ELSEVIER

Chemometrics and Intelligent Laboratory Systems 48 (1999) 21–34

Chemometrics and intelligent laboratory systems

Optimization of signal denoising in discrete wavelet transform

L. Pasti [a], B. Walczak [a,1], D.L. Massart [a,*], P. Reschiglian [b]

[a] *ChemoAC, Pharmaceutical Institute, Vrije Universiteit Brussel, Laarbeeklaan 103, B-1090 Brussels, Belgium*
[b] *Department of Chemistry 'G. Ciamician', University of Bologna, V. Selmi, 2 I-40126 Bologna, Italy*

Received 31 July 1998; received in revised form 12 November 1998; accepted 10 December 1998

Abstract

A method to optimize the parameters used in signal denoising in the wavelet domain is presented. The method, which is based on cross-validation (CV) procedure, permits to select the best decomposition level and the best wavelet filter function to denoise a signal in the discrete wavelet domain. The procedure was validated by using computer generated signals to which white noise was added. Signals having different features and a range of signal to noise ratios were explored. The method was shown to give reliable results for all cases studied. The proposed method was applied to experimental gravitation field flow fractionation records, and the results were compared with classical low pass filtering in the Fourier domain.

Keywords: Discrete wavelet transform; Denoising; Cross-validation; Gravitational field flow fractionation

1. Introduction

All signals obtained as instrumental response of analytical apparatus are affected by noise. The noise degrades the accuracy and precision of an analysis, and it also reduces the detection limit of the instrumental technique. Signal denoising is therefore highly desirable in analytical response optimization. There are different possible approaches to signal denoising and one of the most recent methods is based on the wavelet transform. Wavelets are a new family of basis functions, well localized in both time and frequency domains [1]. Due to their local character, the representation of a signal in the wavelet domain is sparse and allows signal compression and denoising [2,3]. Donoho and Johnstone [4] and Donoho [5] have developed a method known as the wavelet shrinkage to estimate an unknown smoothed signal from data with noise. The methodology can be applied to the discrete wavelet transform (DWT), as well as to the wavelet packet transform (WPT). In the present article, we only focus on the DWT.

* Corresponding author. Tel.: +32-2-477-4737; Fax: +32-2-477-4735; E-mail: fabi@vub.vub.ac.be

[1] On leave from the Silesian University, 40-006 Katowice, 9 Szkolna street, Poland.

0169-7439/99/$ - see front matter
PII: S0169-7439(99)00002-7

The main steps of signal denoising are:

1. decomposition of the signal,
2. thresholding (elimination of small coefficients), and
3. reconstruction of the signal

Two different approaches are usually applied to denoise: hard thresholding or soft thresholding. The hard thresholding method consists in setting all the wavelet coefficients below a given threshold value equal to zero, while in soft thresholding the wavelet coefficients are reduced by a quantity equal to the threshold value [5]. The denoising procedure requires the estimation of the noise level. There are many possible approaches to do so. The most popular one, known as 'universal thresholding' (ThU), is based on statistical properties of white Gaussian noise. The threshold level is calculated from the standard deviation of the detail coefficients of the first level of decomposition. This definition of threshold value can lead to overestimation of the noise level. To overcome this problem, the cross-validation (CV) procedure proposed by Nason [6], can be applied. Other factors that strongly influence the signal denoising are the wavelet function, (i.e., the filter used to decompose the signal), and the level of DWT decomposition (e.g., see Ref. [2]). In WPT, different algorithms have been proposed to select the best basis to represent the signal whereas until now there was no way to optimize the decomposition in DWT. Fang and Chen [7] proposed to optimize the level of decomposition in signal denoising by using an adaptive wavelet filter. In the field of signal compression, an algorithm called minimum descriptor length (MDL) was proposed by Saito [8] to optimize the efficiency of compression in the wavelet domain. The algorithm permits one to select the best basis and the best number of wavelet coefficients to be employed for estimating the signal component in the data. It is the aim of our study to optimize the denoising procedure in DWT by optimization of the wavelet transformation (filter and level) and of the denoising parameter (threshold). To accomplish this task, the CV theory, which was originally developed to optimize the soft threshold value, is applied here to select both the wavelet filter function and the level of decomposition. The possibility of using the MDL criteria to optimize the selection of the wavelet filter function in a denoising procedure is also explored. Application of the CV method to denoising an experimental signal consisting of two components of different frequency is presented.

2. Theory

2.1. Discrete wavelet transform

The wavelet transform of a discrete signal, f, can be presented as a linear transformation involving an orthonormal matrix $\mathbf{W}$

$$\boldsymbol{w} = \mathbf{W}f \tag{1}$$

where $\boldsymbol{w}$ is a vector containing wavelet transform coefficients and $\mathbf{W}$ is the matrix of the wavelet filter coefficients. Each row of the matrix $\mathbf{W}$ represents one basis vector out of a collection of possible basis vectors. The basis vectors for a given wavelet function can be derived from a common function called mother wavelet via two operations: translation and dilation [9].

Signal reconstruction in the case of the orthonormal $\mathbf{W}$ matrix is described by the simple formula

$$f = \mathbf{W}^t\boldsymbol{w} \tag{2}$$

For the decomposition and reconstruction of the signal whose length equals an integer power of two ($N = 2^p$), the fast pyramid algorithm by Mallat [10] can be applied (Fig. 1). Each row represents a level of decomposition.

2.2. Denoising

An instrumental signal, f, can be presented as the sum of two components:

$$f = t + \mathrm{ns} \tag{3}$$

where t is the ideal signal and ns is the noise.

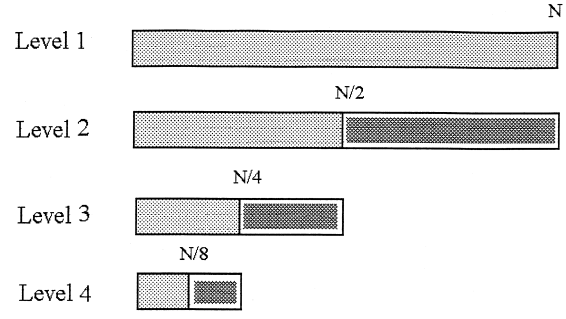

Fig. 1. The pyramid algorithm in DWT.

To obtain the ideal component of the signal (t) from its noisy counterpart (f), one possible approach requires the estimation of the parameters to describe the noise component. Once the parameters are obtained, the two components of the signal (i.e., noise and analytical signal) are separated from each others by comparison of the coefficients that describe the signal and the parameters that describe the noise.

