The Analyst

Wavelet Denoising of Infrared Spectra

Bjørn K. Alsberg*[a], Andrew M. Woodward[a], Michael K. Winson[a], Jem Rowland[b] and Douglas B. Kell[a]

[a] *Institute of Biological Sciences, University of Wales, Aberystwyth, Ceredigion, UK SY23 3DA. E-mail: bka@aber.ac.uk*

[b] *Department of Computer Science, University of Wales, Aberystwyth, Ceredigion, UK SY23 3DA*

The application of wavelet denoising to infrared spectra was investigated. Six different wavelet denoising methods (SURE, VISU, HYBRID, MINMAX, MAD and WAVELET PACKETS) were applied to pure infrared spectra with various added levels of homo- and heteroscedastic noise. The performances of the wavelet denoising methods were compared with the standard Fourier and moving mean filtering in terms of root mean square errors between the pure and denoised spectra and visual quality of the denoised spectrum. The use of predictive ability as a possible objective criterion for denoising performance was also investigated. The main conclusion is that for very low signal-to-noise ratios (S/N) the standard denoising methods (Fourier and moving mean) are comparable to the more sophisticated methods. At higher S/N levels the wavelet denoising methods, in particular the HYBRID and VISU methods, are better. Wavelet methods are also better in restoring the visual quality of the denoised infrared spectra.

Keywords: *Wavelets; wavelet packets; denoising; infrared spectra; homoscedastic noise; heteroscedastic noise; chemometrics; rapid screening*

The rapid quantitative and qualitative information obtained from applying multivariate methods to spectra (*e.g.*, IR, Raman, UV) has been a very popular way of replacing slow wet chemical analyses. We are interested in particular in the determination of the concentrations of important compounds produced by industrially relevant bacteria and yeasts. In order to extract reliable information from any kind of data vector such as those in IR spectra,[1,2] unwanted noise has to be dealt with.[3–6] There are several ways of reducing the effect of noise in spectra, taking the mean of several co-added spectra being the most common. Unfortunately, co-adding of spectra is a time-consuming method when a large number of spectra are to be recorded. In particular, we are interested in cases of high-throughput analyses where FTIR spectra are recorded at numerous locations on 2D surfaces, *e.g.*, TLC plates, in biological tissues and in bacterial colonies. It is true that when a very small number of FTIR spectra are to be recorded, it is relatively rapid to use the appropriate number of co-adds to ensure a satisfactory signal-to-noise ratio, (S/N). However, when only a small number of co-adds is desired, mathematical techniques for removing the noise after recording of the spectra are more attractive.

The theory of convolutions and filter theory have been applied to a wide range of problems. However, there are certain types of problem where the frequency-domain methods are not optimal for removing noise. The traditional filtering methods in most cases rely on the frequencies obtained in the power spectrum being stationary. The theory of wavelets promises several improvements to the traditional filtering methods. The major difference is the fact that wavelet methods can much better resolve frequencies varying in time (or along a pseudo-time axis such as that representing wavenumbers).

Theory

Introduction to Wavelets and Wavelet Packets

Wavelets are becoming an increasingly important tool in image and signal processing.[7–10] Wavelets are effective in extracting both time- and frequency-like information from a time-varying signal. The short-time Fourier transform performs a constant bandwidth splitting of the signal whereas the wavelet transform has a proportional (octave) bandwidth splitting of the frequency domain. Unlike the Fourier transform, the wavelet transform can use a variety of different basis functions with different properties. Non-orthogonal wavelet bases are referred to as *frames*,[11–13] but will not be discussed here.

The more popular orthogonal wavelet bases have several interesting properties that make them a suitable basis for tools in signal analysis and compression. One important property of many wavelet basis functions is their localisation of both time and frequency domains simultaneously.

A *continuous* wavelet decomposition can be written as

$$w(s,b) = \int_{-\infty}^{\infty} \Psi_s(t-b)\, f(t)\mathrm{d}t \qquad (1)$$

where $\psi_s(x)$ is the wavelet function at a particular *scale s*, *i.e.*, the same wavelet function is dilated or contracted according to the scale, and $f(t)$ is function to be analysed; b signifies the translation of the wavelet at scale s. Eqn. (1) can be expressed as a convolution:

$$w(s,b) = \psi_s(t) \otimes f(t) \rightarrow F[w(s,b)] = F[\psi_s(t)]F[f(t)] \qquad (2)$$

Eqn. (1) is really a convolution of the signal f with the wavelet function in the time domain and we have emphasised this in eqn. (2) by using the convolution operator symbol $\otimes$. The equation to the right of the arrow is the convolution in the time domain expressed in the *frequency domain* by straightforward application of the convolution theorem;[14] F indicates the Fourier transform operator.

The *scale* can be interpreted as a measure of frequency. A short scale contains high-frequency components whereas a long scale contains low-frequency components. An intuitive way of looking at the wavelet transform is to interpret it as a sequence of combinations of bandpass filters. The wavelet function $\psi(t)$ (also referred to as the 'mother wavelet') can be interpreted as a high-pass filter acting on the original signal; the *scaling function* $\phi(t)$ (also referred to as the 'father wavelet'), on the other hand, behaves as a low-pass filter. The wavelet function $\psi(t)$ can be written as a linear combination of the scaling function. The scaling function has the property that it can be written in terms of scaled versions of itself:

$$\phi(x) = \sum_{k=0}^{N} c_k \phi(2x-k) \qquad (3)$$

In this paper, only the *discrete* wavelet transform will be used. This means that we restrict the choice of scale s and translations b; in general, we set

$$s = a_0^j,\ b = kb_0a_0^j \tag{4}$$

where $a_0 = 2$ and $b_0 = 1$ is the most common choice; j is an index that can be any natural number.

