Methods Development and Applications of Chemometrics Techniques in Chemical and Biochemical Studies

by

LEUNG Kai-man, Alexander

Submitted in Partial Fulfillment of the
Requirements of the Degree of

DOCTOR OF PHILOSOPHY

in

CHEMISTRY

at the

The Hong Kong Polytechnic University

August, 1998

ABSTRACT

Abstract of thesis entitled

"Methods Development and Applications of Chemometrics Techniques in Chemical and Biochemical Studies"

submitted by LEUNG Kai-man, Alexander

for the degree of Doctor of Philosophy in Chemistry

at The Hong Kong Polytechnic University in August, 1998.

Owing to the rapid development of instrumentation in analytical chemistry, new and efficient data analysis techniques are required for data interpretation. Chemometrics is a new discipline in chemistry that applied mathematical, statistical and other logic-based methods to analyze chemical data in particular in analytical chemistry. From chemical literature, chemometrics techniques have been applied successfully in different areas. The aim of this project is to apply chemometrics techniques such as wavelet transform and factor analysis to analyze data from chemical and biochemical system.

Recently, wavelet transform (WT) was applied successfully in chemistry for data compression and denoising. In this project, WT was employed to process and enhance analytical data in two major areas, they were infrared (IR) spectroscopy and high

performance liquid chromatography coupled with a diode array detector (HPLC-DAD). WT was selected as a new compression method to reduce the size of a small IR spectral library. The wavelet compressed library was employed for spectral library searching. The performance of WT in data compression and library searching were compared with fast Fourier transform. A data length of 2^p with p being any integer is usually required for wavelet computation. A new method called coefficient position retaining (CPR) method was developed in this work to handle experimental data with a length not equal to 2^p.

Derivative spectroscopy is another signal processing technique that commonly used in analytical chemistry for data analysis owing to its popularity on the apparent higher resolution of the differential data when compared with the original data. Derivative of an analytical signal is usually derived from numerical differentiation. However, this method has a major drawback in increasing the noise level in computing higher order derivatives. In order to solve this problem, WT was proposed as a new method for approximate derivative calculation. Results indicated that the proposed method is much better than the conventional numerical differentiation method.

The development of hyphenated instrument such as HPLC-DAD leads to the development of new type chemometrics technique for data interpretation. Heuristic evolving latent projections (HELP) algorithm is one of the most popular method for analyzing data from HPLC-DAD. This method can extract most of useful information from the raw experimental data. FWT was proposed as a pre-processing step of the HELP algorithm. Both compression and denoising properties of FWT could enhance the

HELP algorithm especially for HPLC-DAD data with low signal-to-noise ratio (SNR). This new algorithm (FWT-HELP) was identified as a potential tool for analyzing chemical constituents in traditional Chinese medicine (TCM). In terms of chemistry, chemical constituents of Chinese herbs are a complex black system. There is no a priori information concerning the chemical composition of the samples. Both FWT and HELP algorithms have its own advantages which can simply and enhance the chemical analysis of TCM. A traditional Chinese medicinal herb, *Cordyceps sinensis* (冬蟲夏草), was selected as an example to test the performance of the proposed algorithm.

Methods Development and Applications of Chemometrics Techniques in Chemical and Biochemical Studies

by

LEUNG Kai-man, Alexander

Submitted in Partial Fulfillment of the
Requirements of the Degree of

DOCTOR OF PHILOSOPHY

in

CHEMISTRY

at the

The Hong Kong Polytechnic University

August, 1998

DECLARATION

This is to declare that this work has been done by the author in the Department of Applied Biology and Chemical Technology of The Hong Kong Polytechnic University and this thesis has not been submitted to this or other institution for any academic qualification.

LEUNG Kai-man, Alexander

Date: 28 September 1998

ACKNOWLEDGMENT

I am grateful to the Research Grant Council (RGC) of Hong Kong (Grant No. HKP 45/94E) and the Research Degrees Committee of The Hong Kong Polytechnic University (Grant No. 350/529) for the kind support in granting this project including the studentship and the support for my attendance at the 213th American Chemical Society (ACS) National Meeting in U.S. and the meetings in Hong Kong. I also wish to express my gratitude to all those have helped in various ways during the course of this work.

I would like to thank Prof. F.T. Chau, my supervisor, for his critical advice and supervision of this project as well as his comments on the final draft of this thesis and the others publications. He also gave me a lot of freedom in my research works and let me to participate in the development of the other projects.

I would like to thank Prof. Y.Z. Liang, from the College of Chemistry and Chemical Engineering, Hunan University, P.R. China, for his valuable discussion on chemometrics. He also let me to attend the chemometrics course when he taught in the Department of Chemistry of The Hong Kong Baptist University from October 1995 to February 1996.

I would like to thank Prof. J.B. Gao, from the Department of Mathematics, Huazhong University of Science and Technology, P.R. China, for his discussion and implementation of program for wavelet transform computation. Special thanks are due

to Prof. P.P.H. But and Dr. I.M.Y. Kwok, from the Department of Biology and Chinese Medicinal Material Research Center, The Chinese University of Hong Kong, for their discussion on traditional Chinese herb medicine and supply of experimental data for testing the proposed methods as developed in this project. Thanks are also due to the Hong Kong Government Laboratory for their generous supply of food dye samples and the laboratory technicians in our department for technical support.

Finally, I am indebted to all the members of my family for their constant encouragement during the course of this work especially in the hard times.

TABLE OF CONTENTS

LIST OF APPENDICES

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS AND SYMBOLS

Abbreviations

ANN	Artifical neural network
AR	Alternating regression
CPR	Coefficient position retaining method
CWT	Continuous wavelet transform
DAD	Diode array detector
DDE	Dynamic data exchange
DLLs	Dynamic link libraries
DWT	Discrete wavelet transform
EFA	Evolving factor analysis
FA	Factor analysis
FFT	Fast Fourier transform
FIA	Flow injection analysis
FID	Free induction decay
FSMW-EFA	Fixed size moving window evolving factor analysis
FT	Fourier transform
FWHM	Full width at half maximum
FWT	Fast wavelet transform
FWT-HELP	Fast wavelet transform coupled with heuristic evolving latent projections
GB	Giga bytes
GC-FTIR	Gas chromatography coupled with a Fourier transform infrared spectrometer
GC-MS	Gas chromatography coupled with a mass spectrometer
HELP	Heuristic evolving latent projections
HPLC	High performance liquid chromatography
HPLC-DAD	High performance liquid chromatography coupled with a diode array detector
IM	Imaginary part of Fourier coefficient
IND	Factor indicator function

(cont.)

(cont. Abbreviations and Symbols)

IR	Infrared spectroscopy
ITRT	Inverse translation-rotation transformation
ITTFA	Iterative target transformation factor analysis
KNN	K-nearest neigbhors
MB	Mega bytes
MRSD	Multi-resolution signal decomposition
MS	Mass spectrometry
NIR	Near infrared spectroscopy
NMR	Nuclear magnetic resonance spectrometry
PCA	Principle component analysis
RE	Real part of Fourier coefficient
RLER	Relative log(eigenvalue) ratio
SIMCA	Soft independent modeling of class analogy
SIMS	Secondary ion mass spectrometry
SNR	Signal-to-noise ratio
SVD	Singular value decomposition
TCM	Traditional Chinese medicine
TRT	Translation-rotation transformation
TTFA	Target transformation factor analysis
UV-VIS	Ultra-violet visible spectrometry
WFA	Window factor analysis
WPT	Wavelet packet transform
WT	Wavelet transform

Symbols

a	Scale (dilation) parameter in wavelet transform
b	Position (translation) parameter in wavelet transform
b_k	Coefficients from translation-rotation transformation

(cont.)

(cont. Abbreviations and Symbols)

c	Number of columns
$\mathbf{c}^{\mathrm{T}}$	Coefficient vector in spline
$c_1^{(0)}$	First scale coefficient at resolution level 1
$c_k^{(j)}$	Individual coefficient of scale coefficient in wavelet transform
c_n	Chromatogram for component n
$c_N^{(0)}$	Nth scale coefficient at resolution level 1
$d_k^{(j)}$	Individual coefficient of wavelet coefficient in wavelet transform
f1, f2 , ...	Component number in fungus part of *Cordyceps sinensis*
$f(x)$, $f(\lambda)$	Analytical signal
$f^{(n)}(x)$	nth order derivative of a signal
g_k	Individual coefficient of the high-pass filter $\boldsymbol{G}$ in wavelet transform
h_k	Individual coefficient of the low-pass filter $\boldsymbol{H}$ in wavelet transform
$\mathbf{h}^{\mathrm{T}}$	Knot vector in spline
i	Running index or $\sqrt{-1}$
j	Running index or resolution level in wavelet transform
k, k'	Running index
l1, l2 , ...	Component number in larva part of *Cordyceps sinensis*
m	Positive integer
$\tilde{m}$	Positive integer
n	Number of factors or running index
p	Number of principal components or positive integer
r	Number of rows
$\boldsymbol{r}_i$	Rotation matrix
$\boldsymbol{s}$	Singular value matrix
$\boldsymbol{s}_n$	Spectrum for component n
t	Number of spectra in a library
$\boldsymbol{u}$	Matrix from singular value decomposition treatment
$\boldsymbol{v}$	Matrix from singular value decomposition treatment

(cont.)

(cont. Abbreviations and Symbols)

x	Individual coefficient for a particular function
A	Concentration of a sample
A_λ	Absorbance at wavelength λ
A_{max}	Maximum absorbance in a spectrum
$\boldsymbol{B}$	Coefficient for translation-rotation transformation
$B_j(x)$	B-spline basis
$\boldsymbol{C}$	Scale coefficient in wavelet transform
$\boldsymbol{C}^{(j)}$	Scale coefficient at resolution level j in wavelet transform
$\boldsymbol{C}^{(1)}_{D_{8,n}}$	Scale coefficients derived from the Daubechies D_8 wavelet function for approximate derivative calculation
$\boldsymbol{C}^{(1)}_{D_{18,n}}$	Scale coefficients derived from the Daubechies D_{18} wavelet function for approximate derivative calculation
$\boldsymbol{C}_{ij}$	Chromatographic data
$C(n)$	Constant for nth order derivative
C_ψ	Constant in inverse continuous wavelet transform
$\boldsymbol{D}$	Wavelet coefficient in wavelet transform
D_{2m}	Daubechies wavelet function
$D_{2\tilde{m}}$	Daubechies wavelet function
D_{absder}	Absolute derivative
$D_{absdiff}$	Absolute difference
D_{corr}	Correlation coefficient
D_{Edist}	Euclidean distance
D_{sqrder}	Square derivative
$D_{sqrdiff}$	Square difference
$\boldsymbol{D}^{(j)}$	Wavelet coefficient at resolution level j in wavelet transform
$\boldsymbol{E}_{ik}$	Noise
F_λ	Fraction of total absorbance in a spectrum

(cont.)

(cont. Abbreviations and Symbols)

$\boldsymbol{G}$	High-pass filter in wavelet transform
$\boldsymbol{H}$	Low-pass filter in wavelet transform
$\boldsymbol{I}$	Identity matrix
J	Resolution level in wavelet transform
L	Length of wavelet function
N	Number of coefficients
$N_{c,j}$	Data size of scale coefficient at resolution j
$N_{d,j}$	Data size of wavelet coefficient at resolution j
$N_m(x)$	mth order basis spline wavelet function
$\boldsymbol{P}$	Loading matrix
P_C	Peak center
P_H	Peak height
P_σ	Peak width
$\boldsymbol{R}$	Transformation matrix
R_{comp}	Compression ratio of a spectrum
S	Constant in universal threshold calculation
$\boldsymbol{S}_{kj}$	Spectral data
$SNR(n)$	Signal-to-noise ratio for nth order derivative
$\boldsymbol{T}$	Score matrix
T_{univ}	Universal threshold value
V	Peak volum
$\boldsymbol{X}$	Data matrix
$\boldsymbol{Y}$	Data matrix in Fourier domain
α_j^{hard}	Coefficient in hard thresholding method
α_j^{soft}	Coefficient in soft thresholding method
γ	Regression coefficient
ε	Cutoff value in wavelet transform
λ	Wavelength

(cont.)

(cont. Abbreviations and Symbols)

λ_j^0	Eigenvalues
λ_{max}	Wavelength with maximum absorbance
$\lambda_{SW}(x)$	Shannon-Weaver entropy
$\phi_{j,k}(\lambda)$	Scale function in wavelet transform
$\psi_{a,b}(\lambda)$	Basis function in wavelet transform
$\psi_{j,k}(\lambda)$	Wavelet function in wavelet transform
$\Psi(\lambda)$	Mother wavelet function

CHAPTER 1

Introduction

1.1 Background

Chemometrics is a new discipline in chemistry that applied mathematical, statistical and other logic-based methods to analyze chemical data in particular in analytical chemistry (Massart, *et. al.*, 1988) and has been developed more than 27 years (Brown, 1995). The term chemometrics was coined by the Swedish physical organic chemist, S. Wold, in 1971 (Brereton, 1990). It is mainly employed in the following areas in chemical studies: (1) calibration, validation and significance of analytical measurement; (2) optimization of chemical measurement and experimental procedures; and (3) the extraction of maximum chemical information from analytical data (Haswell, 1992). One objective of chemometrics is to provide tools to assist in the conversion of raw data into useful information (Meglen, 1988). The applications of chemometrics have found considerable success in principal component and factor analysis, experimental design and optimization, statistical evaluation, signal processing and data analysis, and pattern recognition (Brereton, 1987).

Over the past decade, the capabilities of analytical instrumentation have increased greatly (Haswell, 1992). Most modern chemical instruments in laboratories are now attached to computers and robots. Many people want to employ simpler and cheaper instrumentation in monitoring their chemical process in real time. Chemometrics techniques can ensure that useful information can be extracted as completely as possible from these simpler and less informative instruments. Besides, there is a great demand in developing of advanced instruments to perform both qualitative and quantitative analyses in analytical chemistry. With the introduction of hyphenated instruments such

as high performance liquid chromatography coupled with diode array detector (HPLC-DAD) and gas chromatography interfaced to Fourier transform infrared spectrometer (GC-FTIR) or mass spectrometer (GC-MS), chemical analysis is changing from two-dimensional to multi-dimensional analysis. The huge increase in the number of chemical instruments drives the development of chemometrics (Brown, 1995). Meanwhile, the rapid development of microcomputer systems facilitates the use of chemometrics in chemistry because chemometrics require many mathematical and matrix computations. Chemometrics can also convert the routine data analysis process to an automatic one with the help of microcomputer.

The objective of this research work is to apply and modify selected chemometrics techniques to analyze experimental data from chemical and biochemical systems with modern analytical instruments. The research work focused on signal processing and factor analysis of data from HPLC-DAD and infrared (IR) spectroscopy. A series of programs were developed and coded in MATLAB® in this work for the required computations.

MATLAB®, which stands for matrix laboratory, is a high-performance language for technical computing (The MathWorks, Inc., 1996) and is widely used in chemometrics computation for processing of analytical signals (Mitra and Bose, 1994). It integrates computation, visualization and programming in an easy-to-use environment (The MathWorks, Inc., 1996). In chemometrics, most calculations are carried out in form of matrix. MATLAB® is a specialized language for matrix mathematics (Marcus, 1992) and is optimized in speed for modern microprocessor. Besides, MATLAB® is

available for various types of computers such as the UNIX workstations that include Sun, HP 9000, DEC Alpha, IBM RS/6000 and SGI, the IBM-PC with Microsoft Windows 3.1, 95, NT and Linux as the operating system and the Macintosh. Self developed MATLAB® programs can be run on different operating systems without any modification. MATLAB® is also an ideal computer language for teaching and learning chemometrics (O'Haver, 1989; Chau and Chung, 1995). It can save a lot of time to develop custom programs when compared with using traditional computer languages such as C and FORTRAN. MATLAB® can also interface to other windows based programs via dynamic link libraries (DLLs) (Vankeerberghen *et al.*, 1996a,b) and perform data exchange to and from the other programs via dynamic data exchange (DDE) (de Cerqueira and Poppi, 1996).

As stated previously, this research work focused on the application of signal processing techniques in analytical chemistry. Modern chemical instruments especially the hyphenated instruments generated huge amount of data in each measurement. Data compression techniques are required to reduce the data size for storage and further analysis. Commonly used data compression techniques that apply to analytical chemistry will be described in Chapter 2. This chapter will review the basic principle of the compression techniques that includes binary encoding, factor analysis, Fourier transform, spline and wavelet transform.

Recently, wavelet transform (WT), a new mathematical technique, is a very hot tool adopted in various fields of science and engineering. It has been applied successfully for signal processing in chemical studies since 1989. The number of

publications related to the applications of WT to manipulate chemical data increased rapidly in the last two years from one paper being published in 1989 to 18 papers in 1996 and 45 papers in 1997. In Chapter 3, applications of wavelet transform and its derivative wavelet packet transform (WPT) in chemistry from January 1989 to August 1998 will be reviewed. It includes the WT applications in analytical chemistry, quantum chemistry and chemical physics.

Wavelet transform is a new technique for data compression and denoising. The fast wavelet transform (FWT) and its derivative wavelet packet transform (WPT) are employed to compress IR spectrum for storage and spectral searching. In wavelet computation, data length is confined to be a power of 2 (Daubechies, 1992). However, chemical instruments cannot generate a data set exactly with a length of power of 2. As a result, it introduces some problems in wavelet computation. A new method called coefficient position retaining (CPR) method will be introduced in this work to handle experimental data with length not equal to 2^p with p being any integer. Details of this work will be described in Chapter 4.

Derivative spectroscopy is another signal processing technique that commonly utilized in analytical chemistry for data analysis owing to its popularity on the apparent higher resolution of the differential data when compared with the original data (Adam, 1995). Although the technique is an useful tool for data analysis, it has a major drawback in increasing the noise level in higher order derivative calculation (Brown, 1992). To perform higher order derivative calculation via numerical differentiation, noise reduction is usually performed between data from each order of successive

derivative calculation (Antonov and Stoyanov, 1996). A novel method based on wavelet transform was proposed in this project for approximate derivative calculation. Details will be presented in Chapter 5.

The second part of this research work is related to the application of factor analysis techniques to analyzing data from hyphenated instrument. Commonly used methods for such purpose include evolving factor analysis (EFA) (Maeder and Zilian, 1988; Keller and Massart, 1992a), alternating regression (AR) (Karjalainen, 1989; Karjalainen and Karjalainen, 1996), target factor analysis (TFA) (Gemperline, 1984), iterative target transformation factor analysis (ITTFA) (Vandeginste *et al.*, 1985), target transformation factor analysis (TTFA) (Hopke 1989), window factor analysis (WFA) (Mailnowski, 1992), fixed size moving window evolving factor analysis (FSMW-EFA) (Keller and Massart, 1991) and heuristic evolving latent projections (HELP) (Kvalheim and Liang, 1992; Liang *et al.*, 1992). In these methods, HELP predominates over the others in the recent studies (Toft and Kvalheim, 1994; Grung and Kvalheim, 1995). In this research work, we found that the HELP algorithm failed to analyze data from HPLC-DAD with low signal-to-noise ratio (SNR). In order to cope with this problem, a new method called fast wavelet transform heuristic evolving latent projections (FWT-HELP) was developed in this investigation. It was found that the performance of the HELP algorithm can be improved with FWT. Details will be discussed in Chapter 6.

On the next century, it is expected that there is a dramatic growth for traditional Chinese medicine (TCM) and its associated products in the world (Wong, 1997). According to a recent study by the Massachusetts Institute of Technology (MIT), Hong

Kong has the potential to be developed into an international center of TCM (Wang *et al.*, 1997). Research works on TCM become an active research area in different universities in Hong Kong. However, the effectiveness of TCM products depends greatly on the quality of the natural materials used (Lee, 1997). Besides, safety on the TCM products is another issue to be concerned by many international organizations such as the Food and Drug Administration (FDA) in U.S. So, both qualitative and quantitative analysis should be performed on raw materials and TCM products before they are sold to the market. From the point of view of a chemist, chemical constituents of Chinese herbs are a complex black system (Liang *et al.*, 1993). There is no a priori information concerning the chemical composition of the samples. We need to develop a suitable method to analyze the major constituents. In Chapter 7, the performance of the proposed FWT-HELP algorithm in the real situation was explored. A traditional Chinese medicinal herb, *Cordyceps sinensis*, was chosen as an example in this work and analyzed with HPLC-DAD.

In conclusion, wavelet transform is the framework of this research work and thesis. Novel applications which were based on WT were studied and developed and applied to analyze data from chemical and biochemical systems.

CHAPTER 2

A Review on Signal Compression of Spectroscopic Data in Analytical Chemistry

2.1 Introduction

Nowadays, most of the chemical instruments are computerized due to rapid development of modern microelectronics technology. A microcomputer is usually connected to an instrument for control of devices, data acquisition, processing and interpretation (Ratzlaff and Ratzlaff, 1992). Besides, there is a growing tendency to combine different chemical devices together to form a hyphenated instrument. This approach greatly enhances information acquisition and even allows experimental works not possible to be achieved before. Hyphenated instruments such as high performance liquid chromatography coupled with diode array detection system (HPLC-DAD) and gas chromatography-mass spectrometry (GC-MS) are widely utilized in chemical laboratories for qualitative and quantitative analyses (Karjalainen and Karjalainen, 1996). Other two-dimensional (2D) techniques include two-dimensional Fourier transform infrared (2D FT-IR) spectrometry and nuclear magnetic resonance (2D FT-NMR) spectrometry. In addition, application of three-dimensional techniques are becoming routine in chemical studies and those of four-dimensional technique has been proposed (Kuo *et al.*, 1991).

The rapid development of multi-dimensional technique and chemical instruments lead to some problems. One of them is that piles of experimental data are generated from each run of measurement. For example, several Mega-bytes (MB) of data can be produced from each run of a GC-MS or HPLC-DAD analysis. Such an enormous amount of data requires a lot of archive space and much longer time for further processing. Though mass storage systems such as CD-ROM (540 MB) and removable

disk (from 100 MB to 4.7 Giga-bytes (GB)) are available in the market, a higher storage density (number of spectra or signals per MB of storage space) would be preferred. One possible way to solve this problem is through signal compression. In chemical analysis, signal compression is very important especially in setting up digitized spectral library (Warr, 1993a). It is because signal compression can diminish the size of the original database and reduce the time for spectral searching. Besides, with the popularity of the Internet, it provides a new way for the exchange of scientific information (Fenyö *et al.*, 1996). Signal compression techniques can save the bandwidth of the network and improve the data transfer rate. In order to apply signal compression techniques in chemistry, the following requirements have to be fulfilled (Alsberg *et al.*, 1994).

1. It is possible to compress the signal using the chosen method.
2. The compression method is not be too computing intensive.
3. The discrepancy between the original data set and that reconstructed from the compressed representation is acceptable.
4. The compressed representation itself should be operational which means that common numerical methods can be applied to both the compressed and original representations in the same manner.

In general, signal compression methods can be classified into two categories. The first one is to lower the resolution of the original spectra by reducing the numbers of data to be retained while the second one is to compress spectra through transformation to other domains. A review on common compression methods for spectroscopic studies in analytical chemistry will be presented in this chapter.

2.2 Data Compression Methods in Analytical Chemistry

2.2.1 Binary Encoding

The most primitive and natural preprocessing of experimental spectral data for qualitative identification is binary encoding (Scott, 1988). The binary encoding method is based on division of the wavelength, time or mass range scale into a number of intervals with equal spacing. The Wyandotte ASTM code or "0/1" code is an example of the binary encoding (Heite *et al.*, 1978). In this approach, if one or more peaks are located in a certain interval having intensity greater than a preset intensity threshold, for instance, 1% of the most intense peak within a spectrum, the code "1" is assigned to this position. Otherwise, the code "0" is assigned. In this encoding scheme, a peak position may be allocated incorrectly next to the true one. To make sure the peak positions being coded correctly, windows with different widths may be applied. The major drawback of this method is that no intensity information is retained during the spectral compression process. This method is mainly used in compressing spectra from mass spectrometry (MS).

Since the Wyandotte ASTM code cannot be used to record complex spectral structure, Rann (Rann, 1972) proposed another scheme to compress infrared (IR) spectrum. In this scheme, an IR spectrum is divided into ten sections with unequal spacing. At the IR finger-print region (400 - 2500 cm^{-1}), more sections are allocated than others since it contains more useful information. Each section is further divided into ten equal spacing subsections. A number 0 to 9 is assigned sequentially to these subsections. Then, the maximum absorbance within each section is located and the

code for its position is equal to the assigned number of the subsection where it located. After the coding process, each IR spectrum is represented by a number with 10 digits that contains more information about the shape of the spectrum.

Both the Wyandotte ASTM and Rann's methods have a major disadvantage of losing large amount of information such as peak intensity in the compression process. Adams and Black (Adams and Black, 1986) proposed an error-free data compression scheme which is based on the Huffman shift method (Amsterdam, 1986) to code IR spectra. A data point is stored as a binary code word rather than a digital value. Since the possible digital levels for a recorded spectrum do not occur equally, a reduction in storage space can be achieved by assigning shorter binary words to data that occurs more frequently. This approach can save up to 70% of memory space with both peak position and intensity of each wavelength being recorded.

Binary encoding is a simple process but not a good way to compress spectrum because it only records the locations for the most significant peaks with no information about the peak shape and intensity. This scheme is mainly used by the electronic data sorting equipment in 1970s such as the punch card or paper tape reader (Rann, 1972). In recent years, researchers tend to employ different advanced mathematical techniques to compress spectra. Some of the important ones including factor analysis and transformation techniques such as Fourier transform, spline and wavelet transform.

2.2.2 Factor Analysis

Factor analysis (FA) is a multivariate technique for reducing an original data matrix to the one with the lowest dimensionality by using the orthogonal factor space (Malinowski, 1991). Predictions and/or recognizable factors can be obtained from the transformed data matrix. Principal component analysis (PCA) is a major technique that widely adopted in FA (Aries *et al.*, 1991; Smith, 1991). Once PCA is applied, the original data are projected into a much smaller space spanned by the first several eigenvectors. The purpose of FA treatment is not only to reduce the dimensionality of the original data but also to preserve the majority of the original information in a much smaller new space (Wang and Isenhour, 1987).

The PCA process starts with a data matrix X with a dimension of $r \times c$, to represent a set of spectra where r denotes the number of observations adopted in FA and c the number of variables in every observation. X is then decomposed by a mathematical process called singular value decomposition (SVD) into different eigenvalues and eigenvectors (Malinowski, 1991; Lupu and Todor, 1995). The magnitude of an eigenvalue indicates the relative importance of the corresponding eigenvectors because not all eigenvectors are useful. Several criteria have been proposed to determine the number of true factors in X. The most popular one is the average eigenvalue criterion as proposed by Kaiser (Wang and Isenhour, 1987). Those eigenvectors corresponding to eigenvalues having value smaller than an average eigenvalue are assumed to contain more noise than the true information and, hence, can be rejected. An empirical function called the factor indicator function

(IND) (Eq. (2.1)) as proposed by Malinowski (Malinowski, 1991) can also be utilized to determine the number of true factors.

$$\mathrm{IND} = \sqrt{\frac{\sum_{j=c+1}^{n} \lambda_j^0}{r(c-n)^5}} \tag{2.1}$$

In the above expression, n, λ_j^0, r and c represent the number of factors, the eigenvalues, the number of rows (or observations adopted in FA) and number of columns (or variables in every observation) respectively. The IND function is calculated repeatedly for different values of n. Then, the number of true factors is equal to n which has a minimum IND value. After finding the number of true factors, the principle components can be extracted from the eigenvectors. To compress a spectral library, it can be accomplished by matrix multiplication of the library matrix (dimension of $t \times n$) by the principal components (dimension $n \times p$, $n > p$) where t and p denote the number of spectra in the library and the number of principal components respectively. The much smaller spectral library (dimension $t \times p$) can then be used for library search and other purposes.

Normally, FA is coupled with different transformation techniques to enhance the spectral compression efficiency. Hangac *et al.* (Hangac *et al.*, 1982) proposed a compression method by applying FA to the experimental spectrum first and followed by Karhunen-Loève transform. In this way, a fivefold reduction in storage was achieved. The aim of utilizing the Karhunen-Loève transform is to reduce the dimensionality of the original data matrix by eliminating redundant variables.

2.2.3 Fourier Transform

Fourier transform (FT) is a mathematical technique that maps information from one domain into another domain (Glasser, 1987a-c; Wang and Isenhour, 1987). It is widely used in chemistry for signal smoothing and compression. Usually, FT converts data from the time domain to frequency domain or vice versa. The information containing in these two domains is essentially the same. However, different properties in the two domains sometimes make the information easier to handle in one domain than in the other (Wang and Isenhour, 1987). In order to perform FT in the most economical manner, the fast Fourier transform (FFT) as developed by Cooley and Tukey in 1965 (Zupan, 1989) is usually employed. Suppose a spectrum with a data matrix $\boldsymbol{X}$ is measured by a chemical instrument at $2N$ different wavelengths with N being a multiple of 2. After performing FFT on the spectrum, the Fourier coefficients obtained are stored in the $\boldsymbol{Y}$ matrix. The elements of $\boldsymbol{X}$ and $\boldsymbol{Y}$ are related to each other by (Oppenheim and Schafer, 1989)

$$Y(k) = \sum_{n=0}^{2N-1} X(n) W_{2N}^{kn} \tag{2.2}$$

with $W_{2N} = \exp\left(-i(2\pi / 2N)\right)$, $i = \sqrt{-1}$, or

$$X(n) = \frac{1}{2\pi} \sum_{k=0}^{2N-1} Y(k) W_{2N}^{-kn} \tag{2.3}$$

where k and n are running indices from 0 to $2N$-1. Owing to the symmetric properties of $\boldsymbol{Y}$ (Oppenheim and Schafer, 1989), the real part (RE) are related to the imaginary part (IM) of the elements of $\boldsymbol{Y}$ by

$$\mathrm{RE}\{Y(k')\} = \mathrm{RE}\{Y(2N - k')\} \tag{2.4}$$

and $$\mathrm{IM}\{Y(k')\} = -\mathrm{IM}\{Y(2N - k')\} \quad (2.5)$$

with $k' = 1$ to N-1. Therefore, only the first (N+1) Fourier coefficients have to be stored. The last (N-1) coefficients can be derived from Eqs. (2.4) and (2.5) in spectrum reconstruction.

In some cases, further reduction on the number of Fourier coefficients for a spectrum may not affect the quality of the original data after data reconstruction. Owens and Isenhour (Owens and Isenhour, 1983) utilized FT with the clipping (single bit quantization) technique to compress IR spectra in the time domain. Since the IR absorbance equals to 0 at 0Hz, the values in the IR time domain oscillate above and below 0. The clipping operation replaces a set of data points with the corresponding set of bits. Data points with magnitudes greater than 0 are assigned to have a bit value of 1, while the remaining points with a bit value of 0. A 16-fold reduction in data storage was reported in this approach. The clipped FT procedure can also be applied to compress mass spectra (Lam *et al.*, 1981) and NMR spectra (Crawford and Larsen, 1977). Recently, Chau and Tam (Chau and Tam, 1994) proposed different methods to reduce the number of Fourier coefficients for ultraviolet-visible spectra as recorded by a photodiode array detector. The simplest way is to store the Fourier coefficients with values greater than 0.1% of the maximum one. Another method involves the selection of a group of high frequency components in the last 15/16 portion of the first half of the FT transform spectrum and determines the number of Fourier coefficients to be retained based on some criteria. They also proposed a new method by fitting the logarithms of the first-half

elements of the power spectrum by a fourth-degree polynomial. A 90% reduction in the file size for data storage was reported.

2.2.4 Spline

In many cases, a spectrum can be represented by a continuous function such as sine waves, Gaussian peaks, Lorentzian peaks or spline (Alsberg, 1993). Spline is a mathematical technique by dividing a curve into a set of segments or intervals (Alsberg and Kvalheim, 1993). A polynomial up to a certain degree of k may be defined within the limits of different intervals that are defined by a set of points called knots. The polynomials between neighboring intervals are connected to form a continuous function. Alsberg *et al.* (Alsberg, 1994) proposed to compress Fourier transform infrared (FT-IR) spectra by using B-splines. B-splines or basis-splines represent a curve as a linear combination of basis functions. Each basis function is described by a vector of real numbers which determines its size and shape. This vector h^{T} is called the knot vector. The B-spline basis set ($B_j(x)$, $j \in [1,\ldots q-k-1]$, where q denotes the number of elements in the knot vector and k the degree of the spline) is completely described by the knot vector as

$$\mathrm{h}^{\mathrm{T}} \rightarrow \left\{B_1(x), B_2(x), \ldots, B_{q-k-1}(x)\right\} \tag{2.6}$$

A signal function f (the FT-IR spectrum) can thus be formulated as a linear combination of the B-spline basis set $B_j(x)$:

$$f(x) = \sum_{j=1}^{n} c_j B_j(x) \tag{2.7}$$

Therefore, the FT-IR spectrum is described by its knot vector h^T and coefficient vector c^T only, resulting in data compression. A 99% compression in file size was reported in their study.

2.2.5 Wavelet Transform

Wavelet transform (WT) is a new mathematical technique (Daubechies, 1992) and has been applied widely in various field of science and engineering for data compression and denoising (Daiguji *et al.*, 1997; Hilton, 1997; Idris and Panchanathan,1997; Provaznik and Kozumplik, 1997). Since 1989, WT was introduced in chemistry as a tool for signal processing. More than 90 papers have been published from January 1989 to August 1998. Applications can be found in the areas of analytical chemistry, quantum chemistry, chemical physics and chemical engineering. A detailed review on the application of WT in chemistry will be presented on next chapter.

2.3 Conclusions

In this chapter, various spectral compression techniques that are commonly applied in analytical chemistry are described. Spectral compression is a very important technique in modern analytical chemistry owing to the rapid development in chemical instrumentation. For the compression methods mentioned above, Fourier transform is the most popular one because it is easy to implement. However, other new mathematical

techniques such as wavelet transform are under development to give more efficient compression procedures.

CHAPTER 3

An Introduction of Wavelet Transform in Chemical Analysis

3.1 Introduction

Wavelet transform (WT) becomes a popular topic in different field of science and engineering for signal and image processing after the publication of an important paper by I. Daubechies (Daubechies, 1988) in 1988. This new mathematical technique has been demonstrated to be fast in computation with localization and quick decay properties in contrast to existing popular methods, especially, the fast Fourier transform (FFT). Theory of WT was developed extensively in 1980's (Daubechies, 1992). Several excellent reference books on WT were published in 1992 and afterwards. Examples include *Ten Lectures on Wavelets* (Daubechies, 1992) which is a comprehensive account containing the lectures by the author at the NSF-CBMS wavelet conference held at the University of Lowell in June 1990 (Kaiser, 1994). Chui also published a series of reference books on WT (Chui, 1992a-b; Chui *et al.*, 1994; Chui, 1997) which provide general information in wavelet theory, algorithms and applications. A number of MATLAB® based wavelet public domain softwares, which can be downloaded from Internet, are available for non-commercial, research and educational users. That includes WaveLab from Stanford University (Buckheit and Honoho, 1995) and the others (Misiti *et al.*, 1995). Commercial product includes the Wavelet Toolbox from The MathWorks, Inc. (Misiti *et al.*, 1996) and WavBox from Computational Toolsmiths (Taswell, 1995).

From 1989 onwards, WT has been applied for signal processing in chemical studies owing to its efficiency, large number of basis functions available as compared with Fourier transform (FT), and high speed in data treatment. One of the main feature of WT is that it may decompose a signal directly according to the frequency and

represent it in the frequency domain distribution state in the time domain. In the transformation, both the time and frequency information of the signal are retained. In chemical studies, the time domain can be replaced by the other domain such as wavelength. With proper identification of the scales with frequency, higher frequency signals can be separated from the lower one, in the sense that it has zoom-in and zoom-out capability at any frequency. Since WT can focus on any small part of a signal, it can be called a mathematical microscope (Daubechies, 1988). Another feature of WT is that the development of signals in the frequency domain can be constituted with a flexible choice of waveforms rather than with only trigonometric ones as a basis. For instance, in FFT, only sine and cosine function can be chosen as the basis. In contrast, a lot of such functions can be selected in WT. Therefore, WT is a more powerful method to depict relationships among different variables.

Since 1989, more than 90 papers (Table 2.1) have been published by applying WT in chemical study. Applications can be found in the area of analytical chemistry, quantum chemistry and chemical physics. In analytical chemistry, WT was applied successfully in the following areas that include flow injection analysis (FIA), high performance liquid chromatography (HPLC), infrared (IR) spectrometry, mass spectrometry (MS), nuclear magnetic resonance (NMR) spectrometry, ultra-violet visible (UV-VIS) spectrometry and voltammetry. In this chapter, research works on the application of WT in chemical studies as mentioned above are reviewed.

Table 3.1 Number of published papers from January 1989 to August 1998 that are related to application of wavelet transform in chemistry.

Year	Number of published papers
1989	1
1990	0
1991	0
1992	2
1993	5
1994	5
1995	4
1996	19
1997	45
August 1998	10
Total	91

3.2 Theory of Wavelet Transform

3.2.1 Wavelet Transform

Wavelet transform involves the decomposition of a signal function or vector (e.g., a spectrum or an analytical signal) into simpler, fixed building blocks at different scales and positions (Dai *et al.*, 1994). Different algorithms were proposed to achieve this purpose. In the mid-1970s, Coifman and his coworker developed the automatic decomposition method which had a very strong influence on the development of wavelet theory (Coifman and Wickerhauser, 1992). In 1984, Grossman and Morlet (Grossman and Morlet, 1984) proposed the continuous WT which allows the decomposition of a signal into contributions from both space and scale domains based on the invariance under the affine group, namely, translation and dilation. In 1989, Mallat (Mallat, 1989a) introduced the multi-resolution signal decomposition (MRSD) algorithm. Daubechies adopted this approach to construct families of compact supported wavelets and coupled it to quadrature mirror filtering. This provides a general way for constructing orthogonal wavelet bases and leads to the implementation of the fast wavelet transform (FWT) algorithm.

Like Fourier transform, WT operates on a signal $f(\lambda)$ and transforms it linearly from its domain, for example, the wavelength for a UV-VIS spectrum, to a different domain. In Fourier analysis, only sine and cosine functions which are localized in frequency domain can be applied to a function. It has difficulty to process a function having components that localized in the time domain (Cody, 1992). As a result, a small frequency change in FT produces changes everywhere in

this domain. On the other hand, wavelet functions are localized both in frequency or scale and in time via dilations and translations of the mother wavelet respectively. This leads to compact representation of large classes of functions and operators in the wavelet domain. Spectrum with sharp spikes, such as IR spectrum, may well be approximated by substantially fewer wavelet basis functions as compared to the sine and cosine functions adopted in Fourier analysis.

In WT treatment, all basis functions $\psi_{a,b}(\lambda)$ can be derived from a mother wavelet $\Psi(\lambda)$ through the following dilation and translation processes (Daubechies, 1992):

$$\psi_{a,b}(\lambda) = a^{-1/2}\Psi\left(\frac{\lambda - b}{a}\right) \quad a,b \in \Re \text{ and } a \neq 0 \tag{3.1}$$

where a and b are, respectively, the scale (dilation) and position (translation) parameters expressed in real number $\Re$. The basic idea of WT is to represent any arbitrary function $f(\lambda)$ as a superposition of wavelets. The continuous wavelet transform (CWT) of $f(\lambda)$ is given by

$$W_f(a,b) = \int_{-\infty}^{+\infty} \psi_{a,b}(\lambda) f(\lambda) d\lambda \tag{3.2}$$

with $a > 0$ and b having arbitrary values. The inverse CWT can be obtained through the following formula (Daubechies, 1992)

$$f(\lambda) = \frac{1}{C_\psi} \int_0^{+\infty} \frac{da}{a^2} \int_{-\infty}^{+\infty} W_f(a,b) \psi_{a,b}(\lambda) db \tag{3.3}$$

where C_ψ is a constant depending only on ψ. The dilation parameter a and the translation parameter b in Eq. (3.1) vary continuously in the case of CWT, but are restricted to a discrete lattice in the case of discrete wavelet transform (DWT)

(Palavajjhala *et al.*, 1994). In DWT, the wavelet function $\psi_{j,k}(\lambda)$ can be expressed as

$$\psi_{j,k}(\lambda) = 2^{-j/2}\Psi\left(2^{-j}\lambda - k\right) \tag{3.4}$$

with $a = 2^j$, $b = 2^j k$ and j being the resolution level. In this expression, the variables k and j have value of $0, \pm 1, \pm 2, \ldots$ and represent, respectively, the extent of the translation and dilation which describes the frequency of $\psi_{j,k}$ near k (Chau *et al.*, 1996). For discrete data, k can be considered as the shift in the unit of the channel number. Like FT, decomposition of $f(\lambda)$ with respect to the wavelet function series $\left\{\psi_{j,k}(\lambda)\right\}$ is described by the following formula (Daubechies, 1992)

$$f(\lambda) = \sum_{j=-\infty}^{+\infty}\sum_{k=-\infty}^{+\infty} c_k^{(j)}\psi_{j,k}(\lambda). \tag{3.5}$$

Therefore, the signal is represented by a set of coefficients $\left\{c_k^{(j)}\right\}$ in the wavelet domain. The fast implementation method as developed by Mallat (Mallat and Hwang, 1992) for the DWT has made the wavelet method as an effective tool for processing chemical data.

Based on the recent works in WT (Vaidyanthan, 1993), DWT can be considered as a filtering technique under the terminology of signal processing. A wavelet basis is characterized by a particular set of numbers, called wavelet filter coefficients. The DWT treatment is to perform two related convolutions on the signal with a low-pass filter $\boldsymbol{H}\left(=\left\{h_k\right\}\right)$ and a high-pass filter $\boldsymbol{G}\left(=\left\{g_k\right\}\right)$. Then, the signal is converted into two bases with equal size, that is

$$c_k^{(J)} = \sqrt{2}\sum_n c_n^{(J-1)} h_{n-2k} \tag{3.6}$$

and

$$d_k^{(J)} = \sqrt{2}\sum_n c_n^{(J-1)} g_{n-2k} \tag{3.7}$$

with $n = -\infty$ to $+\infty$ and k being a running index which has a variable length depending on the type of wavelet filter and the data length used. The variables h_k and g_k in these equations denote the coefficients of the low-pass and high-pass filters, respectively, with the following properties (Mallat, 1989b)

$$g_k = (-1)^k h_{1-k} \tag{3.8}$$

where

$$\sum_k h_k = 1 \text{ and } \sum_k g_k = 0. \tag{3.9}$$

In the above formulae, the outputs of $\boldsymbol{H}$ and $\boldsymbol{G}$ are referred to, respectively, the smoothed or scale coefficients (major information), $\boldsymbol{C}^{(J)}\left(=\left\{c_k^{(J)}\right\}\right)$, and the differential or wavelet coefficients (detail information), $\boldsymbol{D}^{(J)}\left(=\left\{d_k^{(J)}\right\}\right)$, of the signal data.

