



Wavelet-based spectral analysis

Vu Dang Hoang *



Department of Analytical Chemistry and Toxicology, Hanoi University of Pharmacy, 13–15 Le Thanh Tong, Hanoi, Vietnam

ARTICLE INFO

Keywords:

Chemometrics
Data compaction
Modeling
Noise removal
Resolution enhancement
Signal processing
Spectra
Spectral analysis
Wavelet
Wavelet transform

ABSTRACT

Wavelets are a topic of pure mathematics. But over the past decade, they have shown great promise and are now being adapted for a vast number of signal-processing applications. One of the main advantages of wavelet analysis is the amount of information that can be extracted from a signal. It has demonstrated unprecedented success in terms of asymptotic optimality, spatial adaptivity and computational efficiency. Applications of wavelet transform and wavelet-packet transform in spectral analysis from 2002 to 2013 are reviewed in this article, clearly stating that wavelet methods significantly outperform other traditional methods of signal processing.

© 2014 Elsevier B.V. All rights reserved.

Contents

1. Introduction	144
2. Theoretical background	144
3. Noise removal and resolution enhancement	145
4. Data compaction	148
5. In combination with chemometrics	149
6. Conclusion	152
Acknowledgements	152
References	152

1. Introduction

Signal processing incorporates all aspects of operations on or analysis of analog and digitized signals, representing time-varying or spatially varying physical quantities. In the past decade, this field has unambiguously become a burgeoning area of research and design.

Wavelet Transform (WT) is one of the recent techniques for processing signals. It is defined as mathematical functions that cut up data into different frequency components, and then study each component with a resolution matched to its scale [1]. Wavelet signal processing is different from other signal-processing methods because of the unique properties of wavelets (i.e., wavelets can be symmetric or asymmetric, sharp or smooth, regular or irregular). It can represent signals sparsely, capture the transient features of signals, and enable signal analysis at multiple resolutions. Interest in

wavelets and their potential application has resulted in an explosion of papers in life sciences and physical sciences.

The goal of this review is to comment on all outstanding and state-of-the-art articles of wavelet-based spectral analysis published in the period 2002–13 in the context of developments and challenges yet to be addressed.

2. Theoretical background

Given a time-varying signal, $f(t)$, WT consists of computing coefficients, which are inner products of the signal and a family of wavelets. In other words, WT decomposes a signal into localized contributions (details and approximations) labeled by a scale and a position parameter. The spatial localization of wavelets makes them more suited to present a large class of signals [i.e., spectra can be represented by far fewer wavelets than sinusoids obtained from Fourier Transform (FT)]. WT is categorized into continuous wavelet tools and discrete wavelet tools. Usually, continuous wavelet tools are used for signal analysis, such as self-similarity analysis and time-frequency analysis. However, discrete wavelet tools are employed for both signal analysis and signal processing, such as noise

* Tel.: +84 438-254539; Fax: +84 438-264464.
E-mail address: hoangvd@hup.edu.vn (Vu Dang Hoang).

reduction, data compression, and peak detection. A more detailed introduction to the theory and potential application of WT is available in books [2–4].

Generally, the Continuous Wavelet Transform (CWT) of a signal, $f(t)$, can be written as:

$$W_f(\tau, s) = \frac{1}{\sqrt{|s|}} \int_{-\infty}^{+\infty} f(t) \Psi\left(\frac{t-\tau}{s}\right) dt$$

where τ and s are the so-called translation (or time location) factor and the scaling (or dilation) factor, respectively. Factor $|s|^{-1/2}$ is for energy normalization across the different scales, whereas $\Psi_{\tau,s}(t)$ can be obtained by dilations and translations of a single function $\Psi(t)$, the so-called “mother wavelet”, as follows:

$$\Psi_{\tau,s}(t) = \frac{1}{\sqrt{|s|}} \Psi\left(\frac{t-\tau}{s}\right).$$

Thus, the original signal can be exactly reconstructed from the wavelet coefficients by Inverse Wavelet Transform (IWT):

$$f(t) = \frac{1}{C_\Psi} \int_0^{+\infty} \int_{-\infty}^{+\infty} W_f(\tau, s) \Psi_{\tau,s}(t) d\tau \frac{ds}{s^2}$$

where C_Ψ is defined as:

$$C_\Psi = \int_0^{+\infty} \frac{|\Psi(\omega)|^2}{\omega} d\omega$$

and $\Psi(\omega)$ is the FT of the mother wavelet.

If $s = s_0^j$ and $\tau = k\tau_0 s_0^j$ ($j, k \in \mathbb{Z}, s_0 \neq 0$), the wavelet can be re-written as:

$$\Psi_{j,k}(t) = s_0^{-j/2} \Psi(s_0^{-j}t - k\tau_0)$$

This is used in the computation of Discrete Wavelet Transform (DWT), where $s_0 = 2$ and $\tau_0 = 1$ (dyadic dilations and translations) are generally used. DWT is very useful in compressing data, because it does not change the amount of data (the total number of coefficients is equal to that of data points of the original signal), but the relevant information is often stored in only very few coefficients.

The basic idea of wavelet analysis is that of multi-resolution signal decomposition introduced by Mallat (i.e., the simultaneous appearance of a signal on multiple scales to study its various features) [5]. This algorithm consists of a series of successive decompositions of the signal (with length 2^n) into two components: “detail coefficients D_j ” and “approximation coefficients A_j ”, both with a reduced size of 2^{n-j} , where j is the decomposition level. At each level, the input signal is decomposed by high-pass filters to record the high-frequency components and low-pass filters to extract the low-frequency components for the next scale. The procedure is repeated with sets of high-pass and low-pass filters until at a prescribed level j is reached ($j \leq n$).

By doing so, signal $f(t)$ can be written as a limit of successive approximations at different approximation subspaces, while each of the approximations at subspaces is a smoother version of $f(t)$. According to the Mallat’s algorithm, in the case of CWT, a signal can be decomposed with wavelets as follows:

$$f(t) = \sum_{k=-\infty}^{\infty} A_{j,k} \phi_{j,k}(t) + \sum_{j=1}^n \sum_{k=-\infty}^{\infty} D_{j,k} \Psi_{j,k}(t)$$

where $\phi_{j,k}(t)$ and $\Psi_{j,k}(t)$ are called scaling functions and wavelet functions, respectively. $A_{j,k}$ and $D_{j,k}$ are the approximation coefficients and the detail coefficients mentioned in the Mallat’s algorithm. At each successive scale (or decomposition level), only high-frequency

information (noise) is retained in the details, while the low-frequency information (signal features) is retained in the approximations. The de-noised signal can then be reconstructed with the new estimated wavelet coefficients:

$$f^*(t) = \sum_{k=-\infty}^{\infty} A_{n_{opt},k} \phi_{j,k}(t) + \sum_{j=1}^{n_{opt}} \sum_{k=-\infty}^{\infty} D_{j,k}^* \Psi_{j,k}(t)$$

where $A_{n_{opt},k}$ are the approximation coefficients at the optimal decomposition level n_{opt} , and $D_{j,k}^*$ are the detail coefficients retained.

3. Noise removal and resolution enhancement

Noise distortion can occur during spectral signal acquisition and transmission. Unfortunately, most optical instruments are often limited in sensitivity and specificity by noise fringes superimposed on the recorded spectra. To remove unwanted components, WT was investigated for automated implementation of baseline removal. It was noted that, with synthetic data sets designed to exemplify vibrational spectroscopic signals, WT could filter out high-frequency noise but was less successful with high signal-to-noise ratio (SNR) spectra and in congested regions [6]. In addition, the applicability of translation-invariant WT in filtering light signals from spectrophotometers was studied [7]. The main advantages of this technique are a substantial increase in the SNR and preservation of the spectral peak location and width, as compared to Gaussian, Wiener and orthogonal wavelet filters using a fixed threshold. On the other hand, laser-induced breakdown spectroscopy signals could be de-noised by an extension of the Donoho’s scheme, which uses a redundant form of WT and an adaptive threshold-estimation method [8]. Adaptive stationary wavelet filtering via variable thresholding offers noise-suppression improvement in parallel with signal preservation, which is superior to that offered by DWT thresholding and Gaussian filtering.

In another study, a methodology based on an improved, second-generation Adaptive Wavelet Transform (AWT) algorithm was presented for Raman spectral denoising and baseline elimination [9]. This methodology uses a spectrally adapted lifting scheme to generate an infinite basis of wavelet filters from a single conventional wavelet, and then finds the optimal one. As a result, it is more efficient than DWT because it enables the custom design of wavelet filters according to the topological characteristics of Raman spectra at hand. CWT was also suitable for removing the variant background of NIR diffuse reflectance spectroscopy in pharmaceutical analysis [10]. Nevertheless, DWT could be successfully used for denoising short-wave NIR reflectance spectra when the preprocessed data were used as the input of the Support Vector Machine (SVM) system, especially when the number of samples was small [11].