A common algorithm for calculating discrete wavelet coefficients is the so-called *Mallat algorithm*.[15–17] At each scale, high- (H) and low- (L) pass filters are applied to the input signal. The actual shapes of these filters are determined by the kind of wavelet function used. The output from the high-pass filter at each scale is recorded as the wavelet coefficients. The low-pass filter extracts the low-frequency components for the next scale where another set of high- and low-pass filters is employed. At each successive scale $(n - 1)$ the length of the vector upon which the filters operate is halved; this is referred to as *decimation*. Thus, the number of scales is $\log_2(n)$. The corresponding vector of wavelet coefficients for a scale j is written as $\boldsymbol{w}^{(j)}$. Wavelet reconstruction, *i.e.*, going from wavelet coefficients back into the original domain, is simply a multiplication of a vector $\boldsymbol{w}$ containing the coefficients for all scales with a matrix consisting of all time shifted wavelet functions for all scales:

$$\boldsymbol{f} = \boldsymbol{w}\boldsymbol{B}^{\mathrm{T}} \tag{5}$$

where $\boldsymbol{B}^{\mathrm{T}}$ is the transposed matrix of wavelet functions for all scales. The structure of the vector $\boldsymbol{w}$ is

$$\boldsymbol{w} = [\boldsymbol{w}^{(0)}\boldsymbol{w}^{(1)}\boldsymbol{w}^{(2)}\cdots\boldsymbol{w}^{(J)}] \tag{6}$$

The tree structure of the Mallat algorithm can be extended such that the filters are also used on the output from the high-pass filter. Such a decomposition of the data is encompased in the theory of *wavelet packets*.[18–30] The wavelet packets form a superset of the traditional wavelet coefficients, which corresponds to the leftmost branch of the tree. Because of this algorithmic tree structure, it is possible to *prune* branches in the tree to optimise some fitting criterion.

Denoising Using Wavelets

Noise is a phenomenon that affects all frequencies, whereas the signal of interest is most likely to occupy a small part of the frequency domain. Since the signal will tend to dominate the low-frequency components, it is expected that the majority of high-frequency components above a certain level are due to noise. This is the underlying philosophy for traditional Fourier filtering where low-pass filters cut off the high-frequency components. Similarly, we can expect small wavelet coefficients at short scales to be mainly noise components. The procedure for wavelet denoising will therefore be as follows:

(i) apply a wavelet transform to signal f^{noisy} and obtain the vector $\boldsymbol{w}$ of wavelet coefficients;
(ii) suppress or remove those elements in $\boldsymbol{w}$ that are thought to be attributed to noise; and
(iii) apply the cognate inverse wavelet transform to $\boldsymbol{w}$ to obtain a function f^{denoised}.

In this paper, we use eight different denoising techniques. Two of them (Fourier filtering and moving average) are used as references for the performance of the wavelet methods. All the methods presented are part of the WaveLab package for MATLAB,[31] which was used in all experiments presented.

Wavelet denoising methods in general[32–36] use two different approaches, hard and soft thresholding. The hard thresholding philosophy is simply to set all the wavelet coefficients below a certain threshold to zero. Soft thresholding, on the other hand, reduces the value of wavelet coefficients towards zero if they are above a certain value (referred to as 'shrinking'). For a certain wavelet coefficient at scale j we have

$$w_k = \mathrm{sign}(w_k)(\,|\,w_k\,| - \lambda)_+ \tag{7}$$

where *sign* returns the sign of the wavelet coefficient w_k and the parentheses represent the threshold value. We will sometimes refer to this function as $SOFT(\boldsymbol{w},t)$, where t is the threshold and $\boldsymbol{w}$ the vector to be thresholded.

Methods for Denoising

SURE

This denoising method is based on Stein's Unbiased Risk Estimate[37] and is applied to the whole wavelet coefficient vector, *i.e.*, the thresholding is performed on each scale j. The SURE method is a hard thresholding approach where the major work is invested in finding the right threshold for the different scales. First, we need to sort the squared wavelet coefficients $\{a_k = [w_k^{(j)}]^2\}$ in ascending order. The cumulative total of a_k is computed:

$$b_i = \sum_{k=1}^{i} a_k \tag{8}$$

Further, a vector $\boldsymbol{c}$ is needed which has the same size as the number of elements in the current scale j. The first element in $\boldsymbol{c}$ is $n_j - 1$, where n_j is the number of elements in the wavelet coefficient vector at scale j, and decreases linearly for successive vector elements to 0 (the last element). A *risk value*, r_i, is computed for every wavelet coefficient:

$$r_i = \frac{(n_j - 2i) + b_i + a_ic_i}{n_j} \tag{9}$$

The wavelet coefficient that has the minimum r_i is selected as the threshold value for that scale j. Note that the absolute value of the coefficient is used as the threshold. If the coefficient $|\,w_i^{(j)}\,|$ is chosen as the threshold, all coefficients with absolute values below will be set to zero. We will refer to this threshold as $t = SURE\ [\boldsymbol{w}^{(j)}]$.

VISU

For some of the denoising methods we need to specify L, which is the longest scale used in the thresholding. In this method we apply the denoising only on the coefficients in the index interval $[2^L + 1, n]$ where n is the number of wavelet coefficients. The L parameter must be much smaller than J where $n = 2^J$. We define the threshold parameter $t = (2 \log n)^{\frac{1}{2}}$, which is used in the soft thresholding scheme described above.

HYBRID

This is a soft threshold method where in some cases the soft threshold $t_A = (2 \log n_j)^{\frac{1}{2}}$ is used, and in other cases $t_B = SURE\ [\boldsymbol{w}^{(j)}]$ is used, depending on the parameter e, defined as follows:

$$e = \frac{\left\|\boldsymbol{w}^{(j)}\right\|^2 - n_j}{n_j} \tag{10}$$

such that

$$\text{if } e < \frac{J^{3/2}}{\sqrt{n_j}} \text{ then } SOFT[\boldsymbol{w}^{(j)}, t_A]$$
$$\text{else } SOFT(\boldsymbol{w}^{(j)}, \min(t_A, t_B)] \tag{11}$$

where J is the total number of scales and n_j is the number of elements in scale j.

MEDIAN ABSOLUTE DEVIATION (MAD)

Here the soft thresholding method is applied to the individual scales j. The threshold used for each scale j is

$$s_j = \frac{\text{median}[|\boldsymbol{w}^{(j)}|]}{0.6745} \quad (12)$$

Each wavelet coefficient scale is divided by s_j, $\boldsymbol{v}^{(j)} = \boldsymbol{w}^{(j)}/s_j$, and we use the threshold $t = (2 \log n_j)^{\frac{1}{2}}$. The thresholding is then done by $SOFT\ [\boldsymbol{v}^{(j)}, t]$.