In practice, DWT is conducted in several steps via the multi-resolution signal decomposition algorithm. A signal with a finite length $N\left(=2^p\right)$ with p being any positive integer are first collected as $\boldsymbol{C}^{(0)}$. In order to use Eqs. (3.6) and (3.7) successfully, $\boldsymbol{C}^{(0)}$ is extended periodically at the two extremes. Then, the filters $\boldsymbol{H}$ and $\boldsymbol{G}$ are applied to $\boldsymbol{C}^{(0)}$ to generate the coefficient $\boldsymbol{C}^{(1)}$ and $\boldsymbol{D}^{(1)}$ at resolution level 1. Afterwards, calculations will be performed at resolution 2 by using the quantities $\boldsymbol{C}^{(1)}$ from the previous resolution level to obtain $\boldsymbol{C}^{(2)}$ and $\boldsymbol{D}^{(2)}$. The iterative process will be continued until the desired resolution level $J\left(\leq \log_2 p\right)$ is attained. Thus, the

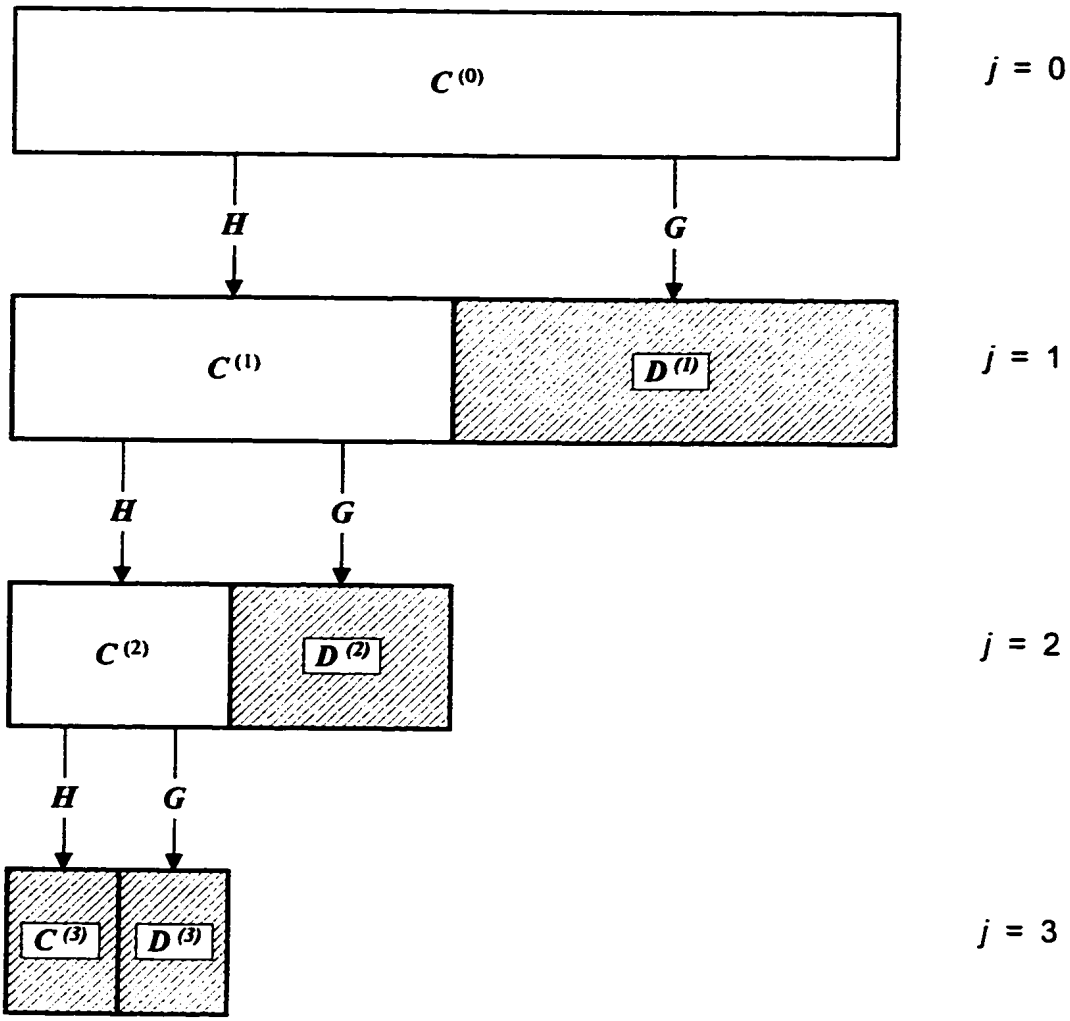
output comprises all wavelet coefficients $\left\{\boldsymbol{D}^{(J)}, \boldsymbol{D}^{(J-1)}, \ldots, \boldsymbol{D}^{(1)}\right\}$ from resolution level 1 to J and the scale coefficients $\boldsymbol{C}^{(J)}$ at the desired resolution level J (Figure 3.1). At each resolution level j, the signal is filtered and half of the samples are kept which give a scale twice as coarse as the previous level. An example on the MRSD calculation can be found in the paper by Chau *et al.* (Chau *et al.*, 1996). The inverse wavelet transform as derived from Eq. (3.3) is defined as follows

$$c_k^{(J-1)} = \sqrt{2}\left(\sum_n c_n^{(J)} h_{k-2n} + \sum_n d_n^{(J)} g_{k-2n}\right) \tag{3.10}$$

from resolution level J down to 0. Details of wavelet theory will not be presented in this chapter and can be found in most of reference books (Chui, 1992a-b; Byrnes *et al.*, 1994; Chui *et al.*, 1994; Wickerhauser, 1994a; Hernández and Weiss, 1996; Hubbard, 1996; Strang and Nguyen, 1996).

Many wavelet functions have been proposed by various workers. Different bases face different tradeoffs on the signal such as how compactly these functions are localized in space and how smooth they are. The most simple one, the Haar wavelet, which is also the first member of the family of Daubechies wavelets (Daubechies, 1992), has been known for more than eighty years in various mathematical fields. The Daubechies family of wavelets D_{2m}, with m being any positive integer from 1 to 10, includes members ranging from highly localized to highly smooth one and it is the most popular one in WT applications. Daubechies' wavelets are orthonormal, compactly supported, having maximum number of vanishing moments for the support and reasonably smooth (Palavajjhala *et al.*, 1994). In addition, there are many other wavelet families such as Meyer wavelet, Coiflet wavelet, spline wavelet,

Figure 3.1 A schematic diagram for the fast wavelet transform computation. The slanting line represents coefficients to be stored.



orthogonal wavelet and local cosine basis and others (Chui, 1992a; Daubechies, 1992). Both Daubechies and spline wavelets are employed commonly in chemical studies.

Spline wavelet is another type of wavelet function that was used in some chemical studies (Lu and Mo, 1996; Zou and Mo, 1996, 1997a-c; Bao *et al.*, 1997a; Fang and Chen, 1997). The *m*th order basis spline (B-spline) wavelet, N_m, is defined as follows (Lu and Mo, 1996):

$$N_m(x) = N_{m-1}(x) * N_1(x) = \int_0^1 N_{m-1}(\lambda - x)d\lambda \tag{3.11}$$

with $m \geq 2$. The * denotes the convolution operation between N_{m-1} and N_1. The *k*th term of N_m is given by

$$N_m(x_k) = \sum_{j=0}^{k} N_{m-1}(x_j) N_1(x_{k-j}) \tag{3.12}$$

with $j \geq 0$ (Hubbard, 1996). The result is equivalent to the sum of product between coefficients in N_{m-1} and N_1 in a shifted manner. The mother wavelet function $\Psi(\lambda)$ may be expressed as

$$\Psi(\lambda) = \frac{1}{2^{m-1}} \sum_{j=0}^{3m-2} (-1)^j N_{2m}(j+1) N_{2m}^{(m)}(2\lambda - j) \tag{3.13}$$

where $N_{2m}^{(m)}(\lambda)$ is the *m*th-order derivative of the function $N_{2m}(\lambda)$:

$$N_{2m}^{(m)} = \sum_{j=0}^{m} (-1)^j \binom{m}{j} N_m(\lambda - j). \tag{3.14}$$

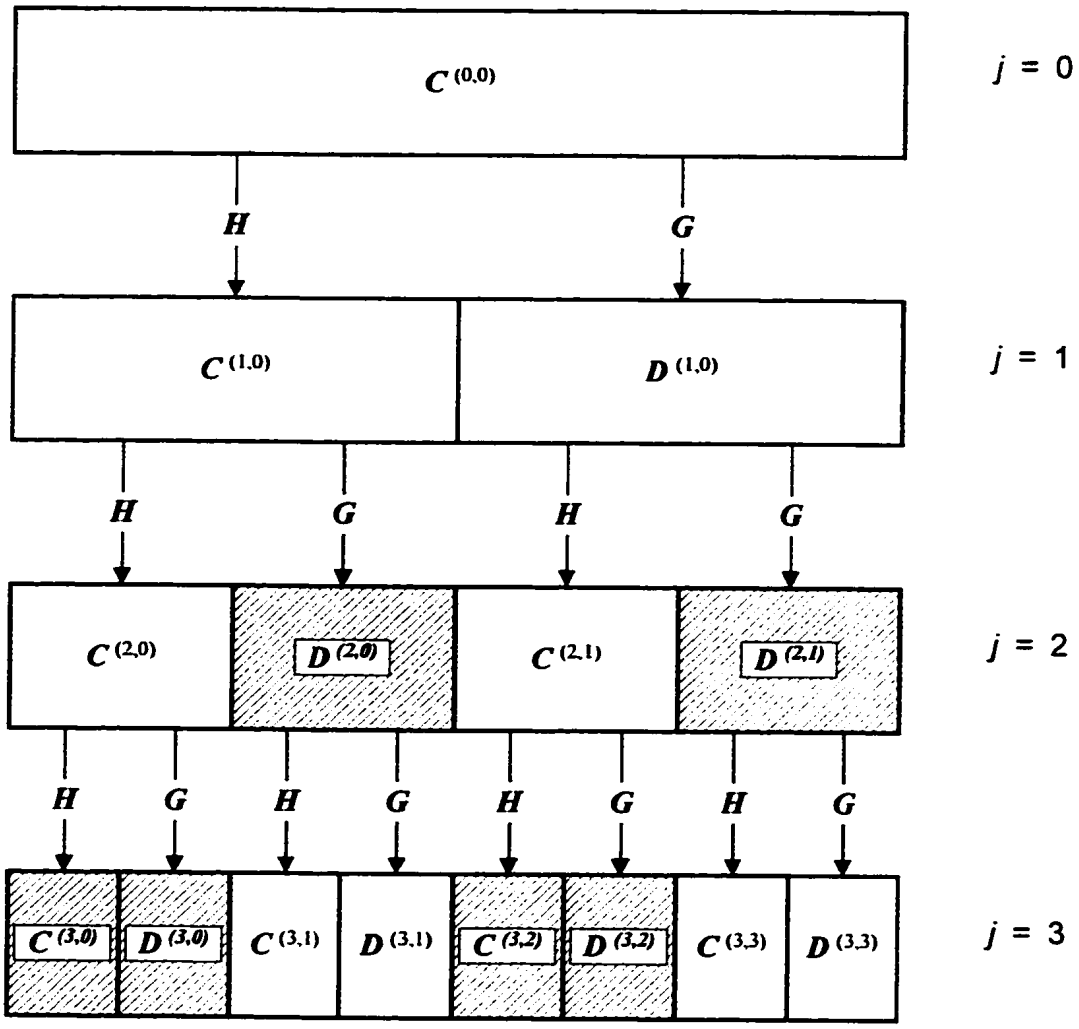
3.2.2 Wavelet Packet Transform

Coifman and Wickerhauser (Coifman and Wickerhauser, 1992) introduced the concept of wavelet packet transform (WPT) for signal processing to generalize the time-frequency analysis of WT (Cody, 1994). In DWT treatment, only outputs from the low-pass filter $\boldsymbol{H}$ are processed by WT. However, in the WPT treatment, both outputs from the low-pass filter $\boldsymbol{H}$ and high-pass filter $\boldsymbol{G}$ are manipulated by WT (Strang and Nguyen, 1996) (Figure 3.2). Wavelet packet is a family of basis functions which allow one to select different orthogonal bases for a given signal vector with finite energies. The standard wavelet basis is an orthogonal basis among the family and each basis is called a wavelet packet basis. Different effectiveness of concentrating energy of a given signal may be attained with different distinct wavelet packet bases chosen. Coifman and Wickerhauser proposed a selection scheme for the best wavelet packet basis with certain selection criteria such as the Shannon-Weaver entropy measure function (Coifman *et al.*, 1994). The output from WPT treatment comprises a suitable combination of wavelet packet bases. For example, in Figure 3.2, $\left\{\boldsymbol{C}^{(3,0)}, \boldsymbol{D}^{(3,0)}, \boldsymbol{D}^{(2,0)}, \boldsymbol{C}^{(3,2)}, \boldsymbol{D}^{(3,2)}, \boldsymbol{D}^{(2,1)}\right\}$ can be chosen as one of the best basis. The total data length of all coefficients is equal to that of the original data.

3.3 General Publications of Wavelet Transform in Chemistry

Two tutorial papers have been published in 1997. Walczak and Massart (Walczak and Massart, 1997a) published a tutorial on the introduction of WPT for noise

Figure 3.2 A schematic diagram for the wavelet packet transform computation. The slanting line represents coefficients to be stored.



suppression and signal compression. The basic principle of WPT calculation and best basis selection was introduced in their paper. Alsberg *et al.* (Alsberg *et al.*, 1997a) presented another tutorial on the introduction of WT to chemometricians. The basic principle and the properties of CWT, DWT and WPT were mentioned in their tutorial. These workers pointed out five major applications of WT in chemistry including signal denoising, baseline removal, zero crossing, signal compression and wavelet regression.

Three more papers have been published in 1997 to provide an introduction of WT to chemists. Wang *et al.* (Wang *et al.*, 1997a) published a paper on the introduction of WT and its applications in chemistry in China. As compared with the work by Alsberg *et al.* (Alsberg *et al.*, 1997a), they pointed out another five major applications of WT in chemistry that include spectral data compression, modification of the quality of chemical signals (denoising), quantum chemistry calculation, chemical dynamic analysis, and chemical fractals. Walczak and Massart (Walczak and Massart, 1997b) also published another paper on the introduction of WT in analytical chemistry. Fundamentals of wavelets and WT were presented in their paper. They pointed out a new application of WT in chemistry for the modeling of multivariate data sets such as calibration and classification of near-IR (NIR) data. An introduction of the mathematical background of FWT on compact intervals was presented by Depcznski *et al.* (Depcznski *et al.*, 1997). They emphasized on the problem of periodization of a general signal that leads to instabilities near the boundary of the interval. In their works, they suggested to use the Sturm-Liouville wavelets, which related to expansions of signals in terms of the eigenfunctions of Sturm-Liouville operators, to overcome this problem on the compact

interval. Wavelets on this interval provide a stable algorithms for non-periodic arbitrary signals calculation which is encountered in most of chemometrics studies.

In a recent paper by Walczak and Massart (Walczak and Massart, 1997c), they developed a new method, which is based on the variance spectrum, for best basis selection with WPT of a set of signals. These workers explained that the existing best basis selection method such as the Shannon-Weaver entropy method are only good for compression and denoising of individual signals but not for a set of signals. By adopting their method, a single best-basis will be chosen for the whole set of data rather than different best-base for an individual signal in a data set.

In chemical study, WT has been mainly utilized for signal smoothing, denoising and compression. A brief introduction on WT denoising and compression will be described in the following sub-sections.

3.3.1 Signal Smoothing and Denoising

One of the main goals in analytical chemistry is to extract useful information from recorded data. However, the achievement of this goal is usually complicated by the presence of noise. Nowadays, many chemical instruments are controlled by computers. It becomes a common practice to reduce the noise by employing digital processing methods such as filtering. In the past decades, a large number of filters were developed in different fields of science and technology. In spite of the existence of diverse filters, only a few, such as Savitzky-Golay, Fourier and Kalman filters

(Brown *et al.*, 1994, 1996), are extensively used by chemists. Nowadays, UV-VIS, IR and NMR spectroscopies are widely used by chemists for sample identification and characterization. In some cases, the true spectral data are masked with noise especially in samples with low concentrations. The most common way of noise suppression is by applying a suitable digital filter on the raw data. Noise is present in spectral data in the form of high frequency signal and causes some kinds of singularities in data treatment. The two most important properties of WT are that (1) singularities of a signal $f(\lambda)$ can be detected by the WT maxima (Mallat and Hwang, 1992) and (2) signal frequencies can be separated at different scales.

In mathematical study, the local singularity of a function is often measured by the Lipschitz exponents (Mallat and Hwang, 1992). Mallat and Hwang established a close relationship between WT and the Lipschitz exponents. In practice, since noise creates singularities with negative Lipschitz regularity, the modulus maxima of noise can be discriminated by looking at the evolution of their amplitudes along the scale level direction. If the modulus maxima have an amplitude which decreases rapidly with an increase in the scale level, this indicates that the corresponding singularities have negative Lipschitz exponents. These maxima are considered to be contributed by the noise and should be removed. Several methods have been proposed for discarding negligible coefficients or noise in the wavelet domain. These include absolute cutoff, relative energy, entropy criterion and decreasing rearrangements and fixed percentages methods (Coifman *et al.*, 1994). In the above methods, only coefficients with values greater than a predefined threshold value are retained. A tutorial on noise suppression with WPT was published by Walczak and Massart

(Walczak and Massart, 1997a). They introduced the universal thresholding algorithm as proposed by Donoho (Donoho, 1995) for coefficients selection. The universal threshold T_{univ} is defined as:

$$T_{univ} = \sqrt{2\log_2 S} \tag{3.15}$$

where $S = N\log_2 N$ for WPT and $S = N$ for DWT calculation. Hard and soft thresholdings were also introduced in their work for selection coefficients. In the former method, the absolute values of all transform coefficients are compared to T_{univ}. If a coefficient has a value less than T_{univ}, it is replaced by zero:

$$\alpha_j^{hard} = \begin{cases} 0 & \text{if } |\alpha_j| < T_{univ} \\ \alpha_j & \text{if } |\alpha_j| \geq T_{univ} \end{cases}. \tag{3.16}$$

In the above equation, the symbols α_j represents coefficients in the wavelet domain. As for the soft thresholding method, coefficients are selected in the following way:

$$\alpha_j^{soft} = \begin{cases} 0 & \text{if } |\alpha_j| < T_{univ} \\ \text{sign}(\alpha_j)(|\alpha_j| - T_{univ}) & \text{if } |\alpha_j| \geq T_{univ} \end{cases}. \tag{3.17}$$

Mittermayr *et al.* (Mittermayr *et al.*, 1996) performed a comparative study in denoising the Gaussian peak with the use of wavelet, Fourier and Savitzky-Golay filters. The Gaussian function was chosen in their work because it is commonly used to represent peaks and bands in chromatographic and spectroscopic studies. The simulated data consists of narrow Gaussian peaks with white noise and low signal-to-noise ratio (SNR). Then, the so-called universal soft thresholding was applied to the simulated data to give the estimated wavelet coefficients. Finally, the inverse WT was employed to reconstruct the smoothed data. The proposed wavelet filters that

include Harr, Daubechies, Symmlets and Coiflets can remove noises without distorting the signal peaks. Their results show that under the chosen conditions, the proposed wavelet method gave superior performance over the classical filter techniques in most cases. They also found that the effectiveness of the optimal denoising methods depends on the magnitudes of the noise level present and the original signal of interest. Barclay *et al.* (Barclay *et al.*, 1997) also performed a similar comparative study in denoising and smoothing of Gaussian peak using wavelet, Fourier and Savitzky-Golay filters. They classified signal denoising and smoothing as two different processes. Smoothing removes high frequency components of the transformed signal regardless of their amplitudes, whereas denoising removes small amplitude components of the transformed signal regardless of their frequencies. Their result indicated that DWT denoising gave better performance than DWT smoothing. Besides, they found that DWT smoothing treatment had a major drawback. Such drawback was apparent when the peak width of the fine structure was in the order of the Nyquist frequency.

In another work by Mittermayr *et al.* (Mittermayr *et al.*, 1997a), the ability of DWT to detect and estimate heteroscedastic noise was performed. Noise can be classified as heteroscedastic and homoscedastic noises. The amplitude of heteroscedastic noise is proportional to the amplitude of the underlying signal while that of homoscedastic noise do not have a direct relationship. They applied a simple F-test to non-overlapping intervals of wavelet coefficients in the highest resolution level and the differences in the variance was detected. Then, by fitting a smoothed

version of the signal with inverse WT versus a smoothed estimated of the local variance, the degree of heteroscedasticity in the signal can be estimated.

Jouan-Rimbaud *et al.* (Jouan-Rimbaud *et al.*, 1997) proposed a new approach which was called relevant component extraction for partial least squares to extract the relevant component from spectral data. WT was adopted to remove noise and irrelevant information from spectral data for multivariate calibration. After WT treatment on the spectral data, coefficients in the wavelet domain were sorted in decreasing order of their absolute values. Then, a smaller group of coefficients were selected by utilizing the minimum description length equation and used for the partial least squares calculation.

3.3.2 Signal Compression

Microcomputer systems are widely used nowadays to acquire and store digital spectral data in chemical study. Computerized database and spectral library are valuable in identification of sample spectra. In general, the more data is available in a library, the better are the search results. However, as the size of a library increases, more space and longer time are required to store the spectra and search the library.

The advancement in microelectronics has greatly enhanced mass storage capacity and processing speed. Archive of information of full spectra rather than only those of absorption peaks becomes more feasible. However, the demand of huge storage capacity is still somewhat inhibited for high resolution spectra. Even if this

problem can be resolved, the computer processing speed and the bandwidth of phone line or network is still a limiting factor. For instance, in order to transfer a spectral data set to a remote analytical site via phone line or Internet with low bandwidth, data compression is necessary to shorten the transfer time and to reduce the transferring cost. On the other hand, if a larger spectral library is used, the searching process is very lengthy and becomes an impractical task. To speed up the process, researchers may use several approaches. One of them is to reduce the resolution of the spectra to be archived in the spectral domain. An one-fold decrease in the resolution of a spectrum leads to only one-halves of the computational time and space being needed. The second method to speed up data processing is to compress the spectra to give smaller data set. The most commonly used compression technique in chemical studies is Fourier transform or its variants. The advantages of FT are frequency localization, orthogonality, and the availability of fast numerical algorithms. This technique has been applied extensively in processing digital data. Some workers (Binkley and Dessy, 1980; Bush, 1974) have pointed out that in addition to noise removal, FT can also be used for data compression. For example, Chau and Tam (Chau and Tam , 1994) applied fast FT (FFT) to minimize the storage of UV-VIS absorption spectra that collected from a photodiode array spectrophotometer.

In recent years, researchers proposed to make use of DWT or FWT and WPT for data compression (Chau *et al.*, 1996, 1997; Leung and Chau, 1997; Walczak and Massart, 1997a). The mathematical treatment for data compression via WT is similar to that for denoising. Data from chemical analysis are transformed to the wavelet

domain by employing either FWT or WPT. Then, a thresholding method is employed to select suitable coefficients in the wavelet domain for storage. By utilizing inverse FWT or WPT treatment, the compressed coefficients can be converted back to its original domain without significant information lost.

3.4 Applications of Wavelet Transform in Chemistry

3.4.1 Analytical Chemistry

3.4.1.1 Chromatography

Ten papers have been reported on adopting WT in chromatographic data processing so far. Pan *et al.* (Pan *et al.*, 1996) applied WT to correct baseline drift in HPLC system for analysis of rare earth elements in China. Baseline drift is always observed in the HPLC system with gradient elution and it affects the result on quantitative analysis such as peak area calculation. Unlike noise which is a high frequency signal, baseline drift is a low frequency signal. So, these workers proposed to transform the raw HPLC data with a Daubechies wavelet function to an optimum resolution level J. Then, zeros are assigned to the corresponding peak position in $C^{(J)}$. After the inverse WT treatment on the signal, the reconstructed data at the 0th level, $\overline{C}^{(0)}$, is obtained which represents the baseline for that HPLC study. Finally, a baseline free chromatogram can be obtained by subtracting the baseline from the raw data. Their result indicated that baseline removal via WT can improve the accuracy and reproducibility of the results in quantitative analysis of HPLC.

Shao *et al.* (Shao *et al.*, 1997a-d; Shao and Cai, 1998), from the University of Science and Technology in China, proposed to use WT in quantitative determination of constituting components in overlapping chromatographic peaks. For unresolved chromatograms, different factor analysis methods such as evolving factor analysis (Maeder and Zuberbuehler, 1986; Maeder, 1987), fixed moving windows evolving factor analysis (Schostack and Malinowski, 1993) and heuristic evolving latent projections (Kvalheim and Liang, 1992; Liang and Kvalheim, 1993a) have been reported for resolving overlapping chromatographic peaks. WT is proposed as a novel method to retrieve separate signals from overlapping chromatographic peaks. In Shao's study, Haar wavelet function was employed to decompose the chromatographic data into localized contributions and to resolve overlapping chromatographic peaks with two and three components respectively. In 1997, the First International Conference on Chemometrics in China (1st ICCC) was held in China. Shao *et al.* (Shao *et al.*, 1997a) presented a paper in the conference which was related to the application of WT in HPLC analysis. WT was proposed for denoising and baseline correction, determination of number of components, and resolving multi-component overlapping chromatogram. Recently, Shao and Cai (Shao and Cai, 1998) proposed to employ WT as a preprocessing step to resolve multicomponent chromatograms of rare earth elements by windows factor analysis (WFA). Again, Harr wavelet function was utilized to denoise the chormatograms with a high level of noise from an RP-HPLC-lactic acid system (Kuroda, 1993). They concluded that the combination of WT and WFA can separate the noise before WFA calculation and can achieve satisfactory results.

Mittermayr *et al.* (Mittermayr *et al.*, 1997b) reported a work on choosing WT as a tool to improve calibration and detection limit on data from gas chromatography coupled with a microwave induced plasma detector system. The Symmlet 8 wavelet function and universal soft thresholding were employed for denoising. The results obtained were compared with the Fourier and Savitzky-Golay filters. They concluded that special care should be taken for the choice of the filter parameters especially when highly different peak widths are present. In such a case, denoising with wavelet filter was better than the other two methods.

Shen *et al.* (1997a-b) developed a new method which was utilized WT to remove the influence of complex backgrounds such as spectral background and chromatographic drift in HPLC-DAD system. The existence of such complex backgrounds would introduce extra factors in the factor analysis calculation as proposed by Maeder and Zilan (Maeder and Zilan, 1988) and Gemperline (Gemperline, 1986). In the first paper (Shen *et al.*, 1997a), a two-dimension wavelet algorithm was adopted to compress, denoise and remove the chromatographic drift in the HPLC-DAD data. The performance of the new method were tested with the constrained background bilinearization method. In the second paper (Shen *et al.*, 1997b), these authors extracted the background from the zero concentration region of the HPLC-DAD data with the help of WT. Then, the background data was subtracted from the original data to generate a new data set and the treated signal was used for the chemical rank analysis. The authors found that the chemical rank can be estimated correctly in factor analysis calculation.

Application of WPT was also found in preprocessing of HPLC data. Collantes *et al.* (Collantes *et al.*, 1997) evaluated several computer-based classifiers as potential tools for pharmaceutical fingerprinting which was based on analysis of HPLC trace organic impurity patterns using WPT. Classifiers chosen in their work included artificial neural network (ANN), K-nearest neighbors (KNN) and soft independent modeling of class analogy (SIMCA). The HPLC data for each L-tryptophan sample was preprocessed by the Haar wavelet function in the WPT treatment. Then, the coefficients thus obtained in the wavelet domain were sorted in descending order. A small portion of sorted coefficients were fed as inputs into the ANN, KNN and SIMCA classifiers to match each sample with its manufacturer. The results were compared with their previous study by using the window preprocessor (Welsh *et al.*, 1996). They concluded that WPT preprocessing provides a fast and efficient way for encoding the chromatographic data into a highly reduced set of numerical inputs for the classification models.

3.4.1.2 Flow Injection Analysis

Bos and Hoogendam (Bos and Hoogendam, 1992) proposed to use WT to minimize the effect of noise and baseline drift in flow-injection analysis. Usually, peak overlapping in FIA is not a problem because it can be easily avoided by adjusting the sample rate. However, as the FIA system is operated near the detection limits, it is difficulty to locate peaks and find the right baseline correction method. It is because weak signals are embedded in the stochastic noise, which is a general problem in FIA. In Bos and Hoogendam's study, the

Morlet wavelet function was utilized to transform the FIA signal into a two-dimensional time-frequency form with both time and frequency information of the signal being retained. After WT treatment, the peak intensity was filtered from noise optimally. The maximum peak position can be searched from the coefficient in the wavelet domain of a well defined peak as obtained for a sample with a relatively high concentration. The position of this maximum on the horizontal axis of the transform conveys the positional information of the peak whereas its position on the vertical axis gives the noise filtering characteristic that can be obtained. Results of their work indicated that for a white noise and a favorable peak shape, a SNR of 2 can be tolerated at the 5% error level. This means that a significant reduction in the detection limit can be obtained in comparison with the conventional signal processing methods.

3.4.1.3 Infrared Spectroscopy

Infrared spectroscopy has been found widespread use in identification and characterization of chemicals. Application of WT in IR spectrometry was reported in fourteen papers. The first one that employed WT in IR spectroscopy was published by Stark *et al.* (Stark *et al.*, 1993). With the aid of WT, they could roughly separate the mineralogical information in the FT-IR absorbance spectrum from noise and the other signals such as absorbance from adsorbed water and organic components. These workers also developed an empirical affine minimax estimators method to estimate the mass fraction of a given mineral in a mixture using the wavelet coefficients.

Bos and Vrielink (Bos and Vrielink, 1994) published the second paper concerning about identification of mono- and di-substituted benzenes utilizing WT in IR spectrometry. Among the many wavelet analyzing functions available, the Daubechies series was chosen for their investigation because it offers different degrees of compactness of support and regularity (Daubechies, 1992). After WT treatment of the IR spectrum, the coefficients obtained were employed as inputs for an identification process that is based on the linear and non-linear neural network classifiers. The relevant information of an IR spectrum is contained in the position and the shape of the absorption peaks. The aim of their work is to show that whether the localization property of WT in both position and scale can be used to extract this information in a concentrated form and to obtain the salient features of an IR spectrum effectively. Using the concentrated information instead of the full spectra, classification of compounds is greatly improved in speed with which the classifiers can be derived. Moreover, it is expected that the quality of the classifiers will improve if they are derived from smaller data sets that still contain all the relevant information. From their study, Bos and Vrielink concluded that WT coupled with Daubechies wavelet functions is a feature extracting method that can successfully reduce IR spectral data by more than 20-fold with a significant improvement in the classification process.

Recently, a new standardization method that based on transferring near-infrared (NIR) spectra in the wavelet domain was proposed by Walczak *et al.* (Walczak *et al.*, 1997). They tried to relate the WT coefficients of the NIR spectra that were obtained from two instruments utilizing an univariate linear model. The

reason to perform standardization on two NIR spectrometers is due to variations between instrumental responses of different spectrometers. The calibration model developed on the first NIR (master) spectrometer may lead to erroneous predictions if it is applied to NIR spectra collected on the second NIR (slave) spectrometer. So, WT was applied on the NIR spectra obtained from both the master and slave spectrometers. Then, an univariate linear model was set up to determine the standardization parameters between the two sets of NIR spectra. Once these parameters were found, NIR spectra in the wavelet domain can be transferred between two different spectrometers for subsequent data analysis. Alsberg *et al.* (Alsberg *et al.*, 1998) proposed a wavelet regression model for variables selection in IR spectrum. They applied various types of variable selection methods to the IR spectrum in the wavelet domain. They found that truncation of weight vectors in partial least squares regression was the most effective method for selecting variables. Ehrentreich *et al.* (Ehrentreich *et al.*, 1998) published a paper on the application of WT for peak recognition in IR spectroscopy. They found that some of the wavelet bases lead to a very good compromise between SNR enhancement and preservation of the real data structures. The wavelet method can help chemist to find the local maxima of the curve corresponding to real data structures.

Different methods have been developed for reducing IR spectral data for library search. These include the simplest approach by lowering the resolution in the spectral domain (Wang and Isenhour, 1987), binary encoding (Heite *et al.*, 1978), factor analysis (Hangac *et al.*, 1982), Fourier transform method (Owens and Isenhour, 1983) and others. All these methods may be used to retain either more spectral information resulting in increasing the library size or less

information leading to poorer searching performance. Qian and Sun (Qian and Sun, 1996) performed an experiment to compress IR spectra of twenty one typical earth resource with WT. A compression ratio between 1:4 to 1:5 and 1:8 to 1:10 were achieved if the compression results were not coded and coded with Huffman coding respectively. Wang *et al.* (Wang *et al.*, 1996) carried out a similar study to compress the IR spectrum of polyethylene film with the Haar wavelet function. A compression ratio of 1:5 was reported. Our research group also published a paper on compressing IR spectral data utilizing WT (Chau *et al.*, 1997). The IR spectrum was transformed into the wavelet domain by using the Daubechies wavelet function. Then, the optimal bit allocation quantization and Huffman coding techniques were applied to the wavelet coefficients obtained to reduce the storage space for each IR spectrum. It was found that a combination of an average bit rate of 7 or 8 and the Daubechies D_{16} wavelet function is good and efficient in compressing IR spectral data significantly with small errors. The results obtained were also compared with those from the FFT and wavelet-based thresholding schemes. It was found that the proposed method outperforms the latter two. Liu *et al.* (Liu *et al.*, 1997a-b) applied wavelet neutral network to compress IR spectrum of polystyrene film. Instead of employing the traditional Sigmoid function, the Morlet wavelet function was employed in a single layer neutral network. Wavenumber and transmittance of the IR spectrum were chosen as the inputs and outputs of the network. A compression ratio of 1:40 was reported in their study. For the same research group in China, Chen *et al.* (Chen *et al.*, 1997) applied WT to improve the peak shape of the characteristic absorption and the spectral resolution in the IR spectra of poly(acrylamide-sodium acrylate)hydrogel

copolymers. They found that WT can resolve the characteristic peaks in the overlapping region and is useful to characterize the composition of copolymer qualitatively and quantitatively.

Application of WPT in IR spectrometry could be found in the works by Wickerhauser (Wickerhauser, 1994b), Walczak *et al.* (Walczak *et al.*, 1996) and Alsberg *et al.* (Alsberg *et al.*, 1997b). Wickerhauser (1994b) published a paper to describe an approximate principal component analysis (PCA) algorithm based on wavelet packet transform. Such algorithm could lower the complexity for finding the principal components or Karhunen-Loève basis eigenvectors from $O(d^3)$ to $O(d \log d)$ with the use of Harr-Walsh wavelet packets. The author applied this approximate PCA algorithm to invert a map that can produce the IR spectrum of the atmosphere from the concentrations of absorbing gases. Walczak's research group chose WPT as a tool for improving pattern recognition based on NIR spectra. Haar wavelet and Coifman's best-basis algorithm (Coifman and Wickerhauser, 1992) were used to construct the full WPT decomposition framework for each NIR spectrum. The preprocessed NIR spectra with WPT treatment could improve the linear discriminant analysis classification when compared to those using either the standard normal variate method or no pre-treatment at all. They also concluded that selecting features from a local discriminant basis instead of those from a full decomposition did not improve the performance of classification. Alsberg's research group did a comparative study in applying wavelet to denoise IR spectra. Six different methods including SURE, VISU, HYBRID, MINMAX, MAD and WPT were applied to pure IR spectra with

different levels of homo- and heteroscedastic noise added. The results from wavelet denoising were compared with those from the standard Fourier and the moving mean filtering methods. They discovered that at very low SNR, the performance of the last two methods were comparable to that of the wavelet methods. Yet, at higher SNR levels, the wavelet denoising methods were better, especially, the HYBRID and VISU method. These authors also found that the wavelet methods were better in restoring the visual quality of the denoised infrared spectra.

3.4.1.4 Mass Spectrometry

In mass spectrometric studies, WT were mainly applied in two areas that include instrumentation design and secondary ion mass spectrometry (SIMS) and five publications have been found. A patent (U.S. Patent Number: 5,436,447) was obtained by Shew (Shew, 1995) in United States. The author invented a new procedure to determine the relative ion abundances in ion cyclotron resonance mass spectrometry by utilizing WT to isolate the intensity of a particular ion frequency as a function of position or time within the transient ion cyclotron resonance signal. The WT intensity corresponding to the frequency of each ion species as a function of time can be fitted by an exponential decay curve. When extrapolating these curves back in time to the end of the excitation phase, accurate values of the relative abundances of different ions within a sample can be determined. An ion cyclotron resonance mass spectrometer with a Haar wavelet analysis module was set up by Shew. The result of the work indicated that WT can

provide high efficiency isolation of individual frequencies in the received signal corresponding to individual species.

Four papers were published by a group of researchers in the Vienna University of Technology for wavelet denoising of SIMS images. SIMS is a type of surface technique for (1) trace analysis, (2) determination of elemental composition, (3) the identity and concentrations of adsorbed species and (4) elemental composition as a function of depth (Strobel and Heineman,1989). This surface technique is capable of measuring the distribution of elements with a lateral resolution of 0.1 μm in the scanning mode. The two-dimensional element distributions generated by scanning SIMS are characterized by Poisson statistics with small integer values. Poor signal statistics is one of the major problems encountered in SIMS measurement especially in trace analysis. As a result, quantification and image processing method such as classification and edge detection are difficult to apply. Nikolov *et al.* (Nikolov *et al.*, 1996) reported an application of a wavelet shrinkage algorithm for denoising SIMS images following the Poisson distribution. Two-dimensional pyramidal wavelet decomposition was introduced in their study. Three different types of wavelet functions including Haar, Daubechies and Coiflet wavelets were tested for denoising SIMS images. The results were compared with those obtained from the optimal mean-square-error Wiener filter. They concluded that wavelet gave comparable mean-square-error to the Wiener filter by improved the SNR. Hutter *et al.* (Hutter *et al.*, 1996) published another paper on the application of wavelet shrinkage algorithm for denoising of SIMS images. These authors found that the

wavelet method could suppress the noise of the images without significant loss of lateral resolution. They also applied neural network for classification of chemical phases on the SIMS images in their work. Similar works were also performed by Wolkenstein *et al.* (Wolkenstein *et al.*, 1997a-b). In the first paper (Wolkenstein *et al.*, 1997a), the Haar, Daubechies and Coiflet wavelets coupled with universal thresholding were applied to denoise the SIMS images of a nickel base soldering alloy. The result was compared with the widely used Savitzky-Golay filter. They found that the wavelet method can suppress noise significantly while sharp edges and corners in the signal remain sharp after processing. In the second paper (Wolkenstein *et al.*, 1997b), Coiflet wavelet function with three vanishing moments was chosen as a tool for SIMS image denoising and this approach was found to improve the Kohonen neural network classification.

3.4.1.5 Nuclear Magnetic Resonance Spectroscopy

Nuclear magnetic resonance spectroscopy is one of the most powerful nondestructive techniques available today for probing structure of matter. Three publications were related to the application of WT in NMR spectroscopy. In 1989, Guillemain *et al.* (Guillemain *et al.*, 1992) were the first research group to propose an application of WT in NMR spectroscopy. In their work, they aimed at investigating how an appropriate use of WT could lead to an excellent estimation of the frequency of spectral lines in a signal and provided direct information on time-domain features of these lines in NMR spectra. These authors reported seven applications of WT in NMR spectroscopy that included estimation of frequency

and amplitude modulation laws in both simple and general cases. spectral line subtraction and re-synthesis, ridge extraction, sum of two sine waves and three exponentially decreasing sine waves.

Recently, Neue (Neue, 1996) published another paper on an application of WT in dynamic NMR spectroscopy which could simplify the analysis of free induction decay (FID) signal. Dynamic NMR spectroscopy is a technique used to measure rate parameters for a molecule (Kemp 1986). The measured resonance frequencies represent the spatial coordinates of spins. Any motion such as bond rotation and other molecular gymnastics may change these frequencies as a function of time. The localization property of WT gives a better picture of the nature of the underlying dynamical process in both the frequency and time domains. The third-order Battle-Lemarié wavelets was employed for crystal rotation and the first-order kinetics with NMR spectroscopy in their study. They concluded that WT will become a routine method in NMR spectroscopy for data analysis. Similar idea can also be found in a book by Hoch and Stern (Hoch and Stern, 1996). They introduced WT as a new data processing technique for smoothing NMR data.

3.4.1.6 Ultraviolet-Visible Spectroscopy

Ultraviolet-visible (UV-VIS) spectroscopy has been used extensively in physical and biological sciences for characterization, identification and quantification of substances (Perkampus, 1992). Five publications have been

found in this area. Two papers were published by Chau and his coworkers on compression and denoising UV-VIS spectra with WT. The first one (Chau *et al.*, 1996) was related to an application of the FWT procedure to compress UV-VIS spectra. Different Daubechies wavelets, threshold values ε and maximum resolution levels J were utilized to study the performance of the proposed method. Translation-rotation transformation (TRT) method was introduced in this study to solve the side-lobe problem in spectrum reconstruction. The problem occurs when there is a sudden data change in the spectrum during periodical extension of the spectral data at both end on WT preprocessing. It was found that the combination of $\varepsilon = 0.003$, $J = 6$ and the Daubechies D_{16} wavelet was good and efficient in compressing UV-VIS spectra with significant reduction in storage space. A compression rate over 96% was achieved. The second paper by Gao *et al.* (Gao *et al.*, 1996) was related to the application of FWT with the use of modulus maxima for denoising UV-VIS spectra. The results obtained were compared with those of the moving window averaging and FFT techniques. In general, the proposed WT method performs better than latter two in spectral denoising.

Both Haar and Daubechies wavelet functions are widely adopted in chemical studies for WT treatment. Since 1996, B-spline wavelet has been proposed as a new function for WT calculation in analytical chemistry. Lu and Mo (Lu and Mo, 1996) published a paper on an application of B-spline wavelet multi-resolution analysis to denoise UV-VIS spectra of Gd^{3+}-Sr^{2+}-chloprophosphoanazo III complex. They studied the effects of the *m*th-order B-spline wavelets, the cutoff frequency value l between the useful signal and noise as well as different

SNR values. They found that the third-order B-spline wavelet function was the most effective one for the purpose. Beside, significant differences were observed for different B-spline wavelets.

Recently, a novel application of WT was developed by Liu *et al.* (Liu *et al.*, 1997c) in China. A wavelet neural network was set up and applied to recognize the UV-VIS spectra of tyrosine, 3,4-dihydroxyphenylalanine and trytophane. The Morlet wavelet and line search conjugate gradient optimization method were used in their neural network. The results indicated that the wavelet neural network had a very good recognition power to differentiate minor differences between similar UV-VIS spectra. They also published another paper on using the wavelet neural network for simultaneous determination of molybdenum and tungsten (Liu *et al.* 1997d). The results were superior to those obtained by back propagation neural network algorithm. Xie *et al.* (Xie *et al.*, 1997) presented a paper on ICCC and proposed to apply wavelet neural network for IR spectra compression, UV spectra recognition and multi-component analysis.

3.4.1.7 Voltammetry

Wavelet transform has been applied successfully in voltammetric data analysis since 1995 in China. Fourteen publications with the use of WT in voltammetry have been found in China. Voltammetry is a popular technique in all fields of chemistry used to study redox reaction. Yan and Mo (Yan and Mo, 1995) were the first research group to introduce WT in processing signal from

voltammetry. They developed a real time continuous wavelet filter for voltammetric signal processing. This filter can improve the SNR of the raw signal and reduce the standard derivation of the processing method. A spline function of order 3 was chosen as an example of real time wavelet filter in their investigation. They applied the new method successfully for real time signal denoising in the study of staircase voltammetry in $ZnSO_4$-K_2SO_4 and $K_3Fe(C_2O_4)_3$-$K_2C_2O_4$-$H_2C_2O_4$ systems. The same research group also introduced spline wavelet in processing voltammetric signal (Zou and Mo, 1996, 1997a). This time, they compared the effect of the order of the spline wavelet function, cutoff frequency *l*, SNR and the number of sampling points in denoising of voltammetric signal. They found that the third-order spline wavelet function with $l = 3$ or 4 was the best combination to denoise staircase voltammmetric signal for Mo(VI)-W(VI)-HAPP-$KClO_3$ and Ti(IV)-$H_2C_2O_4$ systems with the SNR down to 0.1. Lu and Mo (Lu and Mo, 1997) also reported a similar work by adopting multi-frequency channel filtering technique employing spline wavelet for denoising voltammetric signal of Cd(II)-Succ-Ox complex system. In 1998, Lu *et al.* (Lu *et al.*, 1998) published a paper on applying spline wavelet to process data from deconvolution voltammetry of $LiClO_4$-CH_3CN system. They found that WT can provide the opportunity to optimise the deconvolution voltammetric signal for further convolution calculation. Zou and Mo (Zou and Mo, 1997b-c) proposed another new method to process staircase voltammetric signal with spline wavelet multiple filtering technique. In order to extract more useful information from the signal, the wavelet coefficients $\boldsymbol{D}^{(j)}$ at each resolution level were also employed for further signal filtering employing the spline wavelet function. The extracted information was

utilized to compensate the scale coefficients $C^{(J)}$ at each resolution level. They found that this method could be applied successfully to a voltammetric signal with a SNR of 0.1 with the parameters reported in their previous paper (Zou and Mo, 1996, 1997a). The relative errors of peak current and peak potential were less 3% and 10% respectively.