2.3. Threshold estimation

There are many possible approaches to the estimation of the noise level, and a systematic investigation about their performance can by found in Ref. [11]. In Ref. [12], a method based on a cut-off value in the wavelet domain is presented. The universal threshold [5] is defined as:

$$\text{th} = \sigma\sqrt{2\log N} \tag{4}$$

where N is the signal length σ and is the standard deviation of the noise. The latter is estimated from the median of the detail coefficients (detail) at the first level (detail^1) of signal decomposition

$$\sigma = \left|\text{median}(\text{detail}^1)\right|/0.674 \tag{5}$$

2.4. Thresholding policy

Once the threshold value has been calculated one can apply a soft or hard modelling policy: in soft thresholding, for each wavelet transform coefficient w_{ij} and threshold th, the soft thresholded value is calculated as

$$w_{ij}^t = \text{sgn}(w_{ij})\left(|w_{ij}| - \text{th}\right) \text{ if } |w_{ij}| \geq \text{th}$$

$$w_{ij}^t = 0 \text{ if } |w_{ij}| < \text{th} \tag{6}$$

w_{ij} is the jth wavelet coefficient at level i of the decomposition. The threshold was applied only to the detail coefficients.

In hard thresholding the hard thresholded value is given by

$$w_{ij}^t = w_{ij} \text{ if } |w_{ij}| \geq \text{th}$$

$$w_{ij}^t = 0 \text{ if } |w_{ij}| < \text{th} \tag{7}$$

2.5. CV method

Theory of the CV approach is described elsewhere [6]. It was developed to estimate the best threshold value to optimize signal denoising by using a soft thresholding policy. In Ref. [6], the denoising procedure was applied by using only one wavelet function (the 6th member of the Daubechies family) and a constant set of decomposition levels in DWT. In the present work, the basic principle of CV is also used to select the wavelet function and the level of DWT.

The goal of data denoising is to minimize the differences between the denoised signal and the ideal (i.e., noiseless) one. The ideal signal can only be used for synthetic data, because in this case the pure data (noise free) is available. It is possible to compute the reconstruction square error (RSE) calculated starting from the denoised signal, f_d, and the ideal, t:

$$\text{RSE} = \sum_{j=1}^{N} \left(f_d(i) - t(i)\right)^2 \tag{8}$$

In practice, the signal t is not known. An estimation of the RSE can be obtained by means of the CV procedure, that is summarized in Fig. 2.

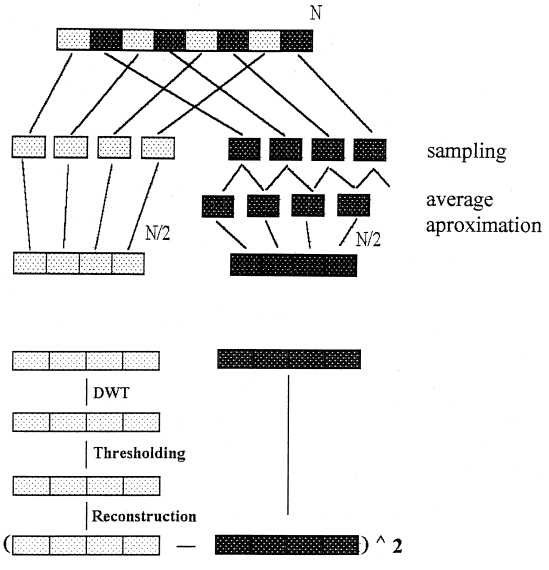

Fig. 2. Construction of odd and even function from the experimental data. Evaluation of ISE value in CV method.

The procedure can be divided into the following steps.

(1) Two new functions—one of odd points (fo) and the other of even points (fe)—are obtained by dividing the original signal and re-indexation of the data points from $i = 1$ to $2^{(p-1)}$.

(2) An estimation of the even function is constructed by interpolation starting from the odd function

$$\mathrm{fe}^*(i) = \frac{(\mathrm{fo}(i) + \mathrm{fo}(i+1))}{2}, \qquad i = 1, \ldots, 2^{p-1} \tag{9}$$

where fe* is the estimated even function.

(3) The signal is transformed in the wavelet domain by applying the transformation matrix to the even function

$$\mathrm{we} = \mathbf{W}\,\mathrm{fe} \tag{10}$$

(4) The transformed signal is filtered by applying a thresholding procedure to the wavelet coefficients

$$\mathrm{we}_t = T(\mathrm{we}) \tag{11}$$

where T is the (soft or hard) thresholding operator.

(5) The reconstructed filtered signal is obtained by applying the inverse transform matrix of the thresholded wavelet coefficients

$$\bar{\mathrm{fe}} = \mathbf{W}^t\,\mathrm{we} \tag{12}$$

where $\bar{\mathrm{fe}}$ is the reconstructed function of fe.

(6) The integrated square error (ISE) is obtained as the sum of the squared differences between the reconstructed function and its estimate

$$\mathrm{ISE}(e) = \sum_{j=1}^{N/2} \left(\mathrm{fe}^*(j) - \bar{\mathrm{fe}}(j)\right)^2 \tag{13}$$

The whole procedure applied to the even function is repeated for the odd function, and one obtains ISE(o). The total ISE is the sum of the two:

$$\mathrm{ISE} = \mathrm{ISE}(e) + \mathrm{ISE}(o) \tag{14}$$

The ISE measures the RSE of the filtered signal with respect to the original one and is used when the shape of the signal is unknown.

More detail about ISE can be found in Ref. [6]. In Ref. [6], ISE was intensively tested as a measure of RSE by using simulated data.

The threshold value obtained by means of the procedure described above, is estimated on the basis of $N/2$ data points (th($N/2$)). The threshold value for the whole function (th(N)) consisting of N points is defined as

$$\mathrm{th}(N) = \left(1 + \frac{\log 2}{\log N}\right)^{1/2} \mathrm{th}(N/2) \tag{15}$$

2.6. MDL method

An interesting approach to data compression and denoising, which is free from any parameter setting or any subjective judgements, has been proposed by Saito [8]. It is based on the assumption that the signal component t, can be efficiently represented by one or more bases, whereas the white noise component, ns, cannot be efficiently represented by any basis. The white noise component in the time domain can be represented as a Gaussian distributed signal having a mean value equal to zero. In the frequency domain, this component gives a constant contribution so that all the wavelet coefficients carry information about the noise whichever wavelet function is used in the transformation. On the other hand, the signal of interest is characterized in the time domain by the shape of the instrumental response to the instrumental input. As a result, one wavelet function will require a smaller number of coefficients to describe the signal of interest than all the others. Saito introduced the following MDL cost function:

$$\mathrm{MDL}(k,n) = \min_{0 \le k < N} \underbrace{(3/2)k \log(N)}_{\text{term 1}} + \underbrace{(N/2)\log\|(\mathbf{I} - \boldsymbol{\Theta}^{(k)})\boldsymbol{w}f\|^2}_{\text{term 2}} \tag{16}$$

where N is the signal length, k is the number of non-zero coefficients present in the vector $\boldsymbol{w}$, $\mathbf{I}$ is the $(N \times N)$ identity matrix, and $\boldsymbol{\Theta}^{(k)}$ is a hard thresholding operator. The operator keeps the largest k coefficients of $\boldsymbol{w}$ and sets all the other N–k coefficients equal to zero.