MINIMAX

This procedure finds 'optimum' thresholds t_n such that the risk $R(f^{\text{denoised}}, f)$, given by

$$R(f^{\text{denoised}}, f) = \frac{1}{n}\sum_{i=1}^{n}(f_i^{\text{denoised}} - f_i)^2 \quad (13)$$

between the estimated wavelet coefficient θ_{est} and the true wavelet coefficient θ satisfy

$$R(\theta_{\text{est}},\theta) \leqslant \Lambda(\varepsilon^2 + R_{\text{opt}}) \quad (14)$$

where Λ is a constant which is related to the optimum risk R_{opt} if we had an oracle that could tell us what wavelet coefficients are larger than the noise level ε. A set of thresholds t_n are used that satisfy $t_n \leqslant (2 \log n)^{\frac{1}{2}}$. For very large n the optimum threshold values will approach $(2 \log n)^{\frac{1}{2}}$. The thresholds are subsequently used in a soft thresholding scheme.

Fourier

This is the classical Fourier denoising approach where the components with high frequencies are assumed to represent noise only and are therefore removed. In this case we use a soft thresholding technique which is dependent on input from the user. A region in the power spectrum of the signal is specified which most likely contains noise; this is usually located in the upper region of the power spectrum. The maximum amplitude value in this region is used as a cut-off level. At the located cut-off frequency a sigmoid function is used to implement a soft threshold.

Moving mean filter

A simple moving average filter was chosen as the 'baseline' method with which to compare the other, more complicated methods. A sliding window of size w is selected. For each step in the sliding process we find the mean of the curve points inside the window. This mean value is used as the output of the filter at each step. The window size is chosen to reflect the frequency content of the true signal by making sure the cut-off frequency of the filter is slightly larger than the bandwidth of the true signal.

WAVELET PACKETS (WP)

The WP denoising method used here has two significant steps: estimation of the best WP basis for denoising followed by proper selection of the denoising threshold. To select a basis, the Coifman–Wickerhauser 'best basis' algorithm is used.[38] This algorithm is based on finding the basis that gives rise to the minimum entropy of the signal energy distribution. The energy of a signal is the sum of the squares of its elements. This energy will be the same for different choices of bases, but the distribution will be different over its coordinates. To quantify this distribution, the entropy of the squares of the coordinates of the signal is used. From a data compression viewpoint it is advantageous to find a distribution with a low entropy. This means the signal can be described by a small number of bits. After finding the best basis the WP method uses Stein's Unbiased Risk Estimate in the calculation of hard denoising thresholds.

Heteroscedastic and Homoscedastic Noise

Definitions and properties

Let us assume we have a signal, $h(t)$, that contains only homoscedastic noise. In general, we write this as

$$h(t) = an(t) + s(t) \quad (15)$$

where a is a scalar that determines the size of the noise $n(t)$ and $s(t)$ is the pure signal. This Fourier transforms to

$$H(w) = aN(w) + S(w) \quad (16)$$

such that the signal is independent of the noise and is concentrated only in the region defined by $S(w)$. For heteroscedastic noise, however, the noise correlates in intensity with the amplitude of the signal

$$h(t) = af[s(t)]\ n(t) + s(t) \quad (17)$$

where $f(s)$ is the dependence of the noise on this signal. A Fourier transform of the noisy signal $h(t)$ produces

$$H(w) = aF[s(t)] \otimes N(w) + S(w) \quad (18)$$

where $\otimes$ is the convolution operator and upper case letters signify the corresponding Fourier transform of the functions in the time domain (written in lower case letters). The spectrum of the heteroscedastic noise can thus be regarded as the convolution of the spectrum of the homoscedastic noise with that of the signal. Accordingly, all frequencies in the spectrum will contain information related directly to the signal, $s(t)$.

Constructing heteroscedastic noise

When constructing heteroscedastic noise for the purpose of assessing different denoising methods, it is necessary to decide on the structure of the function f described in the previous section. The easiest choice is to let it be a constant such that the heteroscedastic noise is

$$h(t) = a\ s(t)\ n(t) + s(t) \quad (19)$$

In this paper, however, we decided to formulate f such that it is in accordance with the type of heteroscedastic noise that is normally present in *absorbance* spectra. It has been demonstrated[39] that the non-linear transform from *transmittance* to *absorbance* spectra itself converts homoscedastic noise into heteroscedasticity. It should be stressed that other phenomena can contribute to the observed heteroscedasticity. For instance, irreproducibility of transmitter offsets may also have an influence. We have, however, in order to simplify, assumed that the heteroscedasticity observed is caused only by the conversion from transmittance to absorbance. We note that the Beer transform is

$$A = \log(I_0/I) = -\log T \quad (20)$$

where A is the absorbance, I_0 is the original intensity of the incident beam, I is the reduced intensity of the beam after passing through the sample and T is the transmittance. Homoscedastic noise in a transmittance spectrum will be converted into heteroscedasticity after the transformation into

absorbance units. This means that our observed signal $s(t)$ is the result of a Beer transform

$$s(t) = -\log[q(t)] \quad (21)$$

where $q(t)$ is the transmittance signal. Adding homoscedastic noise $n(t)$ to the transmittance signal $q(t)$ now gives

$$z(t) = -\log[a\ q(t) + n(t)] = \log\{1/[a\ q(t) + n(t)]\} \quad (22)$$

In order to see that the noise becomes related to the size of the signal, we will use an intuitive rather than a mathematical argument. Assume that we have a large peak in the transmittance region with a small perturbation from noise. The −log (large number + small perturbation) corresponds to a region in the −log function that is relatively flat, *i.e.*, the small perturbations from the noise will not change the output significantly (which now becomes an absorbance). If we have a small signal that has a size comparable to the noise, then we will have a situation where the fluctuation in the signal + noise will be comparable in size with the noise. Now, the −log (small number + small perturbation) will be in a very steep region of the non-linear function and therefore any small changes will correspond to a large change in the output (absorbance). This means that the standard deviation (which is a measure of change in value) will be higher for regions containing small transmittance intensities (*i.e.*, high absorbance) and low for regions containing high transmittance intensities (*i.e.*, low absorbance).

In this paper we make use of this fact and therefore convert from absorbance to transmittance units and add *homoscedastic noise* with a certain S/N. The S/N values in all the experiments are obtained from the following equation

$$\mathrm{S/N} = \sqrt{\frac{\sum_{j}^{n} s_j}{\sum_{j}^{n} (s_j - n_j)^2}} \quad (23)$$

where s_j and n_j are the jth true signal and noise element, respectively. The noisy transmittance spectrum is subsequently transformed back into the absorbance domain where the noise is now heteroscedastic. In general, any non-linear transform of a signal will convert homoscedastic into heteroscedastic noise.