A wavelet-based denoising technique was applied to a quantum-cascade laser spectrometer for *in situ* and real-time atmospheric trace-gas measurements [12]. The wavelet digital-filter technique in post-signal processing proved to give better measurement precision and higher detection sensitivity without reducing the fast temporal response, as compared to other commonly used digital-filter techniques (i.e., Kalman filter, Wiener filter and moving average). Because both wavelets and Kalman filters could handle non-stationary signals, a more promising technique based on incorporating both WT and Kalman filter (i.e., a wavelet-based Kalman filter) would be more effectively applied for trace-gas sensors. It was also found that CWT has better space-time resolution and is relatively simpler to perform, as compared to DWT, for the elimination of the fluctuating background in NIR spectra [13]. WT produces better calibration models of NIR single-beam spectra, with improvements in concentration prediction of the order of 30% being realized relative to models based on second-derivative spectra or spectra preprocessed with simple additive and multiplicative scaling correction

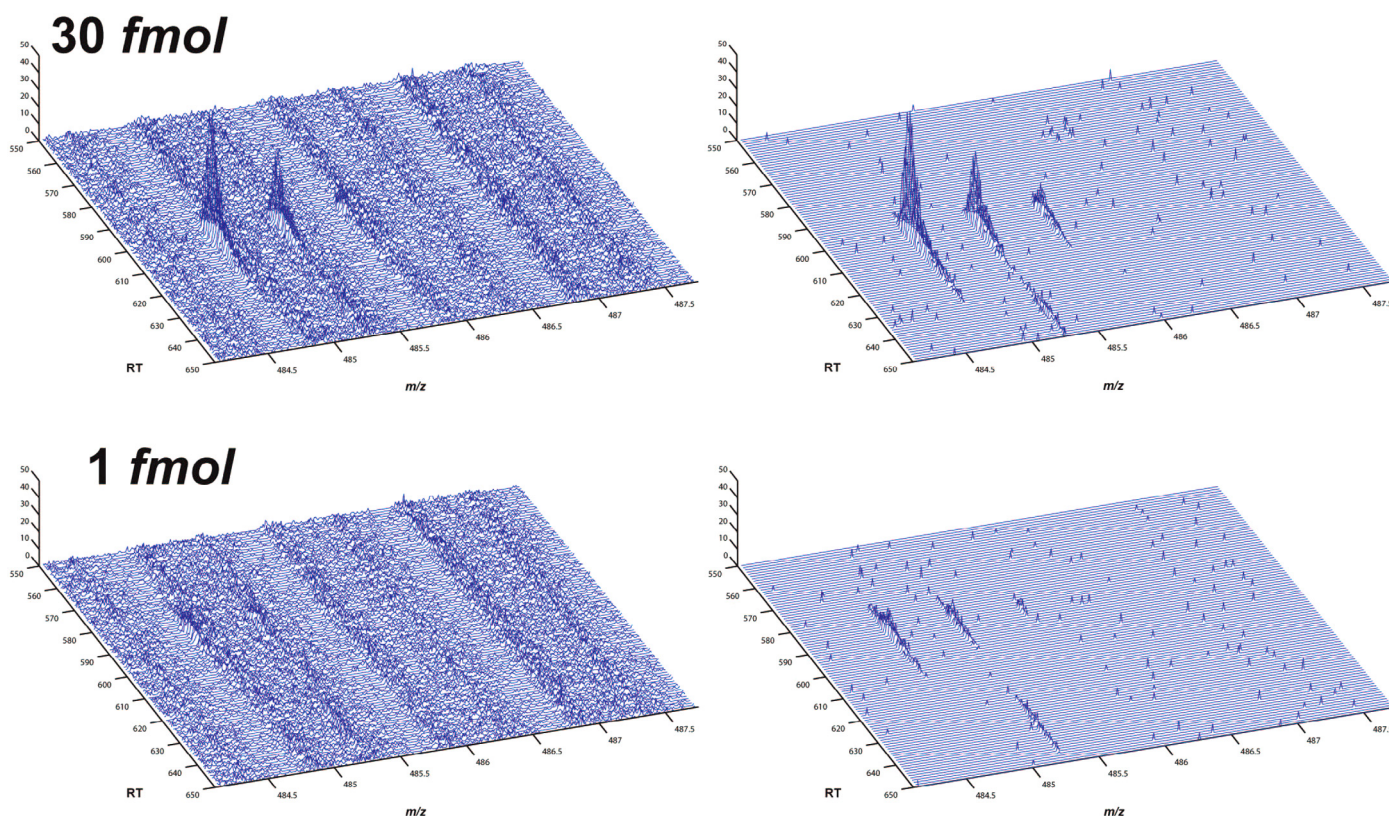


Fig. 1. Stochastic and chemical noise rejection for the same peptide ($m/z = 484.7454$) at different concentrations by a wavelet-based approach (raw signals on the left and filtered signals on the right). No detection problem is seen at the high-concentration peaks (e.g., 30 fmol). In contrast, the peptide peaks can be completely masked under the elution traces of the chemical noise at the low concentrations (e.g., 1 fmol) [18].

[14]. When dealing with femtosecond transient absorption spectroscopy data, where artifacts and spectra only slightly overlap in the frequency domain, iterative wavelets may be efficiently used for baseline correction [15].

Chromatographic datasets of the metabolomics and quality control of herbal medicines usually contain peak shifts that render the acquired signals too complex to extract meaningful information easily. To align shift peaks among one-dimensional chromatograms accurately and rapidly, multi-scale peak alignment (MSPA) was presented [16]. By transforming the chromatogram into the wavelet space using CWT with Haar wavelet as the mother wavelet, peaks are detected and aligned against a reference chromatogram from large to small scale gradually. The aligning procedure is accelerated by fast FT cross correlation. This method could preserve the shapes of peaks, perform well with non-linear retention-time (RT) shifts, and be robust and not sensitive to noise and baseline. In another study, a practical, handy peak-alignment method (alignDE) was proposed with wavelet-pattern matching and differential evolution optimizing [17]. The peak-matching algorithm based on CWT with the Mexican Hat and Haar wavelets as its mother wavelets is robust and accurate to detect peak position by finding ridges over the 2D CWT coefficient matrix of chromatogram in three steps:

- (1) identify the ridges by linking the local maxima;
- (2) identify the peaks based on the ridges lines; and,
- (3) refine the peak-parameter estimation.

Noise in LC-MS data sets is mostly random (small spikes uniformly distributed in both the chromatographic and the mass domains) or chemical (a periodic background in the mass domain and a slowly varying baseline in the chromatographic domain). A

wavelet-based algorithm was proposed to reject both LC-MS chemical and random noise while preserving low-abundance peptides (Fig. 1) [18]. To remove the noise in the mass spectra optimally by processing the signal in the chromatographic domain, spectrographic data are presented as two-dimensional maps. In an effort to unveil low-intensity peptides originally masked by the chemical noise and to reduce false-positive identification in label-free LC-MS differential analysis, the inclusion of a preprocessing step of background subtraction in a common laboratory pipeline was also investigated [19]. This works iteratively by first extracting all single-ion chromatograms (SICs) in a whole LC-MS map and by then processing independently each SIC by means of wavelet decomposition. CWT was further shown to be simpler and easier than a numerical differentiation method in processing noisy signals of capillary electrophoresis (CE) in quantitatively determining co-eluted compounds (i.e., benzoic acid and salicylic acid) [20]. Noisy experimental observations in gas chromatograms could be suppressed by geometric mean filtering and subsequently the standard Cohen-Daubechies-Feauveau WT [21]. Compared with previous techniques in which parameters were adjusted by trial and error, the method designed requires much less effort to tune algorithm parameters and its false detection rate is significantly lower.