The first term (1), known as penalty term, takes into account of the number of non zero coefficients of transform vector. The second term (2) is related to the discrepancy between the original and the reconstructed signal.

3. Experimental

3.1. Simulated signals

In order to illustrate the performance of the proposed method, different signals were considered. In Fig. 3, four simulated signals are presented. Different levels of white noise (with mean zero and standard deviation scaled to achieve different signal to noise ratios) were considered.

The simulated signals are (see Fig. 3): signal 1: a sum of Gaussian functions having different peak shape parameters (width, height); signal 2: the first derivative of signal 1; signal 3: a sum of Lorentzian functions; and signal 4: a sum of spikes having negative and positive values. The full peak width of each spike is equal to one.

3.2. Experimental signals

The experimental signals studied concern gravitational field flow fractionation (GrFFF) of silica particles. The GrFFF system here employed was described in previous papers [13–15]. The channel dimensions were 60 cm from tip to tip, 2 cm breadth and 0.0175 cm thick. A UV–Vis spectrometer Mod 2550 (Varian, Walnut Creek, CA, USA) operating at variable wavelength was used as the detector and the output signal was recorded by a 12-bit I/O DAQ board Mod. Lab PC + (National Instruments, Austin, TX, USA). Sample acquisition rate was set at 0.2 Hz. The mobile phase was a 0.003 M solution of sodium dodecylsulphate (SDS) in Milli-Q (Millipore, Bedford, MA) water, to which NaN_3 0.003808 M was added as bactericide. The flow was generated by a HPLC pump Mod. 2150 (Varian) and was set at 1 ml/min. The samples were bare silica particles (LiChrospher Si-60, Merck, Darmstadt, Germany) of nominal size 5 m (nominal distribution percentiles: $d_{10} = 3.7$ μm, $d_{50} = 5.0$ μm, $d_{90} = 6.8$ μm). Sample dispersions were prepared in Milli-Q water. The obtained fractograms contain 512 data points.

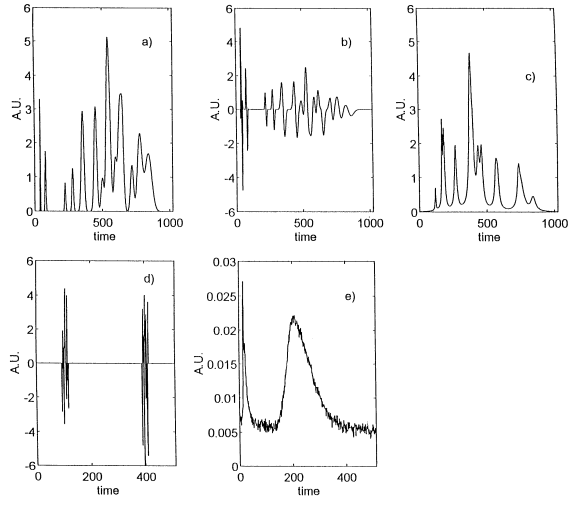

Fig. 3. (a) Signal 1: sum of Gaussian functions. (b) Signal 2: derivative of signal 1. (c) Signal 3: sum of Lorentzian functions. (d) Signal 4: spikes signal. (e) Experimental GrFFF fractogram.

3.3. Filters

Although there are many types of wavelets [16], we restrict ourselves in this study to the Daubechies, Coiflet and Symmlet families of filters.

3.4. Computation

The computations were performed with Matlab for Windows version 4.0 [17], using our own programs. The Matlab Toolbox for wavelets, WavBox3 [18], was used for the library of wavelet filter coefficients.

All the simulated signals contain 1024 points. In signals 1 and 3, the number of peaks was set equal to 17. The peak parameters: peak position, peak width and peak height were uniformly distributed random numbers in the interval 100–900, 10–30 and 1–3, respectively.

Signal 2 was obtained as the first derivative of the signal 1. Signal 4 consists of two sets of 10 consecutive spikes located at 195–205 and 795–805 having alternating positive and negative height selected as uniformly distributed random number in the interval −6–6.

The SNR were 0.065, 0.043, 0.052 and 0.054 for signals 1, 2, 3 and 4, respectively.

4. Results and discussion

4.1. Filter optimization

Each of the wavelet functions has different characteristics. The wavelet filter which is optimal for a

given signal, is not necessarily the best for another type of signal. Filtered signals are usually employed to obtain a quantitative analytical information, which depends strongly on the shape of the signal (e.g., the area and height of a spectroscopic or chromatographic response). Improper choice of filter can cause distortions and artefacts in the reconstructed signal. To illustrate the influence of the wavelet function in signal denoising, the reconstructed and original signals are reported in Fig. 4. The reconstructed signals were obtained by using (1) the same threshold and (2) the same decomposition level. Specifically, soft thresholding with universal threshold and the last level of decomposition was used for all of the signals. The RSE values computed from the reconstructed signals are shown in Table 1. The results clearly demonstrate the effect of filter choice on the RSE.

In our study, we decided to use both the MDL and CV approaches to optimize filter choice, and to compare their performance for the first three signals studied.

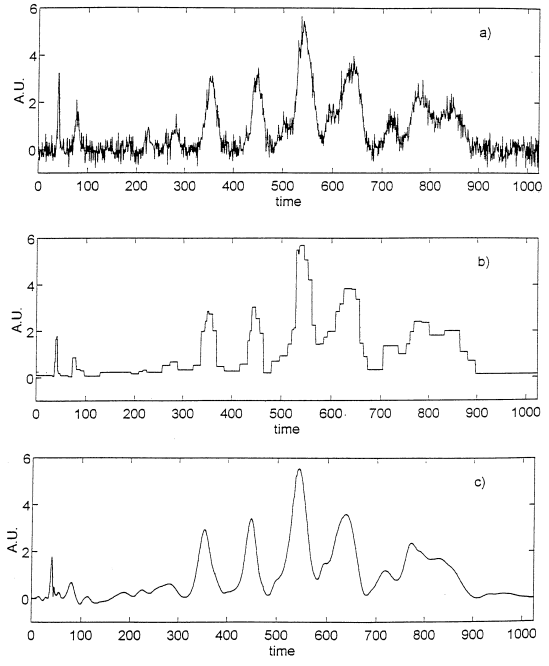

Fig. 4. Influence of filter in signal denoising: signal 1, SNR = 0.065, soft thresholding, universal threshold, and last decomposition level. (a) Noisy data with SNR = 0.4. (b) Reconstructed function Haar filter. (c) Reconstructed function Daubechies 14 filter.