Bao *et al.* (Bao *et al.*, 1997a-b; 1998) also carried out a similar study in voltammetry with WT. They performed a comparison between spline wavelet smoothing and Fourier smoothing in processing differential pulse stripping voltammetric (DPSV) signal of Pb^{2+} ion in KCl solution. They concluded that WT had certain advantages such as simple procedure, short operation time, little memory required, high precision and good reproducibility in processing analytical chemistry signals. In a recent paper, Bao *et al.* (Bao *et al.*, 1998) discussed the advantages and disadvantages of spline wavelet and Fourier transformation methods in processing DPSV signal. They designed a combined algorithm of the above methods to process DPSV signal with SNR value of 0.1. The combined algorithm can minimize the shift of peak current position after processing with spline wavelet. But, in a recent study by Bao and Mo (Bao and Mo, 1998), they concluded that WT cannot be applied to process voltammetric signal with low SNR because there is a shift of voltammetric peak after wavelet denoising.

Chen *et al.* (Chen *et al.*, 1996) utilized the Gaussian difference wavelet function for the treatment of differential pulse voltammetric (DPV) data of Cu^{2+} in Cu-KNO_3 system. The function is defined as:

$$\Psi(\lambda) = \exp\left(-\frac{\lambda^2}{2}\right) - \frac{1}{2}\exp\left(-\frac{\lambda^2}{8}\right). \quad (3.18)$$

In DPV quantitative analysis, it is very difficult to measure the peak height in a sample with low concentration. As a result, it affects the linear detection range of the DPV system. So, these workers employed the Gaussian difference wavelet function to transform the DPV signal obtained from a highly concentration Cu^{2+} ionic system and to determine the scale parameter *a*. Then, the DPV signals at other concentrations were transformed with a predetermined scale parameter. A new calibration curve was plotted by using the results obtained from with WT treatment. They found that a new linear relationship could be obtained in this approach, and hence, extended the detection range to sample with lower concentration.

Application of WPT in voltammetry was found in a recent work by Fang and Chen (Fang and Chen, 1997). They investigated the feature of WPT for white noise which is caused by random and irregular process and developed an adaptive wavelet filter for the selection of optimum critical level of related frequency band automatically in voltammetric study. Their outcomes showed that the adaptive wavelet filter could be applied to a system with interference originating from existing power supply which is useful for the study of fast electron transfer process.

Yu *et al.* (Yu *et al.*, 1997) presented a paper on ICCC which was related to chronopotentiometric stripping voltammetry with wavelet analysis. They authors

reported that the sensitivity of their proposed method was four times higher than that without wavelet analysis. Zheng and Mo (Zheng and Mo, 1997) presented a paper with an application of B-spline wavelet and RLT filtration in staircase voltammetry. The result indicated that B-spline wavelet method could filter the random noise down to SNR of 0.8. In 1998, Zheng *et al.* (Zheng *et al.*, 1998) published another paper with an application of both spline wavelet and Riemann-Liouville transform filtration method to filter random noise and capactive currents from a faradaic signal. This method have been tested with voltammetric data of Cd^{2+} in Cd-KNO_3 system. The authors found that voltammetic signal with SNR of 0.8 can be filtered with the proposed method. Besides, the errors of the peak current and peak potential were less than 5.0% and 1.0% respectively.

Zhong *et al.* (Zhong *et al.*, 1998) applied both Harr and Daubechies D_6 wavelet functions successfully to recover useful information from different kinds of oscillographic chronopotentiometric signal. Oscillographic analysis is a new type of electrochemical analysis method which is based on the observation of signal change from a cathode ray oscilloscope. Effects of the transform frequency, white noise and kinds of wavelet function and the oscillographic signal on the transform results were discussed in their work.

3.4.1.8 Others

Wolkenstein *et al.* (1997c) published a paper on the application of WT on denoising the images from electron probe micro analysis with an energy

dispersive X-ray analyser. Various wavelet bases including Haar, Coiflet, Symmlet and Daubechies were employed in their investigation. The results from wavelet methods were comparable to the Gaussian weighted moving average filter, Savitzky-Golay filter, Fourier filter and Wiener filter in most cases. Ratton *et al.* (Ratton *et al.*, 1997) performed a comparative study of signal processing techniques on a sensor for simultaneous chemical detection. Four test gases of methanol, ethanol, formaldehyde and acetone were examined in their studies. The signal from a Pd-dosed SnO_2 film microsensor was analyzed with the Gram-Schmidt approach, fast Fourier transform and Haar wavelet transform. They aimed to compress and extract the useful information with these three methods. Their result indicated that the highest overall performance was obtained with the Haar wavelet transform. It also compressed information efficiently and removed the noise and drift effects of the signal successfully in their studied range of the sensor.

Suzuki *et al.* (Suzuki *et al.*, 1996) published a paper on the application of WT to process acoustic emission (AE) signals. They found that WT was useful in the recognition of AE signal features through two-dimensional contour maps and 3-dimensional projections of wavelet coefficient. The Gabor wavelet was chosen as the mother wavelet in their work and applied to analyze the AE signal for glass fiber reinforced composites. Mao *et al.* (Mao *et al.*, 1997) presented a paper on ICCC which was related to the application of WT in photoacoustic (PA) spectroscopy. PA spectroscopy is a newly developed analytical method for characterization of spectrum of solid samples with high reflection and high

diffusion. WT provides a convenient and objective way to decompose real PA signal from noise and disturbance arising from baseline. On the same conference, Bao *et al.* (Bao *et al.*, 1997c) presented a paper with the use of spline wavelet and Fourier transforms in analytical chemistry. The authors found that the combined method is a powerful tool to process sudden- and gradual-change signal and signal with SNR of 0.1.

A novel application of WT in potentiometric titration was developed by Wang *et al.* (Wang *et al.*, 1997b) in China. They made use of the edge detection property of WT to determine the end-point in potentiometric titration. There are two major methods for end-point determination in potentiometric titration. The first one is achieved through the direct graphical interpretation of the titration curve, such as Behrend's method, Brötter's method and Tubbs' method (Ren and Ren-Kurc, 1986). The second one is through mathematical interpretation of co-ordinates of the recorded points such as Gran's method (Gran, 1952; Chau 1990; Chau *et al.*, 1990) and derivative method (Strobel and Heineman, 1989). WT can be regarded as a new mathematical interpretation method for end-point determination. In Wang's work, they proposed to use the maximum absolute value of the first-order differential function to determine the end-point in potentiometric titration. The wavelet function chosen is a continuously differential spline function of order 2 which is a third-order spline function with unity integral value of the first order differential function. After WT treatment on the titration curve, the maximum absolute value among the wavelet coefficients $\boldsymbol{D}^{(j)}$ represents the end-point of that titration. For a polyprotic acid system, the local maximum

absolute value represents the corresponding end-point during titration. They concluded that the accuracy of their method can reach the experimental limit automatically.

3.4.2 Quantum Chemistry and Chemical Physics

Fourteen papers have been published to report works on applying WT to different quantum chemistry and chemical physics calculations. Two of them were published by Permann and Hamilton (Permann and Hamilton, 1992, 1994). The first one (Permann and Hamilton, 1992) was related to wavelet analysis of time series for the Duffing oscillator. The Duffing oscillator was developed as a simple model to account for the hardening spring effect in many mechanical systems. It is one of the most common examples of a nonlinear oscillator. These authors intended to show that WT is a robust tool that could be utilized to obtain qualitative information for highly non-stationary time series. It may be used to detect a small-amplitude harmonic forcing term when the dynamics is chaotic. The second paper (Permann and Hamilton, 1994) was related to wavelet analysis of the time series of the position of a Morse oscillator which is weakly forced and weakly damped. The Morse oscillator is commonly used for studying diatomic molecule. In their investigation, they wanted to show that wavelet analysis can be applied to determine multiple forcing frequencies quantitatively owing to localization property of the orthogonal wavelet basis functions in both frequency and time domain.

Fischer and Defranceschi (Fischer and Defranceschi, 1993) developed a new method to visualize both the position and momentum characteristics of atomic orbitals on the same drawing with the use of WT. They applied the proposed method to hydrogen atom with Gaussian atomic orbitals in the position space, momentum space and position-momentum space. In order to simplify the computation, wave function in one-dimensional space was considered in their work. The first derivative of the Gaussian function was chosen as the mother wavelet function. WT provides an alternative way to view the atomic orbitals when compared with the Fourier transform. The new method can provide more information on the analysis of the oscillating character of the wave function. They also proposed to employ WT to solve Hartree-Fock equations of hydrogen atom in a later study (Fischer and Defranceschi, 1994a). Fischer and Defranceschi (Fischer and Defranceschi, 1994b) published another paper to introduce WT as a new mathematical tool for quantum chemistry. They concluded that WT can play a role of a magnifying glass to scan wave functions at different scales over the whole position space. Besides, it can be adapted to improve the Gaussian approximation. The authors (Fischer and Defranceschi, 1994c) also published a paper that is concerned about representing the atomic Hartree-Fock equations in a wavelet basis by means of the Beylkin, Coifman, Rokhlin (BCR) algorithm. These workers tested the method using four wave functions of the Slater type which comes from the exact solution of the Hartree-Fock equation and three Gaussian approximations constructed with one (STO-1G), two (STO-2G) and three (STO-3G) Gaussian functions. They found that the total energies computed with these new algorithms were relatively close to the analytical values. Recently, the authors (Fischer and Defranceschi, 1998) published another paper on

decomposing the Schrödinger operator in a non-standard form with the BCR algorithm. Their method has been used in an iterative method to solve the corresponding eigenvalue problem for hydrogen-like atoms. Saji and Konno (Saji and Konno, 1998) published a paper to discuss the usefulness of WT for avoiding difficulties associated with the Wigner transform. They also demonstrated the suitability of WT using the hyperbolic analyzing wavelet in analyzing the amplitude modulation of the localized two-soliton solution and the localized three-soliton solution for the nonlinear Schrödinger equation.

Cho *et al.* (Cho *et al.*, 1993) developed a three-dimension wavelet analysis for electronic structure calculations. The wavelet analysis provided a systematically improvable and tractable description of electronic wave functions and overcame limitations of conventional basis expansions. These authors demonstrated their new method for *ab initio* electronic structure calculations by computing the $1s$ states for all the naturally occurring nuclei from hydrogen to uranium on the periodic table and the interaction energies of the hydrogen molecule ion. The Mexican-hat wavelet function was employed in their study. In a later study, Wei and Chou (Wei and Chou, 1996) proposed to employ Daubechies wavelets in self-consistent electronic structure calculations of hydrogen and oxygen molecules. They found that the number of basis functions needed for the computation was much smaller than the conventional method and efficient of computation was improved. Calais (Calais, 1996) proposed two possible applications of wavelets in quantum chemistry. The author suggests to employ wavelets as an instrument for simultaneous visualization of position and momentum densities and as alternative basis functions in quantum computation.

Brewster *et al.* (Brewster *et al.*, 1997) proposed a one dimensional prototype which was based on wavelet theories for electronic structure calculation. The results from their work can lead to the development of multi-resolution algorithms for the actual three dimensional problems of electronic structure calculations.

Li *et al.* (Li *et al.*, 1993), von Kitzing and Schmitt (von Kitzing and Schmitt, 1995), Askar *et al.* (Askar *et al.*, 1996) and Tymczak and Wang (Tymczak and Wang, 1997) applied WT in molecular dynamics and molecular mechanics analysis. Molecular dynamics simulation is an important theoretical in studying equilibrium and non-equilibrium processes in condense matter systems. Li *et al.* (Li *et al.*, 1993) used WT to characterize data generated from condensed phase molecular dynamics simulation. WT was performed on the particle velocities using a multi-resolution decomposition procedure based on orthogonal wavelets of compact support. The results from the wavelet treatment provided a detailed view of the time-dependent energy transfer between different scales of motion. On another paper, von Kitzing and Schmitt (von Kitzing and Schmitt, 1995) proposed to employ WT as a tool for molecular mechanics calculations on biopolymers such as DNA or RNA molecules. In such calculation, large amount of computer resources such as CPU time and memory are required. For example, a calculation on a DNA molecule involves a system with 1,000 to 100,000 atoms and requires a huge computer resources for the energy minimization calculation. Wavelets were chosen as an approximation function to reduce the computer resources during calculation via data compression. Askar *et al.* (Askar *et al.*, 1996) presented a paper on wavelet analysis of two time series that generated by numerical integrations in molecular dynamics and suggested

WT as a possible tool for constructing the solutions. This approach was applied to the study of the deterministic dynamics of a three-dimension polymer model and the stochastic dynamics of a one-dimension chain subjected to random forces within the Langevin equation formalism. These workers found that the WT was a useful tool for analyzing the numerical data representing the dynamics of chain molecules as well as for reconstructing them systematically from a subset of data. Tymczak and Wang (Tymczak and Wang, 1997) reported on the use of Daubechies wavelet bases in the quantum dynamics (Car-Parrinello) algorithm. They found that WT can preserve the advanced features of the Car-Parrinello method and the transform between the real and wavelet spaces can be efficiently carried out using DWT.

Modisette *et al.* (Modisette *et al.*, 1996) reported the use of Daubechies' compact support wavelets in solving quantum mechanical eigenvalue problems. Daubechies wavelet was chosen as general and flexible bases for the variational solution of the time-dependent Schrödinger equation. They identified three major advantages in using WT in quantum chemistry calculation. Firstly, the simultaneous localization of wavelets in both coordinates and momenta allows one to customize the basis set to provide resolution locally. Secondly, the orthonormal wavelets can be used to eliminate the need to solve a generalized eigenvalue problem. Thirdly, the basis sets are easily extensible to multiple dimensions by taking the tensor products of one-dimensional wavelets. These authors also suggested several applications of WT in quantum chemistry such as density functional theory, large-amplitude variational calculation of molecular vibrations, and the solution of time-dependent quantum problems.

3.4.3 Others

Permann and Teitelbaum (Permann and Teitelbaum, 1993) developed a wavelet fast Fourier transform (WFFT) to denoise the millivolt signal for a transient oscillating chemical reaction. They applied this technique to the kinetic study of a common refrigerant CCl_3F. The millivolt signal was collected by employing a moderately fast data acquisition system with a sample rate of 40 million samples per second. As a result, the output signal was very noisy. The Daubechies wavelet function was applied to the signal and the noise components were resolved by FFT on selected parts of the wavelet coefficient coupled with the threshold method. After data processing, the signal was reconstructed to the original domain.

Wickerhauser (Wickerhauser, 1995) also published another paper about time localization techniques with WT. The author examined two problems that related to orthonormal compactly supported wavelet expansions: (1) Given a wavelet coefficient with its nominal scale and position indices, find the precise location of the transient signal feature which produced it; and (2) Given two collections of wavelet coefficients, determine whether they arise from a periodic signal and its translate, and, if so, find the translation which maps one into the other. The author proposed two alternative algorithms to solve these problems which rely on the wavelet coefficients themselves without performing inverse WT. The first problem was solved by locating the center of energy of a wavelet while the second one by a fast algorithm introduced by Beylkin (Beylkin, 1992).

Mallet *et al.* (Mallet *et al.*, 1996) reported their recent development in discriminant analysis on high dimensional spectral data. DWT was classified as a feature extraction technique in discriminant analysis. Feature extraction is a linear or non-linear transformation by mapping a data set with high dimension data to that with a lower dimension. The work from Bos and Vrielink (Bos and Vrielink, 1994) is a good example to demonstrate the advantage in adopting WT for discriminant analysis.

Dohan and Whitfield (Dohan and Whitfield, 1997; Whitfield and Dohan, 1997) have applied WT to identify and characterize of transient water quality from two streams near Vancouver, Canada. Transients are important indicators of immediate impact and are a significant feature of some environmental problems (Whitfield and Wade, 1992). They found that WT can help them to process the water quality data by separating the periodic components from the transient components and separating the seasonal components from the diurnal parts of the periodic components. Besides, the properties of each transient, that is location, duration and magnitude, can be identified from different resolution level. In their work, the orthogonal Daubechies S8 wavelet function was employed.

3.5 Conclusions

Applications of wavelet transform and wavelet packet transform in various fields of chemistry from January 1989 to August 1998 were reviewed. Major research works

that related to the application of WT in China are also included in this chapter. WT is mainly used in data compression and noise removal in chemical studies owing to its efficiency, large number of basis functions available and high speed in data treatment. In most cases, the performance of WT treatment is much better than the Fourier transform. Up to now, one patent, two tutorial papers and more than 90 papers in application of WT to chemistry were published. Signal processing of chemical data via wavelet transform is still under development and the mathematical technique is expected to be one of the most popular method on the future.

CHAPTER 4

Application of Wavelet Transform in Infrared Spectrometry Spectral Compression and Library Search

4.1 Introduction

The development of information technology in chemistry is very important because the type and volume of chemical data increases dramatically in recent years. In modern laboratories, most chemical instruments are driven with a computer for control of the device, data acquisition, signal processing, interpretation and reporting (Ratzlaff and Ratzlaff, 1992). Recently, there is a growing trend in combining different chemical devices together to give hyphenated instruments such as high performance liquid chromatography coupled with diode array detection system (HPLC-DAD) and gas chromatography-mass spectrometry (GC-MS) for multi-dimensional studies. This approach greatly enhances information acquisition and even allows experimental work not possible to be achieved before. However, the analytical chemists always face a dilemma on the use of hyphenated instruments (Karjalainen and Karjalainen, 1996). Hyphenated instruments can produce a huge amount of raw data that occupy a lot of hard disk space in a very short time. But, such data cannot be removed from the hard disk of the instrument until the required analysis is performed. In some cases, the acquired data is needed to achieve for further usage. Processing time is another problem that faced by analytical chemists. In order to extract the most useful information from such huge amount of raw data, data preprocessing steps should be performed. Nowadays, although the processing speed of computer is very fast, it is impossible to process pile of data within a short time. One possible way to reduce the storage space and processing time is through signal compression. In chemical analysis, signal compression is very important especially in setting up digitized spectral library (Warr,

1993a-b) to diminish the size of the original database and to reduce the time for spectral searching.

Rapid development in computer technology leads to a lot of electronic spectral libraries and database available in the market (Warr, 1991). Most of them cover compounds fewer than 100,000 out of 10 million known chemical compounds in different formats such as digitized IR, NMR and mass spectra. In order to identify the spectrum of an unknown compound from a reference library, spectral library search techniques are required. The output of a search usually comprises spectra that are most similar to that of the compound under study, together with an indication of the degree of similarity in each case (Warr, 1993a). So, a fast and highly accurate library search algorithm is desirable. Common algorithms of this kind adopted in chemistry include peak position searching method, direct matching method, spectral simulation method as well as artificial intelligence and pattern recognition techniques such as neural network (Klawun and Wilkins, 1995), fuzzy theory (Blaffert, 1984; Otto and Bandemer, 1986) and expert system (Warr, 1993a; Zhu and Stillman, 1996).

To carry out any search, the spectral library must be constructed first from a set of reference spectra and accompanied by additional information such as structure, name, connection data, and molecular mass of individual compound (Hobert, 1995). In order to reduce the storage space of spectral data, different compression techniques have been developed. Data compression methods that are commonly used in chemistry include binary encoding (Rann, 1972; Adam and Black, 1986; Scott, 1988), spline (Alsberg, 1993; Alsberg and Kvalheim, 1993; Alsberg *et al.*, 1994), Fourier transform (Crawford and Larsen, 1977; Lam *et al.*, 1981; Owen and Isenhour, 1983; Chau and Tam, 1994)

and factor analysis (Wang and Isenhour, 1987; Malinowski, 1991; Hangac *et al.*, 1982). Basic theory of the above mentioned methods can be found in Chapter 2. Currently, Fourier transform (FT) and the inverse FT are the major tools adopted to process experimental data in chemical studies. As mentioned in the previous chapter, wavelet transform (WT) has been introduced in chemical studies for data compression and denoising since 1989 and applied successfully in various fields of analytical chemistry, quantum chemistry and chemical physics.

Different methods have been reported in the literature to facilitate searching of IR spectral database. Some of them are based on peak width and intensity (Penski *et al.*, 1974; Rasmussen and Isenhour, 1979), principal component analysis (Wang and Isenhour, 1987) and Fourier transform and interferogram (Small *et al.*, 1979; de Haseth and Azarraga, 1981; Azarraga *et al.*, 1981). In this work, new data compression procedures have been developed to manipulate infrared (IR) spectra by utilizing the fast wavelet transform (FWT) and wavelet packet transform (WPT) techniques. In this approach, an IR spectrum is converted into the wavelet domain through FWT and WPT treatment. WT was chosen in this work because its performance dominates over FT in many aspects. For example, more number of basis functions are available for WT computation and it has higher efficiency and speed in data treatment when it is compared with FT. In order to minimize the storage space of the spectrum, an absolute cutoff method was utilized to select the coefficients to be archived. Besides, the Shannon-Weaver entropy calculation was also employed to choose the best basis in WPT computation. After data compress process, individual compressed IR spectra were employed to set up a small scale spectral library for testing the efficiency of our

approach in improving the spectral library search. The results thus obtained are compared with those from the fast Fourier transform (FFT) treatment.

4.2 Method of Investigation

In this work, two WT techniques, namely FWT and WPT, were employed to process a selected set of IR spectra. After the FWT or WPT treatment, the coefficients obtained were compressed by using appropriate methods to reduce the storage size. Selected coefficients were utilized to construct a wavelet compressed spectral library for future use. The library consists of compressed IR data in the wavelet domain, compound names of the compressed spectra, and selected parameters such as the types of wavelet function used, the original data length and the assigned resolution level J for spectral reconstruction. The scale coefficients of each compound obtained at resolution level J in the spectral library were employed for the preliminary library search which was performed by comparing the scale coefficients of the unknown spectrum with the reference spectra at a particular resolution level via the direct matching method. A small group of IR spectra were thus extracted from the library. Then, a detail searching was performed by comparing the unknown spectrum and reference spectra at resolution level 0. This search may not be required for IR spectra with very different spectral structures. However, compounds with similar molecular structures are not easy to distinguish using the scale coefficients obtained at higher J levels. Therefore, a second search in the original domain was also carried out based on spectral information obtained at resolution level 0. In order to solve the side-lobe problem in data reconstruction, the

translation-rotation transformation (TRT) method (Chau and Tam, 1994; Chau *et al.*, 1996) was also applied. Besides, the coefficient position retaining (CPR) method has also been developed to handle spectrum with data length in odd number. Details of the above mentioned methods will be discussed in the following sub-sections.

4.2.1 Fast Wavelet Transform

Wavelet transform is a tool that can be utilized to convert data, functions or operators into different frequency components. Then each component is studied with a resolution matched to its scale (Daubechies, 1992). There are different types of wavelet functions available in the literature such as Haar wavelet, Meyer's wavelet, Mexican hat wavelet, spline wavelet, and Daubechies wavelets (Palavajjhala *et al.*, 1994). The later one is now the industry standard for signal compression especially in chemical studies (Borde, 1995). The fast wavelet transform algorithm was developed by Daubechies (Daubechies, 1988, 1990). She adopted the multi-resolution signal decomposition (MRSD) algorithm, which was developed by Mallat (Mallat, 1989b 1992), to construct families of compact supported wavelets and coupled them to quadrature mirror filtering (QMF).

In IR spectrometry, an IR spectrum is represented by a set of coefficients in discrete interval and can be collected as $\boldsymbol{C}^{(0)}$. In FWT, $\boldsymbol{C}^{(0)}$ can be expressed as a linear combination of the data obtained at different resolution levels of the original spectrum through the following formula (Daubechies, 1988; Mallat, 1989b)

$$C^{(0)} = \sum_k c_k^{(J)} \sqrt{2^J} \phi_{J,k}(\lambda) + \sum_{j=1}^{J} \sum_k d_k^{(j)} \sqrt{2^j} \Psi_{j,k}(\lambda) \tag{4.1}$$

with $\phi_{j,k}(\lambda)$ and $\Psi_{j,k}(\lambda)$ representing the scaling and wavelet function respectively, J the highest resolution level assigned in the WT calculation and k being a running index which has a variable length depending on the type of wavelet filter and the data length used. In the above expression, $\boldsymbol{C}^{(j)}\left(=c_k^{(j)}\right)$ and $\boldsymbol{D}^{(j)}\left(=d_k^{(j)}\right)$ are the scale and wavelet coefficients at the jth resolution level, respectively. A signal is usually transformed by a low-pass filter $\boldsymbol{H}$ and a high-pass filter $\boldsymbol{G}$ and is represented, respectively, by a series of scale $\boldsymbol{C}^{(j)}$ and wavelet $\boldsymbol{D}^{(j)}$ coefficients at the jth resolution level. The scale coefficients $\boldsymbol{C}^{(j)}$ denote the approximation of the raw signal $\boldsymbol{C}^{(0)}$ with a resolution of one point per every 2^j point of the original one. The wavelet coefficients $\boldsymbol{D}^{(j)}$ represent details of the original signal at different resolution levels (Coifman *et al.*, 1994). These quantities can be deduced through Eqs. (3.6) to (3.9). In this work, coefficients for filters $\boldsymbol{H}$ and $\boldsymbol{G}$ were derived from the Daubechies wavelet, D_{2m}, with m being any positive integer from 1 to 10 and the filter length is equal to $2m$. Algorithm for computing individual coefficient of a particular Daubechies wavelet filters can be found in any references (Daubechies, 1992; Palavajjhala *et al.*, 1994; Borde, 1995; Press *et al.*, 1996) and not state in here.

In the general case, FWT is applied to a signal with a length of $N\left(=2^p\right)$ where p equals to any positive integer (Mallat, 1989b). Suppose an IR spectrum is expressed in the digital form as $\boldsymbol{C}^{(0)} = \left\{c_1^{(0)}, c_2^{(0)}, \ldots, c_N^{(0)}\right\}$ and a D_4 Daubechies wavelet filter with $\boldsymbol{H} = \left\{h_1, h_2, h_3, h_4\right\}$ and $\boldsymbol{G} = \left\{g_1, g_2, g_3, g_4\right\}$ is employed in the

FWT treatment. In the first step, $\boldsymbol{C}^{(0)}$ is first extended periodically on both sides as the following manner

$$\boldsymbol{C}^{(0)}_{extend} = \left\{\ldots, c^{(0)}_{N-1}, c^{(0)}_{N}, c^{(0)}_{1}, c^{(0)}_{2}, \ldots, c^{(0)}_{N-1}, c^{(0)}_{N}, c^{(0)}_{1}, c^{(0)}_{2}, \ldots\right\}. \tag{4.2}$$

Then, the scale and wavelet coefficients at the $j = 1$ resolution level are determined via Eqs. (3.6) and (3.7) respectively as follow:

$$\begin{bmatrix} c^{(j+1)}_{1} \\ c^{(j+1)}_{2} \\ \vdots \\ c^{(j+1)}_{N/2^{j+1}} \end{bmatrix} = \begin{bmatrix} c^{(j)}_{1} & c^{(j)}_{2} & c^{(j)}_{3} & c^{(j)}_{4} & c^{(j)}_{5} & c^{(j)}_{6} & \ldots & c^{(j)}_{(N/2^{j})-1} & c^{(j)}_{N/2^{j}} \end{bmatrix} \begin{bmatrix} h_1 & 0 & \ldots & h_3 \\ h_2 & 0 & \ldots & h_4 \\ h_3 & h_1 & \ldots & 0 \\ h_4 & h_2 & \ldots & 0 \\ 0 & h_3 & \ldots & 0 \\ 0 & h_4 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & h_1 \\ 0 & 0 & \ldots & h_2 \end{bmatrix} \tag{4.3}$$

and

$$\begin{bmatrix} d^{(j+1)}_{1} \\ d^{(j+1)}_{2} \\ \vdots \\ d^{(1)}_{N/2^{j+1}} \end{bmatrix} = \begin{bmatrix} d^{(j)}_{1} & d^{(j)}_{2} & d^{(j)}_{3} & d^{(j)}_{4} & d^{(j)}_{5} & d^{(j)}_{6} & \ldots & d^{(j)}_{(N/2^{j})-1} & d^{(j)}_{N/2^{j}} \end{bmatrix} \begin{bmatrix} g_1 & 0 & \ldots & g_3 \\ g_2 & 0 & \ldots & g_4 \\ g_3 & g_1 & \ldots & 0 \\ g_4 & g_2 & \ldots & 0 \\ 0 & g_3 & \ldots & 0 \\ 0 & g_4 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & g_1 \\ 0 & 0 & \ldots & g_2 \end{bmatrix} \tag{4.4}$$

For example, the first coefficient of the scale and wavelet coefficients at resolution level 1 is expressed as:

$$c^{(1)}_{1} = c^{(0)}_{1} h_1 + c^{(0)}_{2} h_2 + c^{(0)}_{3} h_3 + c^{(0)}_{4} h_4 \tag{4.5}$$

and

$$d^{(1)}_{1} = d^{(0)}_{1} g_1 + d^{(0)}_{2} g_2 + d^{(0)}_{3} g_3 + d^{(0)}_{4} g_4 . \tag{4.6}$$

After FWT treatment, the numbers of elements of $\boldsymbol{C}^{(1)}$ and $\boldsymbol{D}^{(1)}$ are the same and are equal to $N/2$. Then, the same decomposition process as shown in Eqs. (4.3) and (4.4)

are applied to $\boldsymbol{C}^{(1)}$ again to obtain the required coefficients at the next resolution level. The process is repeated until the desired Jth resolution level is reached. Finally, the original spectrum is expressed as a collection of the scale and wavelet coefficients in the form of $\left\{\boldsymbol{C}^{(J)}, \boldsymbol{D}^{(J)}, \boldsymbol{D}^{(J-1)}, \ldots, \boldsymbol{D}^{(1)}\right\}$. The total length or number of coefficients must be equal to the length of the original spectrum. Figure 4.1 shows a schematic diagram for the FWT treatment with data length equals to $1024\left(=2^{10}\right)$.

The data reconstruction process is similar to Eqs. (4.3) and (4.4) and is performed via inverse FWT (IFWT). The scale coefficients at the (J-1)th resolution level is rebuilt by the following equation:

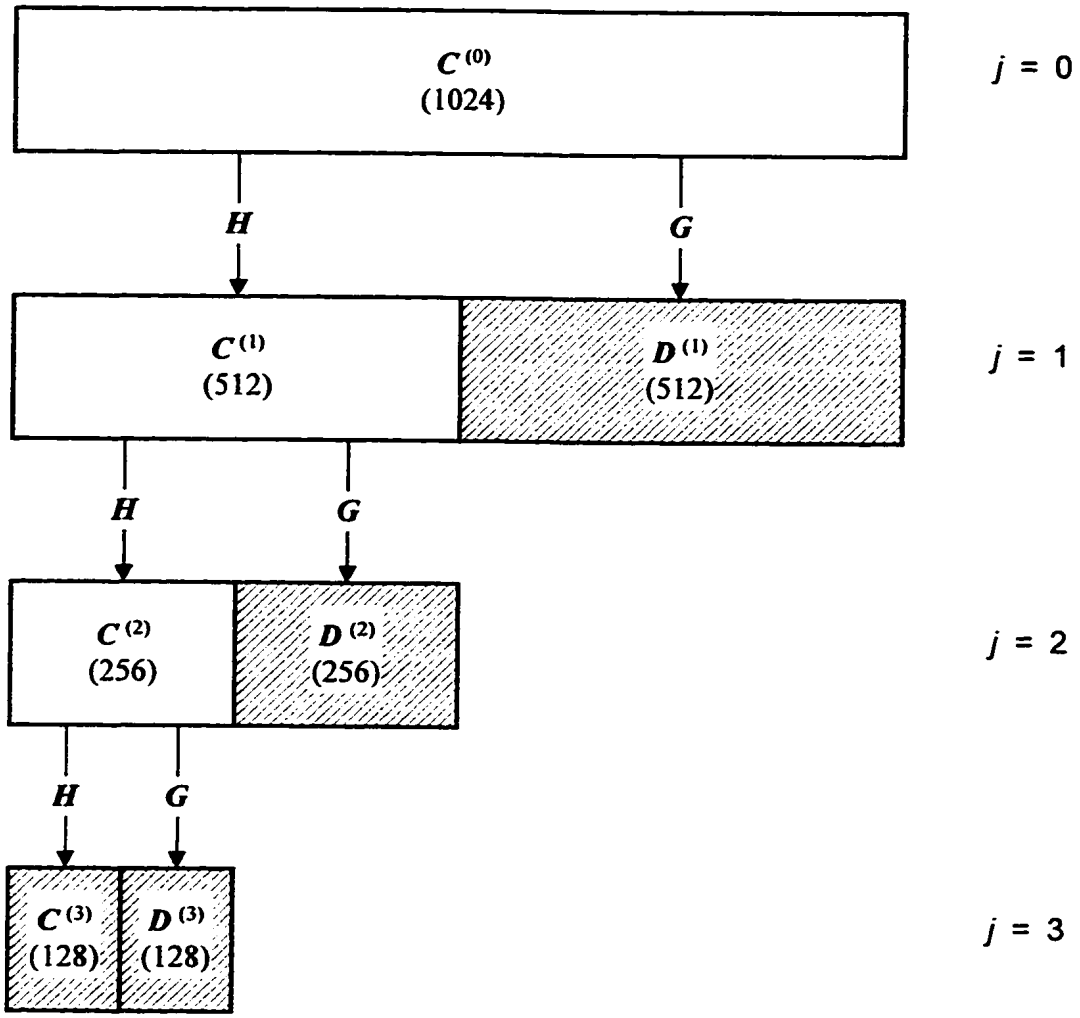
$$\begin{bmatrix} c_1^{(j-1)} \\ c_2^{(j-1)} \\ c_3^{(j-1)} \\ c_4^{(j-1)} \\ c_5^{(j-1)} \\ c_6^{(j-1)} \\ \vdots \\ c_{(N/2^{j-1})-1}^{(j-1)} \\ c_{N/2^{j-1}}^{(j-1)} \end{bmatrix} = \begin{bmatrix} c_1^{(j)} & c_2^{(j)} & \ldots & c_{N/2^j}^{(j)} & d_1^{(j)} & d_2^{(j)} & \ldots & d_{N/2^j}^{(j)} \end{bmatrix} \begin{bmatrix} h_1 & h_2 & h_3 & h_4 & 0 & 0 & \ldots & 0 & 0 \\ 0 & 0 & h_1 & h_2 & h_3 & h_4 & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ h_3 & h_4 & 0 & 0 & 0 & 0 & \ldots & h_1 & h_2 \\ g_1 & g_2 & g_3 & g_4 & 0 & 0 & \ldots & 0 & 0 \\ 0 & 0 & g_1 & g_2 & g_3 & g_4 & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ g_3 & g_4 & 0 & 0 & 0 & 0 & \ldots & g_1 & g_2 \end{bmatrix} \quad (4.7)$$

For example, the first coefficient of the scale coefficient at the (j-1)th resolution level is equal to

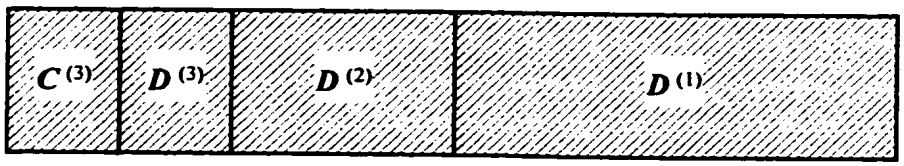
$$c_1^{(j-1)} = c_1^{(j)} h_1 + c_{N/2^j}^{(j)} h_3 + d_1^{(j)} g_1 + d_{N/2^j}^{(j)} g_3 . \quad (4.8)$$

Eq. (4.7) is repeated until resolution level 0 is attained.

Figure 4.1 A schematic diagram showing the operation of the FWT method with a data length of $N = 1024$. The slanting line represents coefficients to be stored.



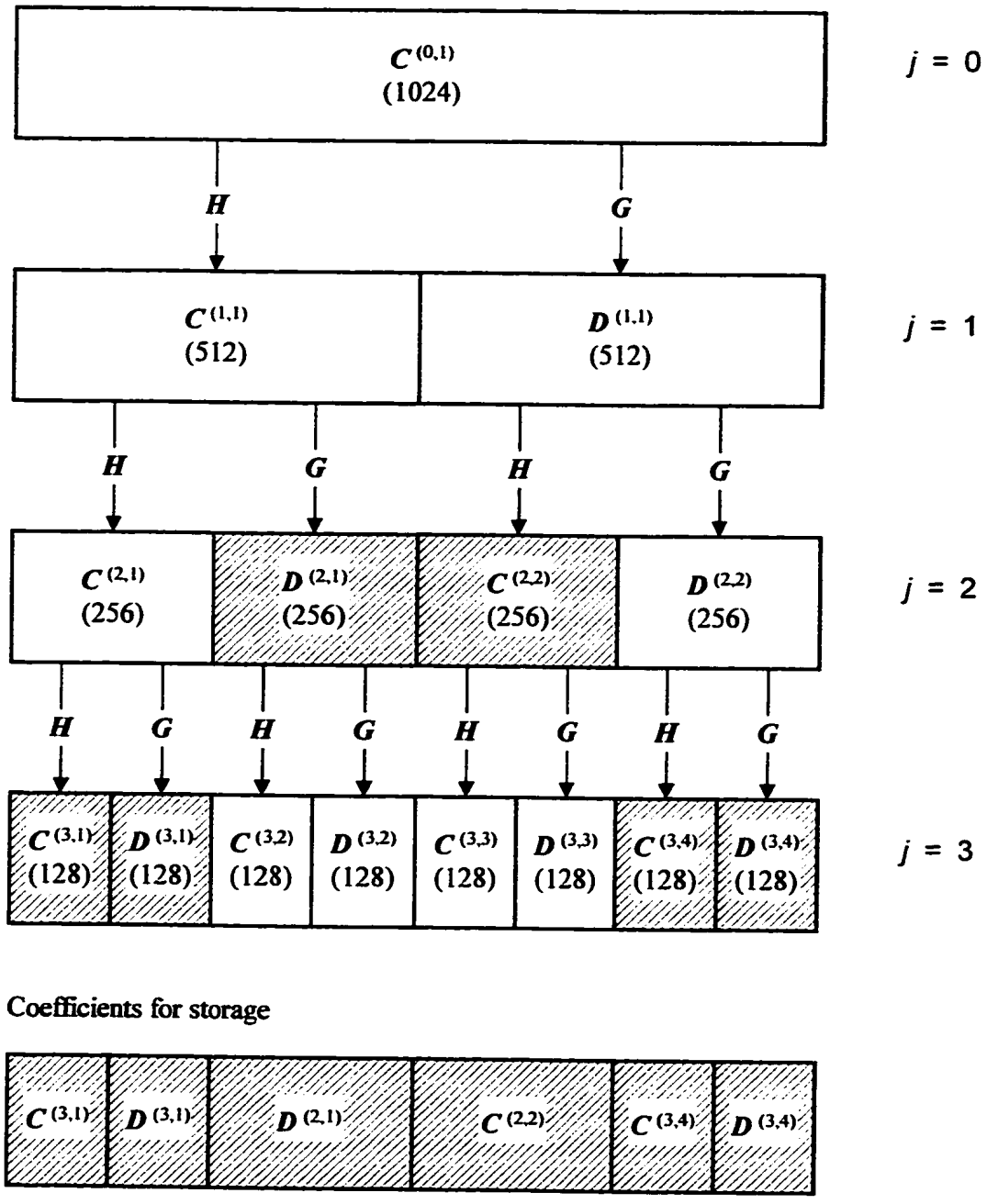
Coefficients for storage



4.2.2 Wavelet Packet Transform

Wavelet packet transform is a derivative of WT that was developed by Coifman and Wickerhauser (Coifman *et al.*, 1994). The discrete WT is generalized in the WPT treatment to provide a more flexible tool for data analysis (Cody, 1994). Applications of WPT in chemical studies can be found in some literature, but this method is less popular than FWT. Only 5 out of 91 papers have been reported with the use of WPT in chemistry (Wickerhauser, 1994b; Walczak *et al.*, 1996; Alsberg *et al.*, 1997; Collantes *et al.*, 1997; Walczak and Massart, 1997). In FWT, a partial multi-resolution analysis is performed. Only $\boldsymbol{C}^{(J)}$ is employed to deduce both scale and wavelet coefficients at the next resolution level. However, WPT allows a full multi-resolution analysis. $\boldsymbol{D}^{(J)}$ is also involved to produce its scale and wavelet coefficients at the same time. As a result, a library of orthonormal bases is obtained. Again, IR spectrum can be represented by a set of coefficients in discrete interval and collected as $\boldsymbol{C}^{(0,1)}$. The first number 0 in the superscript indicates the resolution level and second number 1 indicates the position of scale and wavelet coefficients at a particular resolution level. In general, 2^{J-1} scale and wavelet coefficients will be obtained at resolution level J except the 0th resolution level. Figure 4.2 shows a schematic diagram for the WPT operation with a data length of 2^P. For instance, scale coefficients $\boldsymbol{C}^{(2,2)}$ and wavelet coefficients $\boldsymbol{D}^{(2,2)}$ can be derived from $\boldsymbol{D}^{(1,1)}$ with the following equations if a D_4 Daubechies wavelet filter is employed. $\boldsymbol{D}^{(1,1)}$ can be derived from $\boldsymbol{C}^{(0,1)}$ with Eqs. 4.3 and 4.4.