One of the most important procedures for the analysis of GC/MS data is chemical-compound identification by assigning an experimental mass spectrum to a compound recorded in a reference spectral library using a spectrum-matching algorithm. With reference to this, novel spectral similarity measures of mass spectra based on the discrete wavelet and Fourier transforms were proposed [22]. These methods are composite similarities that are composed of weighted intensities and wavelet/Fourier coefficients using cosine correlation. Their performance in compound

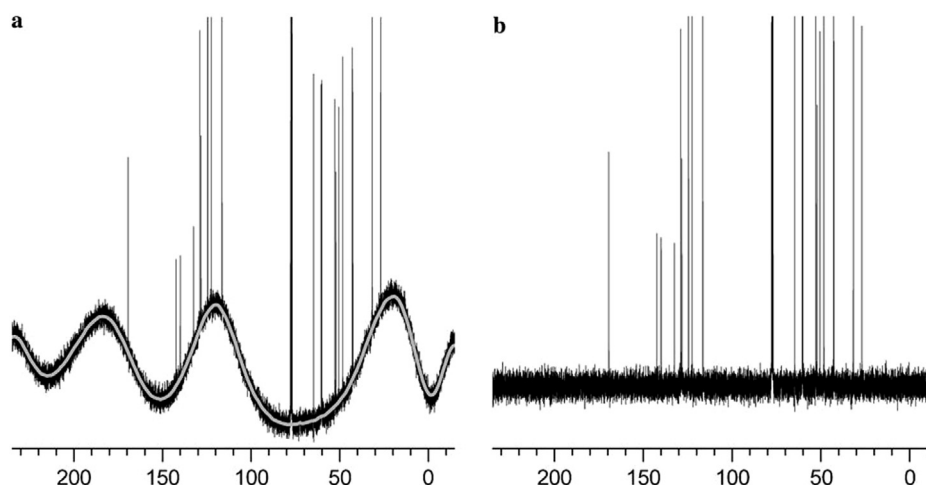


Fig. 2. ^1H -decoupled ^{13}C spectrum of strychnine in CDCl_3 acquired on a Varian Unity Inova 400 NMR spectrometer using a sweep width of 25,157 Hz. The FID was processed with an exponential function (line broadening = 1 Hz) and a FT of 64 Kb (a) and corrected spectrum (b) after applying the proposed algorithm ($k = 1000$, $a = 50$) [26].

identification along with the existing similarity measures was evaluated using the NIST Chemistry WebBook mass database maintained by the National Institute of Standards and Technology (NIST) as a library of reference spectra and repetitive mass spectral data as a set of query spectra. This study suggests that DFT/DWT-based transforms have more potential to reduce the false-positive rate than the literature-reported composite measures.

In MS-based proteomic data analysis, peak detection is an essential step for subsequent protein identification, quantification and discovery of disease-related biomarkers. According to a comprehensive survey done by Chao Yang et al., CWT provides the best performance among all peak-detection algorithms reported so far [23]. The explanation for this is two-fold. First, CWT optimally characterizes the shapes of peaks in mass spectra. Second, the concept of forming ridge lines in CWT effectively removes false-positive peaks.

In CE coupled to NMR detection using small solenoidal coils, spectral interpretation could be subject to interference by electrophoretic currents causing substantial distortion in the NMR spectral line-widths and peak heights. To restore, at least partially, spectral lines from the distorted NMR spectrum, a two-step signal processing method was suggested [24]. First, a gradient-descent method is used to estimate the broadening function of a reference signal as a combination of several Lorentzian functions. Multi-resolution wavelet analysis is then applied to the distorted spectrum to determine an initial estimate of the frequencies of the spectral lines. The gradient-descent method is re-used to deconvolute the estimated broadening function for convergence to the final spectrum, a second set of Lorentzians.

Given the fact that an NMR spectrum comprises a number of local bumps and peaks with different scales, the use of Gabor wavelets for the analysis of multiple NMR spectra was studied [25]. This complex WT can handle multi-scale information efficiently and has an energy-shift-insensitive property enabling direct comparison among multiple NMR spectra. Classification models constructed with the complex wavelet coefficients selected by the false-discovery rate (FDR)-based procedure yield lower rates of misclassification than models constructed with original features and conventional wavelet coefficients. In another approach, CWT-derivative calculation was employed to improve automatic recognition of signal-free regions of NMR data sets [26]. It was followed by a baseline-modeling procedure based on the Whittaker smoother algorithm. This combination could automatically flatten 1D and 2D NMR spectra

with large baseline distortions without loss in signal-to-noise ratio or smoothing artifacts (Fig. 2).

NMR has been widely used as a powerful tool to determine the 3D structures of proteins *in vivo*. This technique, however, requires a tremendous amount of time and expert knowledge for peak-picking, chemical-shift assignment and structure-calculation steps. Recently, WaVPeak, a fully automatic peak-detection method based on wavelet-based smoothing and volume-based filtering, was introduced [27]. This Open Source program has two major advantages over the state-of-the-art peak-picking methods:

- to detect weak peaks embedded in the noise level; and,
- to filter the false positives by estimating the volume of the peaks.

A key issue for NMR analysis is the suppression of the water signal. It could be done by using WAVEWAT, a new processing algorithm based on a multi-resolution analysis (MRA) of the free induction decay using a dyadic DWT [28]. This algorithm is suitable for large data sets and allows the recovery of signals close to that of the solvent without distortion of the signal shape and intensity. Furthermore, the calculation of the area under the water signal could be performed by the reconstruction of the free induction decay using only low-order levels representing the water signal.

To enhance the resolution of composite spectra, wavelet-based derivative spectrometry was considered [29]. Within the framework of this approach, a numerical differentiation algorithm with continuous wavelets was developed. As opposed to conventional derivative spectrometry, it results in the best contrast in differential curves and stable estimations of derivative in the wavelet domain without using regularization (Fig. 3). Furthermore, knowledge of a number of overlapping bands in complex spectra could be obtained with the help of CWT [30]. Unlike the fractional derivative spectrometry method [31], the CWT-based derivative spectrometry requires fewer special points to determine a band shape and can be successfully utilized in the case of complex spectra corrupted with a white and/or high-frequency noise. These findings are understandable because WT can be regarded as a smoothing and differentiation process, and the n^{th} -order derivative can be simply obtained through just one transform procedure, CWT or DWT.

Biorthogonal spline wavelet is superior to Daubechies, Haar and Gaussian wavelet, especially for low SNR signals [32]. With analytical peaks that resemble Gaussian distribution curves, CWT (i.e., Symlet4) could improve the limit of detection (LOD) by factors of

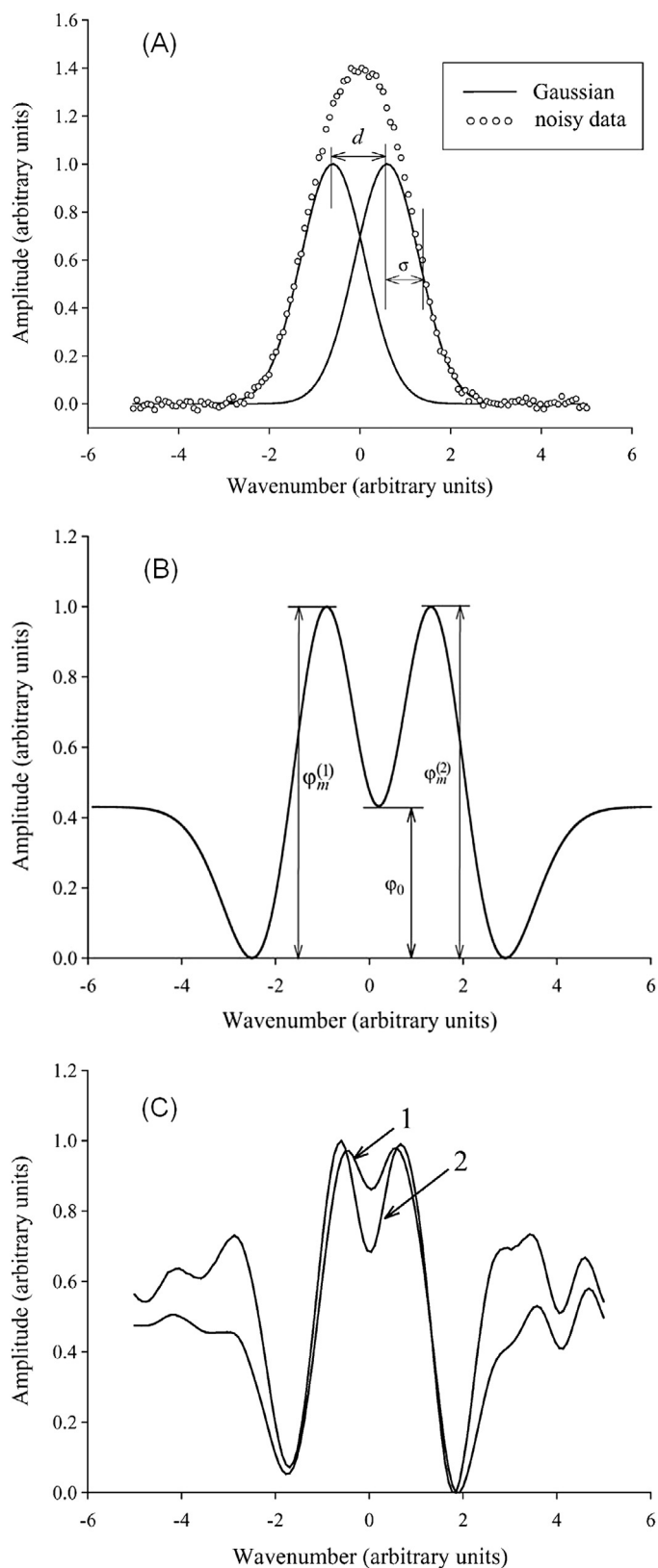


Fig. 3. (A) A synthetic composite spectrum consisting of two unresolved Gaussians spaced at d with half-widths of σ and equal intensities, corrupted by white Gaussian noise; (B) the normalized second derivative of a noise-free synthetic composite spectrum similar to (A) for interpretation of a resolution measure χ ; and, (C) the normalized second derivative of (A) by (1) Derivative Spectrometry (DS) and (2) Wavelet-based Derivative Spectrometry (WDS) [29].