Table 1
RSE obtained using soft thresholding, universal threshold and the last level of decomposition

	Filter	RSE		
		Signal 1	Signal 2	Signal 3
1	Daubechies 2	103.822	78.9043	49.0480
2	Daubechies 4	79.2668	57.3952	38.2101
3	Daubechies 6	87.6230	63.1820	29.5685
4	Daubechies 8	79.9815	57.0649	32.1946
5	Daubechies 10	85.2309	48.9182	41.1821
6	Daubechies 12	79.0661	52.3289	34.2275
7	Daubechies 14	70.4693	47.7125	34.5491
8	Daubechies 16	74.5797	51.3603	32.2474
9	Daubechies 18	74.5323	51.8706	36.4452
10	Daubechies 20	79.9821	50.2982	35.4498
11	Coiflet 1	83.6234	54.3511	38.9230
12	Coiflet 2	76.6838	54.6076	29.9158
13	Coiflet 3	75.9368	47.3356	32.7261
14	Coiflet 4	78.9589	47.5750	37.2475
15	Coiflet 5	80.8651	48.1325	39.9081
16	Symmlet 4	78.6656	51.6501	30.8710
17	Symmlet 5	69.1833	51.1983	33.0240
18	Symmlet 6	81.6876	48.4039	32.4407
19	Symmlet 7	77.6033	57.6496	37.1817
20	Symmlet 8	78.3433	48.3032	36.8043
21	Symmlet 9	75.3502	52.2411	38.7663
22	Symmlet 10	91.1816	52.0670	36.4037

Plots of the MDL cost function as function of the RSE are reported in Fig. 5. The results obtained by means of the CV method, applied to filter selection, are shown in Fig. 6.

As one can see from these data, the CV method is slightly better than the MDL method. Of course, for real signals, RSE cannot be estimated. The optimal filter must be selected using the MDL or CV criterion values (minima). For signals 1 and 3, it was found that the minimum RSE and ISE correspond to the same filter while for signal 2, most of the filters yield comparable RSEs and ISE. When using the MDL method, we found large differences in the MDL values which correspond to a rather stable RSE value. Both these functions reach their minima for the same filter only in the case of signal 2. The CV method allows optimization of the threshold value, decomposition level and filter. It can be applied to both hard and soft thresholding procedures, whereas in the MDL approach only hard thresholding can be used. The evaluation is focused on a proper balance between efficient signal compression and denoising.

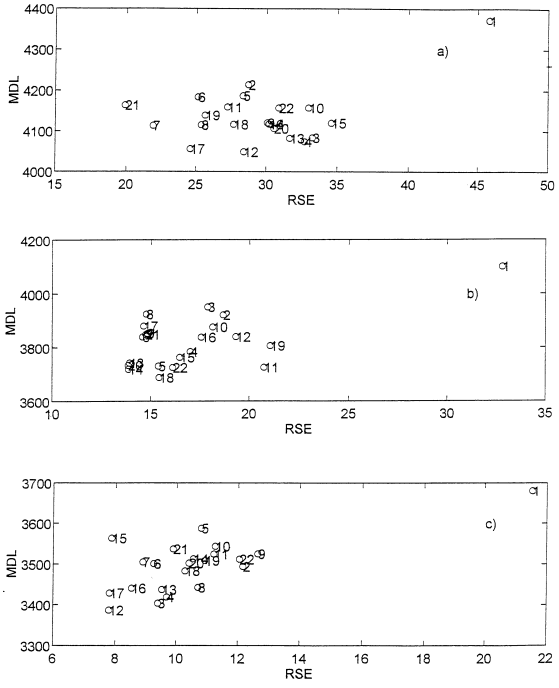

Fig. 5. Filter optimization by using Saito method. The numbers from 1 to 10 represent the Daubechies filter family, from 11 to 15 to Coiflet filter family and from 16 to 22 to Symmlet filter family (see Table 1). (a) MDL vs. RSE signal 1. (b) MDL vs. RSE signal 2. (c) MDL vs. RSE signal 3.

4.2. Optimization of decomposition level

No procedure for choosing the best level of DWT decomposition has yet been proposed and, in most cases, the last decomposition level has been used. In Fig. 7, the differences between the original and the reconstructed signals are plotted for signal 1. In this example, denoising was performed using two different decomposition levels, the soft threshold operation with a constant threshold value and a fixed wavelet filter. The RSE results, reported in Tables 2 and 3, show that the RSE strongly depends on the decomposition level.

For all the analyzed signals, minima are present in both RSE and ISE vs. the decomposition level. The best level found by the CV method is the same as that for which the minimum error, between the noisy and ideal data (RSE), is observed. Moreover, it can be seen that for the cases studied, level 3 or 4 generally yields the lowest RSE. To demonstrate the dependence on filter selection and decomposition level, the ISE values vs. decomposition level plots for a few selected filters are presented in Fig. 8.

It can be seen that for the signals studied, the influence of a filter on RSE is less important than the decomposition level.

4.3. Threshold optimization

Finally, to compare two different denoising policies, the hard and soft thresholding approaches were applied to the noisy signals. The denoised signals are plotted in Fig. 9.

As one can see from this comparison, the RSE obtained by using the universal threshold value is lower when applying the hard rather than the soft thresholding (both with the universal threshold value). This trend is more evident when the spike signal was considered.

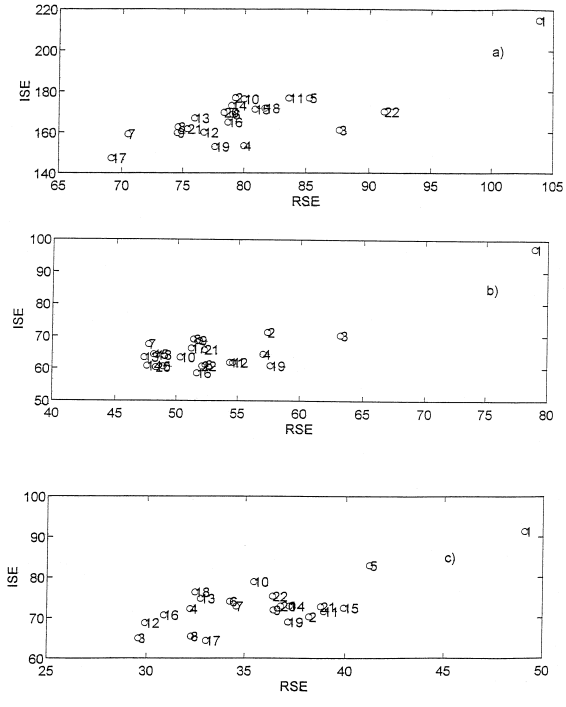

Fig. 6. Filter optimization by using CV method. The numbers from 1 to 10 represent the Daubechies filter family, from 11 to 15 to Coiflet filter family and from 16 to 22 to Symmlet filter family (see Table 1). (a) ISE vs. RSE signal 1. (b) ISE vs. RSE signal 2. (c) ISE vs. RSE signal 3.

