deviation. Note log scale. (**Bottom**) Mean time and standard deviation to perform one baseline removal calculation. Note log scale. Note also that Fig. 13 includes results from methods that were evaluated and are included here for completeness, but are not explicitly described in the text due to space considerations. These are the feature-based classification method,[96] the robust estimation method,[1] and the polynomial and spline variants of the signal removing method (see SRM). Detailed results, similar to those reported for the other methods, for these methods not included here are available from the authors.

sivity is effected by applying the (adjusted) wavelet filter coefficient matrix to the smooth wavelet coefficients only. This results in another vector with interleaved 'smooth' and 'detail' coefficients, with the new number of smooth coefficients now only one quarter of the size of the original spectrum vector. This procedure is repeated until only two smooth (i.e., the smoothest) wavelet coefficients remain. At each step, the detail wavelet coefficients are retained and added to those from previous steps. Using the inverse wavelet filter coefficient matrix, the original spectrum then can be reconstructed from its smooth and detail wavelet coefficients. Note that the coefficients of the wavelet itself are called the *wavelet filter coefficients* while the elements of the filtered 'spectra' are called *wavelet coefficients*. As described above, the wavelet filter coefficients are repeatedly applied to the (smooth) wavelet coefficients until only two remain. The last two wavelet coefficients are the *mother function coefficients*.

At some point during the iterated application of the wavelet filter coefficients, the resulting smooth coefficients may closely approximate the baseline. If the process is terminated at this level of resolution, the baseline could then be estimated from the smooth wavelet coefficients.

Implementation. We implemented wavelet transforms using the Daubechies[122] wavelets of order 4 and terminated the process at resolution 2^6 to obtain the baseline estimate. The baselines were then estimated by linear interpolation from the smooth coefficients obtained at this resolution level, which were fewer (32) than the number of points in the padded spectra (2048).

Results and Discussion. The quantitative results are shown in Fig. 12a and comparative examples of baseline correction on two spectra (SNR 10, SBR 0.1, slope 0.1; SNR 100, SBR 10, slope 1.0) are shown in Fig. 12b. The results indicated that baseline estimation was less successful with high SNR spectra and in congested regions

of spectra. Computation times were comparatively moderately long as shown in Fig. 13.

The drawbacks of this method are the need to establish the appropriate resolution level to obtain the best baseline estimate, the need to pad the data, and the need to obtain extraneous information such as a suitable set of wavelets (i.e., filter coefficients). Padding generally increases the computation time. This method has the advantage that the first few applications of the wavelet filter coefficients will filter out high frequency noise. In addition, thresholding the wavelet coefficients permits both denoising and baseline estimation to be accomplished.

Modes of Failure. Inappropriate wavelet and resolution level choices are detrimental to baseline estimation. Some wavelets may be better suitable for some types of spectra; however, we have here only implemented the Daubechies wavelets. The selection of resolution level involves a trade-off between the degree to which the curvature of the baseline could be approximated and the degree to which residual effects of peaks, especially wider and/or overlapping peaks, can be countered. Finally, the selection of thresholds may affect the results.

Suitability of Method for Automation. This method has further potential for automated implementation (i.e., different from that related to the SRMs[104]). Difficulties are the mentioned selection of a suitable wavelet, automation of the stopping criterion (i.e., the proper level of resolution to represent the baseline well), and thresholding for the wavelet coefficients.

GENERAL DISCUSSION

In an attempt to provide the reader with an estimate of the relative performance of any one of these methods, and hence to facilitate initial method selection, we have compared them on the bases of their overall consistency, accuracy, reproducibility, processing times, and suitability for automation. This overall comparison should benefit those wishing to automate baseline removal in data sets spanning a wide range of possible baseline topographies or where the possible range of topographies is unknown. Where the spectral attributes such as noise level and baseline shape vary within narrower ranges, a better selection of a suitable method could be made by consulting the results for each individual method and comparing these results based on the most relevant spectral attributes. Examples are also shown that facilitate a direct comparison of each method's performance using the same test spectra for each. Although most methods tend to have difficulty under conditions of high noise and spectral congestion, the reasons why and how different methods fail under these (and other) conditions are often specific to the method. The reader is therefore referred to the individual method sections for discussions of the particular difficulties encountered by the different methods as well as their ease of automated implementation. Nevertheless, a general discussion of modes of failure is also given below.