Figure 4.2 A schematic diagram showing the operation of the WPT method with a data length of $N = 1024$.



$$\begin{bmatrix} c_1^{(2,2)} \\ c_2^{(2,2)} \\ \vdots \\ c_{N/4}^{(2,2)} \end{bmatrix} = \begin{bmatrix} d_1^{(1,1)} & d_2^{(1,1)} & d_3^{(1,1)} & d_4^{(1,1)} & d_5^{(1,1)} & d_6^{(1,1)} & \cdots & d_{N/2-1}^{(1,1)} & d_{N/2}^{(1,1)} \end{bmatrix} \begin{bmatrix} h_1 & 0 & \cdots & h_3 \\ h_2 & 0 & \cdots & h_4 \\ h_3 & h_1 & \cdots & 0 \\ h_4 & h_2 & \cdots & 0 \\ 0 & h_3 & \cdots & 0 \\ 0 & h_4 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & h_1 \\ 0 & 0 & \cdots & h_2 \end{bmatrix} \tag{4.9}$$

and

$$\begin{bmatrix} d_1^{(2,2)} \\ d_2^{(2,2)} \\ \vdots \\ d_{N/4}^{(2,2)} \end{bmatrix} = \begin{bmatrix} d_1^{(1,1)} & d_2^{(1,1)} & d_3^{(1,1)} & d_4^{(1,1)} & d_5^{(1,1)} & d_6^{(1,1)} & \cdots & d_{N/2-1}^{(1,1)} & d_{N/2}^{(1,1)} \end{bmatrix} \begin{bmatrix} g_1 & 0 & \cdots & g_3 \\ g_2 & 0 & \cdots & g_4 \\ g_3 & g_1 & \cdots & 0 \\ g_4 & g_2 & \cdots & 0 \\ 0 & g_3 & \cdots & 0 \\ 0 & g_4 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & g_1 \\ 0 & 0 & \cdots & g_2 \end{bmatrix} \tag{4.10}$$

The first coefficient of the scale and wavelet coefficients, $\boldsymbol{C}^{(2,2)}$ and $\boldsymbol{D}^{(2,2)}$ at resolution level 2 can be expressed by Eqs. (4.11) and (4.12) respectively

$$c_1^{(2,2)} = d_1^{(1,1)}h_1 + d_2^{(1,1)}h_2 + d_3^{(1,1)}h_3 + d_4^{(1,1)}h_4 \tag{4.11}$$

and

$$d_1^{(2,2)} = d_1^{(1,1)}g_1 + d_2^{(1,1)}g_2 + d_3^{(1,1)}g_3 + d_4^{(1,1)}g_4 . \tag{4.12}$$

In Figure 4.2, the original IR spectrum can be expressed as a suitable combination of bases to form a wavelet packet table. For examples, one possible combination of the bases subset to represent the original spectrum is $\left\{\boldsymbol{C}^{(3,1)}, \boldsymbol{D}^{(3,1)}, \boldsymbol{D}^{(2,1)}, \boldsymbol{C}^{(2,2)}, \boldsymbol{C}^{(3,4)}, \boldsymbol{D}^{(3,4)}\right\}$. Another possible combination of the bases subset is $\left\{\boldsymbol{C}^{(3,1)}, \boldsymbol{D}^{(3,1)}, \boldsymbol{D}^{(2,1)}, \boldsymbol{C}^{(3,3)}, \boldsymbol{D}^{(3,3)}, \boldsymbol{D}^{(2,2)}\right\}$. Again, the total number of all these coefficients must be equal to that of the original spectrum. In order to choose

the best basis subset that represents the original data in the most effective way from a huge number of possible combinations of bases, the Shannon-Weaver entropy method was employed in this work to search the best basis (Dai *et al.*, 1994). The Shannon-Weaver entropy of a sequence $x = \{x_j\}$ can be expressed as:

$$H_{SW}(x) = -\sum_j q_j \log q_j \tag{4.13}$$

where $q_j = |x_j|^2 / \|x_j\|^2$ and set $q \log q = 0$ if $q = 0$. The symbols $|x_j|$ and $\|x_j\|$ represent the absolute value and root-mean-square norm of x_j respectively. Since Eq. (4.13) does not obey the theorem of additive measure of information (Coifman *et al.*, 1994), Eq. (4.14) instead of Eq. (4.13) is utilized for entropy calculation.

$$\lambda_{SW}(x) = -\sum_j |x_j|^2 \log |x_j|^2 . \tag{4.14}$$

Once the wavelet packet table is set up, the entropy of each basis is determined by Eq. (4.14). Then, a comparison of the entropy values between two adjacent levels is performed in the following manner for the selection of the best basis. For example, in Figure 4.2, if the total sum of entropy of $\lambda_{SW}(\boldsymbol{C}^{(3,1)})$ and $\lambda_{SW}(\boldsymbol{D}^{(3,1)})$ is less than their parent $\lambda_{SW}(\boldsymbol{C}^{(2,1)})$, then both $\boldsymbol{C}^{(3,1)}$ and $\boldsymbol{D}^{(3,1)}$ are chosen as part of the best basis. On the other hand, if the total sum of entropy of $\lambda_{SW}(\boldsymbol{C}^{(3,2)})$ and $\lambda_{SW}(\boldsymbol{D}^{(3,2)})$ is greater than their parent $\lambda_{SW}(\boldsymbol{D}^{(2,1)})$, then $\boldsymbol{D}^{(2,1)}$ is selected as another part of the best basis. This comparison process is repeated from resolution level (J-1) to 1. This selection process is called the best basis method (Saito and Coifman, 1994). Data reconstruction can be performed by inverse WPT (IWPT) with Eq. (4.7) from resolution level J to 0.

4.2.3 Translation-Rotation Transformation Method

In order to apply MRSD to FWT and WPT calculations, the spectral data vector $C^{(0)}$ (or $C^{(0.1)}$) needs to be extended periodically at the two extremes. However, if $c_1^{(0)}$ and $c_N^{(0)}$ at the two extremes do not have the same value, a small delay which is due to discontinuity of the spectral data at the boundary will be observed at both ends of the reconstructed IR spectrum. Such phenomenon is known as side-lobe problem and deteriorates the quality of the reconstructed data (Chau *et al.*, 1996). To solve such a problem, the translation-rotation transformation (TRT) method was adopted (Hayes *et al.*, 1973). The TRT algorithm involves subtraction of the data vector $C^{(0)}$ by selected quantities $B\{= b_1, b_2, \ldots, b_k\}$ to give the rotated array by

$$c_{k.TRT}^{(0)} = c_k^{(0)} - b_k \tag{4.15}$$

with

$$b_k = c_1^{(0)} + \frac{\left(c_N^{(0)} - c_1^{(0)}\right)(k-1)}{N-1} \tag{4.16}$$

where k is a running index from 1 to N. In the spectral reconstruction process, the inverse TRT (ITRT) algorithm (Eq. (4.17)) is employed to generate $C'^{(0)}$ in the original domain.

$$c_k'^{(0)} = c_{k.TRT}^{(0)} + b_k \tag{4.17}$$

4.2.4 Coefficient Position Retaining Method

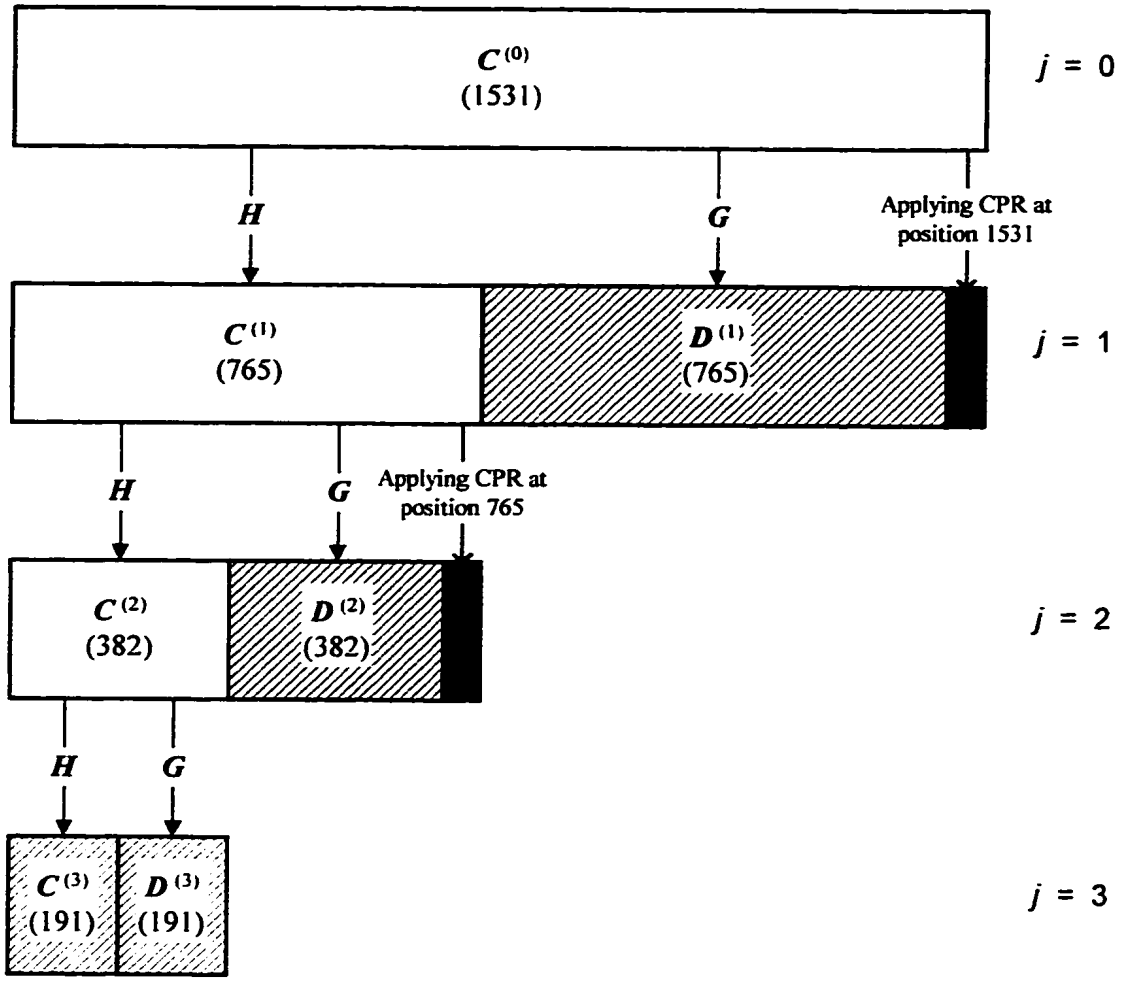
In WT treatment, the data length for a basis to be processed at resolution level j must be an even number. If an odd number data set is encountered at a particular resolution level, the WT calculation will be stopped automatically and cannot be processed to the next higher resolution level. It is because the data length of the scale and wavelet coefficients must be the same after WT treatment. Hence, in an ideal case, the number of spectral data must equal to 2^p where p equals to any positive integer. In Figure 4.1, if the original spectral data length is equal to $1024\left(=2^{10}\right)$, the data length of all bases at each resolution level j can be guaranteed to be an even number. In practice, it is not easy for a chemical instrument to generate 2^p data exactly. To cope with this problem, a series of zeros is usually appended to one end or both ends of the original data set in order to bring the total length to the next power of 2. This method is called zero padding method and is widely used in fast Fourier transform calculation (Verdun *et al.*, 1988; Morrison, 1994). Besides, truncation of data at the end or both ends of the original data to the previous power of 2 can also be adopted in some cases.

A new method, called Coefficient Position Retaining (CPR) method, was developed in this work to process spectrum with odd number of data in the FWT and WPT calculations. In this approach, if the data length $N_{c,j}$ of a scale coefficient $\boldsymbol{C}^{(j)}$ is an even number, FWT is applied as usual by using Eqs. (4.3) and (4.4). The scale and wavelet coefficients obtained at resolution level (j+1) will have the same number of coefficients $N_{c,j+1}\left(=N_{c,j}/2\right)$ and $N_{d,j+1}\left(=N_{c,j}/2\right)$ respectively. On the other

hand, if $N_{c,j}$ is an odd number, FWT is adopted without using the last coefficient of $C^{(j)}$ in the calculation. This coefficient is retained and transferred downward to the same position at the next resolution level. Then, it becomes the last coefficient of $D^{(j+1)}$ at the next resolution level. As a result, the scale and wavelet coefficients will have $N_{j+1,c}\left(=\left(N_{j,c}-1\right)/2\right)$ and $N_{j+1,d}\left(=\left(\left(N_{j,c}-1\right)/2\right)+1\right)$ elements respectively. Figures 4.3 and 4.4 show a schematic diagram for applying FWT to a data set with 1531 and 1023 data points, respectively, with the use of CPR algorithm. Figure 4.4 shows a special case that the total data length of both scale coefficients at all resolution levels are odd number. CPR algorithm can process this data without any problem.

In WPT calculation, the scale coefficient $C^{(j)}$ with odd number of data length is handled in the same way as that for FWT as described above. However, for the wavelet coefficient $D^{(j)}$, the treatment is slightly different. Again, the last coefficient in $D^{(j)}$ is retained and transferred downward to all resolution levels that is below the present one at the same position. All these coefficients selected are not involved in FWT calculation. In WPT, the wavelet coefficient $D^{(j)}$ may consist more than one coefficients rather than only one coefficient retained as in FWT treatment. Such retained coefficients are derived from the (j-2)th or above resolution level. Figure 4.5 shows a schematic diagram for the WPT calculation on a spectrum with 1531 data with CPR treatment. In the first cycle of WPT computation, the last coefficient in $C^{(0,1)}$ is transferred to position 1531 from levels 2 to 4. This coefficient is not involved in the WPT computation for $D^{(1,1)}$ and $D^{(2,2)}$. In the second cycle, since the data length of $D^{(1,1)}$ is an odd number, the last coefficient of $D^{(1,1)}$ at position 1530 is

Figure 4.3 A schematic diagram showing the operation of the FWT method with a data length of N = 1531 coupled with CPR treatment. The slanting line represents coefficients to be stored and the black region shows the position of the coefficient(s) to be archived in using the CPR method.



Coefficients for storage

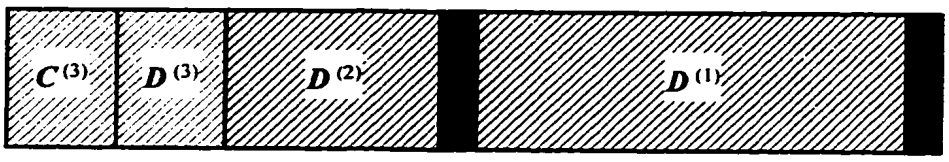


Figure 4.4 A schematic diagram showing the operation of the FWT method with a data length of N = 1023 coupled with CPR treatment. The slanting line represents coefficients to be stored and the black region shows the position of the coefficient(s) to be archived in using the CPR method.

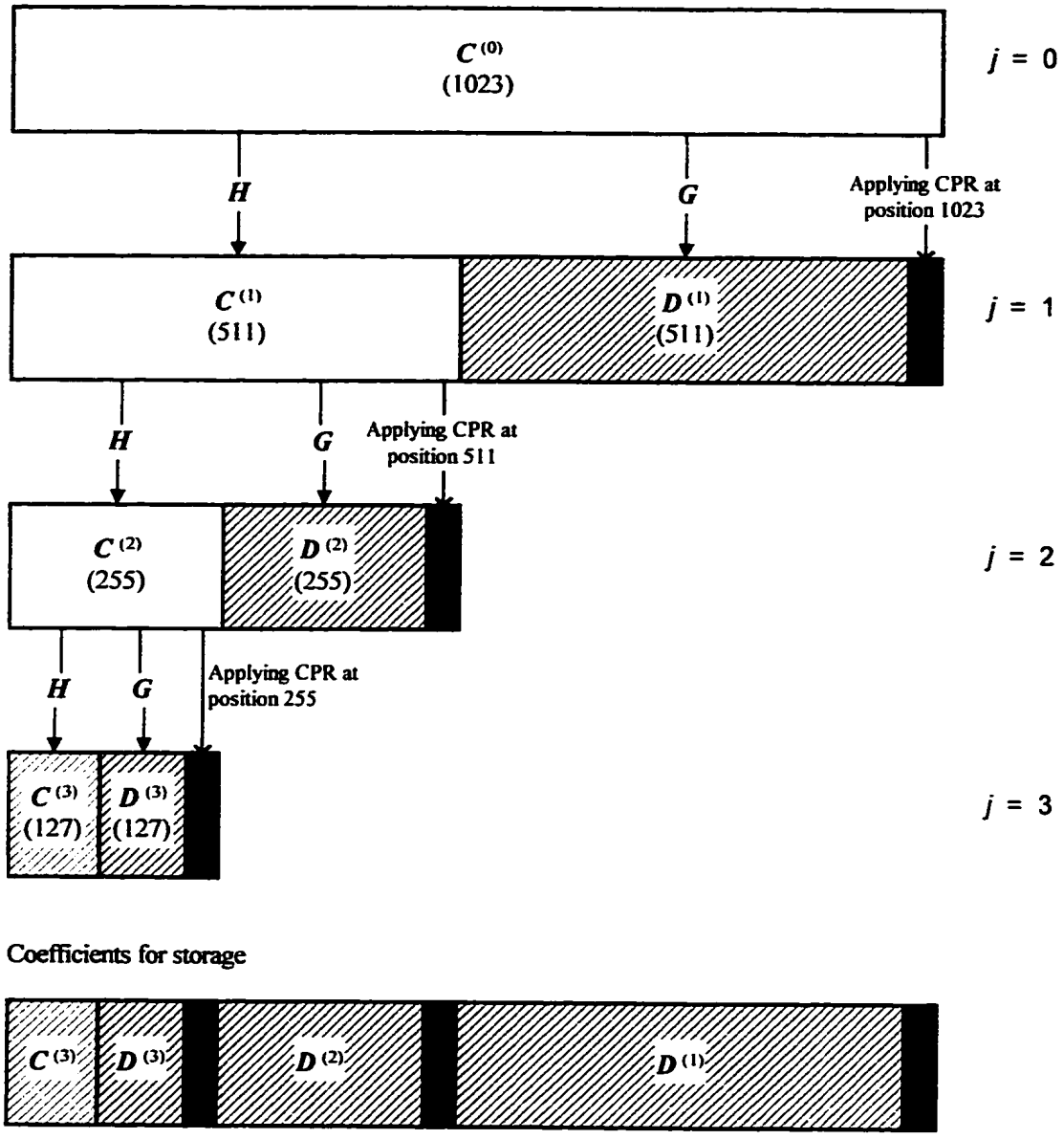
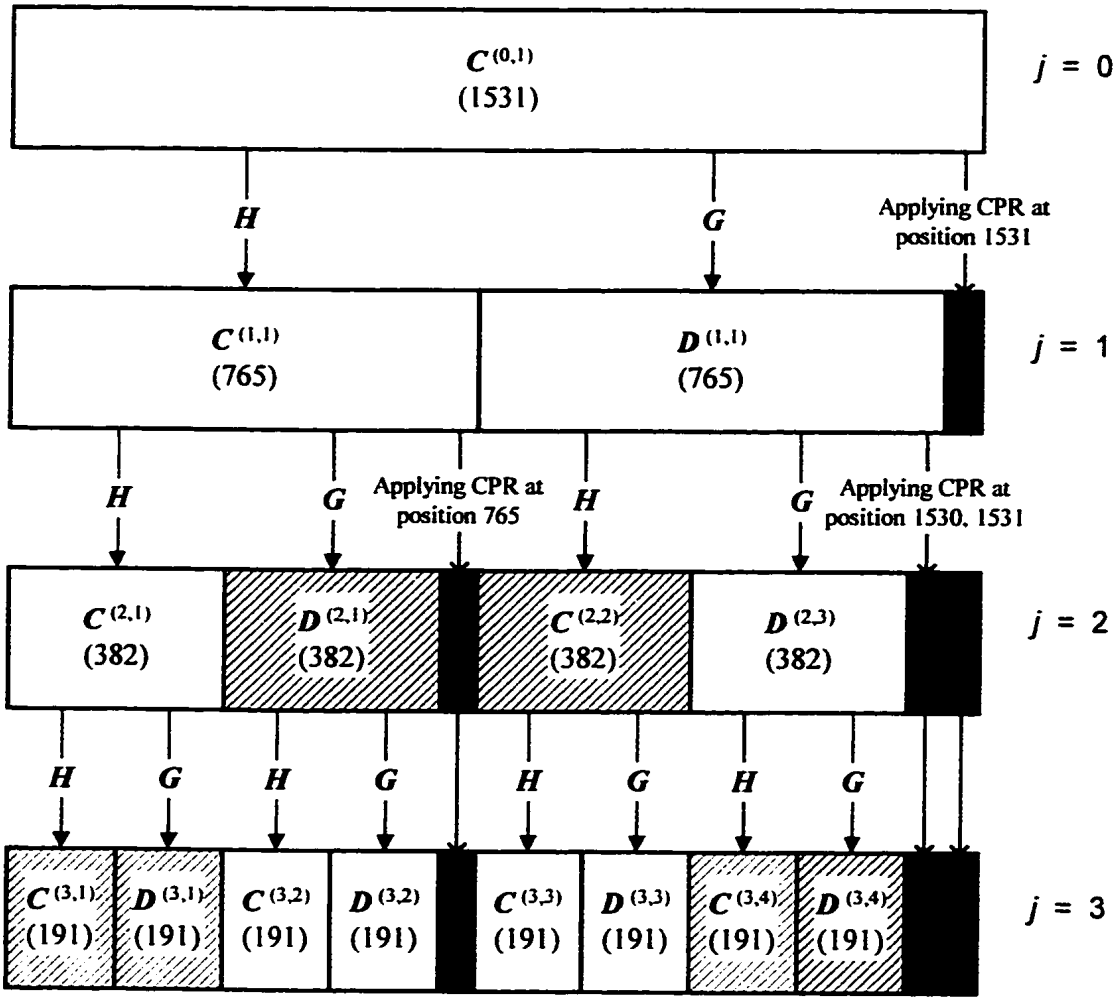
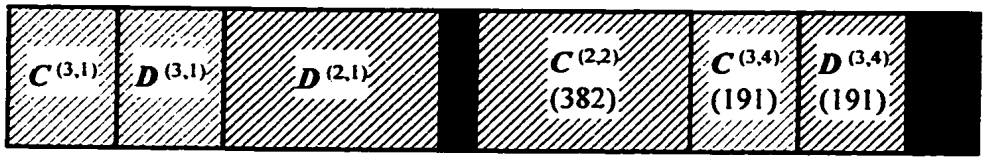


Figure 4.5 A schematic diagram showing the operation of the WPT method with a data length of N = 1531 coupled with CPR treatment. The slanting line represents coefficients to be stored and the black region indicates the position of the coefficient(s) to be archived in using the CPR method.



Coefficients for storage



retained and transferred to position 1530 from levels 3 to 4. As a result, $\boldsymbol{D}^{(2,2)}$ contains two retaining coefficients. Again, these coefficients are not involved in the WPT calculation for $\boldsymbol{D}^{(2,2)}$. In general, the retaining coefficients are not considered in WT calculation.

4.2.5 Criteria for Data Compression

After FWT or WPT processing, the total data length of all $\boldsymbol{C}$ and $\boldsymbol{D}$ coefficients does not change. To reduce storage space, some of these coefficients must be discarded. Several methods have been proposed to select which coefficients in an optimal basis are negligible. They include the absolute cutoff, relative energy and entropy criterion methods (Coifman *et al.*, 1994) and are described as follows. The absolute cutoff method is the simplest one. A cutoff value ε with magnitude greater than zero is assigned. Then, the absolute value of any coefficient c or d with a value less than ε is rejected. In the relative energy method, a cutoff value is defined in the range $0 < \varepsilon \leq 1$. Any coefficient in the absolute-square form ($|c|^2$ or $|d|^2$) that is less than $\varepsilon\|x\|^2$ is discarded. In the entropy criterion method, an average energy of a significant coefficient is defined as $\exp\left(-\lambda_{SW}(x)/\|x\|^2\right)$. Also, the cutoff value is defined in the range of $0 < \varepsilon \leq 1$. Any coefficient in the absolute-square form that has value less than that of $\varepsilon\exp\left(-\lambda_{SW}(x)/\|x\|^2\right)$ is neglected. In this work, the absolute cutoff method is employed to select the required coefficients for storage. It is because the cutoff value ε can be assigned very easily by visual inspection.

In both FWT and WPT treatment, the transformed IR spectra are represented by the scale coefficients $C^{(J)}$ in FWT and $C^{(J,1)}$ in WPT at the required resolution level J. As usual, one cannot remove part of these coefficients for the compression purpose because this will cause loss of useful information and will lead to serious error during signal reconstruction. So no compression were performed on the scale coefficients. As a result, the proposed compression method was only applied to the wavelet coefficients.

4.2.6 Direct Matching Method for Library Search

Direct matching method is the simplest way for spectral library search. It employs a point-by-point comparison calculation to match sequentially an unknown spectrum against the reference spectra available in a library (Lowry *et al.*, 1985). Both the unknown and reference spectra are represented as points in a multidimensional space in a fixed range of wavelength. The degree of similarity between two spectra is measured by the sum of separations between data points. Six different search routines have been proposed for the purpose. They include absolute difference $D_{absdiff}$, absolute derivative D_{absder}, square difference $D_{sqrdiff}$, square derivative D_{sqrder}, correlation coefficient D_{corr} and Euclidean distance D_{Edist}. Mathematical expressions of these routines are given as follows (Rasmussen and Isenhour, 1979; Lowry *et al.*, 1985; Harrington and Isenhour, 1987; Hobert, 1995; Penchev *et al.*, 1996)

$$D_{absdiff} = \sum_{i=1}^{N} |x_i - y_i| \tag{4.18}$$

$$D_{absder} = \sum_{i=1}^{N-1} \left| (x_i - x_{i+1}) - (y_i - y_{i+1}) \right| \tag{4.19}$$

$$D_{sqrdiff} = \sum_{i=1}^{N} (x_i - y_i)^2 \tag{4.20}$$

$$D_{sqrder} = \sum_{i=1}^{N-1} \left[(x_i - x_{i+1}) - (y_i - y_{i+1}) \right]^2 \tag{4.21}$$

$$D_{corr} = \frac{N\sum_{i=1}^{N} x_i y_i - \left(\sum_{i=1}^{N} x_i\right)\left(\sum_{i=1}^{N} y_i\right)}{\sqrt{\left[N\sum_{i=1}^{N} x_i^2 - \left(\sum_{i=1}^{N} x_i\right)^2\right]\left[N\sum_{i=1}^{N} y_i^2 - \left(\sum_{i=1}^{N} y_i\right)^2\right]}} \tag{4.22}$$

and

$$D_{Edist} = \sqrt{\sum_{i=1}^{N} (x_i - y_i)^2} \,. \tag{4.23}$$

In these expressions, x_i and y_i denote, respectively, the data points in the reference and unknown spectra in a particular domain. N represents the total number of data points in the compressed or uncompressed spectrum. A perfect match of any two spectra would give one by using correlation coefficient method (Eq. (4.22)) and a zero value by using the other methods. A lower D_{corr} value or a higher D value for the other methods indicates higher dissimilarity between the spectra. Both the absolute derivative and square derivative routines were applied when the base-line offset was observed. These two routines can discriminate the correct spectrum from the other spectra with similar D value in the library. However, if the unknown spectrum is not available in the library, they may give poor match values for spectra that have similar spectral features. In this study, D_{corr}, $D_{absdiff}$ and D_{absder} were employed in the spectral search. Both $D_{sqrdiff}$ and D_{sqrder} method will have similar results from the $D_{absdiff}$ and D_{absder} methods. So, such methods were not applied in this work.

Some workers did not carry out the direct matching method in the original domain because it involves large number of calculation. They suggested to transform the IR spectrum by Fourier transform before spectrum matching (Glasser, 1987). In this investigation, both FWT and WPT were applied to treat IR spectra for setting up a small scale spectral library. These two methods provide a way to minimize the search time for direct matching of spectra because it involves only minimal number of coefficients represented in the wavelet domain instead of using the whole spectrum.

4.3 Experimental

Twenty organic compounds in AR grade (see Tables 4.2 and 4.3) as utilized in this work were obtained from Aldrich (Gillingham, U.K.) and were used without further purification. Their IR spectra were recorded by using the Nicolet Magna-IR™ Model 750 Fourier transform infrared spectrometer (Nicolet Instrument Corporation, Wisconsin, U.S.) in the range of 400 to 4000 cm^{-1} in percentage of transmittance at 4 cm^{-1} resolution with 16 times of scan and Happ-Genzel apodization. In this setting, each IR spectrum contains 1868 data points. ASCII data in comma-separated value (CSV) format were exported from the ONMIC Version 2.1 (Nicolet Instrument Corporation, Wisconsin, U.S.) which was used to control the FT-IR spectrometer. Data in CSV format were further converted with Microsoft® Excel for Windows™ 95 Version 7.0 (Microsoft Corporation, Washington, U.S.) to tab-delimited format.

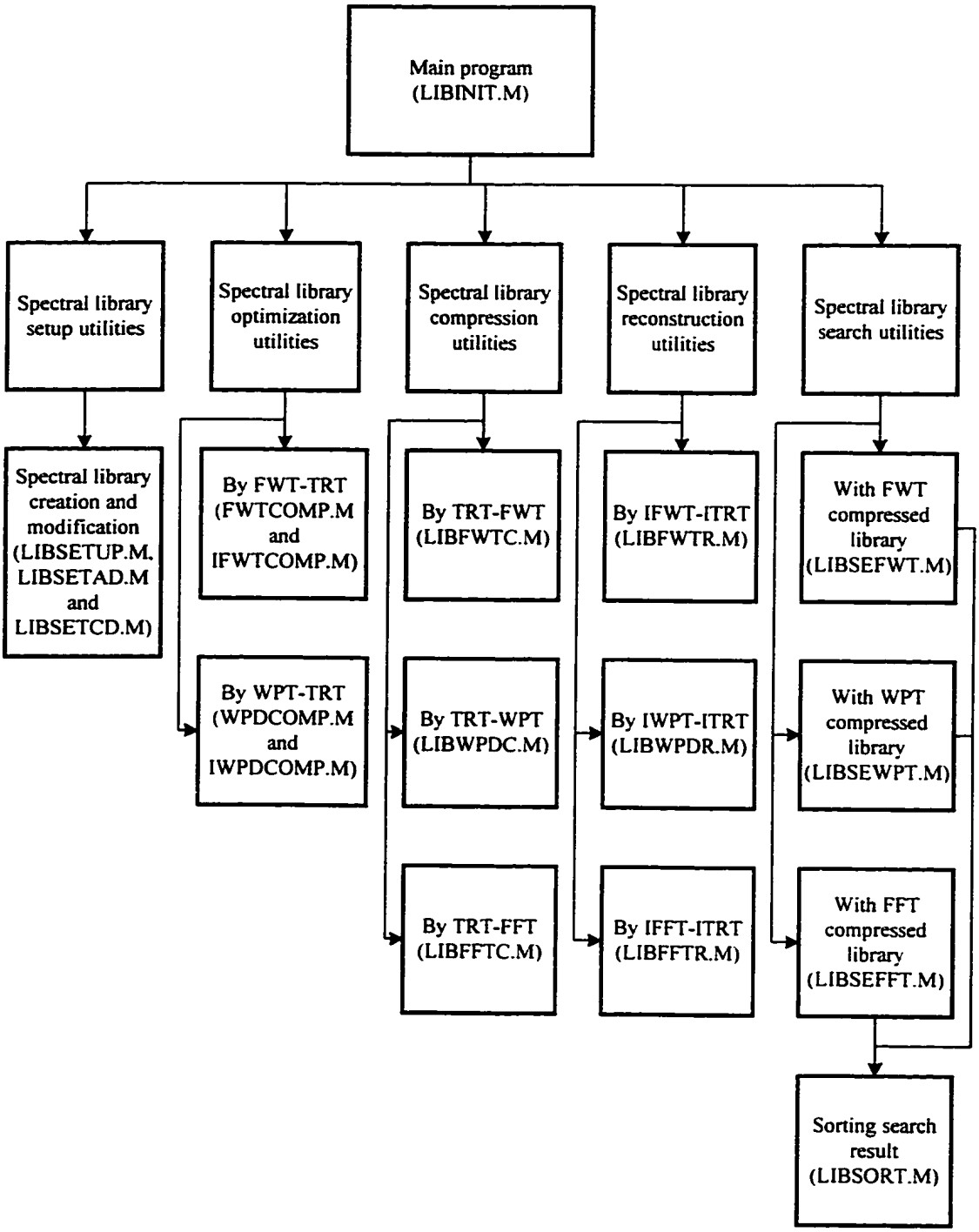
Solid samples were recorded by using KBr pellet method while liquid samples were recorded by liquid film method using NaCl plates. The organic compounds were chosen purposely with similar structures or isomers of one another and their IR spectra are quite similar. The data set comprises the IR spectra of the mono-, di- and tri-substituted benzene with Cl, Br, I, NO_2, OH, CH_3 and CH_2Cl as substituents. All IR spectra were background corrected and normalized to unit transmittance first (Eq. (4.24)) before further treatment through the following formula:

$$F_\lambda = \frac{A_\lambda}{A_{max}} \tag{4.24}$$

where F_λ and A_λ are the fraction of total absorbance and the absorbance at wavelength λ, respectively, and A_{max} is the maximum absorbance in the spectrum.

All WT computations were carried out using the Spectral Library Compression and Search (SLCS) Toolbox Version 1.0 as developed in this study. It was coded in MATLAB® for Windows Version 4.2c (The MathWorks, Inc., 1992) under Microsoft® Windows™ 95 environment on an AST® Advantage!® 818 PC (AST Research, Inc., California, U.S.) with a 133MHz Pentium® processor. The toolbox can be employed to optimize the required parameters for FWT and WPT calculation as well as to perform spectral library compression, construction and reconstruction, and spectral library searching. Figure 4.6 shows a block diagram of the SLCS toolbox and Appendices 4.1 to 4.2 show a program description of individual program and screen capture from the SLCS toolbox. Appendix 4.3 shows the file formats of spectral library that were generated by FWT, WPT and FFT compression. Appendix 4.4 shows a list of file extensions as generated from the SLCS toolbox.

Figure 4.6 A block diagram of the Spectral Library Compression and Search (SLCS) Toolbox Version 1.0.



4.4 Results and Discussion

4.4.1 Choice of Parameters for FWT and WPT Calculations

In both FWT and WPT calculations, the Daubechies wavelet function D_{2m}, resolution level J, and cutoff value ε have to be selected to give optimal performance. In this work, the compression ratio R_{comp} (Eq. (4.25)) measures the compression efficiency of the selected method while D_{corr} (Eq. (4.22)) measures the similarity between the original and reconstructed IR spectrum.

$$R_{comp} = \frac{\text{No. of bytes of the original data - No. of bytes of the compressed data}}{\text{No. of bytes of the original data}} \times 100\% \quad (4.25)$$

The best performance is considered when R_{comp} and D_{corr} have values close to 100% and 1 respectively. In computing R_{comp} for data obtained from WT and FT methods, the file sizes of the compressed data were used. Each file consists both of intensity and position information. In the compression process, a fixed cutoff value ε is assigned. The absolute value of any coefficient c or d with a value less than ε is rejected. So, only the position and the intensity value of the retained coefficients in the wavelet domain that obtained from the cutoff treatment are stored in the compressed IR spectral file.

Table 4.1 shows the results of compressing the IR spectrum of benzoic acid with different ε values and D_{2m} functions at resolution levels of J = 3, 4 and 5 by utilizing FWT. As the value of J increases, R_{comp} increases with a slight decrease in D_{corr} at a fixed cutoff value. The same trend is observed when both J and D_{2m} are fixed with different cutoff values. At a fixed resolution level and cutoff value, a

Table 4.1 Results of applying the proposed FWT scheme to compress IR spectrum of benzoic acid with the use of different cutoff values ε and Daubechies wavelets function D_{2m} at resolution levels of $J = 3$, 4 and 5.

	$J=3$		$J=4$						$J=5$	
	$\varepsilon=0.2$		$\varepsilon=0.1$		$\varepsilon=0.2$		$\varepsilon=0.3$		$\varepsilon=0.2$	
	R_{comp}	D_{corr}	R_{comp}	D_{corr}	R_{comp}	D_{corr}	R_{comp}	D_{corr}	R_{comp}	D_{corr}
D_2	86.51	0.9927	87.37	0.9968	91.38	0.9905	92.61	0.9844	93.42	0.9892
D_4	86.62	0.9968	90.42	0.9980	91.92	0.9955	92.83	0.9912	94.27	0.9943
D_6	86.51	0.9975	90.58	0.9984	91.81	0.9962	92.88	0.9909	94.00	0.9956
D_8	86.78	0.9969	90.26	0.9986	92.24	0.9956	92.99	0.9917	94.33	0.9950
D_{10}	86.83	0.9972	90.69	0.9984	92.08	0.9958	92.67	0.9931	94.27	0.9949
D_{12}	86.94	0.9964	90.15	0.9986	92.13	0.9955	92.72	0.9929	94.27	0.9945
D_{14}	86.67	0.9975	90.47	0.9985	92.02	0.9962	92.88	0.9921	94.16	0.9953

(cont.)

(cont. Table 4.1)

D_{16}	86.83	0.9972	90.96	0.9983	92.34	0.9958	92.56	0.9943	94.27	0.9952
D_{18}	86.83	0.9966	90.15	0.9983	92.08	0.9952	92.88	0.9913	93.90	0.9949
D_{20}	86.88	0.9966	90.10	0.9984	92.08	0.9954	92.83	0.9920	93.90	0.9949

maximum values of R_{comp} and D_{corr} have been found by employing the Daubechies wavelet functions D_{14} and D_{16}. Table 4.2 shows the results of compressing IR spectra of the twenty compounds under study with the proposed FWT and WPT treatments with $J = 4$, $\varepsilon = 0.20$, and using the D_{14} and D_{16} functions. Based on the results of error plots, values of R_{comp} and D_{corr} for 20 compounds, we recommend to use the D_{16} wavelet with $\varepsilon = 0.20$ and $J = 4$ to compress the IR spectrum in the range of 400-4000 cm^{-1} for data storage by FWT and WPT. It is because the Daubechies D_{16} wavelet function has slightly higher compression ratio. Figure 4.7 shows the reconstructed IR spectra from the compressed data as obtained from both FWT and WPT treatments on the IR spectrum of benzoic acid with D_{16}, $\varepsilon = 0.20$ and $J = 4$.

4.4.2 Comparison of Performance of FWT and WPT with That of Fast Fourier Transform

The performance of FWT and WPT in spectral compression were compared with that by using FFT. R_{comp} and D_{corr} were chosen as the key parameters for comparison. As mentioned before, FFT and its inverse are the major techniques adopted currently to compress spectral data in chemical studies. Three methods namely the absolute cutoff (Binkley and Dessy, 1980), least-square fitting (Krauss *et al.*, 1994) and power spectrum (Kirmse and Westerberg, 1971) methods coupled with FFT were employed for selecting Fourier coefficients. The chosen Fourier coefficients represent the compressed form of the original spectral data. In the absolute cutoff method, only those coefficients greater than a threshold value are

Table 4.2 Results of applying the proposed FWT and WPT methods to compress IR spectra of the twenty compounds under study with cutoff value $\varepsilon = 0.20$ at and $J = 4$ by using the Daubechies wavelet functions D_{14} and D_{16}.

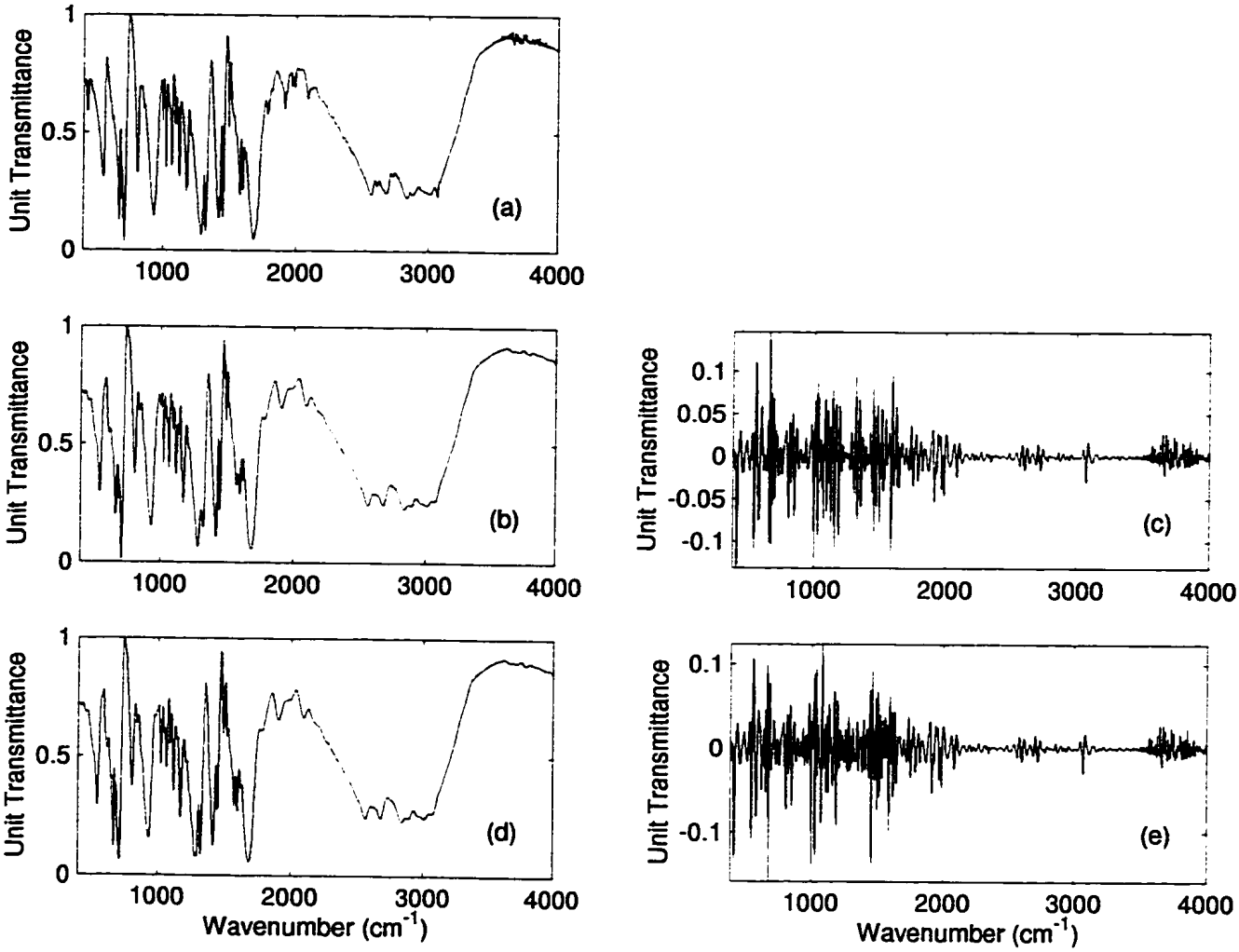
Compound Name	FWT				WPT			
	D_{14}		D_{16}		D_{14}		D_{16}	
	R_{comp}	D_{corr}	R_{comp}	D_{corr}	R_{comp}	D_{corr}	R_{comp}	D_{corr}
1,4-dichlorobenzene	92.99	0.9759	93.09	0.9730	93.04	0.9758	93.09	0.9750
1-bromobutane	93.63	0.9958	93.52	0.9967	93.68	0.9958	93.58	0.9966
1-chloro-2,4-dinitrobenzene	93.74	0.9692	93.79	0.9636	93.79	0.9672	93.79	0.9636
1-chlorobutane	92.93	0.9965	93.20	0.9957	92.99	0.9967	93.25	0.9962
1-iodobutane	92.18	0.9975	92.51	0.9970	92.13	0.9980	92.45	0.9978
2,4-dichlorophenol	93.04	0.9739	92.93	0.9769	93.09	0.9738	92.99	0.9769
2-chlorobutane	92.24	0.9963	92.24	0.9963	92.45	0.9960	92.51	0.9961
2-chlorophenol	92.02	0.9975	92.08	0.9972	92.29	0.9974	92.18	0.9973

(cont.)

(cont. Table 4.2)

2-nitrophenol	92.40	0.9823	92.61	0.9825	92.40	0.9823	92.67	0.9814
3-chlorophenol	92.88	0.9982	93.09	0.9983	92.93	0.9983	93.15	0.9984
3-nitrophenol	93.20	0.9838	93.09	0.9866	93.25	0.9853	93.09	0.9866
4-nitrobenzyl chloride	93.20	0.9720	92.93	0.9693	93.25	0.9716	92.99	0.9700
4-nitrophenol	91.49	0.9957	92.02	0.9944	91.54	0.9957	92.02	0.9946
4-nitrotulene	92.45	0.9865	92.08	0.9869	92.45	0.9871	92.40	0.9859
benzoic acid	92.02	0.9962	92.34	0.9958	92.13	0.9963	92.45	0.9959
bromobenzene	92.08	0.9954	92.08	0.9941	92.40	0.9951	92.45	0.9946
cyclohexyl bromide	91.70	0.9962	91.97	0.9943	91.18	0.9961	92.08	0.9948
cyclopentyl chloride	93.52	0.9976	93.42	0.9979	93.52	0.9978	93.42	0.9982
iodobenzene	91.92	0.9943	91.92	0.9939	92.02	0.9943	91.92	0.9945
toluene	92.67	0.9956	92.72	0.9957	92.72	0.9960	92.88	0.9958
Average	92.62	0.9898	92.72	0.9893	92.70	0.9780	92.77	0.9895
Standard Derivation	±0.66	±0.0100	±0.63	±0.0106	±0.69	±0.0102	±0.54	±0.0106

Figure 4.7 The experimental IR spectrum (a) and reconstructed IR spectra of benzoic acid with the use of compressed data as obtained from FWT (b) and WPT (d) by utilizing the D_{16} function, $\varepsilon = 0.20$, and $J = 4$ with the corresponding error plot from FWT (c) and WPT (e).



stored. In the least-square fitting method, the first $N/2$ Fourier coefficients from the power spectrum are extracted and the logarithms of these quantities are least-square fitted with a fourth-degree polynomial to determine the number of Fourier coefficients to be stored. In the power spectrum method, the number of coefficients for storage is determined from the power spectrum with some criteria. Details of the above three methods can be found in the work by Chau and Tam (Chau and Tam, 1994).