6–7 [33]. A remarkable step from convolution with Gaussian distribution curves via FT to WT (i.e., the detection power improved by factors of ~ 6) was also observed for HPLC-ICP-MS analysis of different arsenic species [34]. The wavelet-smoothing approach proved to be efficient and convenient in removing intense noise from the weak signals of As, Pb, Sb and Se in ICP spectra of both synthetic and real water samples [35].

Due to its power of resolving overlapped peaks, CWT was also applied for spectrophotometric simultaneous kinetic determination of binary mixtures (i.e., phosphate and silicate) based on the formation of phospho- and silico-molybdenum blue complexes in the presence of ascorbic acid; Co^{2+} and Ni^{2+} based on their complexation reactions with 1-(2-pyridylazo)-2-naphthol (PAN) in micellar media at pH 6.0 [36].

CWT filters were also used for analysis of X-ray fluorescence (XRF) spectra as rapid limit tests for the determination of toxic metals (e.g., arsenic, lead, mercury, and chromium) in pharmaceutical materials using hand-held XRF spectrometers [37].

Photoacoustic (PA) spectroscopy is a promising tool for analyzing nearly all phases of matter without any sample preparation. A drawback of conventional PA is that some unresolved shoulders from the spectral overlap of adjacent peaks cannot be discriminated even with the highest performance spectrophotometers. Regarding this, a practical PA method based on the combination of WT and first-derivative PA spectrum was presented to retrieve the weak-intensity signals [38].

For voltammetric signals, more symmetric and regular wavelets (e.g., Symmlet8 and especially Coiflet2) are able to reduce noise as well as LOD by factors up to more than 300 [39]. The combination of CWT and inverse CWT was proposed as a numerical algorithm to separate voltammetric overlapping peaks for the determination of antimony in the presence of copper and the parallel determination of these two elements [40].

Moreover, CWT with a specially defined dedicated mother wavelet proved to be a useful tool for precise detection of the endpoint in a potentiometric titration. The proposed algorithm does not require any initial information about the nature or the type of analyte and/or the shape of the titration curve. But, in the case of noisy or badly-shaped curves, this approach works better than the first-derivative computation preceded by Savitzky–Golay smoothing with respect to relative error and coefficients of variability [41].

4. Data compaction

Identification of every component in a complex NMR spectrum of a mixture always remains a difficult task. When the size of a data set is large and some parameters (e.g., peak position) need to be optimized, the time for computation is too long for practical use. In principle, the application of WT-based compression algorithms to NMR spectra should be straightforward because NMR data are considerably resistant to the random distortions (noise) generated by these algorithms. Thus, a novel algorithm for fast resolution of multi-component overlapping NMR spectrum using wavelet compression and the Immune Algorithm (IA) was proposed [42]. This procedure is done by compressing antigen (NMR spectrum of mixture) and antibodies (standard NMR spectra of estimated components) to the same scheme before being input into the immune network. It was recognized that the speed of WT-IA for resolution of overlapping multi-component NMR spectra was much higher than that of IA alone.

Nowadays, high-resolution multi-dimensional NMR spectroscopy has developed into a powerful method for the determination of the 3D structure of biological macromolecules. Despite the technical advances in computer hardware, there is still a considerable burden on the storage capacity of computers and low processing efficiency when handling, transmitting and storing these multi-dimensional data. To make multi-dimensional NMR anal-

ysis a routine task, a lossy data-compression algorithm based on WT was applied to 2D NMR spectra [43]. This algorithm could afford rapid and extreme compression ratios (e.g., 800:1) depending on the number of signals and the difference between the most and the least intense peaks (dynamic range). High-quality reconstructed 2D spectra could reasonably be used for qualitative and quantitative purposes with a compression ratio of 80:1.

GC-differential mobility spectrometry (GC/DMS) has been widely used for chemical characterization and identification of potentially complex biological or environmental sample types. The GC/DMS data output can be considered as a two-dimensional (2-D) image, consisting of retention time (RT), compensation voltage (CV) and the corresponding signal intensities. To process such large data sets, the conversion of 2-D signals to 1-D signals by summing intensities across RT or CV can result in losing important signal information in one data dimension. To overcome this obstacle, the feasibility of classifying GC/DMS samples by applying 2-D wavelet analysis directly to the original 2-D sample data was investigated [44]. Different from conventional methods, this characterization process keeps a similar 2-D data structure to the original data while significantly reducing the data size, so it not only keeps the information of both RT and CV dimensions but also shows a more robust feature-extraction effect than the *t*-test selected pixels from the original data. Furthermore, this process does not need to be coupled with a specific pattern-recognition method, which would ensure its wide applications to 2-D spectrometry data classification.

A common, important problem of NIR quantitative analysis in multi-component systems is how to predict accurately concentrations of each component with a small-scale calibration set while not impairing its predictive precision. To cope with this problem, the combination of orthogonal discrete Wavelet Packet Transform (WPT), least correlation design and correlation coefficient test (*r*-test) was studied [45]. In the proposed strategy, the calibration set should be prepared according to the least correlation design at first. Their NIR spectra are then decomposed with WPT according to the best decomposition tree, and the *r*-test is performed to select suitable wavelet coefficients as variables for constructing PLS models.

The use of WT as a data-compression method is also applicable to Maximum Likelihood Principal Component Regression (MLPCR), an errors-in-variables method used to accommodate measurement-error information when building multi-variate calibration models [46]. A hindrance of MLPCR is the substantial demand on computational resources (i.e., it is memory intensive and time consuming). It was shown that the error-covariance matrices in the wavelet and spectral domains are related through a two-dimensional WT, which allows the user to account for any effects of the WT on spectral and error structures.

Although DWT is an effective data-compression and noise-suppression tool, an acceptable level of compression of data using wavelets (i.e., safe, non-destructive data reduction) is still questionable. With regard to this, a simple statistical test was studied on the basis of estimation of the Singular Values (SVs) from a data matrix and the SVs at each level of compression followed by the application of Median Absolute Deviation (MAD) of the correlation between original SVs and compression SVs [47].

In practice, signal correction and data compression in multi-variate calibration could be lumped together in a pre-processing method, Orthogonal WAVElet Correction (OWAVEC) [48]. The OWAVEC method operates in two steps. First, each individual signal is decomposed by DWT and subsequently subjected to an orthogonalization algorithm in order to remove irrelevant information (signal correction). Second, the corrected wavelet coefficients are ranked by their variance or their correlation coefficient with the response variable to select finally the subset providing the most stable, reliable calibration model (data compression) [49]. Fig. 4 gives an overview of the steps involved in the OWAVEC procedure.

Parallel factor analysis (PARAFAC) is one of the most popular methods for evaluating multi-way data sets typically acquired by coupled measurement techniques, such as LC-UV, GC-MS, and excitation-fluorescence spectroscopy [50]. This method is popular because it can extract directly interpretable chemometric models with little *a priori* information and handle unknown interferences and missing values. Utilizing multi-dimensional wavelets to compress data down to a few percent of the original size, the analyses of 3-way and 4-way PARAFAC models could be accelerated with factors up to 50.

In the 1980s, the method “comparison analysis using restructured infrared and constituent data (CARNAC)” was developed with the idea of performing quantitative analysis via databases containing NIR and analytical data rather than through some form of regression analysis from that database. The realization of CARNAC involved database compression via Fourier transform, database modification and similarity analysis. Recently, this method was re-developed to incorporate the use of wavelet compression, modern programming techniques to make it more transportable and some refinements of the technique [51].

The efficiency of WT for data compression was also confirmed by using artificial neural network Fuzzy ARTMAP (FAM) to classify cyclic voltammograms according to underlying reaction mechanisms. FAM can be successfully applied to the classification into reversible, quasi-reversible electron-transfer mechanisms and the same mechanism but with a subsequent chemical reaction [52].

5. In combination with chemometrics

By definition, chemometrics is the use of mathematical and statistical methods to improve understanding of chemical information and to correlate quality parameters or physical properties to analytical instrument data. It helps spectroscopists solve the calibration problem for analysis of spectral data in many efficient ways.

The efficient use of a quantitative NIR method depends on the prediction ability and robustness of multi-variate calibration model. In NIR spectral calibration, a multi-variate regression is possibly set up where data arise as curves and where the number of predictors greatly exceeds the number of observations. In this case, wavelets could be used to transform curves into sets of coefficients describing local features of the spectra. A Bayesian decision theory approach is then applied to select those coefficients that predict the response well. A cost function must be specified for the selected coefficients [53].