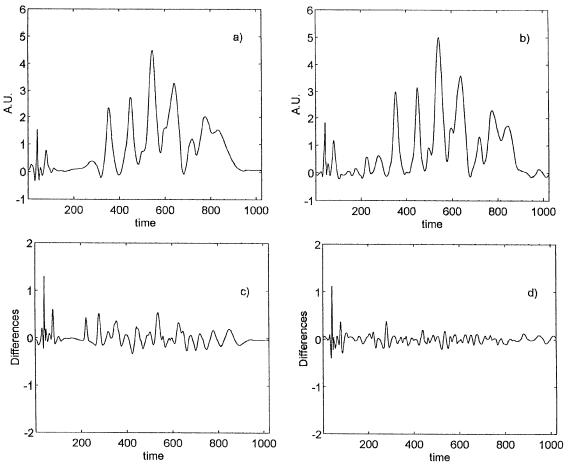

Fig. 7. Level of decomposition: signal 1, SNR = 0.065, Symmlet 5, TrU. (a) Level = 8 maximum level of decomposition. (b) Level = 3. (c) Differences between the reconstructed signal after TrU soft thresholding at decomposition level 8 and the pure data. (d) Differences between the reconstructed signal after TrU soft thresholding at decomposition level 3 and the pure data.

It is obvious that the threshold value affects the quality of the reconstructed signal. For this reason, many algorithms based on statistical properties of the noise were developed to estimate the best threshold value. In this study, we focus on the optimization of the threshold value that can give the best compromise between the denoising and RSE. To do this, we applied the CV procedure in soft thresholding. The ISE vs. the threshold value are reported in Fig. 10.

According to Nason [6], this procedure can improve the denoising operation, and is more conservative for the signal reconstruction. The results given in Fig. 11 act in support of the above statement.

Table 2
RSE as a function of the level of decomposition. Parameter setting: soft thresholding and universal threshold

Decomposition level	Signal 1	Signal 2	Signal 3
	Symmlet 5	Symmlet 4	Daubechies 4
Level 2	28.8564	14.7022	10.0008
Level 3	19.1984	14.6975	6.2813
Level 4	26.7803	28.9324	12.3708
Level 5	50.3305	41.8609	18.5164
Level 6	60.1936	47.3039	23.7150
Level 7	65.5675	49.5522	26.8353
Level 8	67.8609	51.2810	29.0246
Level 9	69.1833	51.6501	29.5685

Table 3
RSE as a function of the level of decomposition. Parameter setting: Hard Thresholding and Universal Threshold

Decomposition level	Signal 1	Signal 2	Signal 3
	Symmlet 5	Symmlet 4	Daubechies 4
Level 1	49.8328	22.5520	21.1474
Level 2	29.3237	13.6732	10.6010
Level 3	17.5134	10.6324	7.4110
Level 4	18.8959	13.8363	9.2841
Level 5	22.2457	15.8403	10.5900
Level 6	24.0170	16.7919	11.9240
Level 7	24.0862	16.9930	12.1607
Level 8	24.5819	17.1992	12.1584
Level 9	24.6093	17.5579	12.1584

As one can see, RSE is higher using the universal threshold value than the value obtained by means of CV procedure. The threshold values selected by CV were applied to the noisy data, and the difference between the reconstructed signal and the ideal one are reported in Fig. 12.

The universal threshold value seems to generate a higher difference between the denoised and ideal signal than the CV threshold.

The CV procedure was applied to each level of the DWT of the signal and a minimum value was always found as shown in Fig. 11. It can be seen that the last decomposition level (noted 1, Fig. 11a) is not strongly dependent on the threshold value and the minimum of

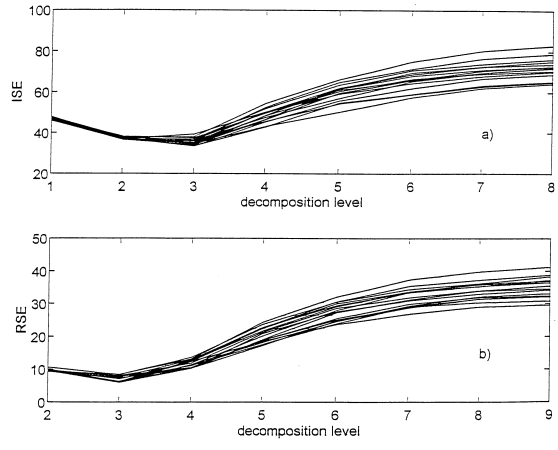

Fig. 8. Effect of the filter choice on the error of reconstruction obtained by using soft thresholding in different decomposition levels. Signal 3, SNR = 0.053, TrU. (a) ISE vs. decomposition level. (b) RSE vs. decomposition level.

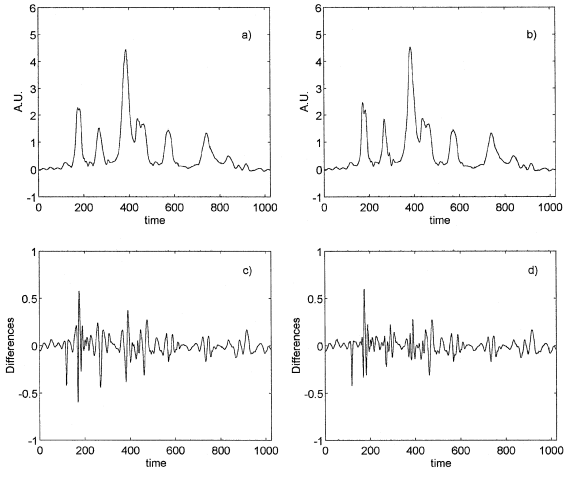

Fig. 9. Hard and soft thresholding: signal 3, SNR = 0.052, Daubechies 4, level = 3, TrU. (a) Reconstructed signal after soft thresholding. (b) Reconstructed signal after hard thresholding. (c) Differences between the reconstructed signal after soft thresholding and the pure data. (d) Differences between the reconstructed signal after hard thresholding and the pure data.

the RSE is higher than for all the other levels. The decomposition level and the threshold value are mutually related, and the CV procedure can be used to optimize both.

The CV method and the hard thresholding approach were applied together to the studied signals.

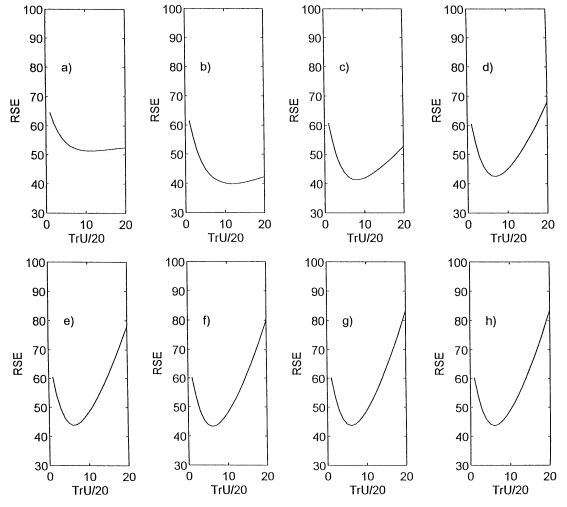

Fig. 10. ISE vs. threshold level, the abscissa values are expressed as unit of TrU/20. Signal 3, SNR = 0.052, Daubechies 4, soft thresholding. Figures from (a) to (h) correspond to the decomposition level from 1 to 8.