Baseline Correction Method Consistency. When collected data exhibit large variability in SNR, SBR, and baseline slopes and shapes, the ability of a BCM to consistently provide an accurate estimate of the baseline given these differing initial data becomes important. An indication of the BCM's ability to consistently estimate the baseline can be obtained from comparing correlation coefficients and χ^2 values for all tested SNR, SBR, and baseline slopes. This comparison has been made for the complete set of test spectra (Fig. 13). Comparisons between methods based on specific spectral attributes, rather than the overall comparison in Fig. 13, can be made by selecting from subfigure (a) of each given method's results figure those spectral attributes of interest (i.e., slope, SBR, SNR) and then comparing these results. Note that Fig. 13 also includes results from methods that were evaluated but are not explicitly described in the text due to space considerations. These are the feature-based classification method,[96] robust linear estimation method,[1] and the polynomial and spline variants of the signal removing method (see SRM). Detailed results for these methods are available from the authors.

Of the methods tested, ANNs and SRMs proved to be the most consistent under the tested conditions. These methods can be expected to perform well in situations where automation is desirable and measured spectra are dissimilar from one another (e.g., two-dimensional Raman spectral mapping). Note that ANNs are adaptive methods and may have to be retrained for data sets where the baseline features are quite different from those encountered in previous sets. The most inconsistent methods tested were composite baseline and robust estimation methods. Care must be exercised with these techniques, as their accuracy depends heavily on initial input spectra.

Baseline Correction Method Accuracy. The overall accuracy of a BCM in estimating a baseline is characterized by an overall high correlation coefficient and low χ^2 (for all SNR, SBR, and baseline slopes, see Fig. 13). Methods that scored high on overall accuracy were, again, ANNs and SRMs. These methods would be suitable for situations where, for example, peak quantification post baseline subtraction is important. The least accurate methods tested were the FDM, MEM, and the FTM. The WTM can produce moderately accurate results depending on the input spectrum. The CBM can produce accurate results in uncongested regions of the spectrum but fails to accurately model the baseline in congested spectral regions.

Baseline Correction Method Reproducibility. A BCM should estimate a baseline with high precision when performing baseline correction repeatedly on the same input data. Furthermore, baseline estimates should be similar for input data that vary only slightly (e.g., similar SNR, SBR, and baseline slopes). Hence, correlation coefficients and χ^2 values for baseline estimates should not vary significantly (e.g., fluctuate between high positive and high negative values) between spectra of similar SNR, SBR, and baseline slopes.

The three peaks in the uncongested region of the test spectra each have independent noise. Therefore, the results shown in the top panel ('Uncongested') of each method's subfigure (a) could be viewed as a representation of an average performance over three runs of an individual peak, each run with independent added noise. Since some of the conditions are highly similar, comparing the performance of different methods on these highly similar spectra gives a good indication of the reproducibility of the method. For example, in the top panel of

Fig. 12a under conditions of slope 0.3 and SBR 10.0, there is very little difference between the correlation coefficients (and χ^2 values, bottom panel) for spectra with SNRs of 2 and 3. This also holds true for spectra with SNR 2 and 3 under other conditions of slope and baseline. In contrast, more variability is evident in the top panel of Fig. 7a for spectra with SNRs of 2 and 3 under conditions of slope 0.3 and SBR 10.0. Indeed, it is evident, by comparing other spectra with similar SNRs, that the CBM is more susceptible to the influence of noise than the WTM. Inspection of Figs. 2–12 indicates that the SSM, FTM, and WTM are highly reproducible methods, whereas the TBC is the least reproducible. Several methods are ''reasonably reproducible'' and have but one or two outlying baseline estimates. These methods include ANNs and SRMs.

Baseline Correction Method Computation Time. If a large number of spectra are to be processed, or the removal of the baseline is but one step in a larger analysis scheme, the required time to process a given spectrum may become important. Figure 13 shows the average time required to estimate the baseline for the tested BCMs. Particularly quick methods include the TBC, SSM, and FTM (all under 1 s per baseline estimate). ANNs are also fast (under 1 s) in estimating the baseline; however, this time does not include the time required to train the network. Training times may vary greatly, depending on the initial data and network architecture (e.g., number of hidden nodes). The method requiring by far the longest computation time is the MEM. In fact, times were sufficiently long (~3 h per spectrum) to render this technique impractical for baseline subtraction purposes alone. This method may achieve greater utility if simultaneous spectral smoothing and line narrowing or resolution enhancement are desired. Other methods requiring moderately long computation times include the CBM (~11 min per spectrum) and the WTM (~3 min per spectrum). Readers interested in those factors that affect the computation times for the different methods should consult the Results and Discussion subsections of the individual methods.