Experimental

Assessment of Denoising Performance

We shall use two different ways of assessing the performance of the different denoising methods. The first method is based on the use of the root mean square (rms) difference between the denoised noisy spectrum and noise-free spectrum as a measure of the performance of the different methods. Since the denoising process can often introduce offsets and scalings that can influence the rms values, we employed the technique of *multiplicative scatter correction* (MSC)[40–45] on the denoised spectra before calculating the rms differences. In MSC it is assumed that the observed spectrum $\boldsymbol{y}$ can be written in terms of the reference spectrum $\boldsymbol{s}$ as $\boldsymbol{y} = a\boldsymbol{s} + b$. The MSC operation on $\boldsymbol{y}$ is therefore $\boldsymbol{y}_{\mathbf{h}} = (\boldsymbol{y} - b)/a$. In this paper $\boldsymbol{y}$ is a spectrum that has been denoised using one of the methods described and $\boldsymbol{s}$ is the noise-free spectrum. Values of a and b are constructed for every denoised spectrum for each denoising method. The rms difference is calculated between $\boldsymbol{y}_{\mathbf{h}}$ and $\boldsymbol{s}$.

The second method of assessing denoising preformance used in this paper is to measure the rms error of *prediction* on an unseen validation set using the denoised data set.

All the data sets used in this paper (data sets 1–4) are such that we know the true underlying noise-free spectrum. This will, of course, not normally be the case in real applications and in general the investigator is left with visual inspection of the denoised signal as a way of determining the appropriateness of the denoising method for the data set. However, in the case of FTIR spectrometry noisy spectra will approach the 'true' noise-free spectra as the number of co-adds (followed by averaging) is increased.

Description of Experimental Conditions

Infrared spectra for data sets 2, 3 and 4 were recorded in the wavenumber interval 4000–600 cm^{-1} using a Bruker IFS28 FT-IR spectrometer (Bruker Spectrospin, Coventry, UK) equipped with a liquid nitrogen cooled MCT (mercury cadmium telluride) detector and a diffuse-reflectance absorbance TLC accessory (4 cm^{-1} resolution, spectra collected at 20 s^{-1}). ASCII data were exported from the Opus software used to control the FTIR instrument and imported into MATLAB.

Data set 1

This data set consists of an artificially generated IR absorbance spectrum to which is subsequently added homoscedastic noise and heteroscedastic noise with S/N = 1, 2, . . ., 30.

Data set 2

Homoscedastic or heteroscedastic noise (S/N = 1,2, . . ., 30) was added to the pure diffuse reflectance spectrum of sodium succinate.

Data set 3

To improve the S/N it is standard procedure in all instruments to take the mean of several spectra of the same sample. The S/N will improve as the square root of the number of co-added spectra that are used in the mean calculation for homoscedastic noise. This square root dependence is approximately true for heteroscedastic noise up to about 300–400 co-adds. Above this number of co-adds the S/N improves almost linearly (not shown).

Unfortunately, taking the mean of a sufficiently large number of spectra is too slow a process when rapid screening of thousands of samples is necessary. Here we want to use denoising methods to improve the accuracy in the estimation of the profile of the 'true' spectrum (*i.e.*, a mean spectrum of many co-adds that has a high S/N) from single co-add spectra (*i.e.*, with low S/N). Our choice of compound in this experiment was glucose at a 20 mM concentration. The reference, and estimate of the 'true' spectrum, was the mean of 300 co-adds. Our data matrix was a set of 300 single co-add spectra. When we plotted the mean vector *versus* the standard deviation vector of this data set we observed a clear linear relationship. This confirms our knowledge about absorbance IR spectra that the standard deviation of the absorbance is proportional to the mean value of the absorbance. Unfortunately, some regions in the spectra were significantly different from the glucose regions and thus removed from the data set. These regions were 4000–3619 and 1186–600 cm^{-1}, which constituted the first and the last part of the spectra.

Each of these spectra was subjected to the eight denoising methods described above and the rms difference was calculated between the denoised spectrum and the 'true' spectrum. The mean value of this rms difference over all the 300 spectra for all the eight methods was used to give an indication of the best method.

Data set 4

Data set 4 consists of 40 diffuse reflectance FTIR spectra of a developed culture of the bacterium *Staphylococcus aureus* containing the antibiotic ampicillin at different concentrations (0.5–20 mM). Infrared spectra (256 co-adds) for each of these samples were recorded. ASCII data were exported from the Opus software used to control the FTIR instrument and imported into MATLAB. The samples were separated into calibration and validation sets, each containing 20 objects, using the DUPLEX method.[46] PLS with leave-one-out cross-validation was used for finding the optimum model for the calibration data. In order to demonstrate the denoising effect at different noise levels, we added heteroscedastic noise to the data set with S/N in the region [1, 20].

Calculation and Presentation Details

All the data were reconstructed using the Symmlet 8 wavelet, which we have found to be a very good wavelet for modelling spectra. One of the reasons for this is that the Symmlet 8 basis resembles to some extent the shape of peaks found in IR spectra.

The low-frequency cut-off for shrinkage was set to $L = 5$. Denoising of the spectra was also performed using a threshold of $L = 0$ but it had a tendency to produce reconstructed spectra that were judged not to be sufficiently smooth. L was also used as the branch depth in the wavelet packet transform.

The rms difference between the reconstructed and the true spectrum (*i.e.*, that with very little noise) was calculated for each method at each S/N.

The resulting matrix for each denoising method contained the rms differences from the true spectrum for each denoised spectrum at each S/N level.

We summarise the results of these matrices by taking the mean and the standard deviation for all the spectra in the data sets. For convenience, we will sometimes refer to the different denoising methods by numbers: 1 = VISU, 2 = SURE, 3 = HYBRID, 4 = MINMAX, 5 = MAD, 6 = FOURIER, 7 = PACKETS, 8 = MOVING MEAN and 9 = NOISY SIGNAL (*i.e.*, the untreated signal containing noise).

Results

Data Set 1, Homoscedastic Noise

Homoscedastic noise (S/N from 1.61 to 30.00) was added to this spectrum and the results obtained by applying the various denoising methods to the noisy spectrum are given in Table 1. The best methods over the whole S/N range are the wavelet HYBRID (3) and Fourier (6) methods. The HYBRID method is slightly better than the Fourier method for low S/N (< 7.5). At very low S/N levels the HYBRID and the Fourier methods together with the moving average converge to almost identical performance.