In another development, correction for the interference of NIR spectra was realized by CWT-OSC (approximate derivative based on CWT and Orthogonal Signal Correction) [54]. This hybrid algorithm is better than CWT, OSC and other pretreatment methods, as it provides the most precise, simplest calibration model whilst decreasing the danger of over fitting of the calibration model. Furthermore, Wavelet Orthogonal Signal Correction-Based Discriminant Analysis (WOSC-DA) was also reported for classification problems [55]. This classification tool combines a wavelet-prism decomposition of a spectral response and OSC so as to reduce both classification errors and model complexity. WOSC-DA performs favorably, as compared to the wavelength-domain-filtering approach used in Partial Least-Squares Discriminant Analysis (PLS-DA) and Orthogonal Partial Least-Squares Discriminant Analysis (OPLS-DA).

The quality of calibration model for NIR analysis could be also greatly enhanced by a combination of CWT to filter spectral interference and a modified Iterative Reweighted PLS (mlRPLS) procedure to select the most informative CWT coefficients [56] as well as to detect the multiple outliers in the CWT domain (Fig. 5) [57].

A combination of NIR spectral wavelength selection by WT and Uninformative Variable Elimination (UVE) was proposed to obtain simple PLS models based on a small subset of wavelengths [58]. To

1) DWT

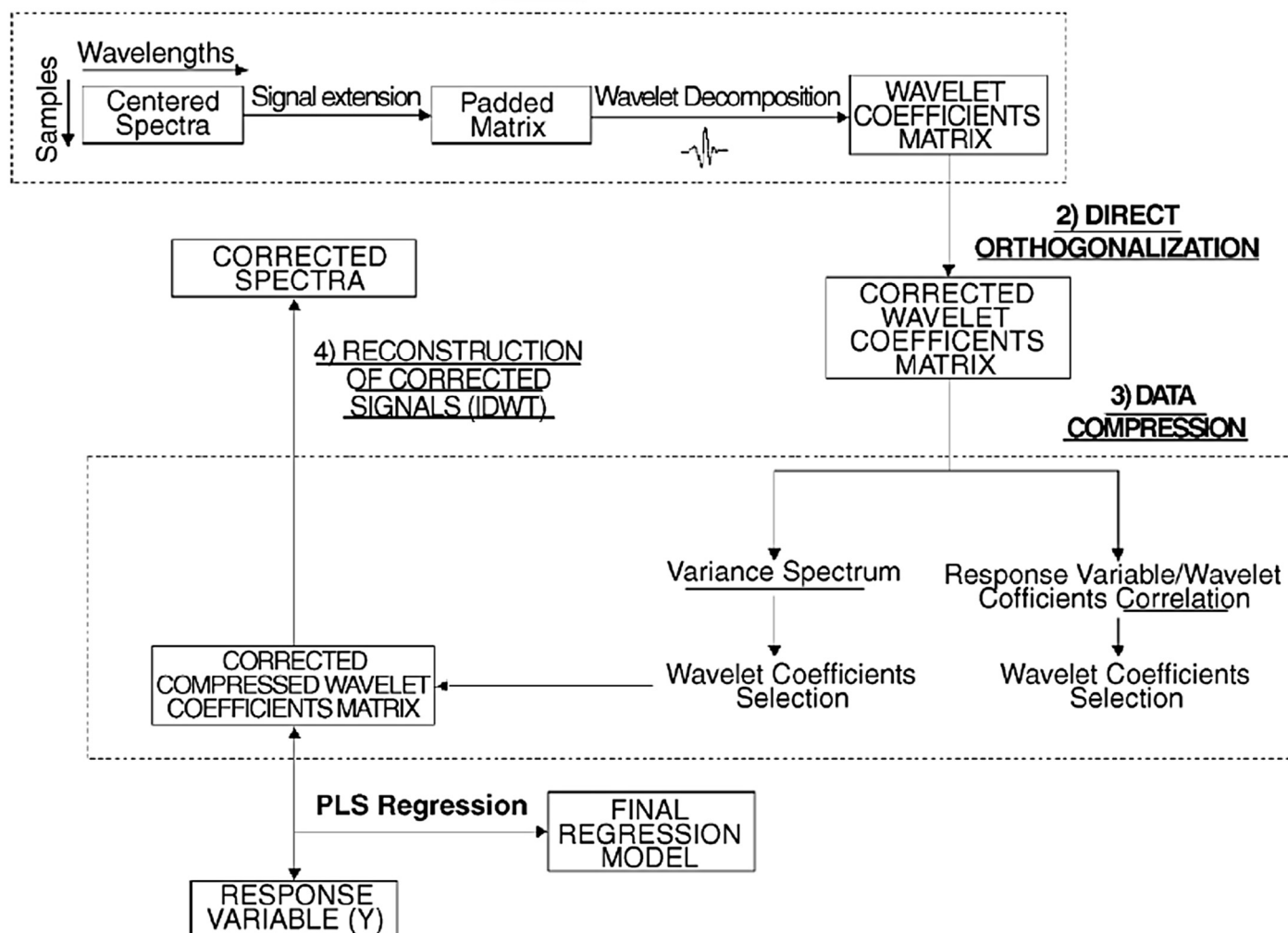


Fig. 4. An overview of the steps involved in the OWAVEC procedure. (1) DWT is applied to previously centered spectra being padded to the nearest length 2^n . (2) Direct orthogonalization is used to correct the wavelet-coefficient matrix by removing a part unrelated to the considered response. (3) Data compression is performed by wavelet coefficients selection based on (i) the variance spectrum or (ii) the correlation coefficients vector. (4) Filtered spectra can be reconstructed using the IDWT. The matrix of corrected and compressed wavelet coefficients is subjected to PLS regression [49].

obtain more parsimonious multi-variate models with higher precision, the use of multi-resolution in conjunction with a modified UVE (mUVE) criterion was tested to remove the low-frequency varying background and the high-frequency noise simultaneously [59]. This combination was exploited for extracting relevant information from NIR spectra for multi-variate calibration of routine components in complex plant samples, which typically consist of broad, weak, non-specific and overlapping bands [60].

More parsimonious calibration models could be obtained with Haar wavelet and Haar scaling function preprocessed NIR spectra rather than single-wavelength-selection algorithms [61]. Undoubtedly, the feature spectra of the original NIR signals can be separated by WT with various scales. For example, with WT preprocessing to remove the background in NIR spectra, the quality of the PLS model for the determination of chlorogenic acid in plant samples was successfully improved [62]. The least uncertainty of prediction difference was realized when analyzing the original NIR spectra of reducing sugar in vinegar by WT [63]. NIR spectra treated with WT showed more effectively the relationship with oil content of instant noodles than original and derivative ones [64]. WT-pretreated NIR spectra could be significantly simplified for the four chemometrics models used [i.e., Least-Squares Support Vector

Machines (LS-SVM), Radial Basis Function Artificial Neural Networks (RBF-ANN), PLS-DA and K-Nearest Neighbors (KNN)] for the classification of samples of *Rhizoma Corydalis*, a common traditional Chinese medicine, according to their geographical origin [65]. Not surprisingly, the input variables for the pattern-recognition method [i.e., KNN, RBF-ANN, LS-SVM and Back Propagation-Artificial Neural Network (BP-ANN)] were best produced when being WT pre-treated, as compared to two other data-pretreatment methods (genetic algorithm-PLS and successive projections algorithm). This method was applied for the discrimination of three kinds of *Indigowoad Root* sample, *Radix Isatidis*, *Rhizoma et Radix Baphicacanthis Cusia* and simulated adulterated samples [66].

Apart from NIR spectra, WT-chemometrics combinations were successfully investigated for: visual-near-infrared reflectance spectroscopy (vis-NIRS) for non-destructive detection of soluble solids contents (SSC) and pH in orange juices [67]; non-destructive determination of ambroxol content in tablets by Raman spectra [68]; NMR screening to analyze a series of HSQC spectra of proteins with different ligands [69]; or, to reveal the meaningful information of internal positional fluctuations of enzymes within NMR structure ensembles [70].