Table 4.3 shows the results of compression for IR spectra of the twenty compounds as obtained by FFT coupled with the compression methods mentioned above. Although the FFT treatment using the least-square-fitting and power spectrum methods give very high D_{corr} values, the compression ratio R_{comp} is very low. Spectral compression by FFT coupled with the absolute cutoff method ($\varepsilon = 0.10$) can attain the same average compression ratio as FWT and WPT (Table 4.2). But, the compression ratios for individual spectra fluctuate in a wide range by using FFT in the range of 85.87% (2,4-dichlorophenol) to 97.22% (1-bromobutane). In this investigation, our aim is to set up a good spectral library for IR spectra. It is desirable to have any method that can provide highly stable compression ratio and minimum distortion in the reconstructed spectra. From the results of Tables 4.2 and 4.3, it is clear that the performance of our proposed FWT and WPT methods in spectral compression is better than that of FFT.

In terms of computational time, the FWT calculation requires shorter time than that of FFT. A FFT computation requires $N \log_2 N$ operations while a single cycle

Table 4.3 Results of applying the FFT method to compress IR spectra of the twenty compounds[a] under study by using the absolute cutoff, the least-square-fitting, and power spectrum methods.

Compound Name	Absolute cutoff method		Least-square-fitting method		Power spectrum method	
	R_{comp}	D_{corr}	R_{comp}	D_{corr}	R_{comp}	D_{corr}
1,4-dichlorobenzene	90.26	0.9903	75.05	0.9997	52.14	1.0000
1-bromobutane	97.22	0.9907	75.05	1.0000	53.69	1.0000
1-chloro-2,4-dinitrobenzene	86.08	0.9976	77.84	0.9995	56.42	1.0000
1-chlorobutane	93.15	0.9981	76.07	1.0000	54.76	1.0000
1-iodobutane	95.61	0.9917	77.09	1.0000	53.21	1.0000
2,4-dichlorophenol	85.87	0.9976	75.05	0.9995	57.98	1.0000
2-chlorobutane	92.24	0.9975	75.05	1.0000	51.28	1.0000
2-chlorophenol	89.72	0.9993	75.05	1.0000	52.41	1.0000
2-nitrophenol	92.93	0.9837	75.05	0.9998	53.69	1.0000
3-chlorophenol	93.52	0.9988	75.05	1.0000	54.01	1.0000

(cont.)

(cont. Table 4.3)

3-nitrophenol	88.92	0.9975	76.93	0.9998	61.46	1.0000
4-nitrobenzyl chloride	89.61	0.9926	75.05	0.9997	52.52	1.0000
4-nitrophenol	87.10	0.9988	77.68	0.9999	53.16	1.0000
4-nitrotulene	93.20	0.9819	76.39	0.9994	51.66	1.0000
benzoic acid	96.84	0.9788	75.54	0.9999	57.07	1.0000
bromobenzene	92.51	0.9903	75.05	0.9999	52.62	1.0000
cyclohexyl bromide	93.25	0.9906	75.05	0.9999	52.89	1.0000
cyclopentyl chloride	96.20	0.9963	77.57	1.0000	52.52	1.0000
iodobenzene	90.04	0.9926	75.70	0.9999	50.70	1.0000
toluene	95.77	0.9831	76.70	0.9999	52.52	1.0000
Average	92.00	0.9924	75.87	0.9998	53.84	1.0000
Standard Derivation	±3.46	±0.0063	±1.04	±0.0002	±2.59	±0.0000

[a] The number of bytes of the original data was used in computing R_{comp}.

of FWT require only N operations for a data length of $N\left(=2^p\right)$ (Cody, 1992; Daubechies, 1994). In FWT, it is expected that the computation time increases when a higher J resolution level is required. Since FFT was specially designed for data length equal to 2^p, a slower algorithm is employed for FT calculation if the data length is not equal to 2^p. For a spectrum with data length equal to 2^p, FFT method is faster than FWT method if similar compression ratio is attained. For example, in compressing an IR spectrum with N = 1024 with similar compression ratio, FFT method requires 10,240 multiplication operations while FWT method needs 30,720 (from Eq. (4.26)) multiplication operations (up to the 4th resolution level with the D_{16} function (L = 16)) with L being the length of the filters $\boldsymbol{H}$ and $\boldsymbol{G}$. But, the computational time would be close for the FFT and FWT methods if the data length of a spectrum is not equal to 2^p. To investigate the computational times for FWT and WPT, the following algorithm is adopted. The total number of multiplication operations after J iterations in FWT treatment given as follows (Chau *et al.*, 1996)

$$L\times N + L\times N/2 + \ldots + L\times N/2^J = 2L\times N\times\left(1-1/2^J\right). \tag{4.26}$$

In WPT calculation, the total number of multiplication operations after J operations is

$$L\times N + 2L\times N/2 + \ldots + 2^J L\times N/2^J = JL\times N. \tag{4.27}$$

For example, to compress an IR spectrum with N = 1868 up to $J=4$ using D_{16} function (L = 16), FWT and WPT require about 56,040 and 119,552 multiplication operations respectively. Hence, the computational time needed for FWT is about 47% shorter than that for WPT calculation.

4.4.3 Performance of the Coefficient Position Retaining Method

The data length of each IR spectrum recorded from our the FT-IR instrument is 1868 which is not equal to power of 2. The FWT calculation will stop at the 2nd resolution level due to an odd number of data length being encountered. It is not possible to discard some data points at either end or both ends of the original spectral data to meet the 2^P requirement. It is because a significant amount of useful information will be lost from the IR spectrum in removing data. As shown in Eq. 4.2, the IR spectrum is required to extended periodically on both sides before wavelet computation. If the zero padding method is utilized to the extend data length to the next power of 2, side-lobe problem will occur in the reconstructed spectrum. The problem occurs when the periodic spectral signal has an abrupt level change within a short interval at the point of zero padding (Morrison, 1994) (Figure 4.8a). Figure 4.8b shows result of the reconstructed IR spectrum of benzoic acid with FWT coupled with zero padding method. It is obvious that side lobes are produced at that position in the reconstructed spectrum (Figure 4.8c-d). These lobes are also observed in the reconstructed IR spectrum when the spectral signal is extended without TRT treatment (Figure 4.9). As pointed out by Daubechies (Daubechies, 1994), there are two things wrong with the use of zero padding method. Firstly, this method would introduce a discontinuity at the point of extension which will be reflected by large wavelet coefficients for fine scales near the two edges. Secondly, this approach uses too many wavelets to represent the original signal. Hence, zero padding method is not a good method to process spectral data via WT when the data length is not equal to the power of 2. The zero padding method also has some effects on data processing with FFT such as distortion of line sharp (Verdun *et al.*, 1988).

Figure 4.8 Periodical extension signal of the IR spectrum of benzoic acid as generated by using the zero padding method (a) with the corresponding reconstructed spectrum (b) and the magnified plots (c) and (d) of the reconstructed spectrum that was produced from the compressed data utilizing D_{16}, $\varepsilon = 0.20$ and $J = 4$.

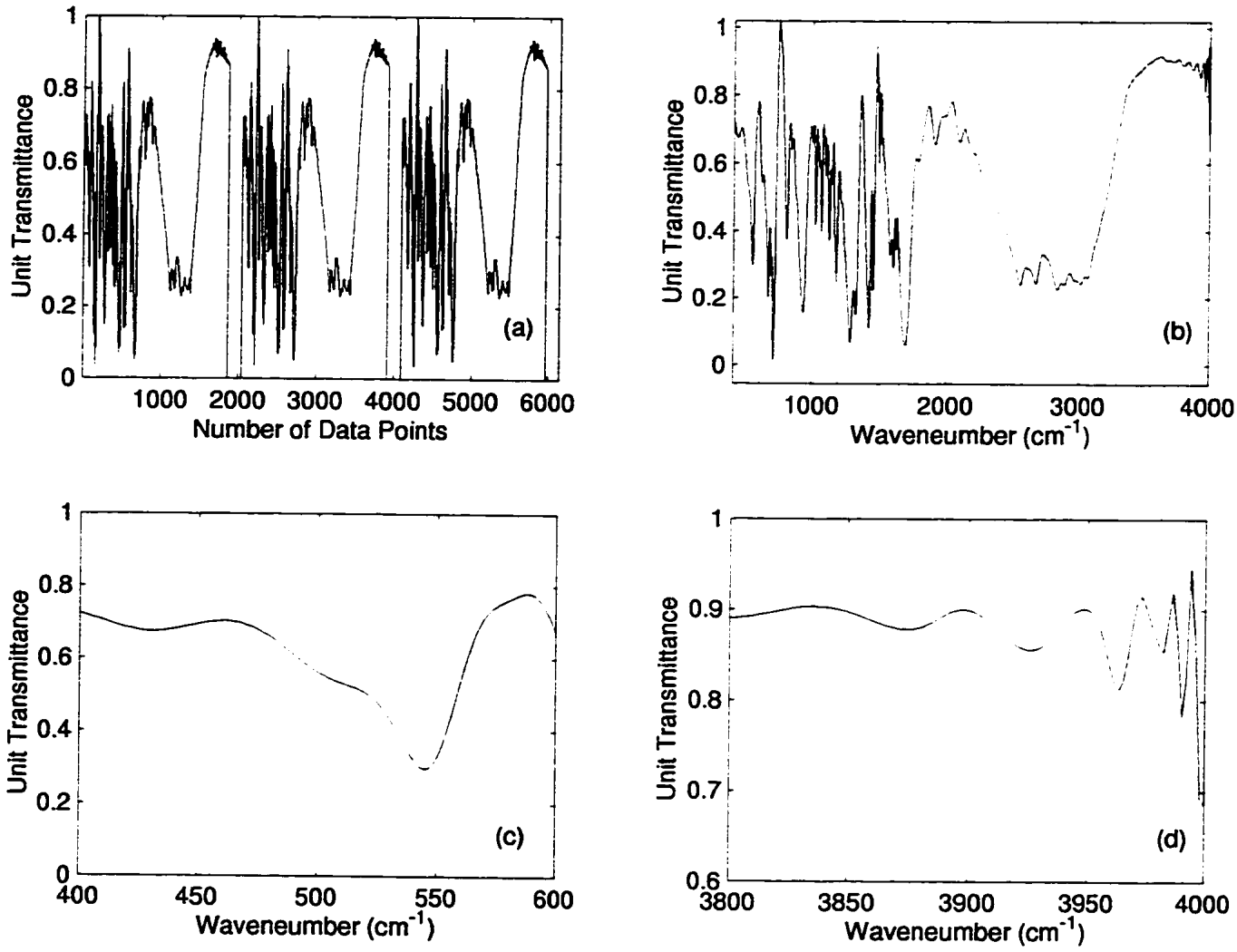
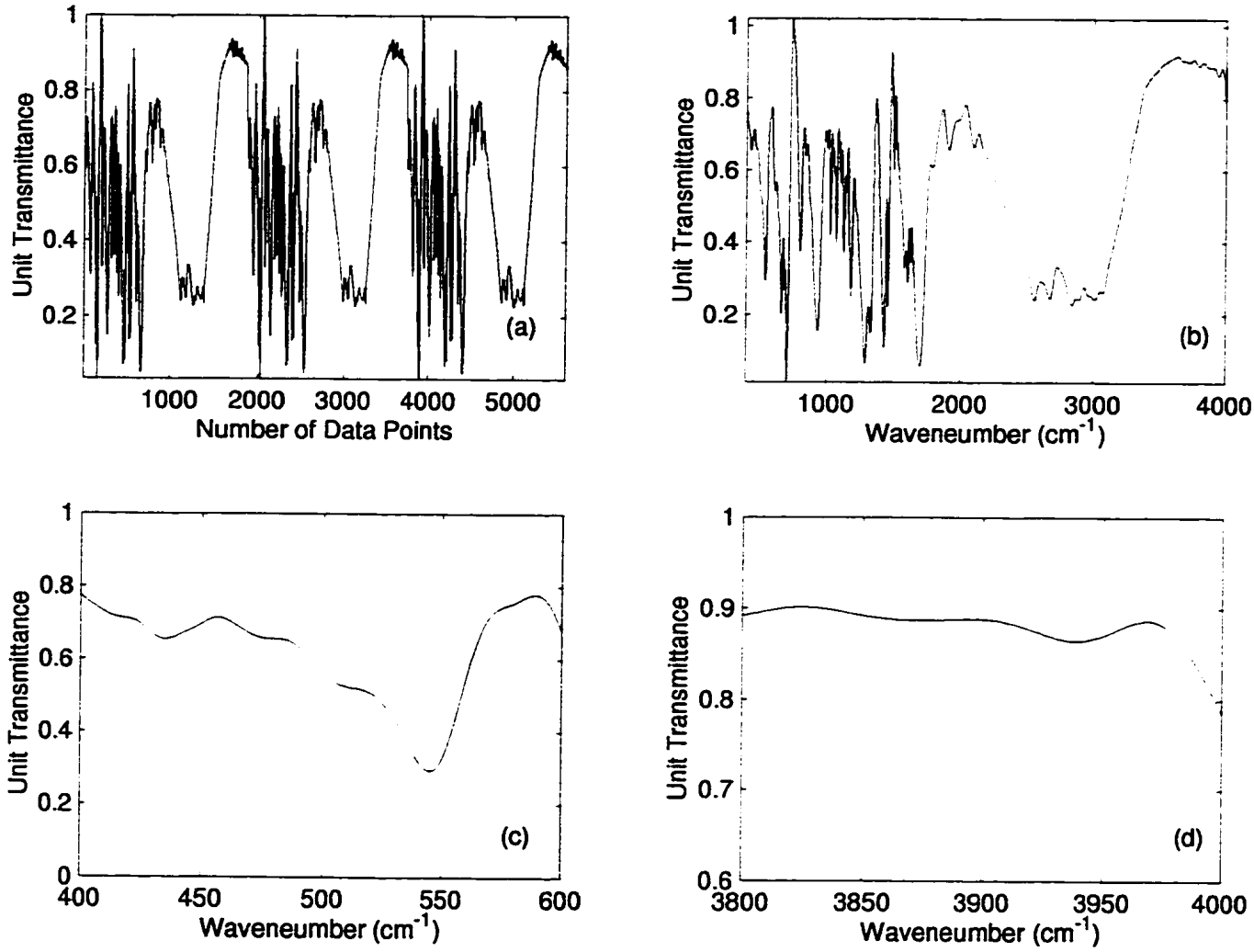


Figure 4.9 Periodical extension signal of the IR spectrum of benzoic acid as generated by using the CPR method without TRT treatment (a) with the corresponding reconstructed spectrum (b) and the magnified plots (c) and (d) of the reconstructed spectrum that was produced from the compressed data utilizing D_{16}, $\varepsilon = 0.20$ and $J = 4$.

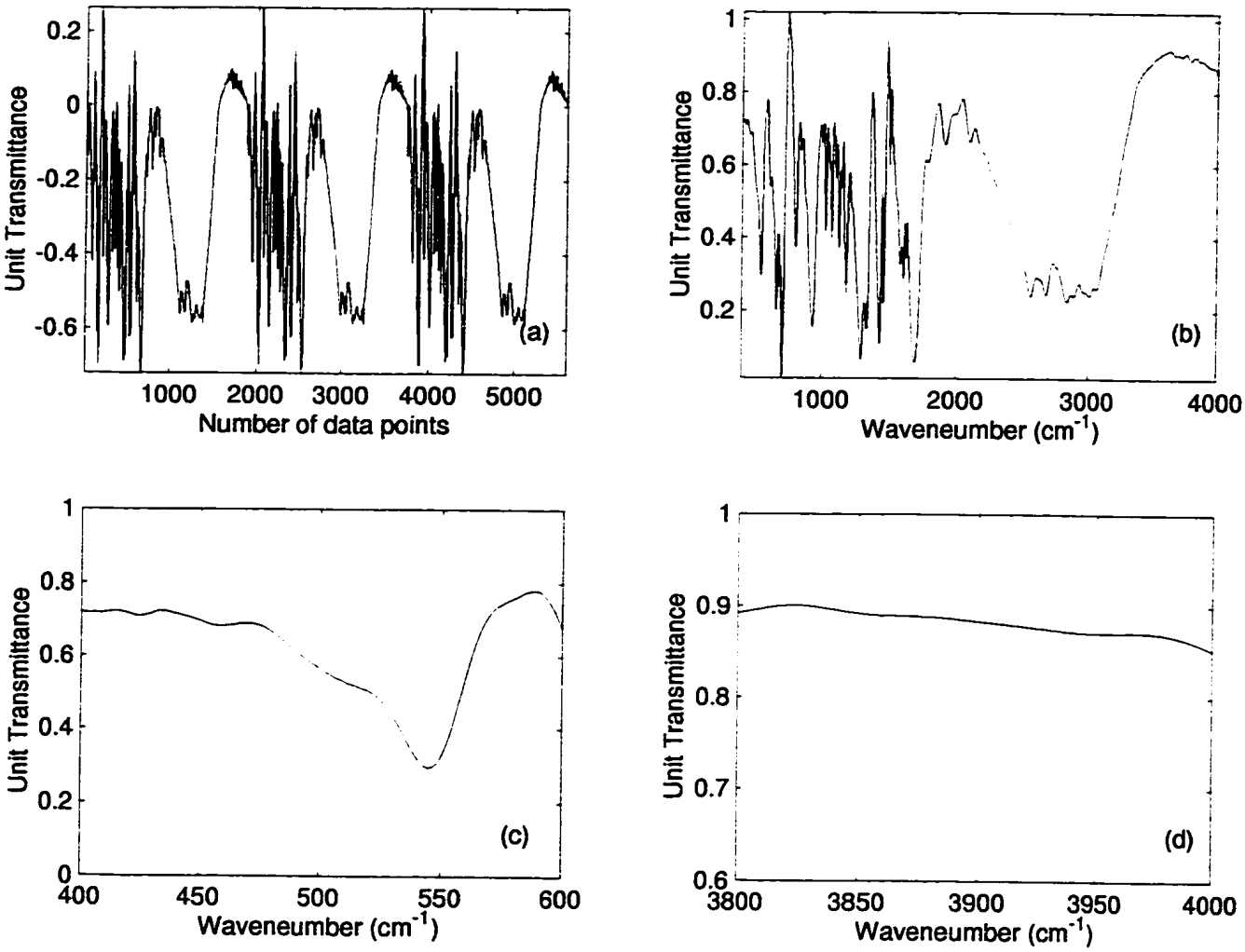


In order to achieve a WT calculation with higher J level, the CPR method was proposed and developed in this study to solve the problem associated with the data set with an odd data number in WT calculation. By comparing the CPR method with the zero padding method, it has two major advantages. Firstly, the former one can process spectral data with any data length. If the length is an odd number, WT calculation can still be performed. The CPR treatment can guarantee that the total data length is unchanged and will not affect the quality of the reconstructed spectrum. It is because data points at the ends of an IR spectrum usually represent the signal background and do not contain any importance information. Secondly, the CPR method does not involve any modification in the original spectral data compared with the zero padding method. Therefore, it can generate a smooth periodic spectral signal for the WT-TRT treatment (Figure 4.10). Hence, the side-lobe problem can be solved.

4.4.4 Setting up of A Small Size Spectral Library

After manipulating individual IR spectra through FWT and WPT, the compressed data can be used to set up a spectral library for future searching. In this work, a small scale spectral library was generated for testing the proposed algorithm. The spectral library as generated by the SLCS software as developed in this work consists of four major types of data. They include the intensity data of a spectrum which is a combination of the suitable bases in the wavelet domain, the wavelength data of each intensity data point, the required data $c_1^{(0)}$ and $c_N^{(0)}$ for ITRT calculation and the compound name. If the spectral library is built up from WPT, extra space is

Figure 4.10 Periodical extension signal of the IR spectrum of benzoic acid as generated by using the CPR-TRT method (a) with the corresponding reconstructed spectrum (b) and the magnified plots (c) and (d) of the reconstructed spectrum that was produced from the compressed data utilizing D_{16}, $\varepsilon = 0.20$ and $J = 4$.



needed for storing the best basis information for individual IR spectrum. Table 4.4 list the compressed file sizes, compression ratios and the computational times for setting up a spectral library containing IR spectra of the twenty selected compounds being processed by FWT, WPT and FFT. The results indicates that both FWT and WPT give higher compression ratios than that by FFT. The later method cannot achieve higher compression ratio because coefficients as obtained in the Fourier domain are complex numbers which require double amount of storage space. If a spectrum can be reduced with a high compression ratio by FFT, most of the high frequency signal will be removed and this will affect the quality of the reconstructed spectrum. However, the computational times for both FWT and WPT (Table 4.4) are longer than FFT because 4 and 15 WT iterations were required to process each IR spectrum, respectively. Since only a single data treatment is required to establish a spectral library, the longer computational time is not a big problem. Moreover, in spectral library searching, coefficients represented in either Fourier or wavelet domain can be employed for identification of an unknown spectrum. In the wavelet domain, the IR spectrum can be represented by a smaller number of coefficients when compared to that in the Fourier domain owing to the higher compression ratios using the proposed WT method. So, the searching time in the wavelet domain is anticipated to be shorter. Therefore, spectral compression using FWT or WPT are a good choice.

In compressing IR data, WPT performs better than FWT. In general, the IR spectrum usually has a lot of peaks in the fingerprint region. After a single cycle of WT calculation, a portion of the high frequency signals are preserved at the wavelet coefficients together with the noise (Figure 4.11). In order to extract such high

Table 4.4 Compression ratios and computational times for setting up a spectral library[a] with twenty IR spectra processed by FWT, WPT and FFT respectively.

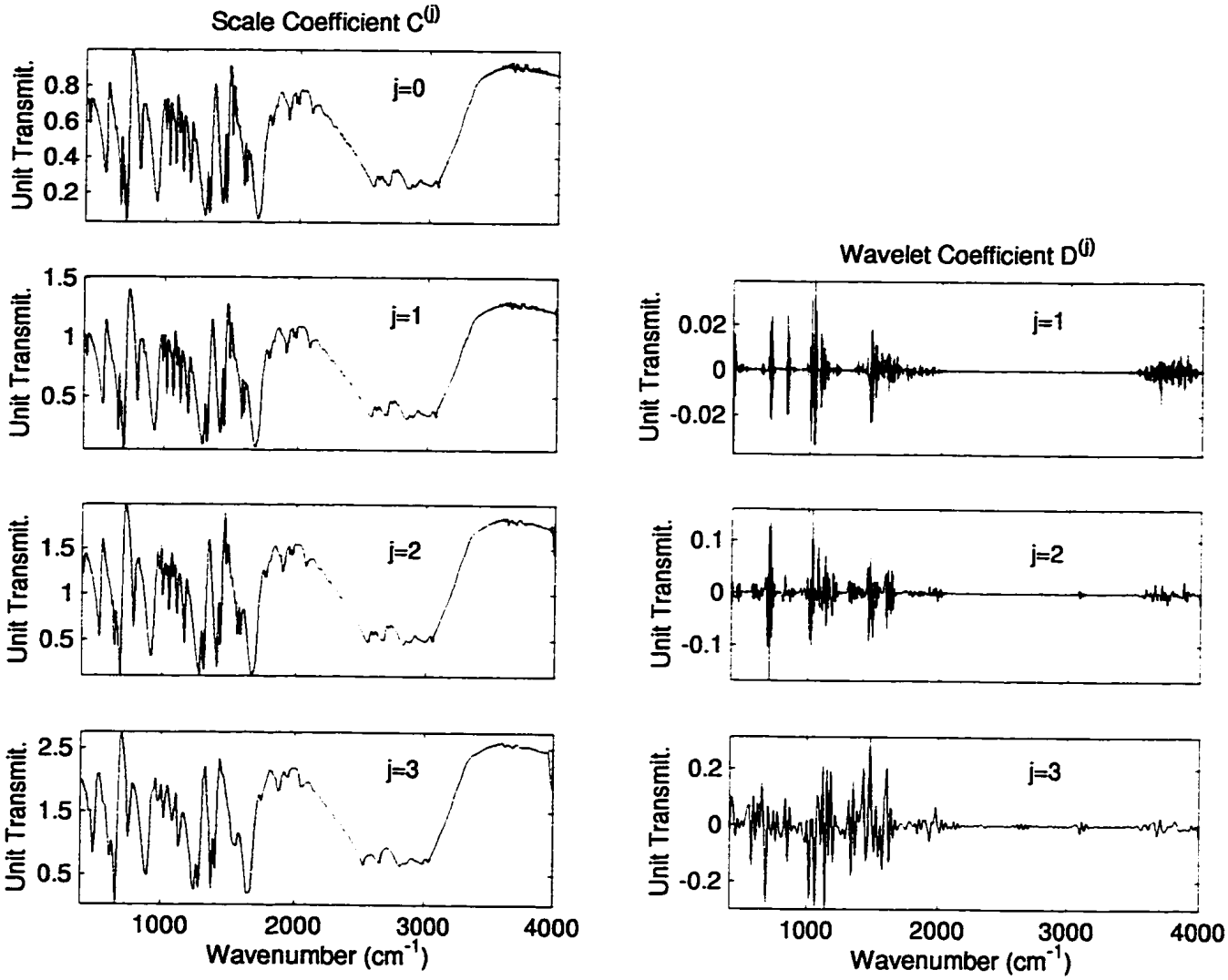
Compression methods	Compressed file size / KB	File compression ratio / %	Total compression time / s	Average compression time / s
FWT[b]	27.3	91.16	24.33	1.22
WPT[b]	30.0	90.30	160.22	8.01
FFT (absolute cutoff method)[c]	67.5	78.14	8.29	0.42
FFT (least-square fitting method)[c]	117.2	62.05	9.89	0.49
FFT (power spectrum method)[c]	229.1	25.81	7.91	0.40

a) The original spectral library size is 308.8 KB in the binary format.

b) Calculation with the use of $\varepsilon = 0.20, J = 4$ and D_{16} function.

c) Calculations were carried out using the FFT built-in function under the MATLAB® environment which is much faster than that by using the script file. So, the FFT computational time will be expected to longer than the listed values if a script file is utilized. Both FWT and WPT were carrying out using the SLCS script files.

Figure 4.11 Plots of the wavenumber against the scale and wavelet coefficients in unit transmittance as obtained from compressing the IR spectrum of benzoic acid at different resolution levels using the D_{16} function.

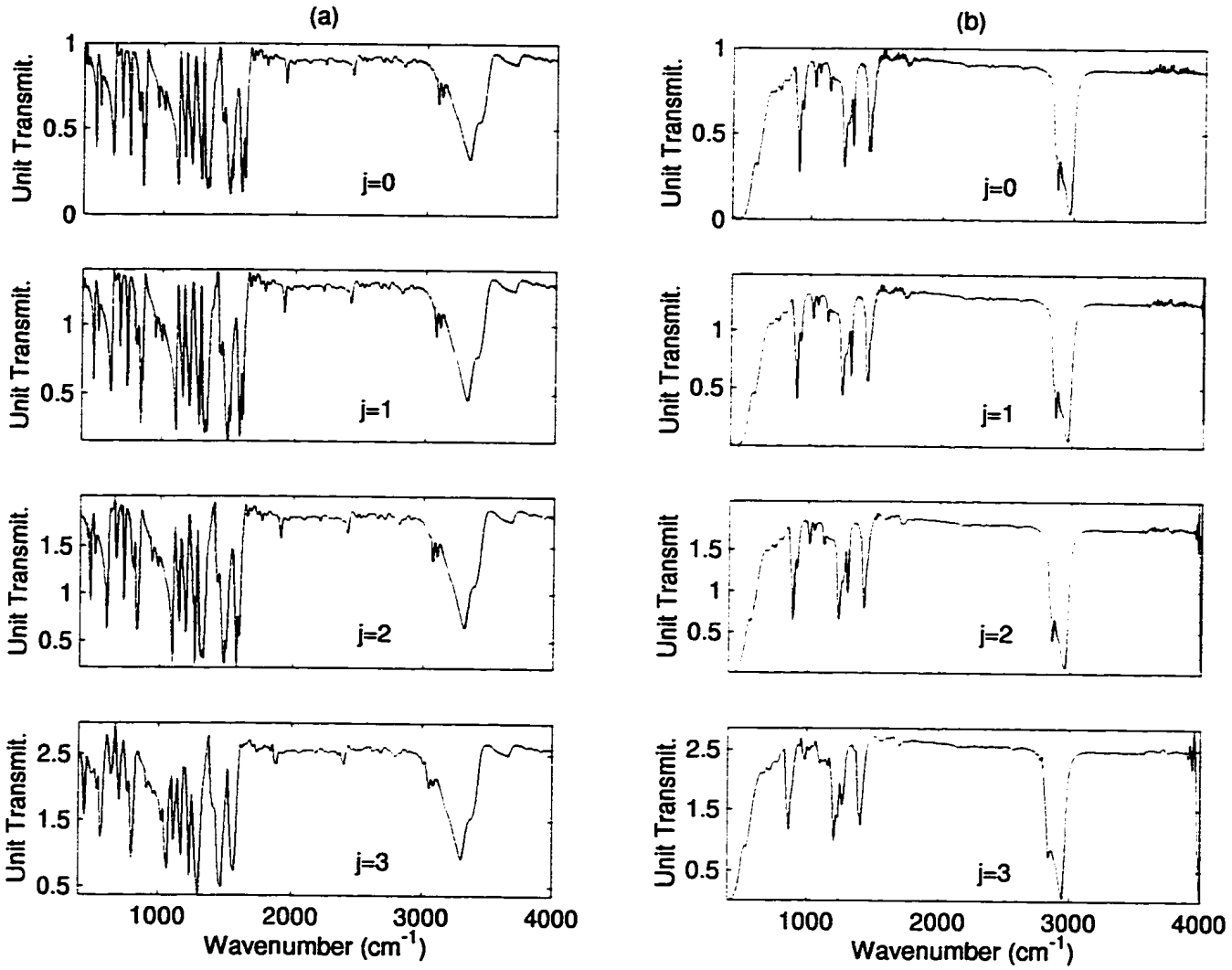


frequency signals, the noise present in $D^{(J)}$ must be removed. In WPT treatment, WT calculation was performed on all bases so that the noise signal can be separated from the required high frequency data. By employing the Shannon-Weaver entropy calculation as mentioned previously, the best basis can be selected out to represent the original signal. However, in FWT treatment, no further WT calculation can be performed on $D^{(1)}$ (Figure 4.1) to extract the required signal. In such case, error would be observed in the reconstructed IR spectrum if a very high cutoff value is employed for noise removal and data compression treatment.

4.4.5 Spectral Library Searching

Spectral compression coupled with prefilter technique is the most common way for searching a large spectral library. By using this technique, a small group of spectra will be extracted from the library for further processing. For example, Lo and Brown (Lo and Brown, 1991) proposed to compress a spectral library by FFT and adopted principal component regression for prefiltering. In this work, the IR spectrum was compressed by WT and represented by a smaller group of scale coefficients $C^{(J)}$ at the assigned resolution level J. Coefficients of $C^{(J)}$ represent the approximation of the original signal $C^{(0)}$ (or $C^{(0,1)}$ in WPT) with a resolution of one point per every 2^J point of the original one. So, most of the characteristic peaks in the original IR spectrum can be retained in this approach. Figures 4.12a and 4.12b depicts, respectively, the scale coefficients $C^{(J)}$ of the IR spectra of 4-nitrophenol and cyclophentyl chloride at various resolution levels. $C^{(J)}$ was utilized as a reference spectrum in the library searching process. In this investigation, each IR

Figure 4.12 A diagram to show the scale coefficient $C^{(j)}$ of the IR spectrum of 4-nitrophenol (a) and cyclophentyl chloride (b) as obtained from resolution levels 0 to 3 via FWT treatment using the D_{16} function.



spectrum in the library was represented by 166 coefficients with the inclusion of 116 coefficients from $C^{(4)}$ and 50 coefficients from $D^{(1)}$ to $D^{(4)}$ after data compression. Up to 90% of the searching time can be saved for using the compressed spectra compared with that for the uncompressed spectra in the direct matching method.

The IR spectrum of 4-nitrophenol was chosen as our unknown one to test the performance of the proposed method for library search. Tables 4.5 and 4.6 show the results of the preliminary and detail search. In the preliminary one, all reference spectra in the library are involved in the direct matching process. After this step, hit indices are generated in descending order to indicate the similarity between reference spectra in the library and the unknown spectrum. Hit index 1 will be assigned to the reference spectrum being found to have the highest D_{corr} and the lowest $D_{absdiff}$ and D_{absder} values from the direct matching calculations, and index 2 for the next highest D_{corr} and lowest $D_{absdiff}$ and D_{absder} values and so on. Thus, a lower hit index value indicates higher similarity between the reference and the unknown spectra. The number of selected IR spectra in the hit index can be adjusted according to the size of the spectral library. In this work, only five hit indices of 1 to 5 of the corresponding reference IR spectra are listed and used in the detail search. Table 4.5 shows the search results of the unknown IR spectrum of 4-nitrophenol of the FWT and WPT compressed libraries. All the three direct matching schemes could identify the unknown spectrum correctly. In the detail search, the suspected reference spectra are reconstructed from the compressed data up to level 0. At this level, the reconstructed spectrum does not exactly represent the spectrum in its original form, but, a smoothed version of the original spectrum. It is because part of the noise signal was

Table 4.5 Results of the preliminary library search[a] of the unknown IR spectrum (spectrum of 4-nitrophenol) from the library being set up using FWT and WPT compressed spectra (see text).

	Correlation coefficient method		Absolute difference method		Absolute derivative method	
Hit index	Compound identified	Value	Compound identified	Value	Compound identified	Value
1	4-nitrophenol	0.9698	4-nitrophenol	5.64	4-nitrophenol	2.24
2	3-nitrophenol	0.7453	3-nitrophenol	11.58	4-nitrotulene	9.64
3	4-nitrotulene	0.6164	1-cholor-2,4-dinitrobenzene	14.36	2-chlorophenol	11.49
4	1-cholor-2,4-dinitrobenzene	0.5712	4-nitrotulene	14.59	1-bromobutane	11.96
5	2-nitrophenol	0.5354	2-nitrophenol	15.72	1-cholor-2,4-dinitrobenzene	12.31

[a] Both FWT and WPT methods give the same result.

Table 4.6 Results of the detail search of the unknown IR spectrum (spectrum of 4-nitrophenol) from the library being reconstructed at resolution level 0 (see text).

Method use	Hit index	Correlation coefficient method		Absolute difference method		Absolute derivative method	
		Compound identified	Value	Compound identified	Value	Compound identified	Value
FWT	1	4-nitrophenol	0.9639	4-nitrophenol	103.26	4-nitrophenol	15.88
	2	3-nitrophenol	0.6868	2-nitrophenol	210.95	1-cholor-2,4-dinitrobenzene	23.93
	3	3-chlorophenol	0.5637	4-nitrotulene	211.84	3-nitrophenol	25.20
	4	4-nitrotulene	0.5611	3-nitrophenol	222.91	1-bromobutane	25.36
	5	1-cholor-2,4-dinitrobenzene	0.5534	2,4-dichloro-phenol	252.68	4-nitrobenzyl chloride	25.47

(cont.)

(cont. Table 4.6)

WPT	1	4-nitrophenol	0.9643	4-nitrophenol	103.18	4-nitrophenol	15.66
	2	3-nitrophenol	0.6864	2-nitrophenol	211.21	1-cholor-2,4-dinitrobenzene	23.93
	3	3-chlorophenol	0.5641	4-nitrotulene	211.66	3-nitrophenol	24.96
	4	4-nitrotulene	0.5600	3-nitrophenol	222.75	1-bromobutane	25.37
	5	1-cholor-2,4-dinitrobenzene	0.5534	2,4-dichloro-phenol	252.64	4-nitrobenzyl chloride	25.55

removed in the WT compression process. The overall results for the preliminary and detail search are slightly different. It is because the resolution of the reconstructed spectra increases for lower resolution levels. Thus, more information is contained in the reconstructed spectrum. Since it is very time consuming to carry out detail search on a huge spectral library with spectra in their original form, the preliminary search provides a fast way to find out a group of reference spectra with high similarity. Then, the selected spectra are reconstructed via either inverse FWT or inverse WPT depending on the compression method used for further confirmation. By adopting this approach, an unknown IR spectrum can be identified within very short time compared to the conventional methods such as full spectrum searching.

Library search by using FFT was also performed in this work for comparison. Similar to that in WT treatment, the IR spectral library was first compressed by FFT with the use of absolute cutoff, least-square fitting and power spectrum methods. After FFT transform, the IR spectrum was converted to a frequency spectrum. Most of the useful information is located at both ends of the frequency spectrum. By using the above mentioned methods, some coefficients in the central part of the frequency spectrum were replaced by 0. Due to the symmetric properties of Fourier transform (Eqs. 2.4 and 2.5), only the non-zero coefficients at the first half of the frequency spectrum were stored in the library. Then, the non-zero coefficients in the Fourier domain were employed for library searching. In this study, the whole frequency spectrum in the Fourier domain was involved for library search. Table 4.7 gives the results of library search by using FFT. The results obtained indicates that coefficients from FFT coupled with absolute cutoff method is not good for spectral searching. Although the absolute cutoff method by discarding Fourier coefficients with values

Table 4.7 Results of the preliminary search of the unknown IR spectrum of 4-nitrophenol through the spectral library being compressed by FFT with the use of absolute cutoff, least-square fitting and power spectrum methods.

Method use	Hit index	Correlation coefficient method		Absolute difference method		Absolute derivative method	
		Compound identified	Value	Compound identified	Value	Compound identified	Value
Absolute cutoff method	1	4-nitrophenol	0.9373	1-cholor-2,4-dinitrobenzene	856	1-cholor-2,4-dinitrobenzene	979
	2	3-nitrophenol	0.8215	3-nitrophenol	898	3-nitrotulene	1059
	3	2-nitrophenol	0.7849	1,4-dichloro-benzene	952	1,4-dichloro-benzene	1125
	4	benzonic acid	0.6898	2-nitrophenol	1000	2-nitrophenol	1219
	5	3-chlorophenol	0.6393	2,4-dinitro-phenol	1016	2,4-dinitro-phenol	1244

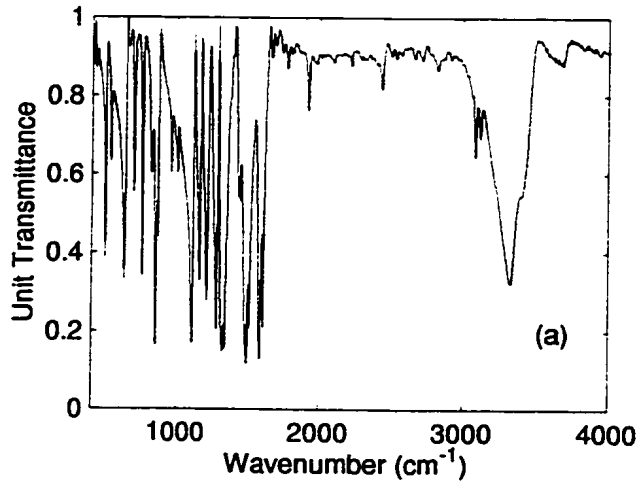
(cont.)

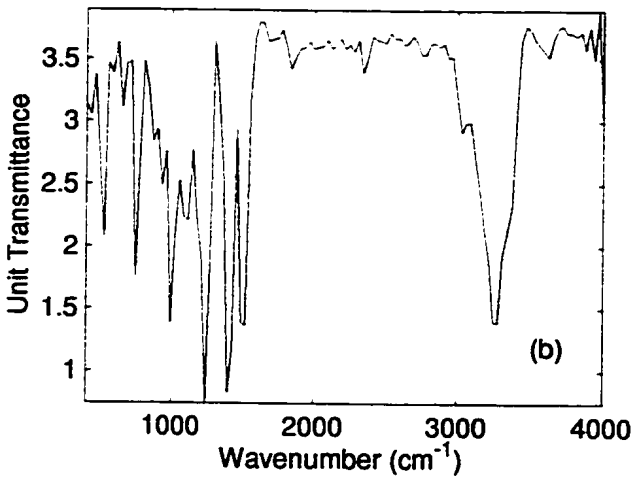
(cont. Table 4.7)

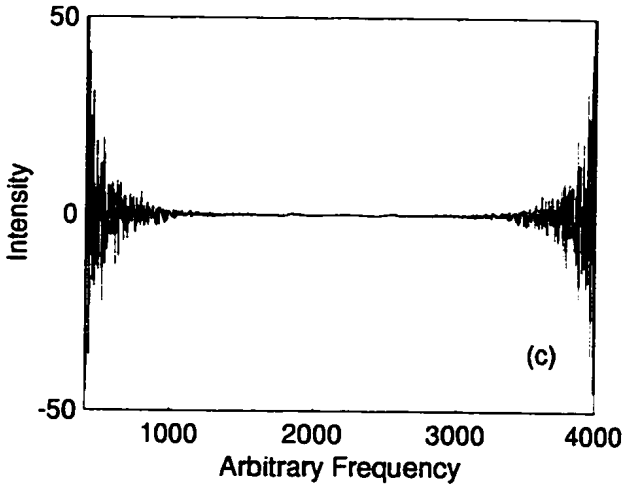
Least square method	1	4-nitrophenol	0.9808	4-nitrophenol	585	4-nitrophenol	657
	2	3-nitrophenol	0.8160	3-nitrophenol	1700	3-nitrophenol	1991
	3	2-nitrophenol	0.7368	1-cholor-2,4-dinitrobenzene	1818	4-nitrotulene	2070
	4	3-chlorophenol	0.5937	2,4-dinitro-phenol	1843	1-cholor-2,4-dinitrobenzene	2121
	5	4-nitrotulene	0.5868	2-nitrophenol	1845	4-nitrobenzyl chloride	2163
Power spectrum method	1	4-nitrophenol	0.9808	4-nitrophenol	624	4-nitrophenol	672
	2	3-nitrophenol	0 8159	3-nitrophenol	1769	3-nitrophenol	2042
	3	2-nitrophenol	0.7367	1-cholor-2,4-dinitrobenzene	1880	4-nitrotulene	2142
	4	3-chlorophenol	0.5937	2,4-dinitro-phenol	1904	1-cholor-2,4-dinitrobenzene	2171
	5	4-nitrotulene	0.5865	2-nitrophenol	1907	4-nitrobenzyl chloride	2209

less that the cutoff value can compress a spectrum with very high compression ratio, some features are removed in the process. Of course the bad performance may also be due to the use of only the real part of the coefficients in the FFT domain for the two library searches. As a result, the compressed reference spectra do not provide enough information for identifying the unknown IR spectrum correctly. For the other two FFT methods, the search results are quite similar to those from FWT and WPT. But, spectral library treated with FWT and WPT is better than that by FFT especially in visual comparison and memory storage. In some cases, the search routine cannot identify the unknown spectrum correctly for compounds selected in our library. Then, users need to compare the reference and unknown spectra by their own experience. One of the advantages in using both FWT and WPT methods for processing spectral library is that they can provide an easier way for visual comparison. Users can just make use of the scale coefficients at a particular resolution level to do so (e.g. Figures 4.13a and 4.13b) because most of the characteristic peaks in the IR spectrum are still retained in the scale coefficients even in higher resolution levels. In FFT, only the real part of the coefficients are utilized for searching. Figure 4.13c shows the real part of the FFT coefficients of the IR spectrum of 4-nitrophenol. These coefficients forms a frequency spectrum that consists of a lot of peaks. It is not easy to identify different functional groups directly by inspecting the frequency spectrum. Hence, the FFT coefficients are not suitable for visual comparison. In order to perform visual comparison via FFT coefficients, inverse FFT must be applied. With regard to our proposed methods, inverse FWT, inverse WPT and inverse CPR are not necessary in the preliminary search process while they are performed only in the detail search process. Besides, FFT calculation requires extra amount of memory space for data storage because all the FFT coefficients are stored in the form of complex number

Figure 4.13 (a) The IR spectrum of 4-nitrophenol, (b) plot of the scale coefficients as obtained from a FWT treatment on the IR spectrum of 4-nitrophenol using D_{16} at $J = 4$ against the wavenumber, and (c) plot of the Fourier coefficients of the IR spectrum of 4-nitrophenol as generated from FFT against frequency.







while WT coefficients in form of real number. Therefore, spectral library search is more efficient for a library that is set up by either FWT or WPT compression and is better than that using FFT compression.