The mean rms for the HYBRID method is 6.21 (median 3.67) and the mean for the Fourier method is 6.70 (median 4.59). To obtain a visual impression of the denoising process, we inspected the reconstruction results of four methods, HYBRID, Fourier, PACKETS and moving mean, at S/N = 4.69. The visualisation (not shown) seems to confirm the rms differences in that the denoised spectrum from the HYBRID method is better than the results from the Fourier and moving average methods. The PACKETS denoised spectrum, however, seems visually to be better than the reported rms values would suggest. In the Fourier denoised spectrum we observe unwanted ringing effects from aliasing. Similar ringing effects can be seen in the wavelet reconstructions at lower frequencies.

Data Set 1, Heteroscedastic Noise

Heteroscedastic noise was added to the noise-free data set 1. The results of applying the eight different denoising methods to the noisy spectrum is shown in Fig. 1. Again, the HYBRID denoising method (3) is slightly better for almost all the S/N levels, with the Fourier (6) next, equalling the HYBRID over most of the S/N range. The PACKETS method performs almost as well in this data set. The mean rms difference produced by the HYBRID method is 6.27 compared with 7.14 for the Fourier method.

Reconstructed denoised spectra for methods 3, 6, 7 and 9 at a noise level S/N = 4.67 are displayed in Fig. 2. The Fourier, HYBRID and moving average methods all show similar ringing effects which are absent in the PACKETS reconstruction.

Table 1 Rms differences between the ideal spectrum and the denoised spectrum for eight denoising methods applied to data set 1 to which has been added homoscedastic noise. The column headed S/N contains the signal-to-noise ratio used for each of the 20 experiments. The methods are indicated by numbers: 1 = VISU, 2 = SURE, 3 = HYBRID, 4 = MINMAX, 5 = MAD, 6 = Fourier, 7 = WAVELET PACKETS, 8 = Moving mean and 9 = NOISY SIGNAL

S/N	1	2	3	4	5	6	7	8	9
1.61	31.34	32.05	30.14	41.44	30.55	27.61	30.34	81.69	27.52
3.27	17.96	17.17	12.74	23.90	17.50	13.22	20.11	40.15	13.12
4.69	17.09	13.69	11.07	16.55	14.59	12.29	14.02	27.97	12.70
6.05	12.79	15.46	9.38	14.96	13.24	13.18	11.79	21.68	10.57
7.98	10.23	10.84	6.84	10.25	11.10	6.64	7.31	16.44	9.38
9.08	10.56	8.53	8.51	9.35	11.96	10.64	9.12	14.46	9.25
10.68	8.84	9.12	5.54	9.03	9.48	5.43	6.94	12.29	9.25
12.50	7.09	6.93	4.78	7.60	8.62	4.89	6.11	10.50	8.54
13.56	6.79	5.76	3.95	6.26	6.52	6.38	5.11	9.68	7.86
15.02	6.46	5.51	3.71	5.90	8.43	5.53	4.83	8.74	8.45
17.43	5.76	4.08	3.64	5.26	7.37	3.73	4.13	7.53	8.08
18.69	5.60	3.46	2.95	4.42	6.99	3.06	3.87	7.02	7.77
20.92	5.39	4.58	3.35	4.45	8.04	4.29	3.65	6.27	8.43
21.45	5.60	3.91	3.09	3.98	7.59	3.06	3.29	6.12	8.17
22.88	5.08	3.46	2.79	3.81	7.17	2.84	3.51	5.74	7.91
24.80	4.50	4.37	2.42	3.81	6.60	2.42	3.20	5.29	7.93
26.79	4.46	3.70	2.47	3.58	6.12	2.33	3.24	4.90	7.99
28.10	4.39	2.96	2.37	3.12	6.30	2.20	3.30	4.67	7.87
29.28	4.16	2.59	2.29	2.97	6.16	2.25	2.76	4.48	7.78
30.00	4.22	2.97	2.24	2.87	5.90	2.10	2.61	4.37	7.88
Mean	8.92	8.06	6.21	9.18	10.01	6.70	7.46	15.00	9.82
Median	6.11	5.05	3.67	5.58	7.81	4.59	4.48	8.13	8.30

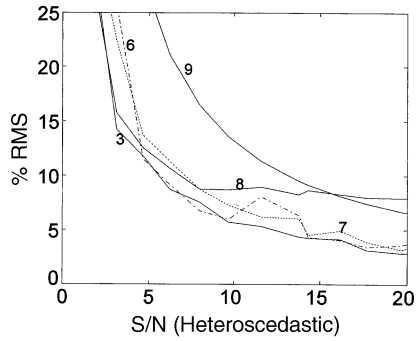

Fig. 1 Result of denoising using six wavelet methods, Fourier and median filter denoising on data set 1. Heteroscedastic noise. Only four of the applied methods are shown.

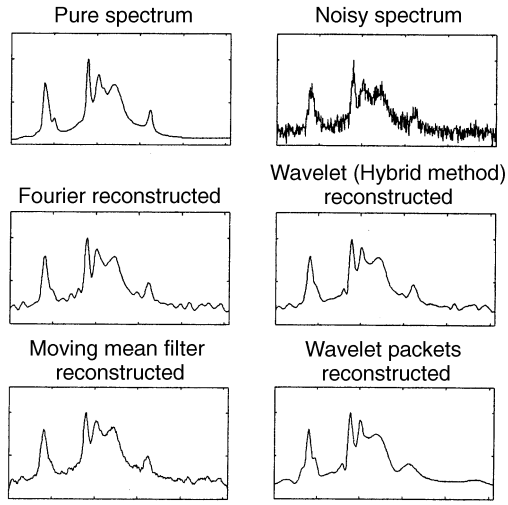

Fig. 2 Reconstructed noisy spectra of data set 1 containing heteroscedastic noise.

Data Set 2, Homoscedastic Noise

The effect of the different denoising methods on data set 2 (to which are added various levels of homoscedastic noise) is shown in Table 2. Again, using the rms difference criterion, HYBRID (3) is the best method overall, followed very closely by both the Fourier and PACKETS methods, which behave very similarly to each other. The denoised spectrum for the four methods HYBRID, moving mean, Fourier and PACKET at noise level S/N = 4.97 was computed (not shown). We observe that the visual quality of the WAVELET PACKET denoised spectrum is better than that of the Fourier method, although this is not reflected in the trend observed in the rms difference values. The PACKET method in general produces smooth spectra but can be erroneous in larger detail.