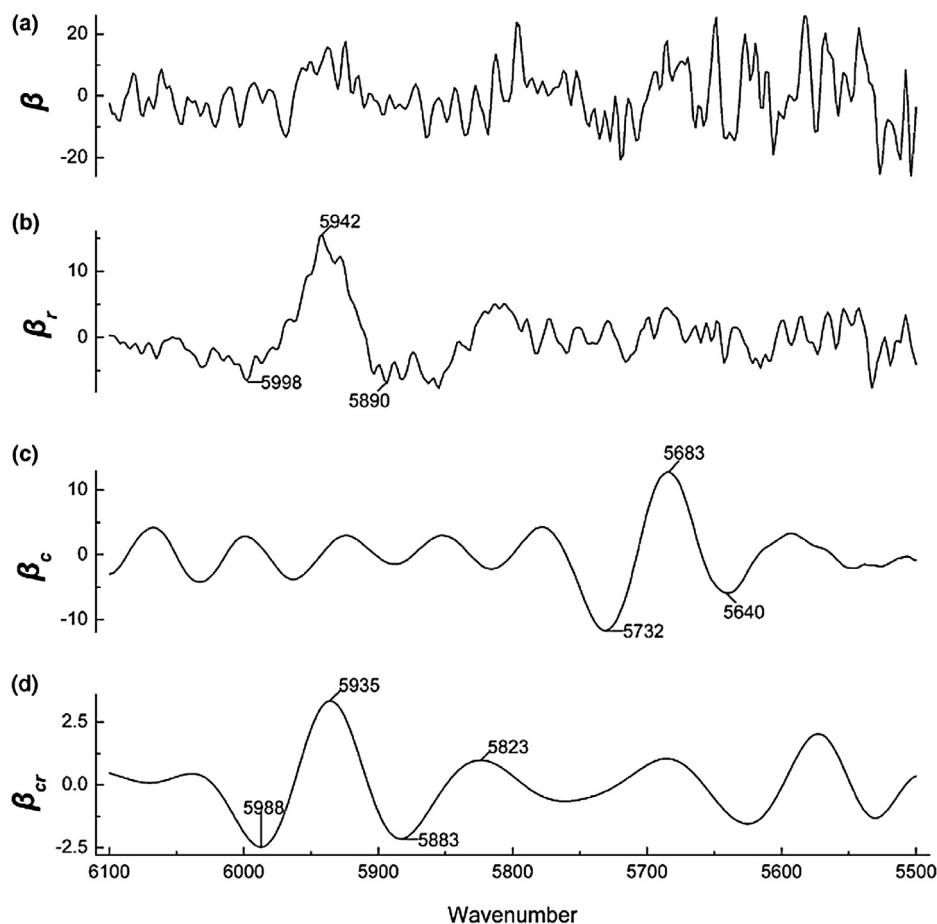


Fig. 5. Vector of regression coefficients obtained by different calibration models: a) PLS (β), b) mlRPLS (β_r), c) CWT-PLS (β_c), and d) CWT-mIRPLS (β_{cr}), indicating that CWT-mIRPLS is the most robust against both spectral interference and multiple outliers [57].

FT-NIR is an effective technique to predict the chemical compositions of biomass in on-line monitoring in the bio-energy industry. In some applications, its high dimensional and collinear spectral data make it difficult to construct reliable prediction models. Regarding this, highly sophisticated non-linear models were built with wavelet-compressed FT-NIR spectral data via reproducing kernel methods [i.e., Kernel Partial Least Squares (KPLS) regression and Kernel Ridge Regression (KRR)]. A Vertical Energy Thresholding (VET) method was also employed for easy-to-implement wavelet compression [71]. In another study, DWT multi-resolution analysis was used to extract the feature vectors representing spectral characteristics of the FT-IR of three sibling plants. They were used to train an SVM for the classification of seeds of green bristle grass, yellow foxtail, and Chinese pennisetum, which were difficult to distinguish by traditional phytotaxonomy [72].

Recently, a methodology named Adaptive Multiscale Regression (AMR) was suggested for reliable Raman quantitative analysis [i.e., Raman spectra could be split in a calibration set into frequency components at different scales using Adaptive Wavelet Transform (AWT) and parallel member models be then constructed for a fusion model straightforwardly estimated by a PLS model [73]]. In an investigation on moisture content of tea by linear and non-linear feature-extraction algorithms, multi-signal wavelet decomposition was realized to expose the internal structure of all the Vis/NIR diffuse reflectance spectra [74].

With reference to simultaneous determination with overlapping spectrophotometric signals, a Wavelet-based Latent Variable Regression (WLVR) method was developed [75]. Using this method,

the latent variables could be made by projecting the wavelet-processed signals onto orthogonal basis eigenvectors. The spectrophotometric analysis of multi-component mixtures could also be realized by an OSC-WPT-PLS approach [76]. To enhance the ability of eliminating noise and unrelated information as well as the quality of regression method, the concept of fusing the information from different wavelet scales to obtain a PLS model was proposed [77]. Based on the concept of data mining in chemometrics, another method named OSC-WPT-RBFN was presented for resolving overlapping spectra. This method combined the capacity to remove noise and unrelated information of both OSC and WPT as well as the capacity to overcome the convergence problem in back-propagation training and facilitating non-linear calculation of radial basis function neural network [78].

For non-stationary signal processing, the stepwise coupling of WT and ANN to yield Wavelet Neural Networks (WNNs) (i.e., using wavelet functions as the activation functions of hidden layer neurons) was proposed (Fig. 6). This strategy was successfully applied to resolve strongly voltammetric overlapping signals in the determination of phenolic compounds [79]. These signal-processing tools that combine chemometrics and electrochemical sensors are of paramount importance for the development of the electronic-tongue concept. WNN was also applied to establish a Quantitative Structure–Electrochemistry Relationship (QSER) model for prediction of the half-wave reduction potential ($E_{1/2}$) of some organic compounds (e.g., aldehydes and ketones and to find structural factors, which affect their potential [80]. WNN is robust and has better predictive power than the conventional ANN and the linear models (MultiLinear Re-

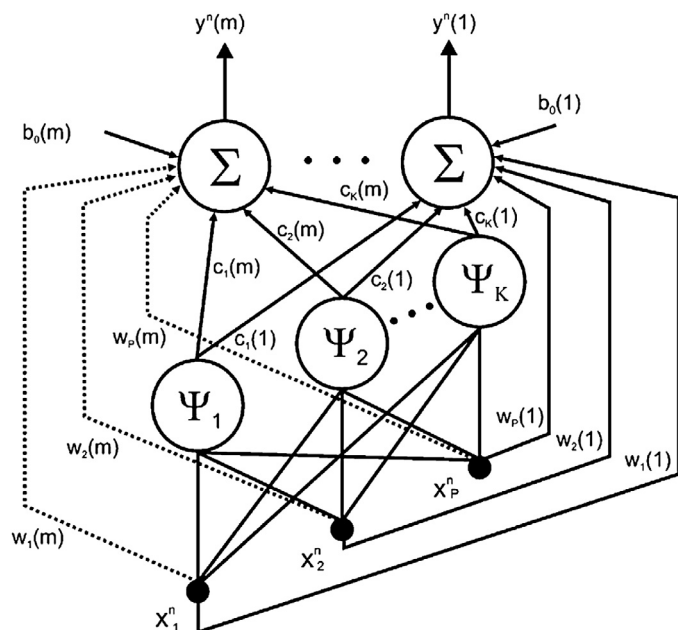


Fig. 6. WNN model: x_i^j denotes the j^{th} current value of the i^{th} voltammogram, and $y^n(m)$ the sought information for each compound concentration. The output $y^n(r)$ (where n is an index, not a power) depends on the connection weights $c_k(r)$ between the output of each neuron and the r^{th} output of the network, the connection weights $w_j(r)$ between the input data and each output, an offset value $b_0(r)$ used to adjust functions that has a mean value other than zero, the n^{th} input vector x^n and the wavelet function Ψ_i of each neuron [79].

gression (MLR) and PLS) when some non-linear interactions exist between selected descriptors in the models. A Fast Wavelet Transform (FWT)-based feature selection algorithm, coupled with different multi-variate calibration methods (MLR and PLS), was also investigated for the resolution of quaternary mixtures of chlorophenols at about their detection-limit levels using differential pulse adsorptive stripping voltammetry on a new polyethylene-glycol-modified sonogel-carbon electrode [81].

6. Conclusion

In the past decade, wavelets have increasingly shown great applicability in the field of analytical chemistry, as reviewed above. WT is preferred over other existing algorithms in noise removal, resolution enhancement, data compression and chemometrics modeling in chemical studies. This is due to its efficiency, large number of basic functions available and high speed in data treatment. Signal processing of chemical data via WT is still under development. With the hope that ongoing research on WT will produce new stimuli for data and signal processing, this mathematical algorithm is likely to go on being one of the most popular techniques in the future.

Acknowledgements

This article is dedicated to the memory of my beloved maternal grandfather, Dang Van Cuong (1928–2014), who endlessly inspired me to learn French and English.