4.5 Conclusions

Fast wavelet transform and wavelet packet decomposition has been proposed in this work for compressing IR spectra of twenty organic compounds with similar structure that were compressed at the 4th resolution level with the use of the Daubechies wavelet function D_{16} and cutoff value $\varepsilon = 0.20$ by using these WT methods. The results indicate that compression by FWT and WPT can reduce the spectral library file size by 90% which is better than that by FFT. Besides, the coefficient position retaining method was developed in this study to process spectral data with odd number data length, which cannot be handled by conventional FWT and WPT methods.

After data compression, the coefficients obtained were selected and employed to build a small scale spectral library for future searching. Spectral library searching of this database was found to be better than that treated by FFT especially in the aspect of visual comparison in some cases. The scale coefficients as obtained from FWT and WPT can be used effectively for preliminary searching in a large spectral library. Our proposed methods can minimize the search time in the search by using the direct matching method. Since only a limited number of IR spectra were involved in the study, we only point out the advantages of library search for IR spectral database that is treated

by the proposed WT methods. Further experiments are required to determine the optimal parameters for compression of spectra and the performance of library search with WT in a large library system. The proposed method can be acted as a temperate for building a larger wavelet based library system.

CHAPTER 5

Wavelet Transform: A Novel Method for Derivative Calculation in Analytical Chemistry

5.1 Introduction

Derivative calculation is a very powerful technique in analytical chemistry for data analysis owing to its popularity on the apparent higher resolution of the differential data when compared with the original data (Adam, 1995). It can be utilized for resolving spectra, sharpening peaks, eliminating unwanted background interference and carrying out quantitative analysis (Rojas *et al.*, 1988; Talsky, 1994). In analytical studies, applications of the derivative technique can be found in ultraviolet-visible (UV-VIS) spectroscopy (Howell and Hargis, 1994; Talsky, 1994; Hargis *et al.*, 1996), chromatography (Strobel and Heineman, 1989) and potentiometric titration (Jeffery *et al.*, 1989). Although the technique is an useful tool for data analysis, it has a major drawback in increasing the noise level in higher order derivative calculation (Brown, 1992). To perform higher order derivative calculation via numerical differentiation, noise reduction is usually performed between data from each order of successive derivative calculation (Antonov and Stoyanov, 1996). Commonly used methods of this kind include sliding average, weighted average, smoothing by polynomial, and time averaging method (Talsky, 1994). Some researchers also proposed their own algorithm to solve this problem (Antonov and Stoyanov, 1993; Barak, 1995; Antonov and Stoyanov, 1996)

Wavelet transform (WT) has attracted recent interest in applied mathematics for signal and image processing (Wickerhauser, 1994). This new mathematical technique has been demonstrated to be fast in computation with localization and quick decay properties in contrast to some existing popular methods, especially, the fast Fourier

transform (FFT). Since 1989, WT has been proposed for signal processing in chemical studies owing to its efficiency, large number of basis functions available, and high speed in data treatment. In this work, WT is introduced as a novel method for approximate derivative calculation in processing analytical data. The major advantage of the proposed method is that it can perform numerical differentiation and noise reduction in the same calculation and can enhance the signal-to-noise ratio (SNR) for even higher-order derivatives.

5.2 Method of Investigation

5.2.1 Derivative Method

In analytical chemistry, derivative method is mainly used for resolving spectrum, locating peak maxima and minima in a spectrum and detecting inflection points in the titration curve. Since it is very difficult to obtain a mathematical expression or function to represent an analytical signal $f(x)$, the simplest method to perform derivative calculation on $f(x)$ is by numerical differentiation (Hecht, 1990; Meloun *et al.*, 1994). The signal $f(x)$ is usually divided equidistantly on the x-axis and the first derivative of $f(x)$, $f^{(1)}(x)$, is evaluated from the slope of two adjacent points by Eq. 5.1:

$$f^{(1)}\left(\frac{(x_{i+1}+x_i)}{2}\right) \approx \frac{f(x_{i+1})-f(x_i)}{x_{i+1}-x_i} \tag{5.1}$$

where i is a running index from 1 to N-1 and N is the data length of $f(x)$. In general, the average of x_{i+1} and x_i is chosen as the new coordinate on the x-axis to minimize the error of signal shifting. To obtain a higher order derivative $f^{(n)}(x)$, a step by step numerical differentiation is adopted:

$$f^{(n)}\left(\frac{(x_{i+1}+x_i)}{2}\right) \approx \frac{f^{(n-1)}(x_{i+1}) - f^{(n-1)}(x_i)}{x_{i+1}-x_i}; \qquad n \geq 1 \tag{5.2}$$

where n represents the derivative order. Other methods found in chemical studies for derivative computation include the use of Fourier transform (Horlick, 1972) and polynomial filters (Barak, 1995) such as the Savitzky-Golay polynomial filter (Savitzky and Golay, 1964; Steinier *et al.*, 1972; Madden, 1978; Gorry, 1990; Rzhevskii and Mardilovich, 1994). The Savitzky-Golay polynomial filter is widely used in analytical chemistry for numerical derivative computation due to its simplicity in calculation.

A differentiated analytical signal has the following properties after the derivative calculation (Talsky, 1994). (1) Peak maxima or minima give extrema in even derivatives and zero crossing in odd derivatives. (2) Peak maxima produce a positive extrema in the 4nth derivative and a negative extrema in the (4n-2)th derivative (n = 1, 2, 3 ...). (3) Inflection points give zero crossing in even derivatives and extrema in odd derivatives. (4) As the derivative order increases, the number of extrema exceeds that of the fundamental curve because each inflection gives an additional extremum by differentiation. As a result, virtual extrema or satellites are observed and they may superimpose extrema of neighbouring analytical peaks in some cases. (5) As the derivative order increases, the full width at half maximum

(FWHM) becomes smaller and the peak becomes sharper (Fell, 1980). (6) Finally, the SNR depends on the width or the number of points in the FWHM of the peak (O'Haver and Begley, 1981) with

$$SNR(n) \approx \frac{C(n)}{FWHM^n} \tag{5.3}$$

where $C(n)$ and n denote a constant and the derivative order respectively. In order to improve the SNR for higher-order derivatives, digital signal smoothing is required to perform between successive derivative order calculations. Commonly used methods include running averaging (O'Haver, 1991), least-squares polynomial smoothing (Willson and Edwards, 1976), spline (Losev, 1990) and Fourier transform (Lee and Wade, 1994). In this work, wavelet transform (WT) is introduced as a new tool for approximate derivative calculation.

5.2.2 Wavelet Transform

As mentioned in the previous chapters, wavelet transform is a very powerful method to break down a signal into simpler, fixed building blocks at different scales and positions (Dai *et al.*, 1994). In WT, a signal $f(x)$ is transformed linearly from its original domain to wavelet domain. The process is similar to apply the signal with Fourier transform (FT). However, in Fourier analysis, only sine and cosine functions can be adopted which are localized in the frequency domain only. It has difficulty in processing a function having components that localized in the time domain (Cody, 1992). As a result, a small frequency change in FT produces changes everywhere in the time domain. Such difficulty can be overcome by wavelet functions, which are

localized both in frequency or scale and in time via dilations and translations of the mother wavelet respectively. This leads to a compact representation of a large classes of functions and operators in the wavelet domain.

Different wavelet functions such as Daubechies wavelet, Meyer wavelet, Coiflet wavelet, spline wavelet, orthogonal wavelet and local cosine basis and others (Chui, 1992; Daubechies, 1992) were proposed by various workers. Different bases face different tradeoffs such as how compactly these functions are localized in space and how smooth they are. In chemical studies, both the Daubechies and spline wavelet functions have been widely used for data compression and smoothing. Here, the Daubechies family of wavelets D_{2m}, with m being any positive integer from 1 to 10, were chosen for approximate derivative calculation. The approximate first derivative of $f(x)$ is assigned to be equal to the difference between two scale coefficients at the first resolution level (Eq. (5.4)) as

$$f^{(1)}(x) \approx C^{(1)}_{D_{2m}} - C^{(1)}_{D_{2\tilde{m}}} \qquad m \neq \tilde{m} \tag{5.4}$$

where D_{2m} and $D_{2\tilde{m}}$ represent any two Daubechies wavelet functions with m and $\tilde{m}$ being any positive integer. Any higher order derivative computation can be achieved by using the result obtained from the lower derivative calculation as an input for WT calculation. Then Eq. (5.4) is applied again to determine the approximate derivative at the next higher-order. To generalize the approximate derivative calculation, such iteration process can be expressed as follow:

$$f^{(n)}(x) \approx C^{(1)}_{D_{2m,n}} - C^{(1)}_{D_{2\tilde{m},n}} \qquad m \neq \tilde{m} \text{ and } n \geq 1 \tag{5.5}$$

with $C^{(1)}_{D_{2m,n}}$ and $C^{(1)}_{D_{2\bar{m},n}}$ obtained from WT treatment of $f^{(n-1)}(x)$ at the first resolution level with Daubechies wavelet functions D_{2m} and $D_{2\bar{m}}$.

5.3 Experimental

All data simulations and computations were carried out by using the wavelet toolbox, WTDeriv as developed in this study. It was coded in MATLAB® for Windows Version 5.1 (The MathWorks, Inc., 1996) under Microsoft® Windows™ 95 environment on an AST® Advantage!® 818 PC (AST Research, Inc., California, U.S.) with a 133MHz Pentium® processor.

In order to test the performance of the proposed algorithm, it has been applied to process a chromatogram for a real sample. The chromatogram were acquired on a HP 1050 HPLC-DAD system equipped with a 20 μL sample loop and a HP LiChrosomb (200 × 4.6 mm i.d., 5μm) C_{18} RP column (Hewlett-Packard, Germany). The mobile phase was a mixture of methanol-water (30:70 v/v). The mobile phase was adjusted to such ratio so that partial overlapping in the chromatograms were observed. The flow rate was 0.5 mL/min. The spectra were collected in the wavelength range from 250 nm to 600 nm with a spectral resolution of 2 nm. The obtained data sets were converted to ASCII text files with a self developed macro program, 3DATA.MAC (Appendix 5.1), under the HP ChemStation Revision A.03.01 environment (Hewlett-Packard, Germany). The macro can extract a particular range of spectro-chromatogram for further data analysis.

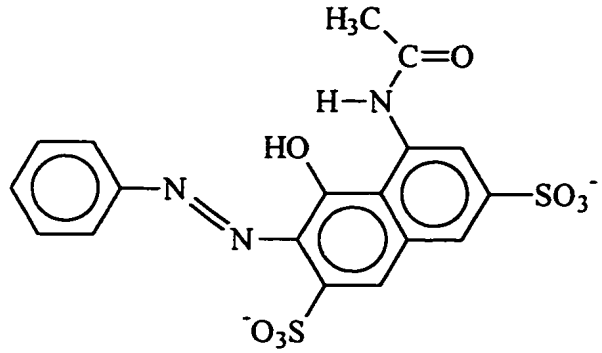
All solvent were of HPLC grade quality (Lab-Scan, Ireland) and double de-ionized water (Millipore-Waters, MA, USA) was used. Food dye samples of Red 2G (λ_{max} = 532 nm) and Allura Red (λ_{max} = 502 nm) (Figure 5.1), which were provided by the Hong Kong Government Laboratory, were used in this studies. A two-component sample solution consisted of 20 ppm of Red 2G and 20 ppm of Allura Red was prepared by dissolving directly in the mobile phase. 10 μL of sample solution was injected into the HPLC-DAD system for analysis.

5.4. Results and Discussion

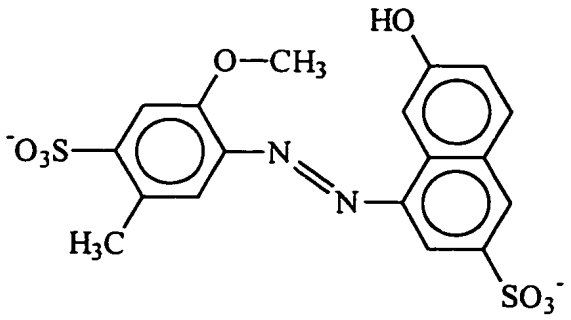
5.4.1 Optimum Daubechies Wavelet Functions Selection

In this work, the proposed method for derivative calculation was tested with different simulated signals such as Gaussian function (Meyer, 1994), Lorentzian function (Fell, 1980) and sigmoid function (Zupan and Gasteiger, 1993). In order to apply fast wavelet transform (FWT) successfully on the analytical data $f(x)$, they needed to be extended periodically at the two extremes. However, in the real situation, the start and end points of the signal do not have a common value. As a result, side-lobe problem will be encountered and affect the signal reconstruction and derivative calculation. To solve such a problem, the translation-rotation transformation (TRT) method was adopted (Chau *et al.*, 1996; Hayes *et al.*, 1973). The TRT algorithm involves subtraction of the data vector $\boldsymbol{C}^{(0)}\left(= f(x)\right)$, with data length N, by selected quantities $\boldsymbol{B}\left\{= b_1, b_2, \ldots, b_k\right\}$ to give the rotated array by

Figure 5.1 Chemical structure of (a) Red 2G and (b) Allura Red.



(a)



(b)

$$c_{k,TRT}^{(0)} = c_k^{(0)} - b_k \tag{5.6}$$

with

$$b_k = c_1^{(0)} + \frac{\left(c_N^{(0)} - c_1^{(0)}\right)(k-1)}{N-1} \tag{5.7}$$

where k is a running index from 1 to N. Besides, FWT was developed to process data with a length of N equals to 2^p with p being any positive integer. In practice, it is not easy for a chemical instrument to generate 2^p data exactly. To cope with this problem, the coefficient position retaining (CPR) method as described in chapter 4 was employed in the wavelet calculations.

In order to apply Eq. (5.5) successfully for approximate derivative calculation, we needed to choose two different Daubechies wavelet functions. It was found that if we compare $C_{D_{2m}}^{(1)}$ and $C_{D_{2\tilde{m}}}^{(1)}$, which were computed from two different Daubechies wavelet functions D_{2m} and $D_{2\tilde{m}}$ respectively, they had a similar shape and magnitude in the y-axes and a very small spatial shift was observed in the x-axes. These shifts of the two signals that followed by the subtraction process gives an approximate first derivative of the original signal (Talsky, 1994). Figures 5.2b-5.2e shows the spatial shifts of the scale coefficients $C_{D_{2m}}^{(1)}$ which were obtained by processing the Gaussian signal (Figure 5.2a) with D_{2m}, $m = 1,2,\ldots,10$. The peak center $\left(P_C\right)$, peak height $\left(P_H\right)$ and peak width $\left(P_\sigma\right)$ have values of 50, 1, and 4 respectively. Daubechies wavelet functions with m greater than 11 is not accurate and are not commonly used in signal processing. Table 5.1 shows the percentage shifts of the peak maxima and their associated values in the y-axis as compared with the original Gaussian function (Figure 5.2a). In most cases, the shifts were not greater

Figure 5.2 (a) A simulated Gaussian signal with 1000 data points which was generated by using the peak center (P_C), peak height (P_H) and peak width (P_σ) having values of 50, 1 and 4 respectively for WT derivative calculation. (b) A spatial delay plot which was obtained by applying the TRT method and the Daubechies wavelet functions D_{2m} with $m = 1,2,\ldots,10$ to the Gaussian signal as shown in (a). The lower part shows the magnified spatial delay plot (c) in the positive slope region, (d) near the peak maxima, and (e) in the negative slope region.

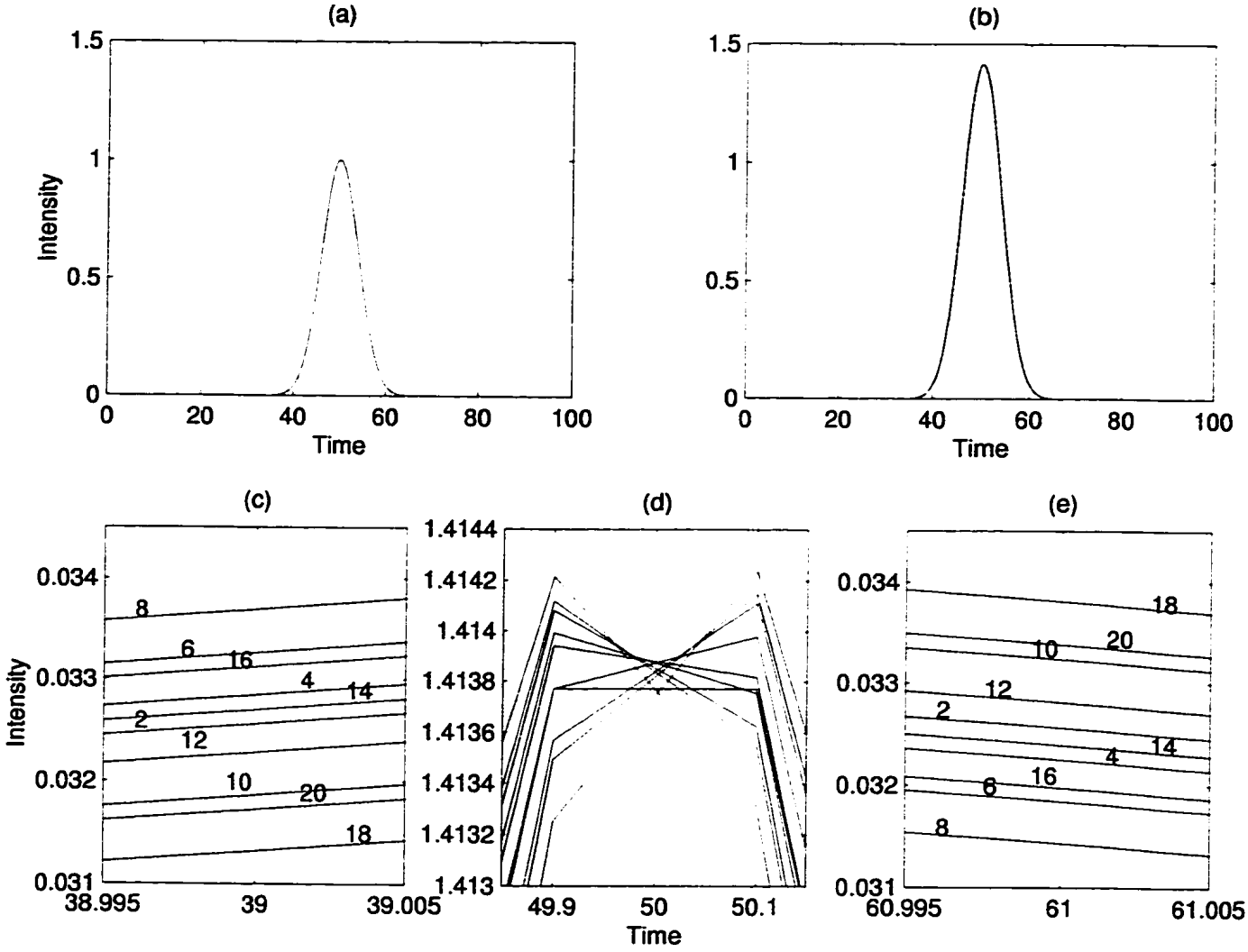


Table 5.1 Percentage of the shift of the peak maximum of the Gaussian signal as shown in Figure 5.1(b) with WT treatment when compared with the Gaussian peak (Figure 5.1(a)). The x-axes were re-scaled to that of the original signal for comparison.

Daubechies wavelet function used	Position of the peak maximum	Percentage of the shift of the peak maximum	Maximum peak height
The original Gaussian peak (Figure 5.1(a))	50.0		1.0000
D_2	50.0	0.0%	1.4138
D_4	49.9	-0.2%	1.4140
D_6	49.9	-0.2%	1.4141
D_8	49.9	-0.2%	1.4142
D_{10}	50.1	0.2%	1.4141
D_{12}	50.1	0.2%	1.4139
D_{14}	49.9	-0.2%	1.4143
D_{16}	49.9	-0.2%	1.4141
D_{18}	50.1	0.2%	1.4142
D_{20}	50.1	0.2%	1.4142

than 0.2%. In Figures 5.2c and 5.2e, a special pattern of Daubechies wavelet functions were observed in the regions with position and negative slopes. In the positive slope region (Figure 5.2c), the D_{2m} functions were arranged in a special order with $2m$ = 8, 6, 16, 4 ,14, 2, 12, 10, 20 and 18 and in decreasing intensity at a given value on the x-axis. The opposite trend was observed in the negative slope region (Figure 5.2e). Besides, the same pattern was also discerned in the position slope regions when the Lorentzian and the sigmoid functions were used (Figure 5.3c and 5.3f). To perform approximate derivative calculation, we need to choose D_{2m} and $D_{2\bar{m}}$ from the upper and lower sections of Figure 5.2c. For example, in the positive slope region,

$$f^{(1)}(x) \approx C^{(1)}_{D_{16}} - C^{(1)}_{D_2} \tag{5.8}$$

can be adopted to evaluate the first derivative of $f(x)$. In Figures 5.2c, 5.3c and 5.3f, the optimal combination of the Daubechies wavelet functions for the derivative calculation was found to be D_8 and D_{18} which give the largest difference or highest values in the first derivative with higher SNR value. In this approach, Eq. (5.5) can be rewritten as follow:

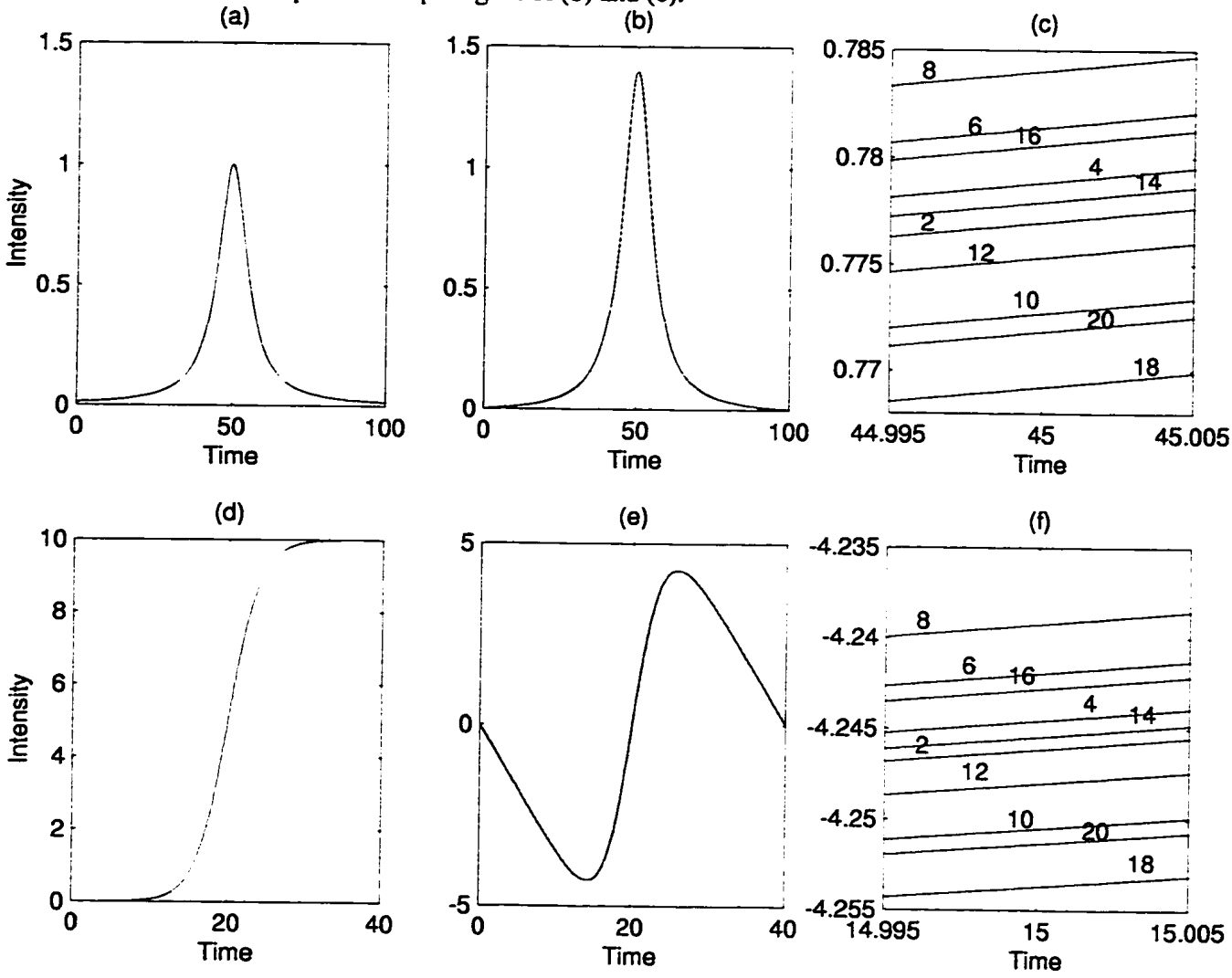
$$f^{(n)}(x) \approx C^{(1)}_{D_{8,n}} - C^{(1)}_{D_{18,n}} \quad \text{and } n \geq 1 \tag{5.9}$$

with $C^{(1)}_{D_{8,n}}$ and $C^{(1)}_{D_{18,n}}$ being obtained from WT for $f^{(n-1)}(x)$ at the lower derivative.

5.4.2 Comparison with the Conventional Derivative Calculation

Traditionally, smoothing is required to perform between each order of derivative calculation because unwanted noise is introduced into the derivative data.

Figure 5.3 (a) A simulated Lorentzian function and (b) a simulated sigmoid function with 1000 data points for WT derivative calculation. The spatial delay plots (b) and (e) were obtained by applying the TRT method and the Daubechies wavelet function D_{2m} with $m = 1,2,\ldots,10$ to the Lorentzian and sigmoid functions respectively with the corresponding magnified spatial delay plots (c) and (f) at the positive slope region of (b) and (e).



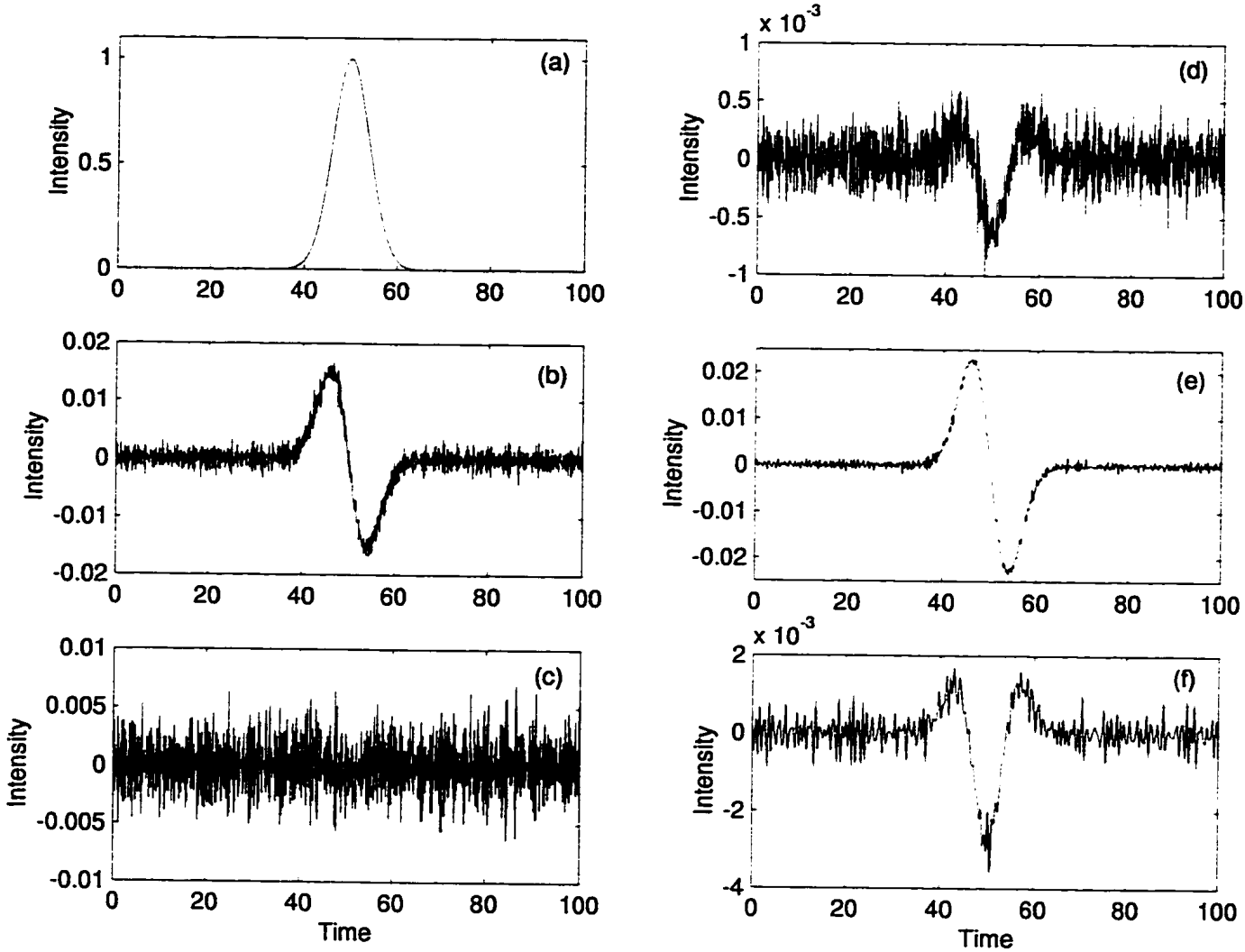
It involves a convolution of the data series with a smoothing function which consists of a set of weighting coefficients to control the increase of noise magnitude or SNR. Various smoothing methods proposed differ only in the way how the coefficients are calculated (Rojas *et al.*, 1988). The Daubechies wavelet function is also a kind of filter. As different values of m are chosen, different sets of weighting coefficients are generated for data smoothing. There are two major advantages by employing WT for approximate derivative calculation. Firstly, both derivative and smoothing calculations can be performed in the same calculation. Secondly, SNR can be improved in higher order derivative calculations as described below.

As stated before, the relative SNR of an unsmoothed derivative signal is inversely proportional to the power n of the width or the number points in the FWHM of the peak (Eq. 5.3). Table 5.2 shows typical values of the relative SNRs of unsmoothed differentiated Gaussian and Lorentzian signals (O'Haver and Begley, 1981). If the width or the number of points in the FWHM of the peak can be reduced, the SNR can be improved. Figures 5.4b-5.4f shows a comparison between the first and second derivatives of a Gaussian signal (Figure 5.4a) with white noise (SNR = 500) from conventional derivative calculation without smoothing and WT treatment with optimum D_8 and D_{18} functions. From the results of conventional derivative calculation, the SNR of the first derivative is about 4.9 (Figure 5.4b). However, if no smoothing is performed on the first derivative data before the second derivative calculation, the SNR of the second derivative is very low and the signal is masked with the noise completely (Figure 5.4c). So, in the conventional derivative calculation, the first derivative data must be denoised or smoothed with a suitable

Table 5.2 A list of $C(n)$ values for derivative orders from 0 to 8 as used for Eq. (5.3) in SNR calculation of the unsmoothed derivative Gaussian and Lorentzian signals.

	Constant $C(n)$	
Derivative order n	Gaussian signal	Lorentzian signal
0	1	1
1	2.02	1.84
2	3.26	4.1
3	8.1	16.6
4	17.7	64
5	52	390
6	141	2204
7	478	17065
8	1675	132275

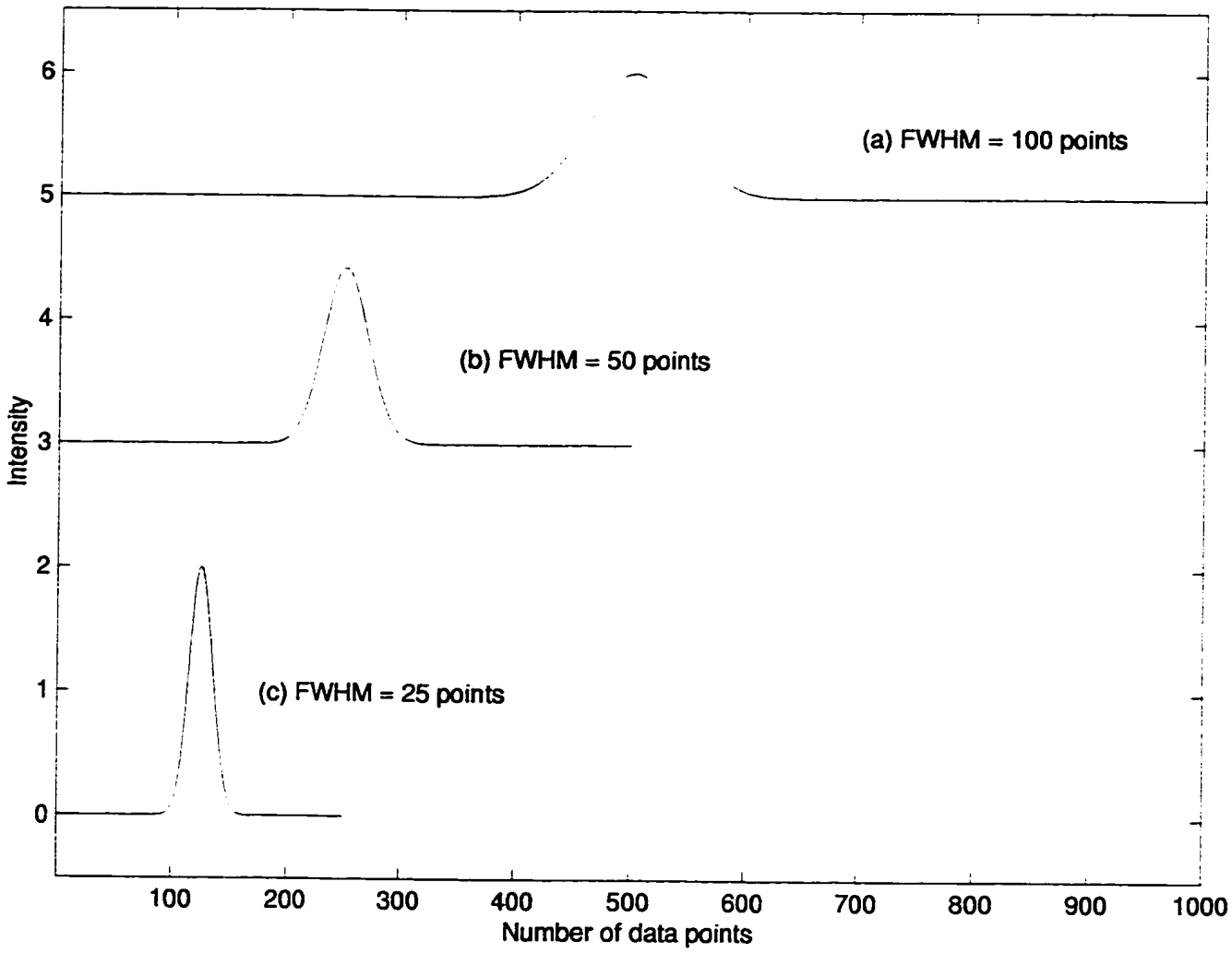
Figure 5.4 (a) A simulated Gaussian signal as shown in Figure 5.1(a) with white noise having a value of 0.001 (SNR = 500) for WT derivative calculation. (b) and (c) show the first and second derivatives plots respectively deduced from the conventional method without smoothing. Plot (d) shows the result of the second derivative derived from the traditional method with smoothing. Plots (e) and (f) show the first and second derivatives obtained from the wavelet transform treatment with the use of the D_8 and D_{18}. functions.



filter prior to the second derivative calculation. Savitzky-Golay polynomial filter and Fourier transform are the denoising filters widely used in chemical studies. But, the former method requires to choose a proper polynomial while the later involves complex number calculation. If the first derivative was smoothed with a Savitzky-Golay 17 points filter, the SNR of the second derivative is about 2.4 (Figure 5.4d). On the other hand, the SNR of the first and second derivatives from our WT derivative calculation are about 11.4 and 3.5 respectively. The x-axes on the first and second derivative plots of Figures 5.4e and 5.4f were re-scaled for visual comparison. When compared with the conventional methods, no complex number was involved and only simple matrix operations were required in our wavelet derivative calculation.

Such an improvement on SNR using WT method over the conventional methods as mentioned above can be explained in view of the number of data points represented for the FWHM of the Gaussian peak adopted in this work (Figures 5.4a). Figures 5.5b and 5.5c show, respectively, the number of points for FWHM at resolution levels 1 and 2 which correspond to two separated WT calculations on the original signal (Figure 5.4a). The FWHM values are reduced by 50% at each WT treatment. Therefore, the SNR in the WT derivative calculation is enhanced. By using WT, the data length of the first derivative is equal to 50% of the original data. It is because the first derivative was derived from the difference of the scale coefficients $C^{(1)}_{D_{1,1}}$ and $C^{(1)}_{D_{1,2}}$ (Figure 5.5b). Their data lengths are equal to 50% of that of the original data. In the second derivative calculation, the coefficients from the first derivative are used as inputs for WT computation. As a result, a further 50% of

Figure 5.5 (a) A simulated Gaussian signal in Figure 5.1(a) with white noise having a value of 0.001 (SNR = 500) for WT derivative calculation. (b) The simulated Gaussian signal which was processed with WT at resolution level 1 with the D_8 wavelet function. (c) The simulated Gaussian signal which was obtain with WT at resolution level 2 with the D_8 wavelet function.



data length is reduced and the data length of the second derivative is equal to 25% of the original data.

Since the proposed WT method was used to perform an approximate derivative calculation in this work, most of the properties in the conventional method are retained. Figure 5.6 shows a comparison between the conventional and wavelet methods for a signal with overlapped peaks. The first derivative obtained from the traditional method was smoothed with the Savitzky-Golay 17 points filter for the second derivative calculation (Figure 5.6c). In the first derivative plots (Figures 5.6b and 5.6d), both methods give the same results on the position of peak maximum and turning point. Moreover, in the second derivative plots (Figures 5.6c and 5.6e), both give similar results on the peak center position. The major difference between the two methods were the SNR on the first and second derivatives and the number of coefficients of each derivative as can be seen from the plots.

Figure 5.7a shows the chromatogram of a mixture of Red 2G and Allura Red at 506.5nm. Figures 5.7b and 5.7c show the first and second derivative plots of the chromatogram by using the proposed approximate WT derivative method. When the results were compared with the Savitzky-Golay 17 points filter method (Figures 5.7f and 5.7g), similar result can be obtained except the resolution of the first and second derivative plots were reduced due to the compression of WT. Besides, the results from the WT are much better than that from the conventional numerical derivative method. As shown in Figures 5.7d and 5.7e, the SNR values of the first and second derivative plots are increased and mask the signals.

Figure 5.6 (a) A simulated signal with 1000 data points which were produced by overlapping peak 1 with $P_C = 50$, $P_H = 1$ and $P_\sigma = 4$ and peak 2 with $P_C = 60$, $P_H = 2$ and $P_\sigma = 4$ with white noise having value of 0.001 (SNR = 500 for peak 1 and SNR = 250 for peak 2 for WT derivative calculation. The first derivative (b) and the second derivative (c) of the signal are obtained by using the conventional method. The first derivative (d) and the second derivative (e) of the signal as obtained by the proposed WT method using D_8 and D_{18}.