Baseline Correction Method Suitability for Complete Automation. When processing large numbers of spectra, it is often beneficial to have a completely automated baseline removal method. Generally, suitable automated methods have few parameters that require optimization or have parameters that are possible to optimize in a straightforward manner. In this regard the NMM, TBC, SRM, and SSM are suitable for full automation. Methods with some potential in this regard, especially in specialized applications, are the ANN, MEM, and WTM. Methods such as the CBM, which have peak edge detection requirements integrated into the BCM algorithm, are more problematic to automate. Difficult parameter estimation also renders the FTM difficult to automate completely. From a practical point of view, any method that produces results within the general range spanned by manual methods and that has the potential for complete automation ought to be acceptable. Specific reasons that aid or hinder the automated implementation of the different methods are given for each method in their respective Suitability of Method for Automation subsection.

Baseline Correction Methods for Best Performance. Some methods, although perhaps not very satisfactory overall, may nevertheless perform exceptionally well under specific conditions. Where the quantitative processing of spectra is required, utilizing such information may be valuable for selecting an appropriate BCM. We have therefore provided individual FOM for each method as an aid to determining the most suitable BCM given specific objectives and characteristics of the data sets to be processed. When selecting an appropriate BCM, keep in mind the issue of reproducibility as discussed above.

Baseline Correction Methods Modes of Failure. Equation 1 indicates that the decomposition of a sum lies at the heart of baseline correction. This implies that any method of baseline correction must have a degree of additional information about either the signal or the baseline component of this sum to succeed. The various methods differ then in how they use, extract, or make assumptions about aspects of this additional information. Therefore, at a fundamental level, failure of baseline correction reflects an inability to tease the sum of signal and baseline apart, regardless of convolution and noise complications. All of the methods considered here use one or more explicit, and sometimes implicit, parameters and hence will fail or succeed, in general, to the extent that these parameters can be tuned or selected to discriminate between signal and baseline. Sometimes these parameters are based on explicit assumptions made regarding the signal and its shape (e.g., fitting to Lorentzian peaks in the FDM), or the baseline and its shape (e.g., fitting a polynomial to the spectrum in the SRM) or on implicit assumptions regarding the signal (e.g., the width of the processing window in the NMM), or the baseline (e.g., the broad blurring function used for the baseline in the MEM). In the foregoing cases, the assumptions are of a global nature, but sometimes assumptions are made about local regions of the baseline (e.g., interpolating under a peak from a nearby region of the baseline in the CBM). Furthermore, assumptions about aspects of the signal or baseline can be made in the time and/or frequency domains.

Empirical parameter selection should therefore be seen as the implementation of prior knowledge, the extraction of knowledge from the data at hand, or the tacit making of assumptions about the signal and/or baseline. Hence, methods with fewer parameters, parameters that can readily be determined, and parameters that can discriminate effectively between signal and baseline, under a broad range of conditions, are the better candidates for automated implementation.

CONCLUSION

We have presented here a number of different baseline-correction methods, discussed their specific approaches to the problem of baseline-removal, their implementations, results, advantages and disadvantages, and suitability for automation. We have also compared them on overall measures of accuracy and efficiency (i.e., processing speed). We hope that this will give readers a sense of the variety of approaches that exist and enable them to expedite the selection of a suitable method for their particular problem. We also hope that these results will enable readers to fine-tune and extend chosen meth-

ods in order to obtain the best possible results on their data.

We have tried to optimize the different methods sufficiently to render them qualitatively compatible on the designed set of test data used here. Their relative performances on other data sets may differ. Nevertheless, we have covered a sufficiently wide range of conditions as to make these comparisons useful and permit extrapolation to relatively similar data sets.

Refinements and Hybrid Methods. Our results revealed particular difficulties encountered with some methods and may suggest ways to overcome these. For example, methods that rely on peak picking (e.g., FDM, CBM) will be difficult to implement in automated procedures unless peaks, and in some cases, their leading and trailing edges, can also be determined automatically. Therefore, a combination of such methods with those that implicitly perform peak identification (e.g., TBC) may prove satisfactory.