References

- [1] A. Graps, An introduction to wavelets, *IEEE Comp. Sc. Eng.* 2 (1995) 50–61.
- [2] J.C. Goswami, A.K. Chan, *Fundamentals of Wavelets*, second ed., John Wiley & Sons, Inc., 2011.
- [3] G. Bachman, L. Narici, E. Beckenstein, *Fourier and Wavelet Analysis*, Springer-Verlag, 2000.
- [4] F.-T. Chau, Y.-Z. Liang, J. Gao, X.-G. Shao, *Chemometrics from Basics to Wavelet Transform*, John Wiley & Sons, Inc., 2004.
- [5] S. Mallat, A theory for multiresolution signal decomposition: the wavelet representation, *IEEE Trans. Pattern Anal. Mach. Intell.* 11 (1989) 674–693.
- [6] G. Schulze, A. Jirasek, M.M.L. Yu, A. Lim, R.F.B. Turner, M.W. Blades, Investigation of selected baseline removal techniques as candidates for automated implementation, *Appl. Spectrosc.* 59 (2005) 545–574.
- [7] C.R. Vargas, C.H. Loaiza, Absorption spectrophotometry signal de-noising using invariant wavelets, *Rev. Ing. Invest.* 31 (2011) 142–154.
- [8] J. Schlenke, L. Hildebrand, J. Moros, J.J. Laserna, Adaptive approach for variable noise suppression on laser-induced breakdown spectroscopy responses using stationary wavelet transform, *Anal. Chim. Acta* 754 (2012) 8–19.
- [9] D. Chen, Z. Chen, E. Grant, Adaptive wavelet transform suppresses background and noise for quantitative analysis by Raman spectrometry, *Anal. Bioanal. Chem.* 400 (2011) 625–634.
- [10] P. Li, G. Du, W. Cai, X. Shao, Rapid and nondestructive analysis of pharmaceutical products using near-infrared diffuse reflectance spectroscopy, *J. Pharm. Biomed. Anal.* 70 (2012) 288–294.
- [11] L. Liang, B. Wang, Y. Guo, H. Ni, Y. Ren, A support vector machine-based analysis method with wavelet denoised near-infrared spectroscopy, *Vib. Spectrosc.* 49 (2009) 274–277.
- [12] J. Li, U. Parchatka, H. Fischer, Applications of wavelet transform to quantum cascade laser spectrometer for atmospheric trace gas measurements, *Appl. Phys. B* 108 (2012) 951–963.
- [13] C. Ma, X. Shao, Continuous wavelet transform applied to removing the fluctuating background in near-infrared spectra, *J. Chem. Inf. Comput. Sci.* 44 (2004) 907–911.
- [14] B. Wan, G.W. Small, Wavelet analysis used for spectral background removal in the determination of glucose from near-infrared single-beam spectra, *Anal. Chim. Acta* 681 (2010) 63–70.
- [15] O. Devos, N. Mouton, M. Sliwa, C. Ruckebusch, Baseline correction methods to deal with artifacts in femtosecond transient absorption spectroscopy, *Anal. Chim. Acta* 705 (2011) 64–71.
- [16] Z.-M. Zhang, Y.-Z. Liang, H.-M. Lu, B.-B. Tan, X.-N. Xu, M. Ferro, Multiscale peak alignment for chromatographic datasets, *J. Chromatogr. A* 1223 (2012) 93–106.
- [17] Z.-M. Zhang, S. Chen, Y.-Z. Liang, Peak alignment using wavelet pattern matching and differential evolution, *Talanta* 83 (2011) 1108–1117.
- [18] S. Cappadona, F. Levander, M. Jansson, P. James, S. Cerutti, L. Pattini, Wavelet-based method for noise characterization and rejection in high-performance liquid chromatography coupled to mass spectrometry, *Anal. Chem.* 80 (2008) 4960–4968.
- [19] S. Cappadona, P. Nanni, M. Benevento, F. Levander, P. Versura, A. Roda, et al., Improved label-free LC-MS analysis by wavelet-based noise rejection, *J. Biomed. Biotech. Volume* (2010) Article ID 131505.
- [20] L. Jiao, S. Gao, F. Zhang, H. Li, Quantification of components in overlapping peaks from capillary electrophoresis by using continuous wavelet transform method, *Talanta* 75 (2008) 1061–1067.
- [21] B. Shim, H. Min, S. Yoon, Nonlinear preprocessing method for detecting peaks from gas chromatograms, *BMC Bioinformatics* 10 (2009) 378.
- [22] I. Koo, X. Zhang, S. Kim, Wavelet- and Fourier-transform-based spectrum similarity approaches to compound identification in Gas Chromatography/Mass Spectrometry, *Anal. Chem.* 83 (2011) 5631–5638.
- [23] C. Yang, Z. He, W. Yu, Comparison of public peak detection algorithms for MALDI mass spectrometry data analysis, *BMC Bioinformatics* 10 (2009) 4.
- [24] Y. Li, M.E. Lacey, J.V. Sweedler, A.G. Webb, Spectral restoration from low signal-to-noise, distorted NMR signals: application to hyphenated capillary electrophoresis-NMR, *J. Magn. Reson.* 162 (2003) 133–140.
- [25] S.B. Kim, Z. Wang, S. Orantara, C. Temiyasathit, Y. Wongsawat, Feature selection and classification of high-resolution NMR spectra in the complex wavelet transform domain, *Chemom. Intell. Lab. Syst.* 90 (2008) 161–168.
- [26] J.C. Cobas, M.A. Bernstein, M. Martín-Pastor, P.G. Tahoces, A new general-purpose fully automatic baseline-correction procedure for 1D and 2D NMR data, *J. Magn. Reson.* 183 (2006) 145–151.
- [27] Z. Liu, A. Abbas, B.-Y. Jing, X. Gao, WaVPeak: picking NMR peaks through wavelet-based smoothing and volume-based filtering, *Bioinformatics* 28 (2012) 914–920.
- [28] U.L. Günther, C. Ludwig, H. Rüterjans, WAVEWAT – Improved solvent suppression in NMR spectra employing wavelet transforms, *J. Magn. Reson.* 156 (2002) 19–25.
- [29] S.S. Kharintsev, D.I. Kamalova, M.K. Salakhov, A.A. Sevastianov, Resolution enhancement of composite spectra using wavelet-based derivative spectrometry, *Spectrochim. Acta [A]* 61 (2005) 149–156.
- [30] S.S. Kharintsev, D.Z. Galimullin, A.Y. Vorob'ev, M.K. Salakhov, Band shape determination with robust estimator based on continuous wavelet transform, *Spectrochim. Acta [A]* 65 (2006) 292–298.
- [31] S.S. Kharintsev, M.K. Salakhov, A simple method to extract spectral parameters using fractional derivative spectrometry, *Spectrochim. Acta [A]* 60 (2004) 2125–2133.
- [32] X. Shao, C. Ma, A general approach to derivative calculation using wavelet transform, *Chemom. Intell. Lab. Syst.* 69 (2003) 157–165.
- [33] S. Prikler, J.W. Einax, Improving detection power in trace analysis using wavelet transform, *Anal. Bioanal. Chem.* 403 (2012) 2563–2567.