Data Set 2, Heteroscedastic Noise

Fig. 3 shows the effect of the different denoising methods on data set 2, to which are added various levels of heteroscedastic noise. The HYBRID method is again best at mid and high S/N but the Fourier method performs better at low S/N. The most interesting feature from the analysis is the far superior performance of the moving mean over all the other methods for mid to low S/N.

Surprisingly, the moving mean (8) method gives much lower rms values than any of the spectral methods. Is this trend also observed in the denoised spectrum for the different methods (S/N = 3.67)? Visual inspection of the reconstructions does not reflect the rms values above (see Fig. 4). This is presumably due to the eye favouring smoothness such that a smooth signal with large scale errors would be seen as better than a slightly noisier trace which is actually more representative of the true signal.

Data Set 3

The eight different denoising methods described above were used on each of the single co-adds and the rms difference was calculated between the denoised co-adds and the 'true' signal. The mean and the standard deviation values for all the rms values were calculated as a measure of ranking the denoising

Table 2 Rms differences between the ideal spectrum and the denoised spectrum for eight denoising methods applied to data set 2 to which has been added homoscedastic noise. The column headed S/N contains the signal-to-noise ratio used for each of the 20 experiments. The methods are indicated by numbers: 1 = VISU, 2 = SURE, 3 = HYBRID, 4 = MINMAX, 5 = MAD, 6 = Fourier, 7 = WAVELET PACKETS, 8 = Moving mean and 9 = NOISY SIGNAL

S/N	1	2	3	4	5	6	7	8	9
1.34	62.95	100.10	51.66	80.51	64.78	62.61	59.25	58.00	121.68
2.52	48.99	39.54	33.86	38.65	51.93	45.73	42.53	33.18	64.58
3.86	36.21	27.41	22.51	26.45	43.20	27.29	28.91	26.50	42.21
4.97	32.82	20.76	19.79	18.61	38.38	24.07	24.40	23.62	32.76
6.71	25.60	15.37	16.42	14.34	29.15	16.47	19.84	20.30	24.24
7.68	20.52	17.87	11.48	14.10	31.78	16.28	14.17	20.10	21.19
8.92	20.62	12.41	10.59	11.89	27.57	13.24	13.47	19.38	18.24
10.60	18.21	10.40	9.47	9.49	26.66	11.16	12.89	18.71	15.36
11.26	16.45	11.36	8.83	9.76	23.77	11.51	10.60	18.68	14.45
12.66	15.34	10.54	8.27	8.72	24.53	10.47	10.80	18.42	12.86
13.77	13.52	10.01	7.88	8.70	23.44	9.59	9.14	18.31	11.82
15.56	13.80	7.88	7.34	7.20	22.58	8.37	8.56	18.12	10.46
16.27	12.20	8.57	7.14	7.48	21.69	8.28	7.73	18.06	10.00
17.55	12.03	8.03	6.42	6.82	22.09	7.59	7.55	17.91	9.27
19.04	11.50	6.95	6.08	5.94	19.97	8.11	7.46	17.86	8.55
20.37	11.31	6.56	6.00	5.76	20.02	6.93	7.14	18.03	7.99
21.48	10.54	6.75	5.59	5.55	20.53	6.66	6.60	17.79	7.58
24.02	9.67	5.30	5.13	5.17	22.67	6.06	5.95	17.72	6.77
23.96	9.55	6.04	5.22	5.14	21.76	6.30	6.05	17.95	6.79
24.98	9.23	5.73	5.02	4.72	21.10	6.15	6.46	17.65	6.51
Mean	20.55	16.88	12.74	14.75	28.98	15.64	15.47	22.67	21.81
Median	14.57	10.21	8.07	8.71	23.60	10.03	9.87	18.37	12.34

methods. The measured S/N of the data set was fairly high, 107 ± 5. The mean rms values for the different methods were 0.85 ± 0.05 for VISU, 1.07 ± 0.05 for SURE, 1.04 ± 0.05 for HYBRID, 1.06 ± 0.05 for MINMAX, 2.8 ± 0.1 for MAD, 1.75 ± 0.03 for Fourier, 1.03 ± 0.05 for PACKETS, 1.07 ± 0.05 for moving mean and 3.88 ± 0.03 for NOISY SIGNAL. All the wavelet methods [except for method 5 (MAD)] have a lower rms than the Fourier and moving mean methods. Again, using visual inspection of the reconstructions (not shown), we see that the Fourier and moving mean methods have a tendency for oversmoothing of certain regions. The wavelet methods seem to be better at capturing significant spikes in the spectra.

Data Set 4

The noise-free and untreated data set was first analysed with the partial least squares (PLS) method. A five factor PLS model was formed using full cross-validation. When this model was applied to an unseen validation set, an rms prediction error of 9.7% was achieved. Further denoising on this data set will improve the prediction error by approximately 2%. In order to observe the performance for different noise levels, heteroscedastic noise was added to the data set (S/N = 1–20). For each noise level, all the denoising methods described earlier were used prior to PLS modelling (five factors extracted). The results of the rms errors of prediction are shown in Fig. 5. As shown before, the traditional methods such as Fourier and moving average perform better than or as well as all the wavelet methods for noise levels lower than S/N = 7. Above this noise level one of the wavelet denoising methods (VISU) performs better. None of the denoising methods can achieve an rms error in prediction lower than 10% in the noise level range.

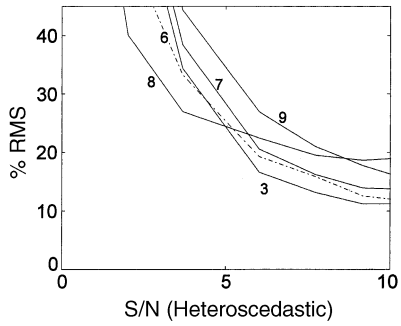

Fig. 3 Result of denoising using six wavelet methods, Fourier and median filter denoising on data set 2. Heteroscedastic noise. Only four of the applied methods are shown.