Another possibility is the generation of hybrid methods where two or more different methods are used successively or iteratively. The iterative application of signal removal and wavelet transforms is an example.[104] Taking the first derivative before applying a signal removal approach may constitute another hybrid method. Clearly, a range of potential refinements and hybrid methods exist and we hope that this comparison of baseline correction methods for automation will spur further development in this area.

ACKNOWLEDGMENTS

We thank all those volunteers who did manual baseline corrections for this work. We thank the Natural Sciences and Engineering Research Council of Canada for financial assistance and the Canadian Foundation for Innovation and the British Columbia Knowledge Development Foundation for funding the resources made available for this work through the UBC Laboratory for Molecular Biophysics.

1. A. F. Ruckstuhl, M. P. Jacobson, R. W. Field, and J. A. Dodd, J. Quant. Spectrosc. Radiat. Trans. **68,** 179 (2001).
2. J. Wilson and C. McInnes, J. Chromatogr. **19,** 486 (1965).
3. F. Bauman and F. Tao, J. Gas. Chromatogr. **5,** 621 (1967).
4. M. O. Deighton, IEEE Trans. Nucl. Sci. **NS16,** 68 (1969).
5. R. N. Hager, Jr., and R. C. Anderson, J. Opt. Soc. Am. **60,** 1444 (1970).
6. I. Fijii, T. Innouye, H. Muto, K. Onodera, and A. Tani, Analyst (Cambridge, U.K.) **94,** 189 (1969).
7. H. R. Ralston and G. E. Wilcox, "A computer method of peak area determinations from Ge(Li) gamma spectra", in *Proceedings of the international conference on modern trends in activation analysis II Gaithersburg, MD,* J. R. deVoe and P. D. LaFleur, Eds. (National Bureau of Standards, Washington, D.C., 1969), p. 1238.
8. B. Goldberg, J. Chromatogr. Sci. **9,** 287 (1971).
9. B. Groszwendt, Nucl. Instrum. Methods **93,** 461 (1971).
10. L. Hedegus and J. Petersen, J. Chromatogr. Sci. **9,** 551 (1971).
11. A. Robertson, W. V. Prestwich, and T. A. Kennett, Nucl. Instrum. Methods **100,** 317 (1972).
12. P. M. A. Shirley, Phys. Rev. B **5,** 4709 (1972).
13. F. Caesar and M. Klier, Chromatographia **7,** 526 (1974).
14. B. G. M. Vandeginste and L. de Galan, Anal. Chem. **47,** 2124 (1975).
15. Y. Kawarasaki, Nucl. Instrum. Methods **133,** 335 (1976).
16. P. Gans and J. B. Gill, Appl. Spectrosc. **31,** 451 (1977).
17. G. A. Pearson, J. Magn. Reson. **27,** 265 (1977).
18. A. K. Atakan, W. E. Blass, and D. E. Jennings, Appl. Spectrosc. **34,** 369 (1980).
19. Z. Hippe, A. Bierowska, and T. Pietryga, Anal. Chim. Acta **4,** 279 (1980).
20. E. F. G. Woerlee and J. C. Mol, J. Chromatogr. Sci. **18,** 258 (1980).
21. W. Dietrich and R. Gerhards, J. Magn. Reson. **44,** 229 (1981).
22. T. J. Kennett, W. V. Prestwich, and R. J. Tervo, Nucl. Instrum. Methods **190,** 313 (1981).
23. W. Kipiniak, J. Chromatogr. Sci. **19,** 332 (1981).
24. S. Steenstrup, J. Appl. Crystallogr. **14,** 226 (1981).
25. W. Westmeier, Nucl. Instrum. Methods **180,** 205 (1981).
26. G. I. Johansson, X-ray Spectrom. **11,** 194 (1982).
27. D. D. Burgess and R. J. Tervo, Nucl. Instrum. Methods, Phys. Res. **214,** 431 (1983).
28. P. M. A. Sherwood, "Data analysis in X-ray photoelectron spectroscopy", in *Practical surface analysis: by Auger and X-ray photoelectron spectroscopy,* D. Briggs and M. P. Seah, Eds. (John Wiley and Sons, New York, 1983), p. 445.