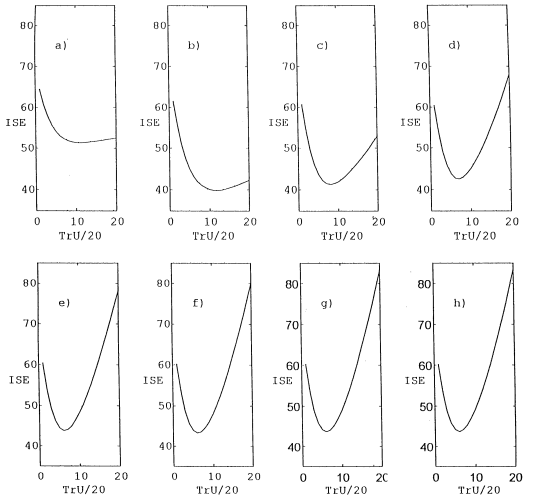

Fig. 11. RSE vs. threshold level, the abscissa values are expressed as unit of TrU/20. Signal 3, SNR = 0.052, Daubechies 4, soft thresholding. Figures from (a) to (h) correspond to the decomposition level from 1 to 8.

As one can see in Fig. 13, using hard thresholding policy, the universal threshold value does not overestimate the level of noise. The minimum of the ISE value is found near to the universal threshold value. In the threshold range near the TrU value, the decomposition level seems to be the more important parameter in the hard thresholding approach.

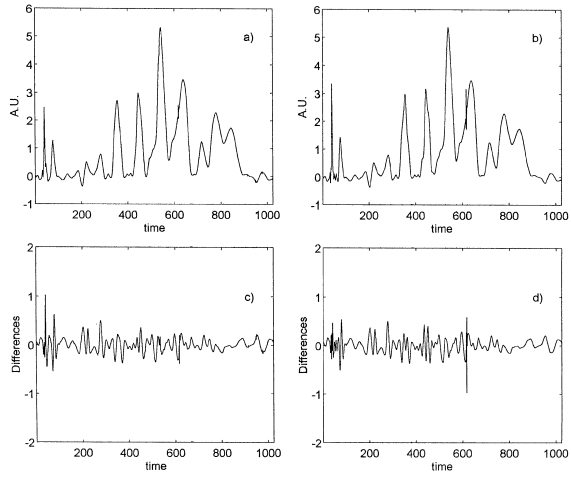

Fig. 12. Threshold optimization. Signal 1, SNR = 0.065, Symmlet 5, soft thresholding. (a) Reconstructed signal after soft thresholding: universal threshold. (b) Reconstructed signal after soft thresholding: CV threshold. (c) Differences between the reconstructed function and the pure data: universal threshold. (d) Differences between the reconstructed function and the pure data: CV threshold.

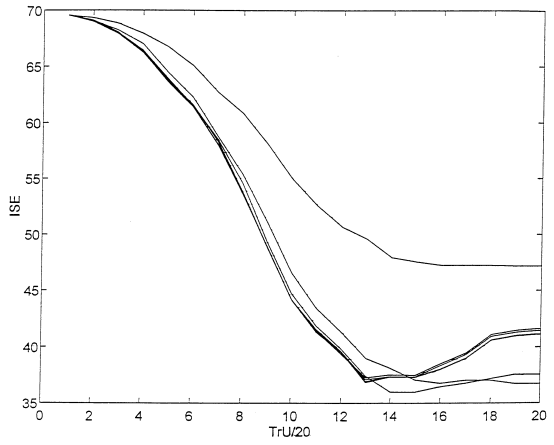

Fig. 13. Hard Thresholding optimization by means of CV·ISE vs. threshold value for different decomposition levels. Signal 3, SNR = 0.052, Daubechies 4.

In Fig. 14 and in Tables 4 and 5, the results obtained with both soft and hard thresholding with optimal parameter settings are reported, and the RSE obtained from these two approaches are shown to be comparable.

The whole procedure was applied to signals having different signal to noise ratio (from 0.01 to 0.1) and the results obtained are reported in Fig. 15. For a particular signal, it is interesting to note that the positions of the minima do not strongly depend on the signal to noise ratios. The abscissa represents the ratios of TrU which depend on the SNR and is computed separately for each signal reported in Fig. 15.

For the cases studied here, it is observed that the best results in terms of RSE correspond to the level three of the DWT decomposition and to a threshold value from 0.4 to 0.6 times the universal threshold value reported in Table 6.

To demonstrate possible limitations of the CV procedure, the case of signal 4 was used. For this type of signal (with very narrow peaks), the corresponding even and odd functions (see Section 2) are considerably different from the original signal and hence the results obtained are meaningless.

4.4. Real data

In Fig. 3e, the experimental fractogram obtained for silica gel particles by GrFFF is reported, it can be seen that the signal is affected by noise and the SNR of the signal was 0.032. GrFFF is a subset of field flow fractionation (FFF) techniques, which employs the gravitational field to develop separation of micron-sized particles [19]. FFF is a separation method that resembles conventional chromatography in many aspects [20]. Therefore, FFF output signals can be

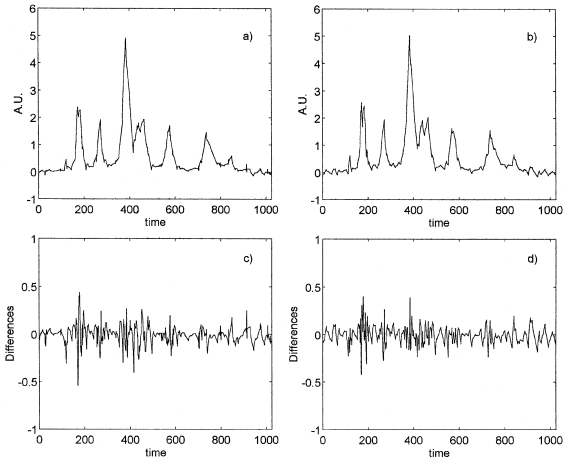

Fig. 14. Reconstructed signals obtained by optimization of the filter parameters: signal 3, SNR = 0.052, Daubechies 4. (a) Soft thresholding, level = 3, threshold = 10 TrU/20. (b) Hard thresholding, level = 4, threshold = 16 TrU/20. (c) Differences between the reconstructed function and the pure data: soft thresholding. (d) Differences between the reconstructed function and the pure data: hard thresholding.

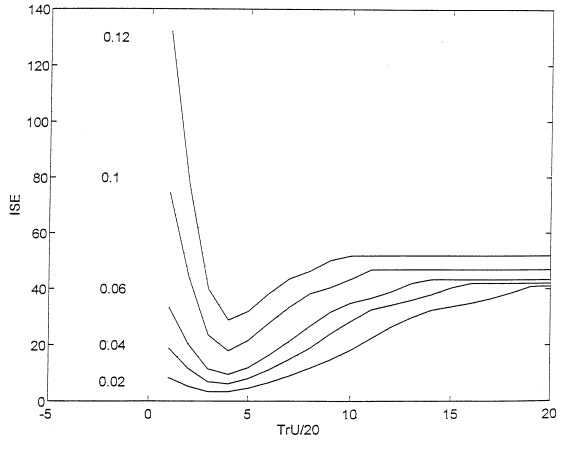

Fig. 15. CV method applied to different amount of noise added to signal 3. ISE vs. the threshold level. Daubechies 4, level = 3.