- [34] S. Prikler, D. Pick, J.W. Einax, Comparing different means of signal treatment for improving the detection power in HPLC-ICP-MS, *Anal. Bioanal. Chem.* 403 (2012) 1109–1116.
- [35] X. Ma, Z. Zhang, Wavelet smoothing applied to the determination of trace arsenic, lead, antimony and selenium in environmental water by ICP-OES, *J. Anal. At. Spectrom.* 19 (2004) 738–742.
- [36] A. Afkhami, M. Abbasi-Tarighat, Application of continuous wavelet transformation to the simultaneous kinetic determination of binary mixtures, *Talanta* 78 (2009) 424–431.
- [37] S. Arzhantsev, X. Li, J.F. Kauffman, Rapid limit tests for metal impurities in pharmaceutical materials by X-ray fluorescence spectroscopy using wavelet transform filtering, *Anal. Chem.* 83 (2011) 1061–1068.
- [38] J. Zheng, Z. Tang, Y. He, L. Guo, Sensitive detection of weak absorption signals in photoacoustic spectroscopy by using derivative spectroscopy and wavelet transform, *J. Appl. Phys.* 103 (2008) 093116.
- [39] S. Prikler, J.W. Einax, Wavelet transform as a new approach to the enhancement of signal-to-noise ratio in anodic stripping voltammetry, *Anal. Bioanal. Chem.* 395 (2009) 1707–1711.
- [40] M. Jakubowska, R. Piech, Dedicated mother wavelet in the determination of antimony in the presence of copper, *Talanta* 77 (2008) 118–125.
- [41] M. Jakubowska, B. Baś, W.W. Kubiak, End-point detection in potentiometric titration by continuous wavelet transform, *Talanta* 79 (2009) 1398–1405.
- [42] X. Shao, Z. Yu, L. Sun, Resolution of multicomponent NMR signals using wavelet compression and immune algorithm, *Spectrochim. Acta [A.]* 59 (2003) 1075–1082.
- [43] J.C. Cobas, P.G. Tahoces, M. Martin-Pastor, M. Penedo, F.J. Sardina, Wavelet-based ultra-high compression of multidimensional NMR data sets, *J. Magn. Reson.* 168 (2004) 288–295.
- [44] W. Zhao, S. Sankaran, A.M. Ibanez, A.M. Dandekar, C.E. Davis, Two-dimensional wavelet analysis based classification of gas chromatogram differential mobility spectrometry signals, *Anal. Chim. Acta* 647 (2009) 46–53.
- [45] C.-B. Cai, L. Xu, Q.-J. Han, H.-L. Wu, J.-F. Nie, H.-Y. Fu, et al., Combining the least correlation design, wavelet packet transform and correlation coefficient test to reduce the size of calibration set for NIR quantitative analysis in multi-component systems, *Talanta* 81 (2010) 799–804.
- [46] M.N. Leger, P.D. Wentzell, Maximum likelihood principal components regression on wavelet-compressed data, *Appl. Spectrosc.* 58 (2004) 855–862.
- [47] M. Kompany-Zareh, F. van den Berg, Determination of an acceptable level of spectral data compression by Discrete Wavelet Transforms, *Anal. Chim. Acta* 668 (2010) 137–142.
- [48] I. Esteban-Diez, J.-M. González-Sáiz, C. Pizarro, OWAVEC: a combination of wavelet analysis and an orthogonalization algorithm as a pre-processing step in multivariate calibration, *Anal. Chim. Acta* 515 (2004) 31–41.
- [49] I. Esteban-Diez, J.-M. González-Sáiz, C. Pizarro, Generalization of OWAVEC method for simultaneous noise suppression, data compression and orthogonal signal correction, *Anal. Chim. Acta* 544 (2005) 89–99.
- [50] J. Cramer, Y.-C. Kim, F. Vogt, K. Booksh, Accelerating the analyses of 3-way and 4-way PARAFAC models utilizing multi-dimensional wavelet compression, *J. Chemometrics* 19 (2005) 593–606.
- [51] A.M.C. Davies, T. Fearn, Quantitative analysis via near infrared databases: comparison analysis using restructured near infrared and constituent data-deux (CARNAC-D), *J. Near Infrared Spectrosc.* 14 (2006) 403–411.
- [52] E.P. Sapozhnikova, M. Bogdan, B. Speiser, W. Rosenstiel, EChem++ – An object-oriented problem solving environment for electrochemistry. 3. Classification of voltammetric signals by the Fuzzy ARTMAP neural network with respect to reaction mechanisms, *J. Electroanal. Chem.* 588 (2006) 15–26.
- [53] M. Vannucci, P.J. Brown, T. Fearn, A decision theoretical approach to wavelet regression on curves with a high number of regressors, *J. Stat. Plan. Inference* 112 (2003) 195–212.
- [54] D. Chen, F. Wang, X. Shao, Q. Su, Elimination of interference information by a new hybrid algorithm for quantitative calibration of near infrared spectra, *Analyst* 128 (2003) 1200–1203.
- [55] W. Ni, S.D. Brown, R. Man, Wavelet orthogonal signal correction-based discriminant analysis, *Anal. Chem.* 81 (2009) 8962–8967.
- [56] D. Chen, B. Hu, X. Shao, Q. Su, Variable selection by modified IPW (iterative predictor weighting)-PLS (partial least squares) in continuous wavelet regression models, *Analyst* 129 (2004) 664–669.
- [57] D. Chen, B. Hu, X. Shao, Q. Su, A new hybrid strategy for constructing a robust calibration model for near-infrared spectral analysis, *Anal. Bioanal. Chem.* 381 (2005) 795–805.
- [58] D. Wua, X. Chena, P. Shia, S. Wanga, F. Fenga, Y. Hea, Determination of α -linolenic acid and linoleic acid in edible oils using near-infrared spectroscopy improved by wavelet transform and uninformative variable elimination, *Anal. Chim. Acta* 634 (2009) 166–171.
- [59] D. Chen, X. Shao, B. Hu, Q. Su, A background and noise elimination method for quantitative calibration of near infrared spectra, *Anal. Chim. Acta* 511 (2004) 37–45.
- [60] X. Shao, F. Wang, D. Chen, Q. Su, A method for near-infrared spectral calibration of complex plant samples with wavelet transform and elimination of uninformative variables, *Anal. Bioanal. Chem.* 378 (2004) 1382–1387.
- [61] C.E.W. Gributs, D.H. Burns, Parsimonious calibration models for near-infrared spectroscopy using wavelets and scaling functions, *Chemom. Intell. Lab. Syst.* 83 (2006) 44–53.
- [62] X. Shao, Y. Zhuang, Determination of chlorogenic acid in plant samples by using Near-Infrared spectrum with wavelet transform preprocessing, *Anal. Sci.* 20 (2004) 451–454.
- [63] X. Fu, G. Yan, B. Chen, H. Li, Application of wavelet transforms to improve prediction precision of near infrared spectra, *J. Food Eng.* 69 (2005) 461–466.
- [64] B. Chen, X. Fu, D. Lu, Improvement of predicting precision of oil content in instant noodles by using wavelet transforms to treat near-infrared spectroscopy, *J. Food Eng.* 53 (2002) 373–376.
- [65] Y. Lai, Y. Ni, S. Kokot, Discrimination of Rhizoma Corydalis from two sources by near-infrared spectroscopy supported by the wavelet transform and least-squares support vector machine methods, *Vib. Spectrosc.* 56 (2011) 154–160.
- [66] Y. Ni, R. Song, S. Kokot, Discrimination of Radix Isatidis and Rhizoma et Radix Baphicacanthis Cusia samples by near infrared spectroscopy with the aid of chemometrics, *Spectrochim. Acta [A.]* 96 (2012) 252–258.
- [67] H. Cen, Y. He, M. Huang, Characterizing the moisture content of tea with diffuse measurement of soluble solids contents and pH in orange juice using chemometrics and VisNIRS, *J. Agric. Food Chem.* 54 (2006) 7437–7443.
- [68] M.-S. Hwang, S. Cho, H. Chung, Y.-A. Woo, Nondestructive determination of the ambroxol content in tablets by Raman spectroscopy, *J. Pharm. Biomed. Anal.* 38 (2005) 210–215.
- [69] N. Trbovic, F. Dancea, T. Langer, U. Günther, Using wavelet de-noised spectra in NMR screening, *J. Magn. Reson.* 173 (2005) 280–287.
- [70] M. Hu, Y. Li, G. Yang, G. Li, M. Li, Z. Wen, Wavelet transform analysis of NMR structure ensembles to reveal internal fluctuations of enzymes, *Amino Acids* 42 (2012) 1773–1781.
- [71] J.I. Park, L. Liu, X.P. Ye, M.K. Jeong, Y.-S. Jeong, Improved prediction of biomass composition for switchgrass using reproducing kernel methods with wavelet compressed FT-NIR spectra, *Expert Syst. Appl.* 39 (2012) 1555–1564.
- [72] P. Yu, C.-Y. Wan, C.-S. Wu, J.-N. Shou, C.-G. Cheng, Classification of green bristle grass, yellow foxtail and Chinese pennisetum seeds via HATR-FT-IR combined with chemometrics, *J. Spectrosc. Volume* (2013) Article ID 531747.
- [73] D. Chen, Z. Chen, E.R. Grant, Adaptive multiscale regression for reliable Raman quantitative analysis, *Analyst* 137 (2012) 237–244.
- [74] X. Li, C. Xie, Y. He, Z. Qiu, Y. Zhang, Reflectance spectroscopy using wavelet transform and multivariate analysis, *Sensors* 12 (2012) 9847–9861.
- [75] L. Gao, S. Ren, Simultaneous multicomponent analysis of overlapping spectrophotometric signals using a wavelet-based latent variable regression, *Spectrochim. Acta [A.]* 71 (2008) 959–964.
- [76] L. Gao, S. Ren, Combining orthogonal signal correction and wavelet packet transform with partial least squares to analyze overlapping spectra of three kinds of metal ions, *Spectrochim. Acta [A.]* 73 (2009) 960–965.
- [77] L. Gao, S. Ren, Multivariate calibration of spectrophotometric data using a partial least squares with data fusion, *Spectrochim. Acta [A.]* 76 (2010) 363–368.
- [78] L. Gao, S. Ren, Combining orthogonal signal correction and wavelet packet transform with radial basis function neural networks for multicomponent determination, *Chemom. Intell. Lab. Syst.* 100 (2010) 57–65.
- [79] J.M. Gutiérrez, A. Gutiérrez, F. Céspedes, M. del Valle, R. Muñoz, Wavelet neural networks to resolve the overlapping signal in the voltammetric determination of phenolic compounds, *Talanta* 76 (2008) 373–381.
- [80] Z. Garkani-Nejad, H. Rashidi-Nodeh, Comparison of conventional artificial neural network and wavelet neural network in modeling the half-wave potential of aldehydes and ketones, *Electrochim. Acta* 55 (2010) 2597–2605.
- [81] J.M. Palacios-Santander, L.M. Cubillana-Aguilera, M. Cocchi, A. Ulrici, I. Naranjo-Rodríguez, R. Seeber, et al., Multicomponent analysis in the wavelet domain of highly overlapped electrochemical signals: resolution of quaternary mixtures of chlorophenols using a peg-modified Sonogel–Carbon electrode, *Chemom. Intell. Lab. Syst.* 91 (2008) 110–120.