29. R. J. Tervo, T. J. Kennett, and W. V. Prestwich, Nucl. Instrum. Methods **216,** 205 (1983).
30. J. M. Daubenfeld, J. C. Boubel, and J. J. Delpuech, J. Magn. Reson. **62,** 195 (1985).
31. R. E. Klevitt, J. Magn. Reson. **62,** 551 (1985).
32. D. W. Osten and B. R. Kowalski, Anal. Chem. **57,** 908 (1985).
33. S. Glaser and H. R. Kalbitzer, J. Magn. Reson. **68,** 350 (1986).
34. G. Otting, H. Widmer, G. Wagner, and K. Wuthrich, J. Magn. Reson. **66,** 187 (1986).
35. H. A. Spaink, T. T. Lub, R. P. Otjes, and H. C. Smit, Anal. Chim. Acta **183,** 141 (1986).
36. K. N. Thomsen, J. N. Pedersen, and N. Pind, Anal. Chim. Acta **184,** 133 (1986).
37. P. Van Espen, K. Janssens, and J. Nobels, Chemom. Intell. Lab. Syst. **1,** 109 (1986).
38. I. L. Barsukov and A. S. Arseniev, J. Magn. Reson. **73,** 148 (1987).
39. D. G. Cameron and D. J. Moffatt, Appl. Spectrosc. **41,** 539 (1987).
40. C. K. Mann and T. J. Vickers, Appl. Spectrosc. **41,** 427 (1987).
41. J. N. Froning, M. D. Olson, and V. F. Froelicher, J. Electrocardiol. **21,** S149 (1988).
42. D. Marion and A. Bax, J. Magn. Reson. **79,** 352 (1988).
43. C. G. Ryan, E. Clayton, W. L. Griffin, S. H. Sie, and D. R. Cousens, Nucl. Instrum. Methods, Phys. Res. B **34,** 396 (1988).
44. S. Tougaard, Surf. Interface Anal. **11,** 453 (1988).
45. D. Marion and A. Bax, J. Magn. Reson. **83,** 205 (1989).
46. D. Marion, M. Ikura, and A. Bax, J. Magn. Reson. **84,** 425 (1989).
47. J. A. Maxwell, J. L. Campbell, and W. J. Teesdale, Nucl. Instrum. Methods, Phys. Res. B **43,** 218 (1989).
48. Z. Zolnai, S. Macura, and J. L. Markley, J. Magn. Reson. **80,** 60 (1988).
49. T. C. O'Haver and G. L. Green, Anal. Chem. **48,** 312 (1976).
50. W. F. Maddams, Appl. Spectrosc. **34,** 245 (1980).
51. D. J. Anderson and R. R. Walters, J. Chromatogr. Sci. **22,** 353 (1984).
52. D. D. Burgess, Nucl. Instrum. Methods, Phys. Res. A **221,** 593 (1984).
53. Y. C. Ling, T. J. Vickers, and C. K. Mann, Appl. Spectrosc. **39,** 463 (1985).
54. B. Vekemans, K. Janssens, L. Vincze, F. Adams, and P. Vanespen, Spectrochim. Acta, Part B **50,** 149 (1995).
55. T. J. Vickers, R. E. Wambles, and C. K. Mann, Appl. Spectrosc. **55,** 389 (2001).
56. D. Skoog, *Principles of Instrumental Analysis* (Saunders, Philadelphia, 1985), p. 65.
57. M. S. Epstein, T. C. Rains, and T. C. O'Haver, Appl. Spectrosc. **30,** 324 (1976).
58. R. K. Skogerboe, P. J. Lamothe, G. J. Bastiaans, S. J. Freeland, and G. N. Coleman, Appl. Spectrosc. **30,** 495 (1976).
59. D. I. Hoult, C. N. Chen, H. Eden, and M. Eden, J. Magn. Reson. **51,** 110 (1983).
60. P. M. Henrichs, J. M. Hewitt, and R. H. Young, J. Magn. Reson. **69,** 460 (1986).
61. P. A. Mosier-Boss, S. H. Lieberman, and R. Newbery, Appl. Spectrosc. **49,** 630 (1995).
62. S. E. J. Bell, E. S. O. Bourguignon, and A. Dennis, Analyst (Cambridge, U.K.) **123,** 1729 (1998).
63. S. E. J. Bell, E. S. O. Bourguignon, A. C. Dennis, J. A. Fields, J. J. McGarvey, and K. R. Seddon, Anal. Chem. **72,** 234 (2000).
64. P. Matousek, M. Towrie, and A. W. Parker, J. Raman Spectrosc. **33,** 238 (2002).
65. J. Padayachee, V. Prozesky, W. von der Linden, M. S. Nkwinika,