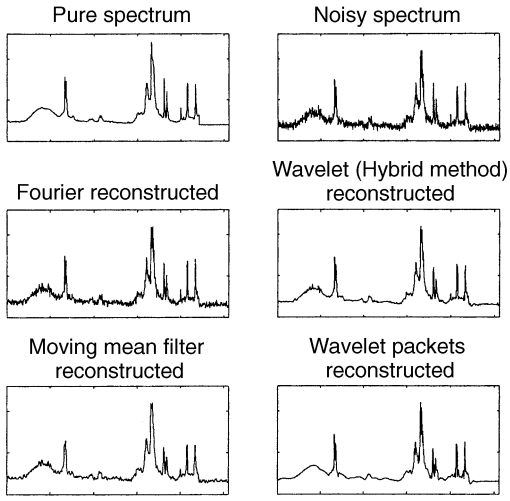

Fig. 4 Reconstruction comparison for data set 2 containing heteroscedastic noise.

Discussion and Conclusions

One important implication of improved denoising techniques in IR spectrometry is that we can improve the S/N with a significantly reduced number of co-adds. This will be of particular importance for experiments in which we wish to record the IR spectra with rapidity, such as in screening for metabolite overproduction and 2D surface mappings. Under such conditions, the benefit of higher throughput is essentially in proportion to the reduction in the number of co-adds. Other experimental methods that may be expected to benefit from wavelet denoising techniques are the group of coupled methods (GC–IR, LC–IR, *etc.*). When the location of smaller peaks in coupled methods is necessary, heteroscedastic noise can have a devastating effect on the correctness of the results.[47–49]

In addition, denoising applied to spectra in general will be of importance in any kind of multivariate modelling performed on the spectra. Examples of popular methods that are often used in the analysis of spectra are PLS[50–60] and neural networks.[61–65] In this case it is possible to construct objective criteria that can be used to optimise denoising of spectra. For instance, it is possible to use the predictive ability or classification error to find the optimum choice of threshold in the wavelet denoising process. The need for such objective criteria is emphasised by the disparity, shown by some of the results in this paper, between the 'quality' of denoised spectra as assessed by rms difference values and by visual assessment of the denoised spectra.

Although the wavelet approach allows greater freedom than traditional filtering methods, this also requires appropriate selection of more parameters by the user in order to optimise the denoising process. The wavelet transform allows a greater degree of compression into a smaller proportion of coefficients than the Fourier transform, and should therefore allow a greater rejection of noise with optimum parameter selection. The major improvement in the wavelet denoising methods compared with the standard filtering methods is the possibility of localising the frequency information to selected parts of the spectrum. For instance, in IR spectra we have the complex fingerprint regions in the 1000–400 cm^{-1} range which will contain sharper peaks than in the 4000–3000 cm^{-1} region. This means that we do not want to apply the same frequency cut-off for these two regions. A standard Fourier filtering approach will look at the whole power spectrum for both regions and apply the frequency cut-off for the whole spectrum. Accordingly, wavelets can be better in such cases because they are localised in both the 'time' (here wavenumber) and frequency domains.

The comparison of the transform methods with our 'baseline' method of a moving mean filter is instructive. It can be seen

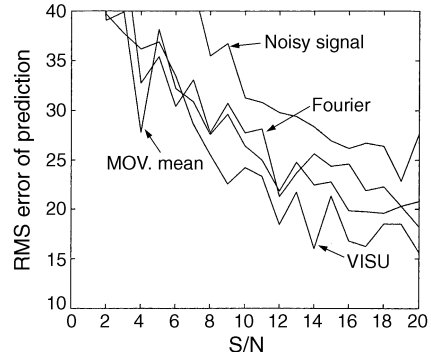

Fig. 5 The rms error of prediction when using the eight denoising methods. Here only the best wavelet method is shown (VISU), together with the comparable classical methods (Fourier and moving mean).

above that, with the filter length chosen as indicated, the moving mean filter is just as efficient as any of the more advanced methods at low S/N, but at the cost of poor performance at high S/N. It should be noted that shortening the length of this filter allows it to perform better than the rest at high S/N, but now at the cost of poor performance at low S/N. The advantage of the adaptive methods such as HYBRID is that they automatically produce performance close to the optimum across the whole range of noise levels studied.

We thank the Chemicals and Pharmaceuticals Directorate of the UK BBSRC, GlaxoWellcome and Bruker/Spectrospin for financial support.