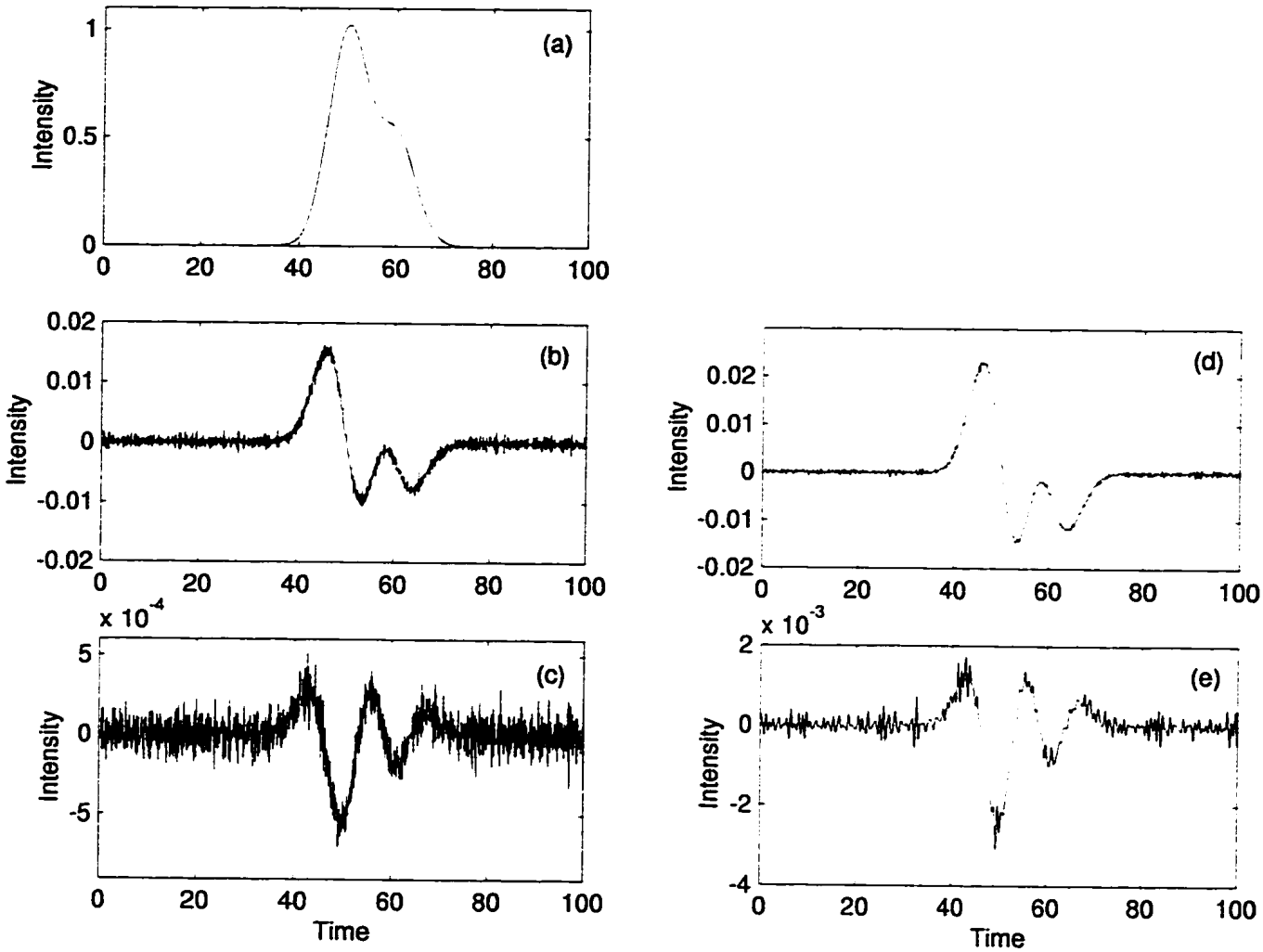
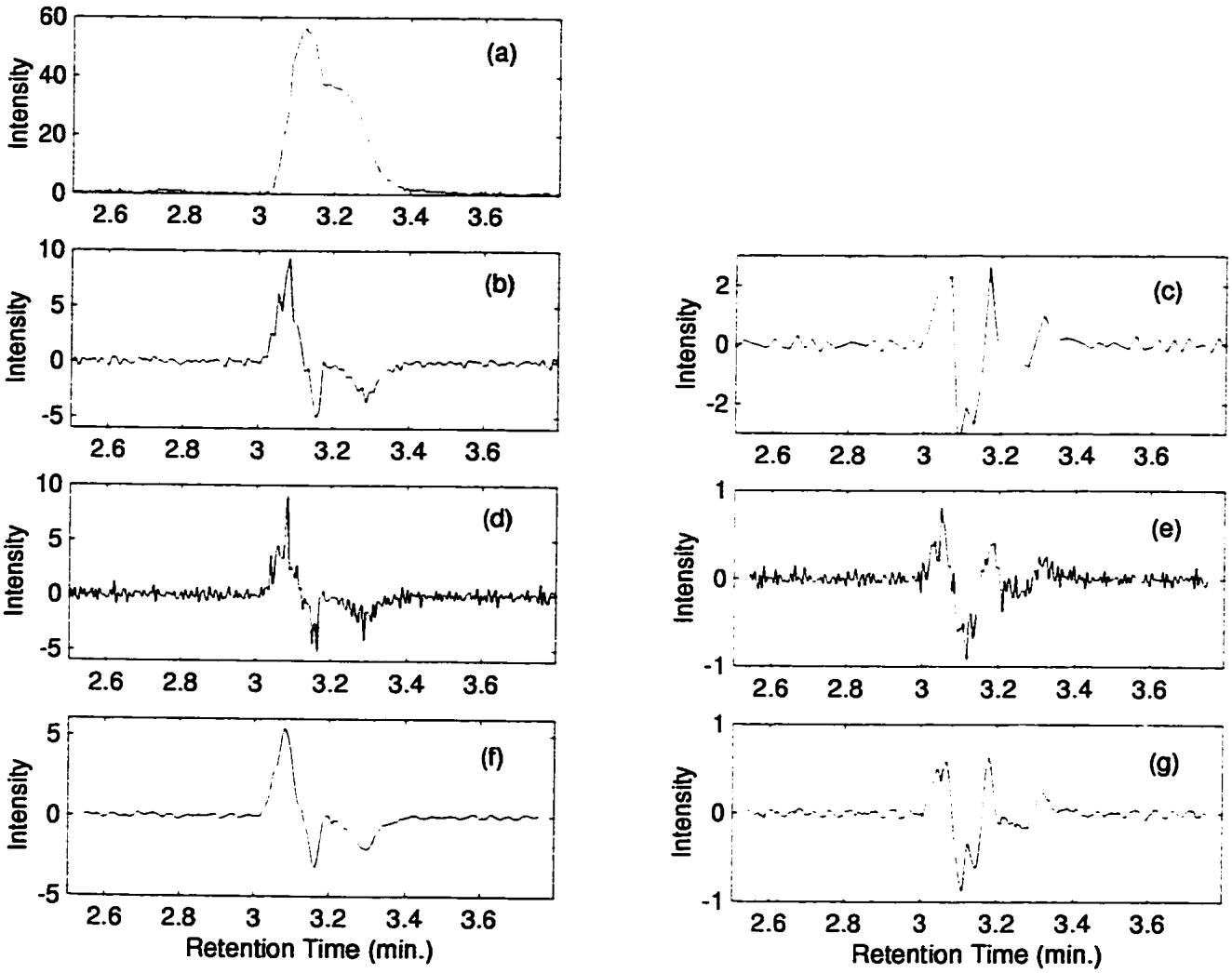


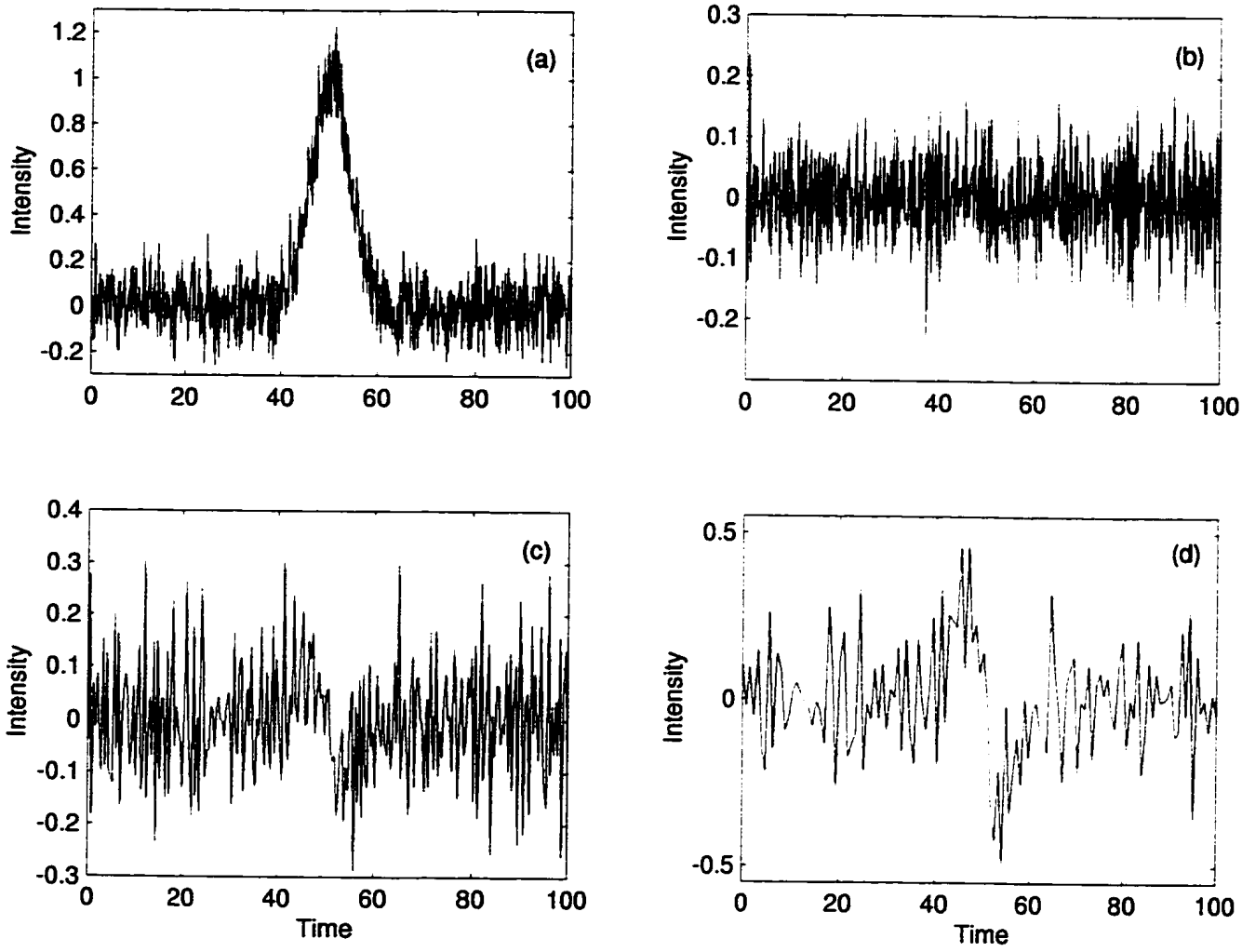
Figure 5.7 (a) An experimental chromatogram of a mixture of Red 2G and Allura Red at 506.5nm. (b) and (c) show the first and second derivatives of the chromatogram by the proposed WT method. (d) and (e) show the first and second derivatives of the chromatogram by the numerical differentiation method. (f) and (g) show the first and second derivative of the chromatogram by the Savitzky-Golay 17 points filter.



5.4.3 Limitations of the Proposed Method

Wavelet transform is a powerful method for approximate derivative calculation as demonstrated in this work. However, the proposed method suffers with some limitations. Firstly, the derivative order depends on the number of data points available in the original signal. If the number of data points are too small, for example 100 points, then only the first two derivatives can be obtained with acceptable results and the higher derivative order calculations are impossible. It is because the number of data point is reduced by 50% for each derivative order computation (Figure 5.5). If the higher order derivative is conducted on this data set, most of the important information will be removed during the computation. In order to tackle this problem, data interpolation such as linear interpolation, polynomial interpolation and spline interpolation methods could be employed to increase the number of data points in each derivative calculation back to the original value (Hecht, 1990; Adams, 1995). However, the interpolation method will inevitably introduce certain amount of noise to the data. Secondly, if the SNR of a signal is very low, the approximate derivative cannot be used with this method (Figure 5.8). In the proposed method, we suggest to use the scale coefficients C at resolution level 1 for derivative calculation. However, certain amount of noise are still retained in C at this resolution level. We cannot obtain the first derivative successfully with the WT method. An alternative approach should be utilized to compute the approximate first derivative. For example, the scale coefficients C at higher resolution level can be employed for the calculation. At the higher resolution level, most of noise signals were isolated and transferred to the wavelet coefficients D.

Figure 5.8 (a) A simulated Gaussian signal in Figure 5.1(a) with white noise having value of 0.1 (SNR = 5) for WT derivative calculation. The first derivatives of the Gaussian signal which were processed with D_8 and D_{18} wavelet functions with WT treatment at resolution levels of 1 (b), 2 (c) and 3 (d) with the number of data points of the first derivatives being 500, 250 and 125 respectively.



5.5 Conclusions

In this work, wavelet transform was proposed as a tool for approximate derivative calculation. The difference between the scale coefficients as obtained from WT treatment at resolution level 1 using the Daubechies wavelet functions D_8 and D_{18} was equal to the approximate first derivative of a particular signal. The result indicated that the new method could enhance the SNR even at higher derivative order calculation without extra effort in data manipulation. Besides, both the derivative and smoothing calculations could be combined in a single step. When compared with the conventional derivative methods, most of the properties that found in these methods are retained in the new method.

CHAPTER 6

Sensitivity Enhancement in Factor Analysis of Two-way Hyphenated Data by Fast Wavelet Transform

6.1 Introduction

Signal processing, one of the important areas in chemometrics, was discussed in previous chapters. From this chapter onward, focus will be put on another area of chemometrics, factor analysis (FA) or principle component analysis (PCA). Factor analysis is a widespread chemometrics technique and has been developed more than 20 years (Brereton, 1995). It is a multivariate methods for reducing data matrices to their lowest dimensionality by the use of the orthogonal factor space and for yielding predictions and recognizable factors through transformations (Malinowski, 1991). Both Auf der Heyde (Auf der Heyde, 1990) and Brereton (Brereton, 1995) have published a tutorial paper on FA.

The rapid development of FA in chemical studies is due to the development of hyphenated instrument in the past decade. There is a great demand in developing advanced instruments to perform qualitative and quantitative analysis in analytical chemistry. With the introduction of hyphenated instruments such as high performance liquid chromatography coupled with diode array detector (HPLC-DAD) and gas chromatography interfaced to Fourier transform infrared spectrometer (GC-FTIR) or mass spectrometer (GC-MS), chemical analysis is changing from two-dimensional to multi-dimensional analysis. The hyphenated techniques provide a suitable mean for multi-component analysis and lead to the development of new methods for data interpretation and evaluation (Toft, 1995). Commonly used methods for such a purpose include evolving factor analysis (EFA) (Maeder and Zilian, 1988; Keller and Massart, 1992a), alternating regression (AR) (Karjalainen, 1989; Karjalainen and Karjalainen,

1996), target factor analysis (TFA) (Gemperline, 1984), iterative target transformation factor analysis (ITTFA) (Vandeginste *et al.*, 1985), target transformation factor analysis (TTFA) (Hopke, 1989), window factor analysis (WFA) (Mailnowski, 1992), fixed size moving window evolving factor analysis (FSMW-EFA) (Keller and Massart, 1991) and heuristic evolving latent projections (HELP) (Kvalheim and Liang, 1992; Liang *et al.*, 1992).

Recently, a mathematical technique called wavelet transform (WT) has been proposed for signal and image processing (Wickerhauser, 1994). One of the main feature of WT is that it may decompose a signal directly according to the frequency and represent it in the time domain. In the transformation, both the time and frequency information of the signal are retained. Since 1989, WT has been applied for signal processing in various fields of analytical chemistry and chemical physics owing to its efficiency, more number of basis functions available, and high speed in data treatment compared to Fourier transform computation. More than 90 papers were published within the period from 1989 to August 1998 as mentioned in Chapter 3.

In this work, fast wavelet transform (FWT) was proposed as a pre-processing step of the HELP algorithm. It was found that if a data set from HPLC-DAD with very low signal-to-noise ratio (SNR), the HELP algorithm failed to determine the minimum number of components and the selective regions for the individual components from the chromatograms of the sample. With the use of FWT algorithm, the data set was transformed with FWT at resolution level 1 on the chromatograms at each wavelength. By using the denoised and compressed data set in the wavelet domain for HELP

analysis, the sensitivity of the HELP algorithm can be improved. Furthermore, the compressed data set can reduce the computational time of the HELP algorithm. Simulated and experimental data were used to evaluate the performance of the proposed method. Details of HELP and FWT algorithms will be discussed in the coming sections.

6.2 Method of Investigation

6.2.1 Principle of Factor Analysis

In factor analysis, data matrices are assumed to be formed as a linear sum of factors with each factor being weighted differently. For instance, experimental data from HPLC-DAD system can be represented by a data matrix $\boldsymbol{X}$ (dimension $i \times k$) and it is considered as the linear sum of the products of the chromatogram $\boldsymbol{C}$ (dimension $i \times j$) and spectrum $\boldsymbol{S}$ (dimension $k \times j$) of the individual component j (Eq. (6.1)) with the noise $\boldsymbol{E}$ (dimension $i \times k$), where the index i denote the retention time (row vector), k the wavelength (column vector), n the total number of components present and t represents the transpose operation.

$$\boldsymbol{X}_{ik} = \sum_{j=1}^{n} \boldsymbol{C}_{ij} \boldsymbol{S}_{kj}^{t} + \boldsymbol{E}_{ik} \tag{6.1}$$

The value of individual $\boldsymbol{X}_{ik}$ is the overall intensity at a particular retention time and wavelength. In principle, FA can be employed to decompose the data matrix $\boldsymbol{X}$ into two abstract matrices. They are called the scores matrix $\boldsymbol{T}$ (dimension $i \times j$) and the loadings matrix $\boldsymbol{P}$ (dimension $k \times j$) which are the row factors and column factors respectively (Eq. (6.2)).

$$X_{ik} = T_{ij,(abstract)} P^{t}_{kj,(abstract)} \tag{6.2}$$

To convert the abstract matrices to information with physical meaning, a transformation matrix R and its inverse R^{-1} are utilized. If the transformed solution can be shown to have physical significance such as chromatograms and spectra in HPLC-DAD data, a real solution to the problem will be found (Eq. (6.3)).

$$X_{ik} = \left(T_{ij,(abstract)} R\right)\left(R^{-1} P^{t}_{kj,(abstract)}\right) = T_{ij,(real)} P^{t}_{kj,(real)} = C_{ij,(pure)} S^{t}_{kj,(pure)} \tag{6.3}$$

After the above calculation, one can determine the number of components n as well as their pure spectra and chromatograms. In the presence of noise, the number of factors that can be deduced from FA is always greater than the true number of constituting factors or components (Malinowski, 1991). It is because the noise E contributes certain number of factors in the data matrix X in the calculation. So, different methods such as EFA, FSMW-EFA and HELP have been proposed to determine the true number of factors.

The basic principle of the EFA, FSMW-EFA and HELP methods is to follow the change or evolution of the rank of the data matrix X as a function of the ordered variable such as retention time (Maeder, 1987). The rank of X is equal to the number of the true components in the sample. In HPLC-DAD, the appearance of each new component at the detector will cause an increase in the rank by one or vice versa. Rank is determined through singular value decomposition (SVD) (Maeder and Zilian, 1988; Lupu and Todor, 1995). SVD can decompose X into a product of three matrices u (dimension $i \times n$), s (dimension $n \times n$) and v (dimension $n \times k$) with the following properties (Eq. (6.4)):

$$X_{ik} = u_{in} s_{nn} v_{nk} \tag{6.4}$$

with $u_{in}^{t} u_{in} = v_{nk} v_{nk}^{t} = I_{nn}$ (identity matrix). u and v are equal to, respectively, the significant eigenvectors $X_{ik}^{t} X_{ik}$ and $X_{ik} X_{ik}^{t}$. s is called the singular value and is a diagonal matrix. The diagonal elements of s are the positive square roots of the corresponding eigenvalues and are arranged in descending order. The magnitude of an eigenvalue indicates the relative importance of that factor. A larger eigenvalue implies that the factor is contributed from a sample component while a smaller one from the noise. The product of u and s equals to the scores T while the transpose of v equals to the loadings P.

6.2.2 Heuristic Evolving Latent Projections

Heuristic evolving latent projections algorithm was proposed by Kvalheim and Liang in 1992 for analyzing data from liquid chromatography with diode array detection (LC-DAD) (Kvalheim and Liang, 1992; Liang *et al.*, 1992). It is classified as a FA or PCA technique for multivariate data analysis. This algorithm is a modification of FSMW-EFA method and was shown to be better than the conventional methods such as EFA, ITTFA and AR (Toft and Kvalheim, 1994; Grung and Kvalheim, 1995). Within the concept of HELP, it has been applied to resolve system containing partial chromatographic selectivity (Grung and Kvalheim, 1995), hidden minor chromatographic peaks (Liang and Kvalheim, 1993b), and baseline drift in the elution profiles (Liang *et al.*, 1993b). HELP has been applied successfully to an LC-UV study involving drug isomers (Liang *et al.*, 1992) and chlorophyll degradation experiments (Liang *et al.*, 1993c), deconvolution of

chromatographic peaks (Hamalainen *et al.*, 1993), assessment of peak origin and purity in one-dimensional chromatography (Liang *et al.*, 1994) and HPLC-DAD study of polycyclic aromatic hydrocarbons (PAHs) content in city air of Hong Kong (Shen *et al.*, 1997, 1998a-b) and others.

In order to apply the HELP algorithm successfully on data analysis of HPLC-DAD, two simple assumptions are made. Firstly, the Beer law is observed with non-zero chromatographic and spectral signal. Secondly, the rule first-in-first-out is obeyed. That is the first substance present in the system will be the first to disappear, the second substance will disappear next and so on (Keller and Massart, 1992a). Suppose the HPLC-DAD system generates a data set $\boldsymbol{X}$ (Figure 6.1) with the rows and columns corresponding to spectra $(\boldsymbol{s}_a)$ measured at different times and chromatograms $(\boldsymbol{c}_a)$ acquired at a different wavelength respectively:

$$\boldsymbol{X} = \sum_{n=1}^{N} \boldsymbol{c}_n \boldsymbol{s}_n^t + \boldsymbol{E} = \boldsymbol{C}\boldsymbol{S}^t + \boldsymbol{E}. \tag{6.5}$$

Here, N and $\boldsymbol{E}$ represents the total number of components and noise in the sample. With the help of PCA treatment, $\boldsymbol{X}$ can be decomposed into a matrix of score $\boldsymbol{T}$ and a matrix of loadings $\boldsymbol{P}$ as follows:

$$\boldsymbol{X} = \boldsymbol{C}\boldsymbol{S}^t + \boldsymbol{E} = \boldsymbol{T}\boldsymbol{P}^t + \boldsymbol{E}. \tag{6.6}$$

In the HELP treatment, the first step involves the construction of the eigenvalues plot (Figure 6.2a) which is a plot of the logarithm of eigenvalues against retention time. Such a plot is obtained by applying FSMW-EFA procedure to $\boldsymbol{X}$ with a particular window size. The window size is defined as the number of rows of data

Figure 6.1 Simulated chromatograms for a two-component system, A (——) and B (– · –), with a peak resolution of 0.37 (a), 0.62 (b) and 0.88 (c) and the corresponding simulated spectra (d).

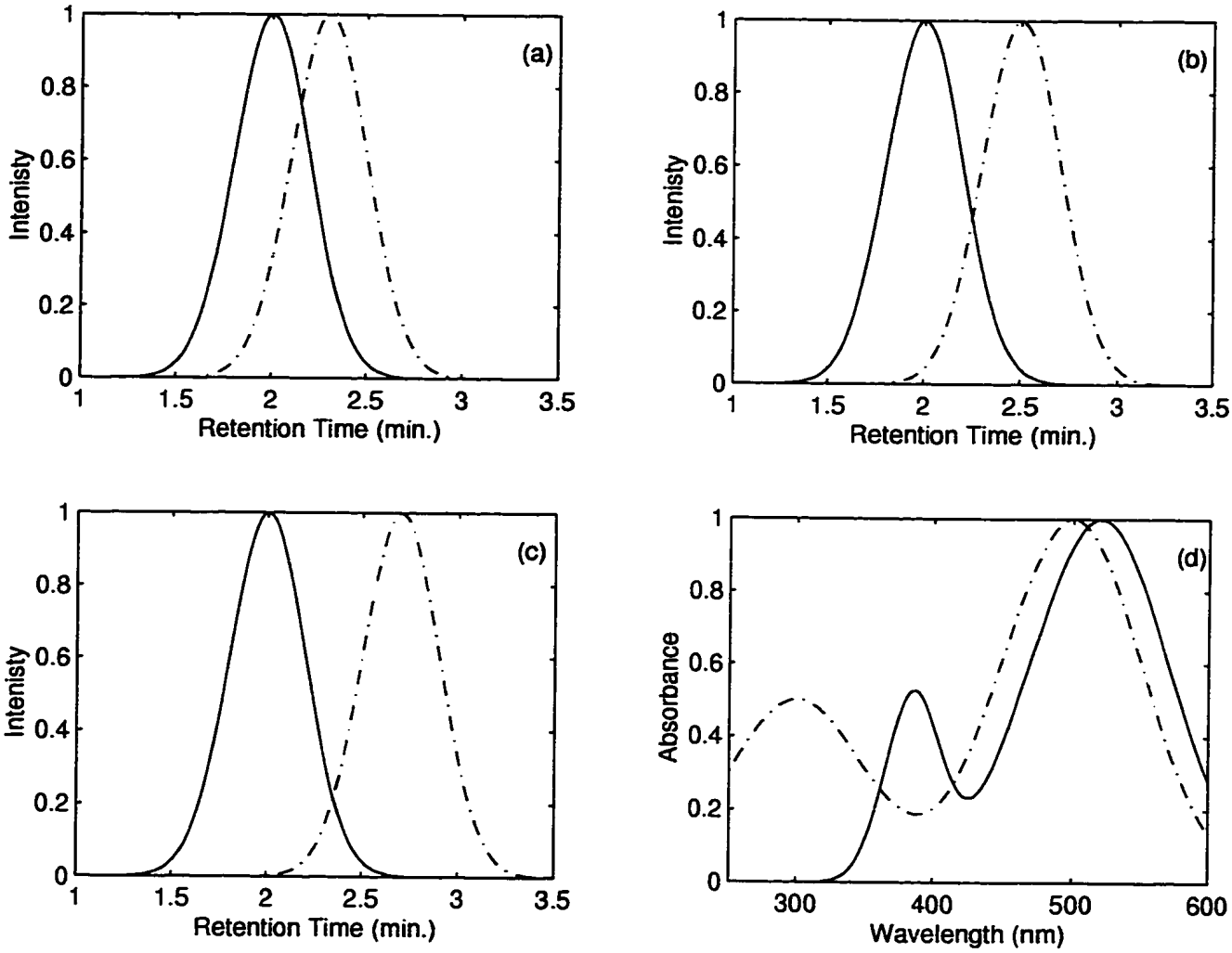
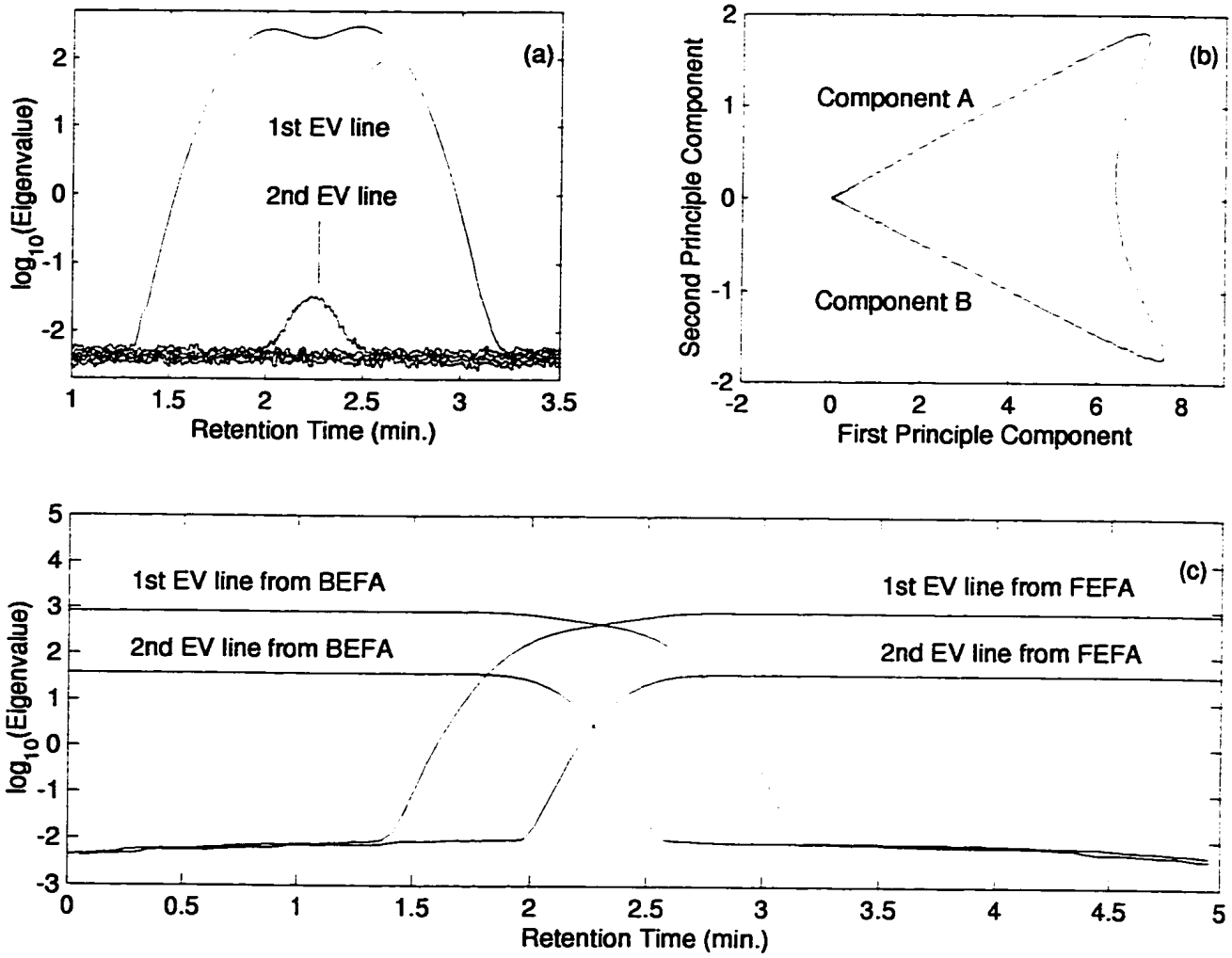


Figure 6.2 (a) Eigenvalue plot obtained from the data set B with a window size of 5. (b) Latent projective graph obtained by plotting the first principle component versus the second principle component of the data set B. (c) Eigenvalue plot obtained from the data set B with evolving factor analysis treatment.



from $\boldsymbol{X}$ for eigenvalues computation. FSMW-EFA is a modified method based on Maeder's EFA algorithm (Keller and Massart, 1991). In EFA algorithm, an expanding window is applied to generate the eigenvalue plot (Figure 6.2c). Firstly, the first row of data from $\boldsymbol{X}$ is selected out to form a sub-matrix and the eigenvalues (EV) are calculated with SVD method. Then, the first two rows of data are extracted and the eigenvalues are obtained again. In this manner, eigenvalues are computed for each addition row of data until the whole data matrix $\boldsymbol{X}$ is extracted. This approach is called the forward evolving factor analysis (FEFA) and represents the left hand side of the eigenvalue plot. The right hand side of the plot is obtained by the backward evolving factor analysis (BEFA) which repeats the EFA calculation on $\boldsymbol{X}$ as described previously but does it in the opposite direction. Instead of adopting an increasing size window to calculate the eigenvalues, FSWM-EFA utilizes a fixed size moving window. An odd number of rows of $\boldsymbol{X}$, which is greater than 1, is selected and their eigenvalues are computed. Then, the window is moved downward by one row and the calculation is repeated until the last row of the matrix $\boldsymbol{X}$ is manipulated. At the end of the computation, all the first eigenvalues as obtained from the above calculations are connected with a line and the same approach is also applied to the other eigenvalues.

In the second step of HELP analysis, the zero concentration, zero component and selective regions of the corresponding components are determined from the eigenvalues plot. The zero concentration region is defined as a chromatographic region where nothing elutes from a particular component (Maeder, 1987; Kvalheim and Liang, 1992). For example, in Figure 6.2a, regions from 2.5 to 3.2 minutes and

from 1.3 to 2.0 minutes are the zero concentration regions for component A and B respectively. On the other hand, the selective region is considered as the region with only one chemical component eluted (Kvalheim and Liang, 1992). Regions from 1.3 to 2.0 minutes and from 2.5 to 3.2 minutes are the selective regions for component A and B respectively (Figure 6.2a). The zero component region is defined as a region with no chemical component involved and is usually contributed from noise of the system. Regions from 0.0 to 1.3 minutes and 3.2 to 3.5 minutes are the zero component regions. In this stage, the latent projective graph (Figure 6.2b) can also provide information on the selective region of individual components. Such regions can be identified from the location of straight lines which pass through the origin in the latent projective graph. For example, the straight line with positive slope represents the selective region for component A while that with negative slope represents the selective region for component B. In the third step, with the information from the zero concentration, zero component and selective regions, chromatograms and spectra of individual components can be resolved through simple matrix computation. In the simplest case such as a simple two components system, if the chromatographic profiles of each components are known, the corresponding spectral profiles can be determined by the following equation:

$$\boldsymbol{S}^{t} = \left(\boldsymbol{C}^{t}\boldsymbol{C}\right)^{-1}\boldsymbol{C}^{t}\boldsymbol{X}. \tag{6.7}$$

Pure chromatogram of a particular component can be determined from the zero concentration region of that component. It can be deduced from the first principle component of a sub-matrix as obtained from the selective region. For example, pure chromatogram of component A can be derived from the SVD on the sub-matrix $\boldsymbol{X}_{1.3-2.0}$ with the subscript being the retention time range of the selective region.

Similarly, the chromatographic profiles can be resolved if the spectral profiles are known:

$$\boldsymbol{C} = \boldsymbol{X}\boldsymbol{S}\left(\boldsymbol{S}^{t}\boldsymbol{S}\right)^{-1}. \tag{6.8}$$

For a more complicated system, the component stripping approach should be used (Liang and Kvalheim, 1993c). Resolution can only be performed by using a stepwise stripping procedure. From Eq. (6.6), we have

$$\boldsymbol{C}\boldsymbol{S}^{t} = \boldsymbol{T}\boldsymbol{P}^{t}. \tag{6.9}$$

Then, a rotation matrix $\boldsymbol{R}$ can be introduced into the above expression so that

$$\boldsymbol{C}\boldsymbol{S}^{t} = \boldsymbol{T}\boldsymbol{R}\boldsymbol{R}^{-1}\boldsymbol{P}^{t}. \tag{6.10}$$

Eq. (6.10) implies that

$$\boldsymbol{C} = \boldsymbol{T}\boldsymbol{R}. \tag{6.11}$$

The nonorthogonal rotation matrix $\boldsymbol{R}$ and its inverse $\boldsymbol{R}^{-1}$ transform the score vector $\boldsymbol{T}$ from the principal component decomposition into concentration profiles of the pure chemical species (Kvalheim and Liang, 1992). In order to determine $\boldsymbol{r}_i$ for the individual component i, both the zero concentration and selective regions are incorporated into Eq. (6.12)

$$\boldsymbol{c}_{sel+zero,j} = \boldsymbol{T}_{sel+zero,j}\boldsymbol{r}_i. \tag{6.12}$$

The subscript $sel+zero,j$ implies the use of the selective region in addition to the zero concentration region for component i. By rearranging Eq. (6.12), $\boldsymbol{r}_i$ can be expressed to

$$\boldsymbol{r}_i = \left(\boldsymbol{T}^{t}_{sel+zero,j}\boldsymbol{T}_{sel+zero,j}\right)^{-1}\boldsymbol{T}^{t}_{sel+zero,j}\boldsymbol{c}_{sel+zero,j} \tag{6.13}$$

$c_{sel+zero,i}$ can be obtained as the score vector of the first principle component in the selective region of component *i*. After determination of r_i for a particular component, its pure chromatogram can be deduced via Eq. (6.11). For a system with more than two components, the component stripping approach is applied in this stage by removing the contribution from resolved component in the following manner

$$X_{new} = X - C_i S_i^t . \tag{6.14}$$

The HELP analysis is again applied to X_{new} until all components are resolved.

6.2.3 Fast Wavelet Transform

In this work, FWT is employed to compress individual chromatogram from HPLC-DAD. Details of FWT computation can be found in the previous chapters and will not be stated in here. The coefficients for filters H and G are derived from the Daubechies wavelet, D_{2m}, with m being any positive integer from 1 to 10 (Daubechies 1992). Besides, the translation-rotation transformation (TRT) (Hayes *et al.*, 1973; *Chau* & Tam, 1994) and the coefficient position retaining (CPR) methods, as proposed in the pervious chapters, are also applied to solve the side-lobe and number of data point problems. The former problem occurs when there is a sudden data change in the spectrum during periodical extension of the spectral data at both ends in FWT preprocessing. The later one occurs when the data length of the spectrum is not equal to 2^p with p for any positive integers.

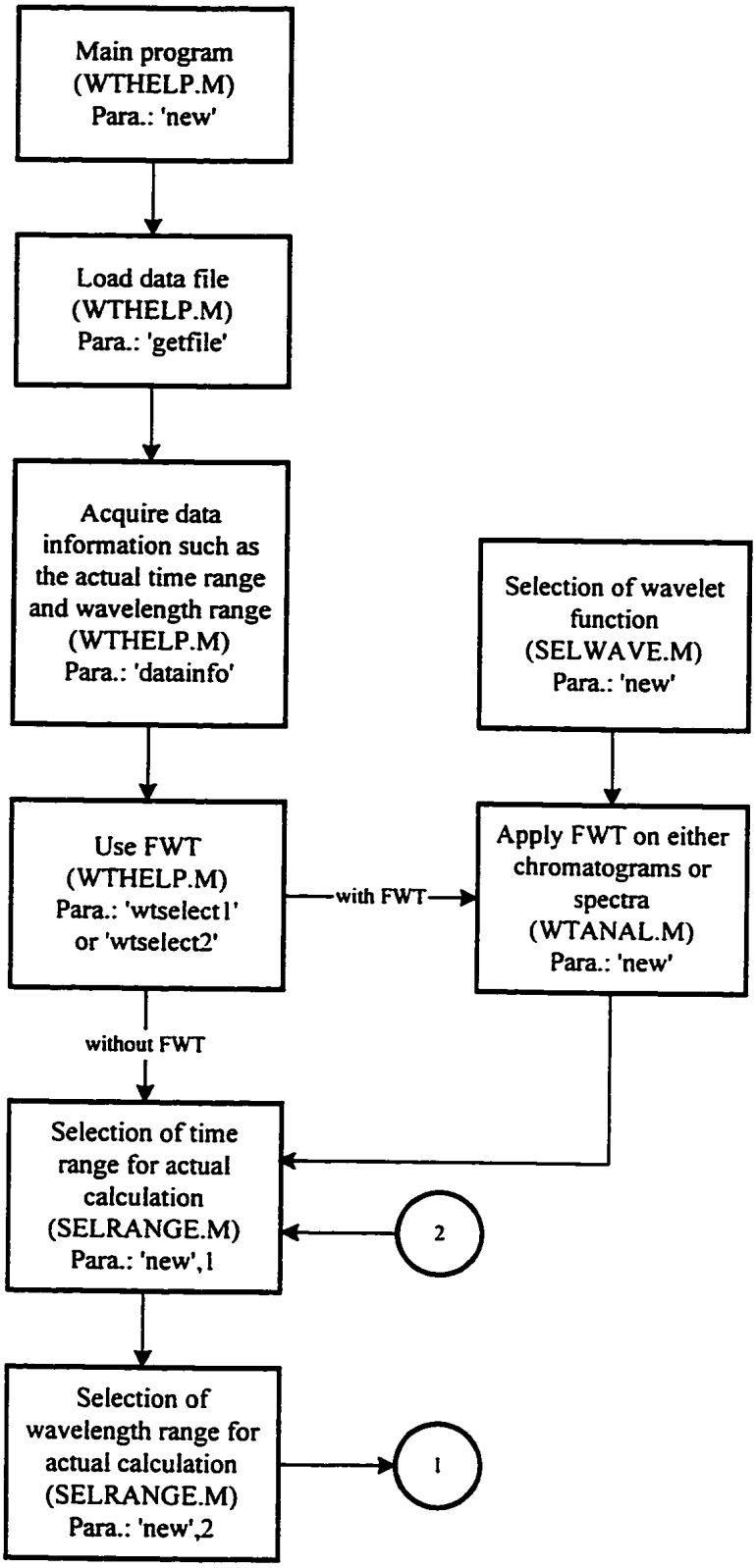
Chromatograms at each wavelength of the HPLC-DAD data are processed with FWT. The scale coefficients $C^{(j)}$ at a particular resolution level j and wavelengths are recombined together to form a new data set. Then, these data are analyzed with the HELP algorithm to determine the minimum number of components present and to resolve pure chromatograms and spectra of the individual components.

6.3 Experimental

All simulations and computations were carried out using the Fast Wavelet Transform Heuristic Evolving Latent Projections (FWT-HELP) Toolbox Version 1.0 as developed in this work. It was coded in MATLAB® for UNIX® Version 5.1 (The MathWorks, Inc., 1996) under IRIX™ 6.2 environment on a SGI Indigo^2 Power Extreme workstation (Silicon Graphics, Inc., CA, U.S.) with a 75MHz MIPS R8000™ processor. The toolbox can be employed for HELP computation with or without FWT treatment. Figure 6.3 shows a block diagram of the FWT-HELP toolbox. Appendices 6.1 and 6.2 give program description and screen captures of the FWT-HELP toolbox. Synthetic data of spectro-chromatograms were used to test the proposed algorithm and they were produced by combinations of Gaussian peaks (Li, 1997; Phillips and White, 1997). Homoscedastic noise with a variance of 0.002 to 0.1 were added to the data sets.

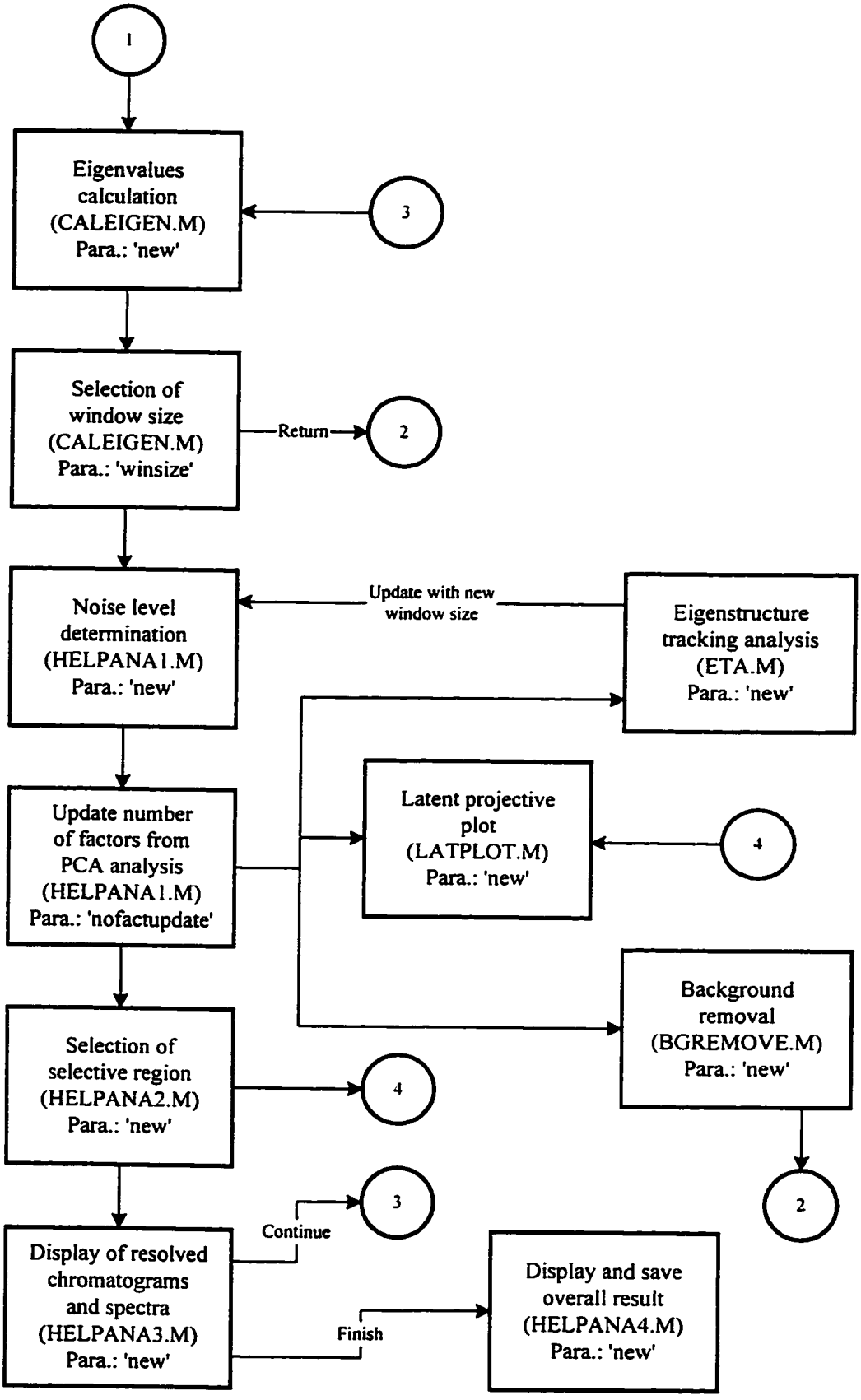
In order to test the performance of the proposed algorithm, real experiments were also performed. The data were acquired on a HP 1050 HPLC-DAD system (Hewlett-Packard, Germany) equipped with a 20 μL sample loop and a HP LiChrosomb (200 × 4.6 mm i.d., 5μm) C_{18} RP column (Hewlett-Packard, Germany). The mobile phase was

Figure 6.3 A block diagram of the Fast Wavelet Transform Heuristic Evolving Latent Projections (FWT-HELP) Toolbox Version 1.0.



(cont.)

(cont. Figure 6.3)



a mixture of acetonitrile-tetrahydrofuran-water (10:35:55 v/v) and was adjusted to such proposed ratio so that partial overlapping were observed in the chromatograms. The flow rate was 0.5 mL/min. The spectra were collected in the wavelength range from 250 nm to 600 nm with a spectral resolution of 2 nm and these spectral data were converted to ASCII text files with a self developed macro program, 3DATA.MAC (Appendix 5.1), under the HP ChemStation Revision A.03.01 environment (Hewlett-Packard, Germany).