and V. Dose, Nucl. Instrum. Methods, Phys. Res. B **150,** 129 (1999).
66. A. Jirasek, G. Schulze, M. M. L. Yu, M. W. Blades, and R. F. B. Turner, Appl. Spectrosc. **58,** 1488 (2004).
67. C. D. Brown, L. Vega-Montoto, and P. D. Wentzell, Appl. Spectrosc. **54,** 1055 (2000).
68. A. J. Phillips and P. A. Hamilton, Anal. Chem. **68,** 4020 (1996).
69. E. J. Hasenoehrl, J. H. Perkins, and P. R. Griffiths, Anal. Chem. **64,** 705 (1992).
70. T. Iwata and J. Koshoubu, Appl. Spectrosc. **48,** 1453 (1994).
71. J. K. Holland, E. K. Kemsley, and R. H. Wilson, J. Sci. Food Agric. **75,** 391 (1997).
72. G. Balcerowska and R. Siuda, Vacuum **54,** 195 (1999).
73. W. Von der Linden, V. Dose, and R. Fischer, "How to separate the signal from the background", in *Maxent96: Proceedings of the maximum entropy conference,* M. Sears, V. Nedeljkovic, N. E. Pendock, and S. Sibisi, Eds. (NMB Printers, Port Elizabeth, South Africa, 1996), p. 146.
74. W. Von der Linden, V. Dose, J. Padayachee, and V. Prozesky, Phys. Rev. E **59,** 6527 (1999).
75. R. Fischer, V. Dose, M. Hanson, and W. von der Linden, "Bayesian background estimation", in *Bayesian Inference and Maximum Entropy Methods,* C. R. Smith, J. T. Rychert, and G. J. Erickkson, Eds. (AIP, Melville, NY, 2001), p. 193.
76. R. Fischer and V. Dose, "Analysis of mixtures in physical spectra", in *Monographs of official statistics: Bayesian methods,* E. I. George, Ed. (Eurostat, Luxembourg, 2001), p. 45.
77. R. Fischer and V. Dose, "Physical mixture modeling with unknown number of components", in *Bayesian Inference and Maximum Entropy Methods in Science and Engineering,* R. L. Fry, Ed. (AIP, Melville, NY, 2002), p. 143.
78. C. G. Ryan, D. R. Cousens, S. H. Sie, W. L. Griffin, G. F. Suter, and E. Clayton, Nucl. Instrum. Methods, Phys. Res. B **47,** 55 (1990).
79. M. A. Raso, J. Tortajada, and F. Accion, Comput. Chem. **15,** 29 (1991).
80. H. Tokutaka, N. Ishihara, K. Nishimori, S. Kishida, and K. Isomoto, Surf. Interface Anal. **18,** 697 (1992).
81. C. G. Ryan, Nucl. Instrum. Methods, Phys. Res. B **181,** 170 (2001).
82. C. G. Ryan, E. van Achterbergh, C. J. Yeats, S. L. Drieberg, G. Mark, B. M. McInnes, T. T. Win, G. Cripps, and G. F. Suter, Nucl. Instrum. Methods, Phys. Res. B **188,** 18 (2002).
83. M. S. Friedrichs, J. Biomol. NMR **5,** 147 (1995).
84. D. M. Lewis and P. C. Chatwin, Boundary-Layer Meteorol. **72,** 53 (1995).
85. L. Keselbrener, M. Keselbrener, and S. Akselrod, Med. Eng. Phys. **19,** 481 (1997).
86. T. R. Griffiths, K. King, H. V. A. Hubbard, M. J. Schwingweill, and J. Muellemeestre, Anal. Chim. Acta **143,** 163 (1982).
87. T. V. Karstang and O. M. Kvalheim, Anal. Chem. **63,** 767 (1991).
88. D. Ebenezer and V. Krishnamurthy, J. Biomed. Eng. **15,** 132 (1993).
89. D. Ebenezer and F. F. Papa, Med. Eng. Phys. **16,** 273 (1994).
90. A. Pullia and G. Ripamonti, Nucl. Instrum. Methods, Phys. Res. A **376,** 82 (1996).
91. G. Ripamonti and A. Pullia, Nucl. Instrum. Methods, Phys. Res. A **382,** 545 (1996).