Table 4
Soft and hard thresholding: the level of decomposition was set equal to 3 and the threshold equal to the universal threshold value

Signal	Filter	ISE		RSE	
		Hard	Soft	Hard	Soft
Signal 1	Symmlet 5	80.5657	86.1442	18.8959	26.7803
Signal 2	Symmlet 4	45.4586	52.8029	13.8363	28.9324
Signal 3	Daubechies 4	37.5393	39.3791	9.2841	13.6150

Table 5
Hard and soft threshold optimization. Parameter setting: decomposition level equal to 3

Signal	Filter	ISE	RSE	Threshold
Signal 1	Symmlet 5	78.5484	17.8813	Soft CV threshold = 8/20 TrU
		78.3221	18.2132	Hard CV threshold = 16/20 TrU
Signal 2	Symmlet 4	41.2889	12.7034	Soft CV threshold = 9/20 TrU
		42.5126	12.6381	Hard CV threshold = 16/20 TrU
Signal 3	Daubechies 4	34.5303	7.4979	Soft CV threshold = 11/20 TrU
		35.8631	7.6132	Hard CV threshold = 14/20 TrU

analyzed by numerical methods that are similar to those used for conventional chromatography. To test our method, we used FFF data rather than signals from more conventional chromatographic techniques because typical FFF responses show quite distinct features that make denoising a challenging task. The first difference, as can be seen in Fig. 3e, is that an FFF experiment yields at least two peaks, the peak widths of which are very different. Here, the two peaks correspond to the void volume (V_0) and to the analyte response and the half-height peak widths of the peaks were 45 s for the void peak and 490 s for the analyte. Both of them are necessary to analyze the substance. In FFF, more information can be extracted from the void volume band than in chromatography. In fact, a retention ratio $R = V_o/V_R$ value higher than 1 is sometimes observed (V_R is the retention volume). This happens for sample species which travel more rapidly than the mobile phase. Peaks corresponding to such components are often only partially resolved from the void peak. The GrFFF experiment shown in Fig. 3e displays a void volume profile that is seen to be composed of at least two bands. In order to extract analytical information from the bands eluting in the retention interval corresponding to the void volume, the signal structure should not be altered by denoising. The second feature which distinguishes FFF bands is that unretained bands have signal frequency values always very much higher (i.e., the peak widths are much more narrow) than those of sample components. Moreover, the polydispersity contribution (always present in partic-

Table 6
Threshold optimization. Parameter setting: soft thresholding, level of decomposition equal to 3

Signal	Filter	ISE	RSE	Threshold
Signal 1	Symmlet 5	78.5484	17.8813	CV threshold = 8/20 TrU
		86.1442	25.7516	TrU
Signal 2	Symmlet 4	41.2889	12.7034	CV threshold = 9/20 TrU
		52.8029	26.9413	TrU
Signal 3	Daubechies 4	34.5303	7.4979	CV threshold = 11/20 TrU
		39.3791	13.0839	TrU

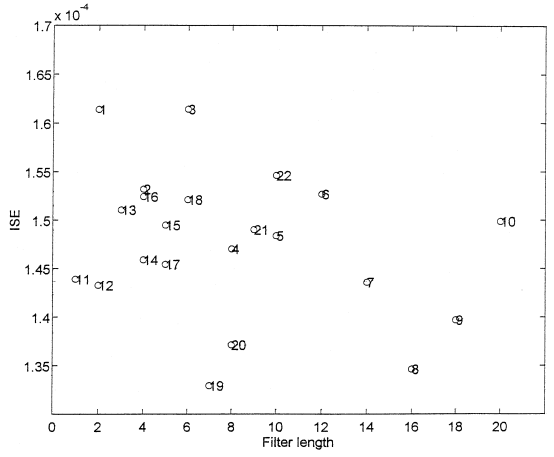

Fig. 16. Filter optimization of the experimental signal (Fig. 3d). ISE vs. filter length. The numbers from 1 to 10 represent the Daubechies filter family, from 11 to 15 to Coiflet filter family and from 16 to 22 to Symmlet filter family (see Table 1).

ulate samples) to total band broadening [21] leads to signal bands which are much wider than those of unretained components. Altogether, for FFF signal denoising, conventional methods like FFT can be inadequate to maintain global signal features.

The proposed CV methodology is here applied to GrFFF data in order to filter the noise component and maintain global signal features. The selection of the wavelet filter was performed as previously described and the results are plotted in Fig. 16.

It can be seen that the second member of the Symmlet family seems to outperform all the other wavelet filters, so it was selected to decompose the signal. Subsequently, the threshold level and the de-

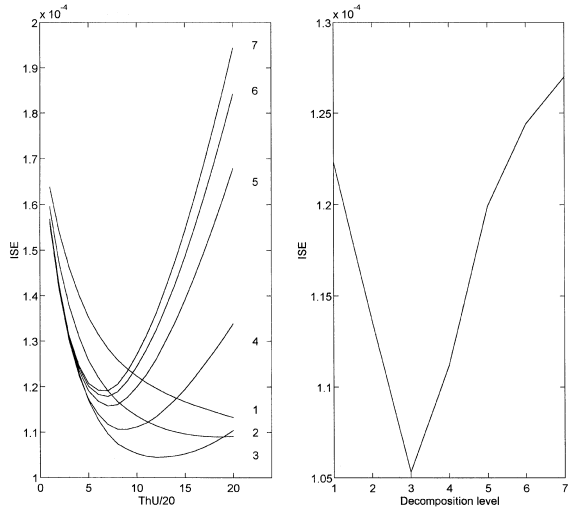

Fig. 17. Threshold value and decomposition level optimization for the experimental signal. (a) Soft thresholding. ISE vs. the threshold value for different level of decomposition. (b) ISE vs. the level of decomposition at constant threshold: 12 TrU/20.

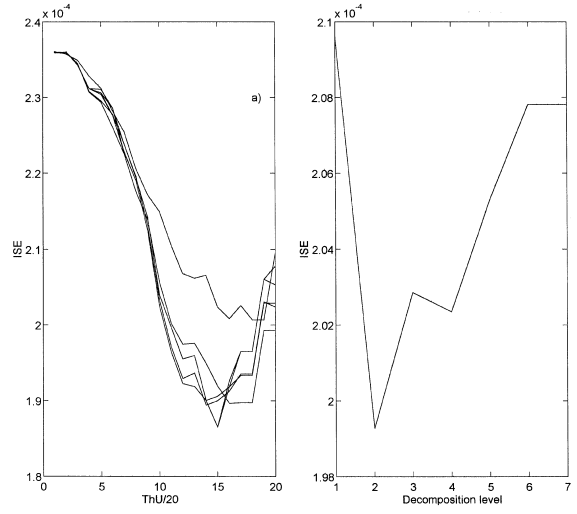

Fig. 18. Level and threshold selection: hard thresholding. (a) ISE vs. the threshold value for different level of decomposition. (b) ISE vs. the level of decomposition at constant threshold: 15 TrU/20.