References

1 Stark, P. B., Herron, M. M., and Matteson, A., *Applied Spectrosc.*, 1993, **47**, 1820.
2 Walczak, B., van den Bogaert, B., and Massart, D. L., *Anal. Chem.*, 1996, **68**, 1742.
3 Berger, J., Coifman, R. R., and Goldberg, M. J., *J. Audio-Eng. Soc.*, 1994, **42**, 808.
4 Coifman, R. R., and Wickerhauser, M. V., *Opt. Eng.*, 1994, **33**, 2170.
5 Donoho, D. L., and Johnstone, I. M., *C.R. Acad. Sci., Ser. I.*, 1994, **319**, 1317.
6 Healy, D. M., Lu, J., and Weaver, J. B., *Ann. Biomed. Eng.*, 1995, **23**, 637.
7 Paternot, X., *Signal Processing via Wavelet and Identification*, Massachusetts Institute of Technology Laboratory for Information and Decision Systems, Cambridge, MA, 1996.
8 *Wavelet Transform Applications to Data, Signal, Image, and Video Processing: September 11–15, 1995, Engineering 867.121*, University of California Los Angeles University Extension Department of Engineering Information Systems and Technical Management Short Course Program, Los Angeles, CA, 1995.
9 *Wavelet Applications in Signal and Image Processing II: 27–29 July 1994, San Diego, California*, SPIE, Bellingham, WA, 1994.
10 Young, R. K., *Wavelet Theory and its Applications*, Kluwer, Boston, 1993.
11 Daubechies, I., *Ten Lectures on Wavelets*, Society for Industrial and Applied Mathematics, Philadelphia, PA, 1992.
12 Daubechies, I., and Grossmann, A., *Commun. Pure Appl. Math.*, 1988, **41**, 151.
13 Teolis, A., and Benedetto, J. J., *Signal Process.*, 1995, **45**, 369.
14 Cartwright, M., *Fourier Methods for Mathematicians, Scientists and Engineers*, Ellis Horwood, New York, 1990.
15 Mallat, S., *IEEE Trans. Pattern Anal. Machine Intell.*, 1989, **11**, 674.
16 Mallat, S., *Trans. Am. Math. Soc.*, 1989, **315**, 69 and 18 215.
17 Davis, G., Mallat, S., and Zhang, Z. F., *Opt. Eng.*, 1994, **33**, 2183.
18 Cody, M. A., *Dr. Dobbs J.*, 1994, **19**, 44.
19 Berger, J., Coifman, R. R., and Goldberg, M. J., *J. Audio-Eng. Soc.*, 1994, **42**, 808.
20 Chu, C. H. H., *Opt. Eng.*, 1994, **33**, 2136.
21 Chui, C. K., and Li, C., *SIAM J. Math. Anal.*, 1993, **24**, 712.
22 Donoho, D. L., in *Different Perspectives on Wavelets*, ed. Daubechies, I., American Mathematical Society, Providence, RI, 1993, pp. 173–205.
23 Jawerth, B., and Sweldens, W., *SIAM Rev.*, 1994, **36**, 377.
24 Laine, A., and Fan, J., *IEEE Trans. Pattern Anal. Machine Intell.*, 1993, **15**, 1186.
25 Lambrecht, C. V., and Karrakchou, M., *Signal Process.*, 1995, **47**, 135.
26 Learned, R. E., and Willsky, A. S., *Appl. Comput. Harmonic Anal.*, 1995, **2**, 265.
27 Park, D., and Lee, M. H., *IEEE Trans. Consumer Electron.*, 1994, **40**, 527.
28 Wickerhauser, M. V., in *Different Perspectives on Wavelets*, ed. Daubechies, I., American Mathematical Society, Providence, RI, 1993, pp. 155–171.
29 Yen, N.-C., *J. Am. Stat. Assoc.*, 1994, **95**, 889.
30 Yoshida, H., Doi, K., Nishikawa, R. M., and Giger, M. L., *Radiology*, 1995, **197**, 393.
31 Buckheit, J., Chen, S., Crutchfield, J., Donoho, D., Gao, H., Johnstone, I., Kolaczyk, E., Scargle, J., Young, K., and Yu, T., http://playfair.Stanford.EDU:80/~wavelab/, 1966.
32 Donoho, D. L., and Johnstone, I. M., *Biometrika*, 1994, **81**, 425.
33 Donoho, D. L., and Johnstone, I. M., *C.R. Acad. Sci., Seri. I*, 1994, **319**, 1317.
34 Donoho, D. L., and Johnstone, I. M., *J. Am. Stat. Assoc.*, 1995, **90**, 1200.
35 Donoho, D. L., *IEEE Trans. Inf. Theory*, 1995, **41**, 613.
36 Donoho, D. L., Johnstone, I. M., Kerkyacharian, G., and Picard, D., *R. Stat. Soc., Ser. B*, 1995, **57**, 301.
37 Donoho, D. L., and Johnstone, I. M., *J. Am. Stat. Assoc.*, 1995, **90**, 1200.
38 Coifman, R. R., and Wickerhauser, M. V., *IEEE Trans. Inf. Theory*, 1992, **38**, 713.
39 Toft, J., and Kvalheim, O. M., *Chemom. Intell. Lab. Syst.*, 1993, **19**, 65.
40 Iversen, A. J., and Palm, T., *Appl. Spectrosc.*, 1985, **39**, 641.
41 Isaksson, T., and Næs, T., *Appl. Spectrosc.*, 1988, **42**, 1273.
42 Sorvaniemi, J., Kinnunen, A., Tsados, A., and Malkki, Y., *Food Sci. Technol. Lebensm.-Wiss. Technol.*, 1993, **26**, 251.
43 Isaksson, T., and Kowalski, B., *Appl. Spectrosc.*, 1993, **47**, 702.
44 Carlsson, A. E., and Janne, K. L. R., *Appl. Spectrosc.*, 1995, **49**, 1037.
45 Helland, I. S., Næs, T., and Isaksson, T., *Chemom. Intell. Lab. Syst.*, 1995, **29**, 233.
46 Snee, R. D., *Technometrics*, 1977, **19**, 415.
47 Keller, H. R., Massart, D. L., Liang, Y. Z., and Kvalheim, O. M., *Anal. Chim. Acta*, 1992, **263**, 29.
48 Kvalheim, O. M., Brakstad, F., and Liang, Y. Z., *Anal. Chem.*, 1994, **66**, 43.
49 Ritter, C., Gilliard, J. A., Cumps, J., and Tilquin, B., *Anal. Chim. Acta*, 1996, **318**, 125.
50 Clementi, S., Alunni, S., Bisani, M. L., Bonelli, d., Chiocchini, D., Cruciani, G., Giulietti, G., Johansson, E., and Wold, S., *Chim. Ind. (Milan)*, 1988, **70**, 78.
51 DeJong, S., *J. Chemom.*, 1993, **7**, 551.
52 Frank, I. E., *Chemom. Intell. Lab. Syst.*, 1990, **8**, 109.
53 Geladi, P., *Chemom. Intell. Lab. Syst.*, 1987, **2**, 257.
54 Geladi, P., *Chemom. Intell. Lab. Syst.*, 1992, **15** R7.
55 Holcomb, T. R., and Morari, M., *Comput. Chem. Eng.*, 1992, **16**, 393.
56 Wakeling, I. N., and Macfie, H. J. H., *J. Chemom.*, 1992, **6**, 189.
57 Wold, S., Martens, H., and Wold, H., *Lect. Notes Math.*, 1983, **973**, 286.
58 Wold, S., Ruhe, A., Wold, H., and Dunn, W. J., *SIAM J. Sci. Stat. Comput.*, 1984, **5**, 735.
59 Wold, S., *Technometrics*, 1993, **35**, 136.
60 Zhu, E. Y., and Barnes, R. M., *J. Chemom.*, 1995, **9**, 363.
61 Bulsari, A. B., *Neural Networks for Chemical Engineers*, Elsevier, Amsterdam, 1995.
62 Bishop, C. M., *Neural Networks for Pattern Recognition*, Clarendon Press, Oxford, 1995.
63 Cheng, B., and Titterington, D. M., *Stat. Sci.*, 1994, **9**, 2.
64 Haykin, S., *Neural Networks*, Macmillan, New York, 1994.
65 Goodacre, R., Neal, M. J., and Kell, D. B., *Zbl. Bakteriol.*, 1996, **284**, 516.

Paper 6/08255F
Received December 9, 1996
Accepted April 21, 1997