92. A. Pullia and G. Ripamonti, Nucl. Instrum. Methods, Phys. Res. A **391,** 301 (1997).
93. A. Savitzky and M. J. E. Golay, Anal. Chem. **36,** 1627 (1964).
94. D. W. Marquardt, J. Soc. Ind. Appl. Math. **11,** 431 (1963).
95. W. Dietrich, C. H. Rudel, and M. Neumann, J. Magn. Reson. **91,** 1 (1991).
96. A. Rouh, M. A. Delsuc, G. Bertrand, and J. Y. Lallemand, J. Magn. Reson. A **102,** 357 (1993).
97. R. P. Lippmann, IEEE Acoust. Speech Signal Process **4,** 4 (1987).
98. D. R. Hush and B. G. Horne, IEEE Signal Process **10,** 8 (1993).
99. P. Guntert and K. Wuthrich, J. Magn. Reson. **96,** 403 (1992).
100. S. Golotvin and A. Williams, J. Magn. Reson. **146,** 122 (2000).
101. R. O. Duda and P. E. Hart, *Pattern classification and scene analysis* (John Wiley and Sons, New York, 1973).
102. H. G. Schulze, Ph.D. Thesis, The University of British Columbia, Vancouver, B.C., Canada (1996).
103. C. A. Lieber and A. Mahadevan-Jansen, Appl. Spectrosc. **57,** 1363 (2003).
104. Y. Wang and J.-Y. Mo, Chem. J. Internet **5,** 052015 (2003).
105. M. A. Kneen and H. J. Annegarn, Nucl. Instrum. Methods, Phys. Res. B **109,** 209 (1996).
106. D. A. McNulty and H. J. H. MacFie, J. Chemom. **11,** 1 (1997).
107. R. Synovec and E. Yeung, Anal. Chem. **57,** 2162 (1985).
108. A. P. Shreve, N. J. Cherepy, and R. A. Mathies, Appl. Spectrosc. **46,** 707 (1992).
109. H. K. Xiao and S. P. Levine, Anal. Chem. **65,** 2262 (1993).
110. G. Della Lunga, R. Pogni, and R. Basosi, J. Magn. Reson. A **108,** 65 (1994).
111. G. Della Lunga and R. Basosi, J. Magn. Reson. A **112,** 102 (1995).
112. R. Koradi, M. Billeter, M. Engeli, P. Guntert, and K. Wuthrich, J. Magn. Reson. **135,** 288 (1998).
113. R. P. Durman and D. J. Wood, "A new method of fluorescence rejection in Raman spectroscopy", in *Analytical Applications of Spectroscopy,* C. S. Creaser and A. M. C. Davies, Eds. (The Royal Society of Chemistry, London, 1988), p. 235.
114. W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical recipes in C* (Cambridge University Press, Cambridge, UK, 1994), p. 823.
115. J. L. Starck, E. Pantin, and F. Murtagh, Publ. Astronom. Soc. Pacific **114,** 1051 (2002).
116. L. S. Greek, H. G. Schulze, M. W. Blades, A. V. Bree, B. B. Gorzalka, and R. F. B. Turner, Appl. Spectrosc. **49,** 425 (1995).
117. E. T. Olejniczak and H. L. Eaton, J. Magn. Reson. **87,** 628 (1990).
118. R. Saffrich, W. Beneicke, K. P. Neidig, and H. R. Kalbitzer, J. Magn. Reson. B **101,** 304 (1993).
119. M. Oku, K. Wagatsuma, and H. Matsuta, J. Electron Spectrosc. Relat. Phenom. **83,** 31 (1997).
120. U. Amato and T. Sapatinas, "Wavelet shrinkage approaches to baseline signal estimation from repeated noisy measurements", *Technical report 243/02* (Istituto per le Applicazioni del Calcolo, Naples, Italy, 2002).
121. W. Chang and B. Vidakovic, Comp. Stat. Data Anal. **40,** 317 (2002).
122. I. Daubechies, Commun. Pure Appl. Math. **41,** 909 (1988).