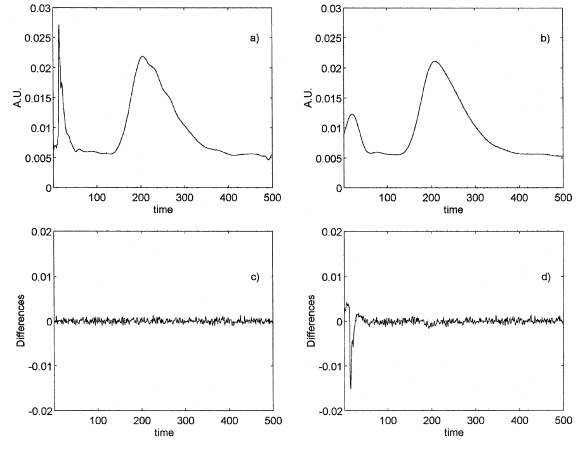

Fig. 19. Filtered experimental GrFFF signal: (a) Wavelet. (b) Low pass filter in Fourier domain. (c) Differences between the reconstructed function and the noisy experimental data: wavelet filtering. (d) Differences between the reconstructed function and the noisy experimental data: Fourier filtering.

composition level were optimized in a single step. The results obtained by applying a soft thresholding policy are reported in Fig. 17.

It is easily seen that the ISE value depends on the threshold value and for a threshold value near to the universal one, the selection of the level is the most important parameter. In the range between 0.5 and 1 ThU, the CV error is larger when different decomposition levels are compared than when small differences in the threshold value are considered. Moreover, it can be seen that for the best decomposition level, the RSE in the considered range is almost constant.

The same results were also obtained when the hard thresholding policy was applied (see Fig. 18).

In this case, the best threshold value corresponds to 3/4 TrU, and the best level of decomposition was found equal to the second decomposition level thresholding. It appears therefore that in filtering the noise, the decomposition level is the most important parameter to optimize.

The filtered signal is reported in Fig. 19 together with the same experimental signal filtered by FFT. The frequency cut-off after forward transform was determined as the frequency beyond which the standard deviation of the power spectrum amplitudes remain constant (i.e., 0.0045 Hz). As one can see, although the noise was removed, only with wavelets using CV no artefact was introduced and the total signal structure was very accurately maintained across both the low and high retention zones.

5. Conclusion

For the cases studied in this article, the MDL and CV approaches seem to be promising tools for filter optimization. The CV approach can also be applied to select the decomposition level in both soft and hard thresholding. The level of decomposition is sometimes more important than the filter. It seems that the last level of the DWT decomposition does not generally give the best performance.

Hard thresholding usually is more robust than soft thresholding, particularly for sharp signals. By means of the CV technique, the threshold value can be optimized, in order to increase the signal to noise ratio without introducing artefacts.

Acknowledgements

We thank the Fonds voor Wetenschappelijk Onderzoek, the DWTC, and the Standards Measurement and Testing program of the EU for financial assistance.

References

[1] C.K. Chen, Introduction to Wavelets, Academic Press, Boston, 1991.
[2] C.R. Mittermayer, S.G. Nikolov, H. Hutter, M. Grasserbauer, Wavelet denoising of Gaussian peaks: a comparative study, Chemom. Intell. Lab. Syst. 34 (1996) 187–202.
[3] B. Walczak, B. van den Bogaert, D.L. Massart, Application of wavelet packet transform in pattern recognition of near-IR data, Anal. Chem. 68 (1996) 1742–1747.
[4] D.L. Donoho, I.M. Johnstone, Ideal spatial adaptation via wavelet shrinkage, Biometrika 81 (1994) 425–455.
[5] D.L. Donoho, Denoising by soft-thresholding, IEEE Trans. Information Theory 41 (3) (1995) 613–627.
[6] G.P. Nason, Wavelet regression by cross-validation, J. R. Stat. Soc. Ser. B: Methodological 58 (2) (1996) 463–479.
[7] H. Fang, H-Y. Chen, Wavelet analyses of electroanalytical chemistry responses and an adaptive wavelet filter, Anal. Chim. Acta 346 (1997) 319–325.
[8] N. Saito, Simultaneous noise suppression and signal compression using a library of orthonormal bases and the minimum description length criterion, in: E. Foufoula-Georgiou, P. Kumar (Eds.), Wavelets in Geophysics, Academic Press, New York, 1994.
[9] B. Vidakovic, P. Muller, Wavelets for Kids, Discussion Papers, Duke University, 1991.
[10] S.G. Mallat, A theory for multiresolution signal decomposition: the wavelet representation, IEEE Trans. Pattern Anal. Machine Intell. 11 (1989) 674–693.
[11] B.K. Alsberg, A.M. Woodward, M.K. Winsor, J. Rowland, D.B. Kell, Wavelet denoising of infrared spectra, Analyst 122 (1997) 645–652.
[12] V.J. Barclay, R.F. Bonner, I.P. Hamilton, Application of wavelet transform to experimental spectra: smoothing, denoising, and data set compression, Anal. Chem. 69 (1997) 78–90.
[13] P. Reschiglian, G. Torsi, Determination of particle size distribution by gravitational field flow fractionation: dimensional characterization of silica particles, Chromatographia 40 (1995) 67.
[14] P. Reschiglian, D. Melucci, G. Torsi, Experimental study on the retention of silica particles in gravitational field flow fractionation: effects of the mobile phase composition, J. Chromatogr. A 245 (1996) 740.
[15] D. Melucci, G. Gianni, G. Torsi, A. Zattoni, P. Reschiglian, Experimental analysis of second-order effects on gravita-

tional field flow fractionation retention of silica particles, J. Liq. Chrom. Rel. Technol. 20 (1997) 2615.
[16] G. Strang, T. Nguyen, Wavelet and Filter Banks, Wellesley–Cambridge Press, Wellesley, MA, 1996.
[17] Matlab, The MathWorks, 1992.
[18] C. Taswell, WavBox 3, available via taswell@sccm.stanford.edu.
[19] J.C. Giddings, Field flow fractionation: separation and characterization of macromolecular–colloidal–particulate materials, Science 260 (1993) 1445.
[20] J.C. Giddings, Unified Separation Science, Chap. 8, Wiley, 1991.
[21] J.C. Giddings, Y.H. Yoon, K.D. Caldwell, M.N. Myers, M.E. Hovings, Evaluation and comparison of gel permeation chromatography and thermal field flow fractionation for polymer separations, Sep. Sci. 10 (1975) 447.