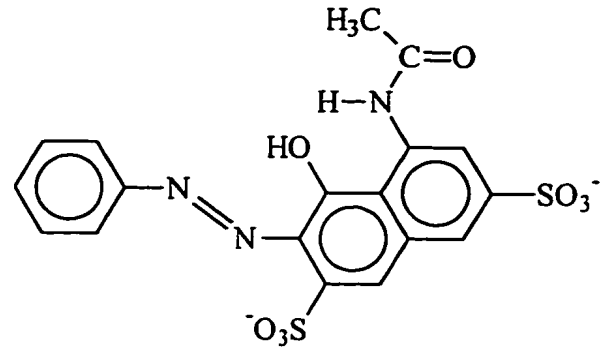
All solvent were of HPLC grade quality (Lab-Scan, Ireland) and double de-ionized water (Millipore-Waters, MA, U.S.) was used. Food dye samples of Red 2G (λ_{max} = 532 nm) and Amaranth (λ_{max} = 520 nm) (Figure 6.4), which were provided by the Hong Kong Government Laboratory, were used in this study. A two-component sample solution mixture was prepared by dissolving directly 1 ppm of Red 2G and 2 ppm of Amaranth in the mobile phase. 5 and 10 μL of sample solutions were injected into the HPLC-DAD system for investigation.

6.4 Results and Discussion

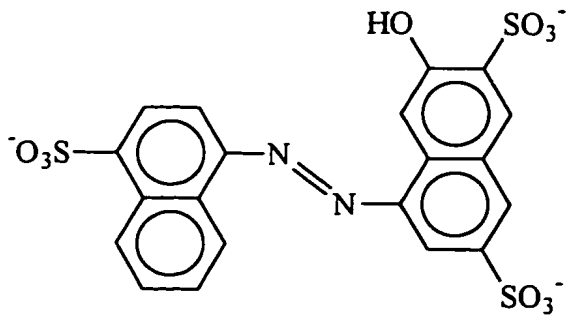
6.4.1 Sensitivity of Heuristic Evolving Latent Projections

In order to investigate the effect of signal-to-noise ratio (SNR) on the sensitivity of the HELP algorithm, simulated HPLC-DAD data sets were generated for comparison. Data Set A, B and C simulates a two-component system with different peak resolutions of 0.37, 0.62 and 0.88 respectively the chromatogram. Both chromatogram and spectrum of individual components were obtained by

Figure 6.4 Chemical structure of Red 2G (a) and Amaranth (b).



(a)



(b)

overlapping a series of Gaussian functions (Figure 6.1). The data consists of 1,000 points on the retention time axis and 177 points on the wavelength axis. Various levels of homoscedastic noise with a variance of 0.002 to 0.100 were added to the data sets. Table 6.1 shows the results from the synthetic data sets. In this table, SNR values were measured from the maximum of the chromatographic peak with the greatest absorbance on the wavelength axes while the relative $\log_{10}(\text{eigenvalue})$ ratio (RLER) were calculated by the following equation:

$$\text{RLER} = \frac{\text{Maximum height of the second } \log_{10}(\text{eigenvalue}) \text{ line}}{\text{Maximum height of the first } \log_{10}(\text{eigenvalue}) \text{ line}} \qquad (6.15)$$

The mean value of the third $\log_{10}(\text{eigenvalue})$ line which represents the noise level of the system were chosen as the baseline for the above calculation. A higher RLER value can facilitate the determination of the selective regions from the eigenvalues plot. Eq. (6.15) is good for a two-component system only. For a case with more than two components, the mean value of the $\log_{10}(\text{eigenvalue})$ line at the noise level is chosen as the baseline for the calculation. In our study, it was found that when the value of RLER is less than 0.050 or SNR is less than around 20, the second $\log_{10}(\text{eigenvalue})$ line becomes flatten and represents the background noise level only (Figure 6.5) if a window size of 5 was adopted in the computation. As a result, the required selective regions for the HELP analysis cannot be located directly from the eigenvalue plot.

To tackle with this problem, two alternative methods were employed. The first one is by increasing the window size in the FSMW-EFA computation. RLER can be improved from 123% to 472% times and 604% to 1,175% when a window size of 11

Table 6.1 Sensitivity of the HELP method at different peak resolutions and noise levels.

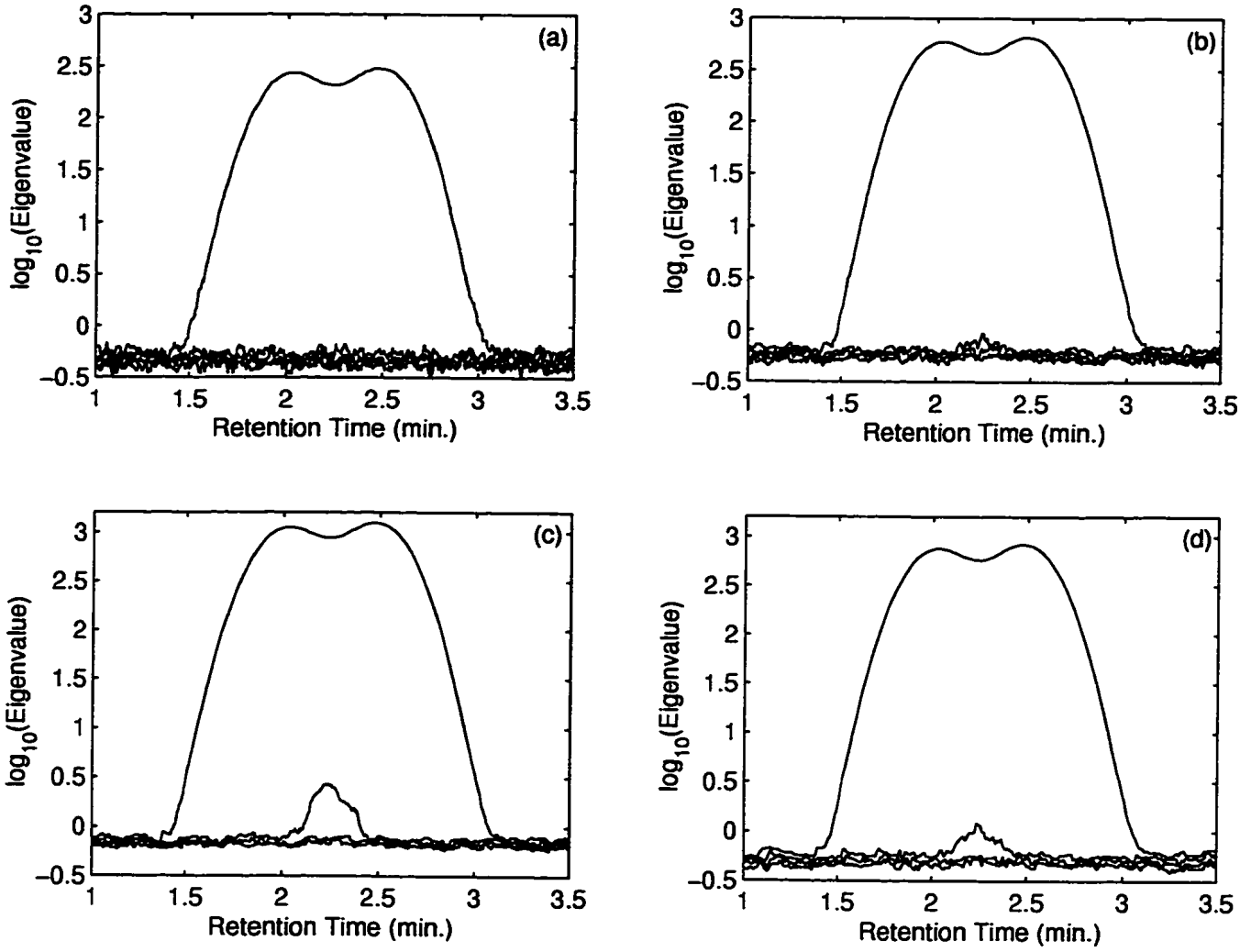
Peak resolution	Noise level	SNR in chromatogram	Windows size used in FSMW-EFA calculation	RLER	Percentage of change in RLER
0.37 (Set A)	0.002	741	5	0.330	
	0.010	149	5	0.163	
	0.020	75	5	0.080	
	0.050	31	5	0.015	
	0.100	17	5	0.011	
	0.100	17	11	0.063	472
	0.100	17	21	0.174	1482
	0.200	9	23	0.079	
0.62 (Set B)	0.002	525	5	0.348	
	0.010	106	5	0.174	
	0.020	53	5	0.086	
	0.050	22	5	0.052	

(cont.)

(cont. Table 6.1)

	0.100	12	5	0.026	
	0.100	12	11	0.058	123
	0.100	12	21	0.183	604
	0.200	7	21	0.073	
0.88 (Set C)	0.002	502	5	0.284	
	0.010	101	5	0.117	
	0.020	51	5	0.062	
	0.050	22	5	0.016	
	0.100	11	5	0.008	
	0.100	11	11	0.024	200
	0.100	11	21	0.102	1175
	0.200	7	31	0.082	

Figure 6.5 Eigenvalue plots of the simulated data set B with a noise level 0.100 by using window sizes of 5 (a), 11 (b) and 21 (c). Eigenvalue plot (d) was obtained by applying the Daubechies D_{16} wavelet function at the first resolution level on the data set. A window size of 7 was used.



and 21 were used compared with the result with a window size of 5 at a noise level 0.100 (Table 6.1). The general trend is that a larger window size is required for a data set with lower SNR value. However, we should keep the window size as small as possible. A larger window size would lead to distortion of the eigenvalue plot. It is because the width of the selective region will be decreased if a larger window size is used. The consequence of such distortion will lead to difficulty in determining the selective regions and affecting the HELP analysis. The second method is by using the latent projective graph. In this graph, the selective regions can be identified from the straight line segment which passes through the origin after prolongation (Kvalheim and Liang, 1992). As the value of SNR decreases, the noise level will be reflected from the latent projective graph (Figure 6.6) and this affect the determination of the selective regions of individual component. In the worst case, the straight line segment will be masked completely with noise.

6.4.2 Improvement by Fast Wavelet Transform

In this study, fast wavelet transform was included as a new method to improve the sensitivity of the HELP algorithm especially at low SNR value. It can also reduce the computational time for the HELP analysis. In the proposed method, FWT was applied to the chromatograms at each wavelength. Then, the scale coefficients at each wavelength were recombined together to form a new data set for the HELP analysis. As mentioned in the previous chapters, various types of wavelet functions are available. We have chosen the most popular one in the chemical studies, the Daubechies wavelet function, in this work. Table 6.2 shows the results of RLER by

Figure 6.6 Latent projective plots of the simulated data with a peak resolution of 0.62 and noise level of 0.002 (SNR = 525) (a), 0.010 (SNR = 106) (b), 0.050 (SNR = 22) (c) and 0.200 (SNR = 7) (d).

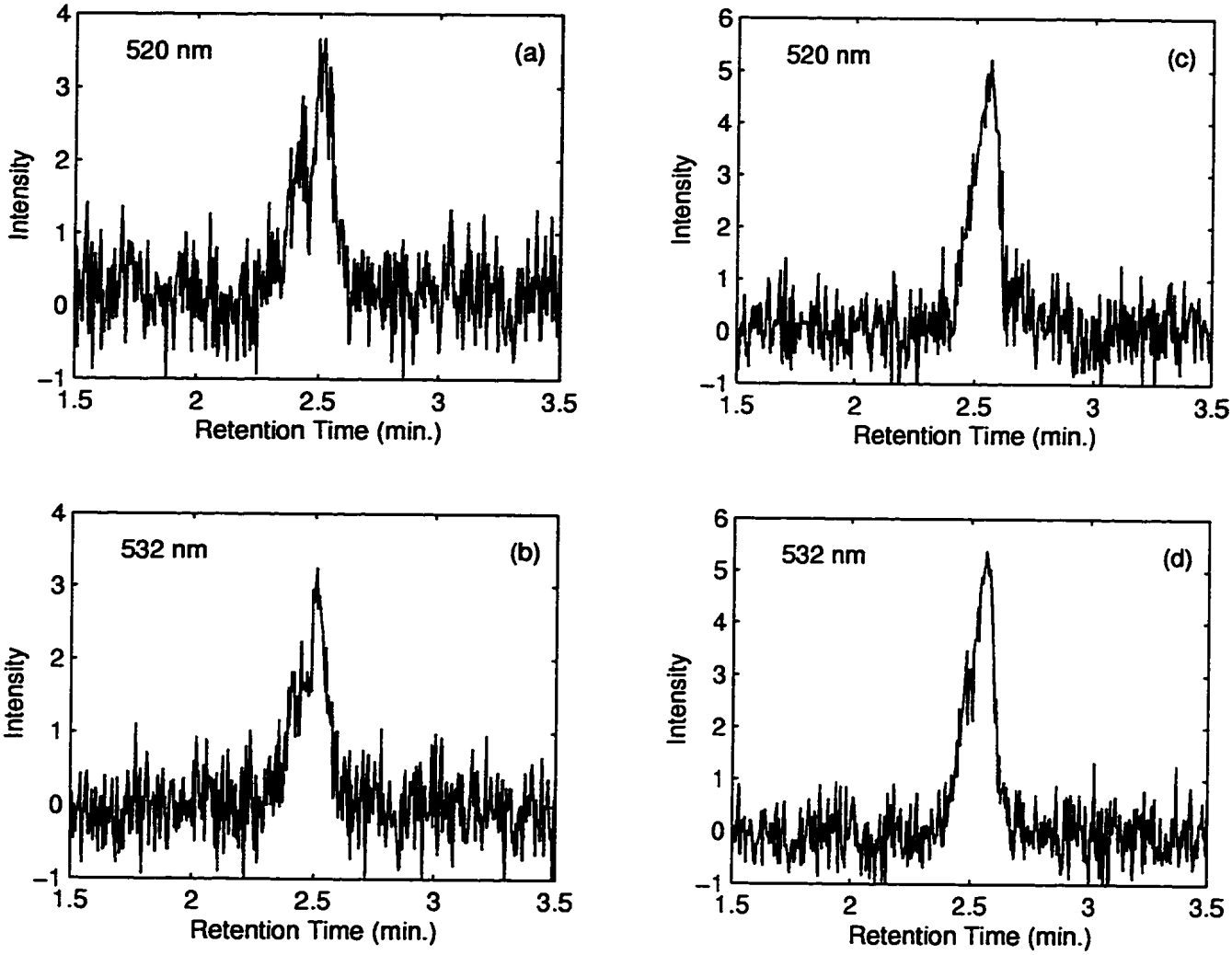


Table 6.2 Sensitivity of the HELP algorithm with different lengths of Daubechies wavelet functions, D_{2m}, at different peak resolutions. The noise level of the data sets is equal to 0.050. FWT was performed to the first resolution level and a window size of 5 was adopted for the HELP algorithm.

Peak resolution	D_{2m}	SNR in chromatogram after FWT	RLER	Percentage of change in RLER [a]
0.37 (Set A)	2	43	0.101	573
	4	43	0.103	587
	6	43	0.099	560
	8	43	0.094	527
	10	43	0.105	600
	12	43	0.099	560
	14	44	0.093	520
	16	43	0.093	520
	18	43	0.104	593
	20	43	0.098	533
0.62 (Set B)	2	31	0.097	87
	4	31	0.094	81
	6	31	0.104	100
	8	31	0.105	102
	10	31	0.093	79
	12	31	0.098	88
	14	31	0.105	102
	16	31	0.105	102
	18	31	0.096	85
	20	31	0.100	92

(cont.)

(cont. Table 6.2)

0.88	2	29	0.066	313
	4	30	0.067	319
	6	30	0.072	350
	8	30	0.075	369
	10	29	0.069	331
	12	29	0.070	338
	14	29	0.074	363
	16	30	0.075	369
	18	30	0.071	344
	20	29	0.071	344

[a] The quantities were obtained by comparing with the corresponding RLER values in Table 6.1.

using different lengths of the Daubechies wavelet functions at the first resolution level on the simulated data sets (Figure 6.5d). As mentioned in the previous section, a proper eigenvalue plot can be obtained by using a window size of 13 or above if the SNR of the data is very low. In general, a small window size is preferred because it can minimize any derivation from the resolved chromatograms and spectra. Figure 6.5d shows the eigenvalue plot as obtained with wavelet treatment. A window size of 7 can recover the missing information from the eigenvalue plot. By comparing RLER values, the Daubechies wavelet D_{16} function is an optimal choice. Table 6.3 shows the results from FWT-HELP analysis. The results indicated that RLER can be improved from 84% to 400% if the same window size is used. Besides, similar RLER can be attained by using a smaller window size. That means we can minimize the error from the determination of the selective regions as mentioned above.

The data size of $C^{(1)}$ as deduced from FWT treatment is half than that of the original data. It represents the original chromatograms at lower resolution and noise level. One may suspect the performance of the calculation on the data set with lower resolution. So, we have performed another computation on the same data set. The resolution of chromatograms were lowered by selecting alternative data points to form a new data set and then it was employed for the HELP analysis. Although this method can generated a data set with similar resolution as $C^{(1)}$, the results as given in Table 6.4 indicate that the percentage change of RLER fluctuated from -15% to 263%. FWT is a linear operation and it can retain the most important information at lower resolution or in the compressed form. This is the intrinsic property associated with wavelet transform. In contrast, compression by another method cannot not retain

Table 6.3 Sensitivity of the HELP algorithm at different peak resolutions with FWT treatment. The noise level of the data sets is equal to 0.100. The FWT was performed with the Daubechies wavelet D_{16} function at the first resolution level.

Peak resolution	SNR in chromatogram after FWT	Windows size used in FSMW-EFA calculation	RLER	Percentage of change in RLER [a]
0.37	22	5	0.037	236
	22	11	0.181	187
	22	21	0.343	97
0.62	16	5	0.048	84
	16	11	0.195	236
	16	21	0.383	109
0.88	16	5	0.032	300
	16	11	0.120	400
	16	21	0.288	182

[a] The quantities were obtained by comparing with the corresponding RLER values in Table 6.1.

Table 6.4 Sensitivity of the HELP algorithm at different peak resolutions by using odd number data points from the original chromatograms. The noise level of the data sets is equal to 0.100.

Peak resolution	SNR in chromatogram	Windows size used in FSMW-EFA calculation	RLER	Percentage of change in RLER [a]
0.37	16	5	0.040	263
	16	11	0.129	105
	16	21	0.281	62
0.62	12	5	0.022	-15
	12	11	0.151	160
	12	21	0.328	79
0.88	11	5	0.022	-8
	11	11	0.080	233
	11	21	0.241	136

[a] The quantities were obtained by comparing with the corresponding RLER values in Table 6.1.

such important information. As a result, data set with FWT treatment has better performance on the HELP analysis.

However, the data size reduction from FWT introduces another problem. In the conventional HELP analysis, the resolved chromatograms and spectra should have the same dimension of $\boldsymbol{X}$. In this case, the resolved chromatograms and spectra should have 1,000 and 177 data points respectively. With the use of FWT, the resolved chromatograms consist 500 data points only. But, such reduction of data points will not affect the result of the HELP analysis. In our calculation, j was limited to the first resolution level. Higher resolution level can further improve the sensitivity of the HELP analysis, but the data lenght in the retention time axes will also be further reduced and affect the accuracy of the analysis especially for the determination of selective regions from the eigenvalues plot. FWT has another role in the enhancing the HELP analysis by reducing the computational time. In practice, the data set from HPLC-DAD consist of around 1,000 to 1,500 data points in the retention time axes and 100 to 200 data points in the wavelength axes. In the computation, most of time is the spent on the calculating of eigenvalues with the above data size (Walczak and Massart, 1997). In such cases, FWT can compress the data set to a smaller size and speed up the eigenvalues calculation. If $\boldsymbol{C}^{(1)}$ is employed for the FSWM-EFA calculation, about 5% to 42% of time can be reduced in the eigenvalues calculation step (Tables 6.5 and 6.6). On the other hand, if the FWT-HELP analysis is performed on resolution level 2, about 3% to 63% of time can be saved.

Table 6.5 Computational time of eigenvalues computation and wavelet transform computation at resolution levels 1 and 2 with different data size.

	Average Computational Time [a] / sec.				
	Window Size			Resolution Level j	
Data Size	7	11	15	1	2
100	1.6469	2.6454	3.6459	1.4038	2.3744
200	3.3738	5.6597	8.3853	1.5418	2.6576
300	5.1407	8.6838	12.6286	1.7622	2.9532
400	6.8856	11.7190	17.1541	2.0221	3.2501
500	8.6534	14.7635	21.6152	2.3570	3.6476
600	10.3912	17.9036	26.1259	2.7548	3.9930
700	12.1858	20.7956	30.6631	3.4840	4.6972
800	13.9418	23.8597	35.1151	4.0710	5.8133
900	15.6348	26.8878	39.6362	4.8520	6.6395
1,000	17.4041	29.8757	44.1360	4.9191	6.5906
1,100	19.1465	32.9595	48.6824	5.8696	7.6055
1,200	20.9677	35.9725	53.1093	7.1782	7.9879
1,300	22.6900	38.9800	57.5784	7.1160	8.9411
1,400	24.7392	42.0418	62.1438	7.9819	9.8778
1,500	26.2108	45.1265	66.5980	9.4059	11.5509
1,600	27.9462	48.3097	77.1397	9.4841	11.6545
1,700	29.6705	51.0956	75.6608	10.3951	12.5377
1,800	31.4585	54.1023	80.1201	11.7067	14.0947
1,900	33.2373	57.1748	85.0874	12.2375	14.3215
2,000	35.1709	60.1927	89.1206	12.7057	15.1866

[a] The computational time was measured by the built-in function in MATLAB®.

Table 6.6 Comparison of computational time of eigenvalue calculation with FWT treatment.

	Percentage of Computational Time Saved					
	Eigenvalue Computation (FWT-HELP at resolution level 1) with Window Size			Eigenvalue Computation (FWT-HELP at resolution level 2) with Window Size		
Data Size	7	11	15	7	11	15
200	5.49	26.02	38.13			
400	21.64	34.45	39.33	3.25	49.69	59.80
600	24.02	36.11	41.12			
800	21.41	33.82	39.56	34.10	51.91	59.57
1,000	22.02	34.12	39.88			
1,200	16.21	28.77	36.27	37.39	53.65	61.18
1,400	18.48	31.55	37.81			
1,600	16.18	30.98	42.18	33.66	51.62	62.65
1,800	13.09	28.66	35.92			
2,000	14.39	29.26	36.22	32.22	50.24	58.71

6.4.3 Analyzes of Real Experimental Data

Figure 6.7 shows the overall chromatograms of the food dye mixture solutions that measured at 520 nm and 532 nm with different injection volumes as mentioned previously. Such injection volumes can show the effect at different SNR values. The SNR value is about 2 and 3 at both wavelengths for the injection volume of 5 and 10 μL respectively. Figure 6.8 shows the overall spectra of the food dye mixture that measured at 2.4 and 2.5 minutes with different injection volumes. Without using FWT treatment, the selective regions were identified when a window size of 13 and 19 was used for the injection volumes of 5 and 10 μL of the sample. Figure 6.9 shows the eigenvalue plots and the latent projective graphs of the sample with different injection volume. On the other hand, when the data was preprocessed with Daubechies wavelet D_{16} function at the first resolution level, a similar result could be attained with a window size of 5 or 7. Figure 6.10 shows the eigenvalue plots and the latent projective graphs of the data with FWT treatment. As compared with Figure 6.11, the second eigenvalue line was observed again on the eigenvalue plots with FWT treatment (Figure 6.10). Figure 6.12 shows the experimental chromatograms of Amaranth and Red 2G measured at 520 nm and 532 nm respectively and their corresponding spectra. Figures 6.13 and 6.14 give the resolved chromatograms and spectra of the sample with and without FWT treatment at resolution level 1 for different injection volumes. The results show that the resolved chromatograms and spectra with FWT treatment are much better than that from the conventional HELP analysis. Both chromatograms and spectra are less noisy than those without FWT treatment. Such improvement can facilitate the quantitative analysis through the

Figure 6.7 Chromatograms of the sample solution of Red 2G and Amaranth measured at 520 nm and 532 nm. (a) and (b) corresponds the chromatographic study with an injection volume of 5 μL while (c) and (d) with an injection volume of 10 μL.

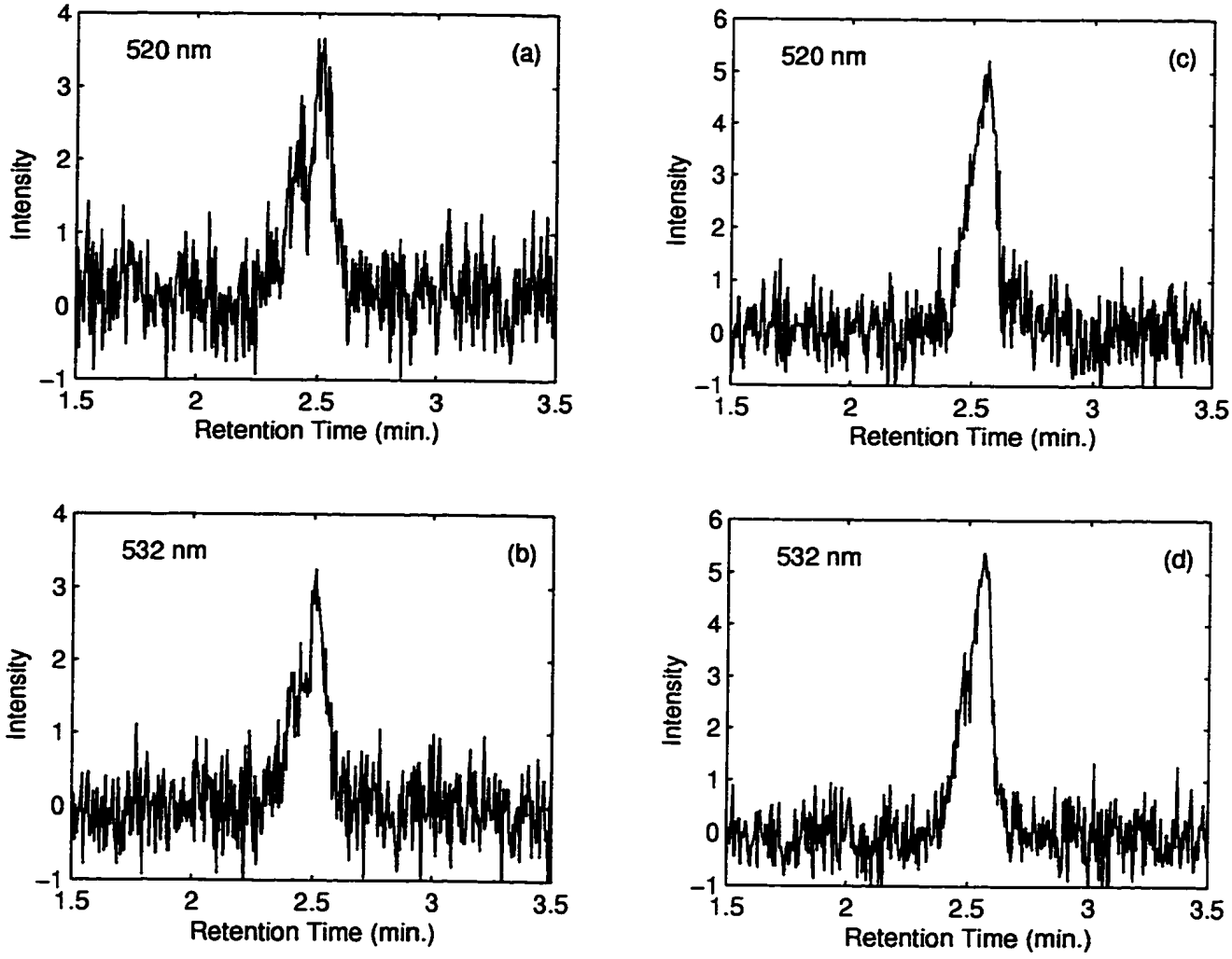


Figure 6.8 Spectra of the sample solution of Red 2G and Amaranth measured at 2.4 and 2.5 minutes with injection volume of 5 μL (a-b) and 10 μL (c-d).

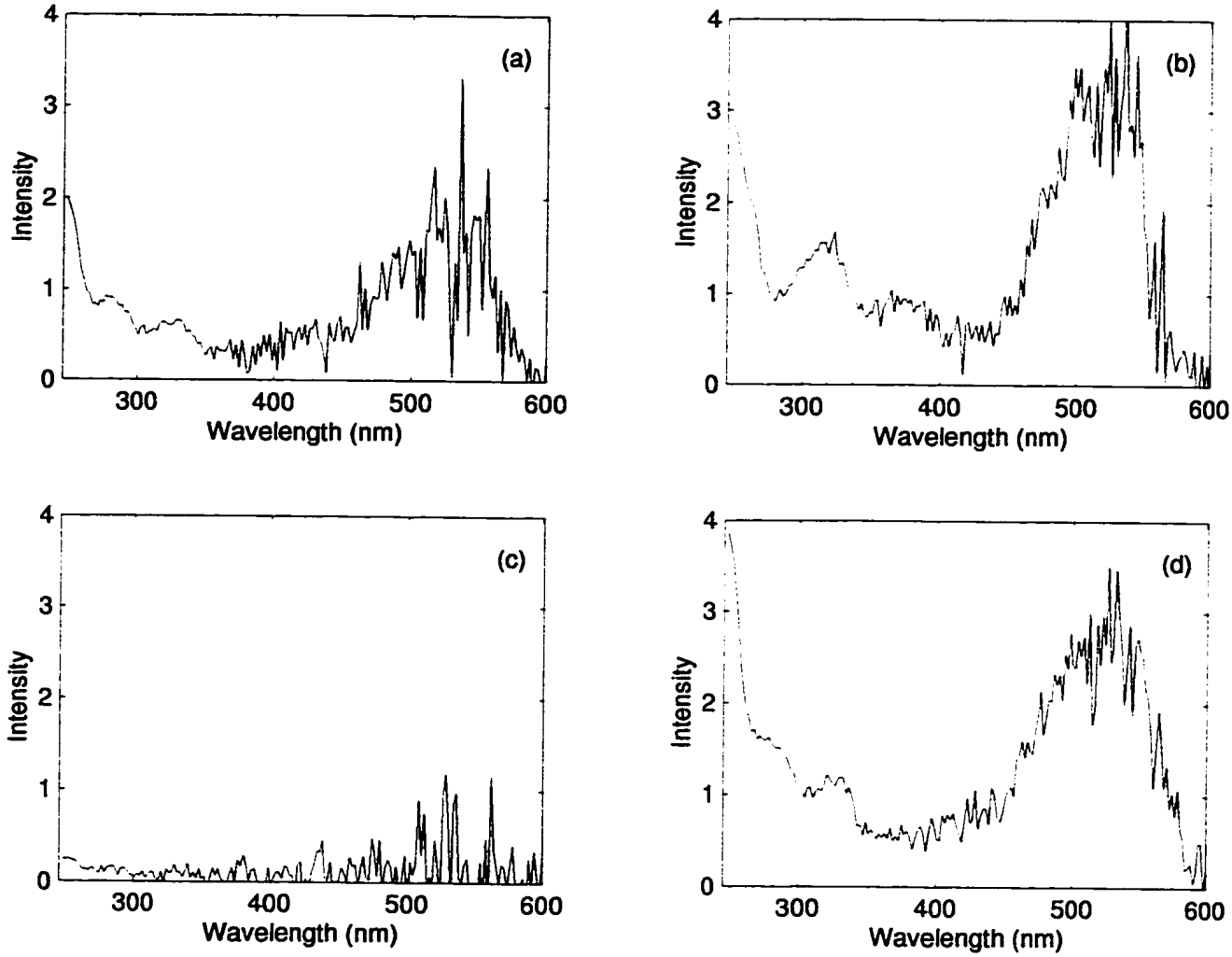


Figure 6.9 (a) The eigenvalue plot with window size of 13 for the sample with an injection volume of 5 μL, (b) and (c) the corresponding latent projective graphs for the wavelength range from 250 nm to 600 nm and from 250 nm to 400 nm respectively. (d) The eigenvalue plot with window size of 19 for the sample with an injection volume of 10 μL, (e) and (f) the corresponding latent projective graphs for the wavelength range from 250 nm to 600 nm and from 250 nm to 400 nm respectively.

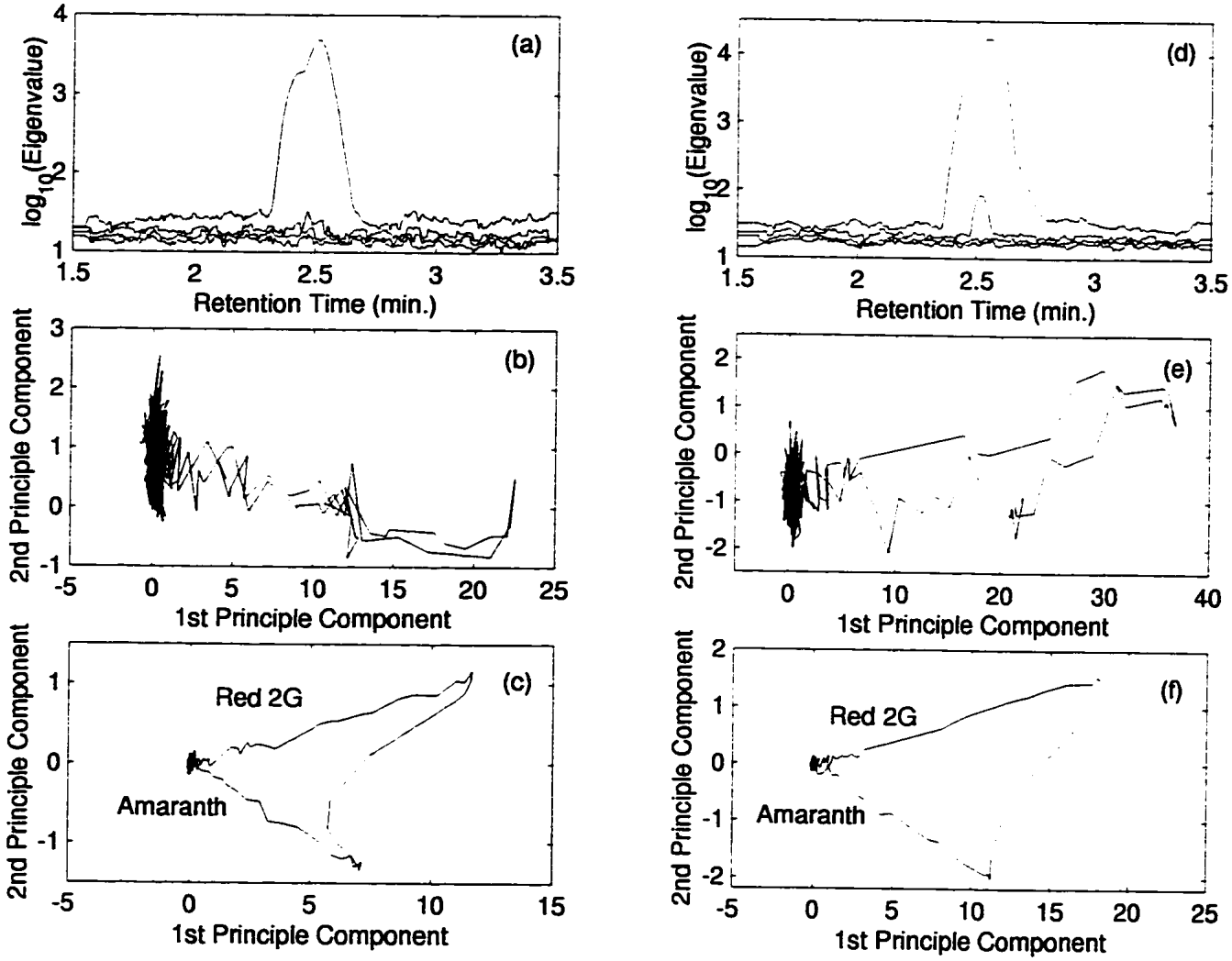


Figure 6.10 (a) The eigenvalue plot with window size of 5 for the sample with injection volume of 5 μL, (b) and (c) the corresponding latent projective graphs for wavelength range from 250 nm to 600 nm and from 250 nm to 400 nm respectively. (d) The eigenvalue plot with window size of 7 for the sample with an injection volume of 10 μL, (e) and (f) the corresponding latent projective graphs for wavelength range from 250 nm to 600 nm and from 250 nm to 400 nm respectively. All plots were obtained by applying a Daubechies D_{16} wavelet function at the first resolution level on the data set.

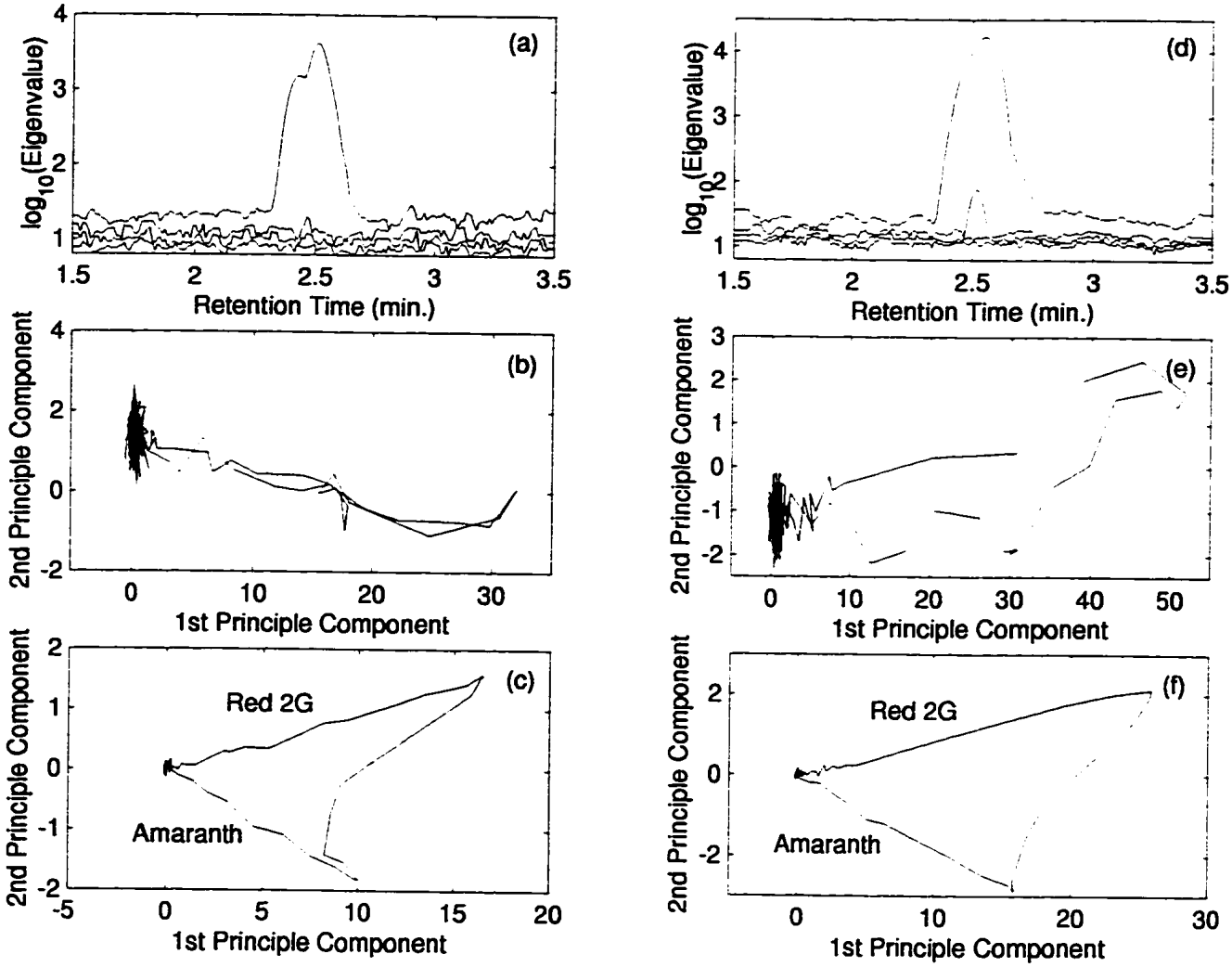


Figure 6.11 The eigenvalue plot with window size of (a) 5 for the sample with injection volume of 5 μL and (b) 7 for the sample with an injection volume of 10 μL. Such plots were generated without FWT treatment.

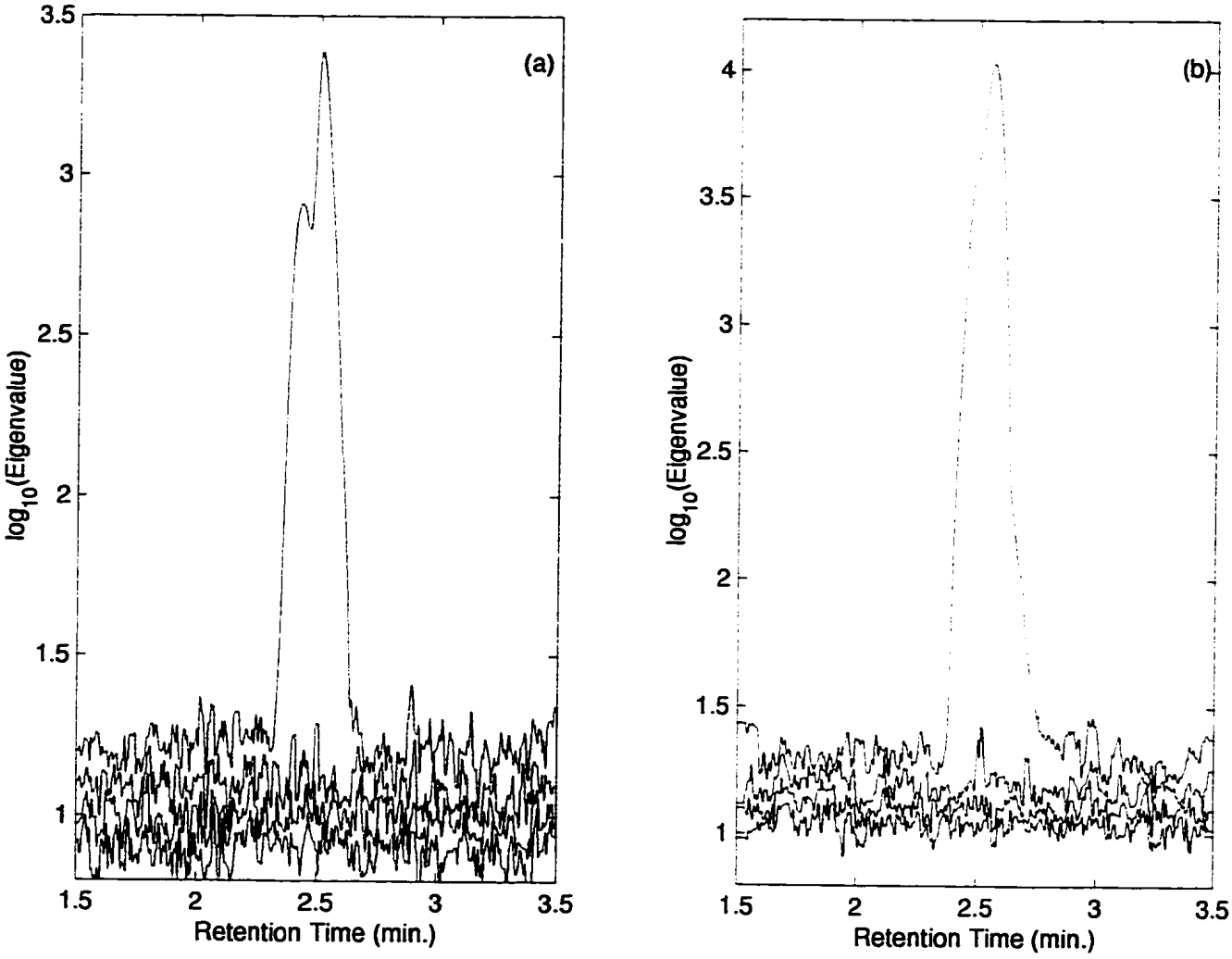


Figure 6.12 (a) Chromatogram of Amaranth at 520 nm and (b) its corresponding spectrum. (c) Chromatogram of Red 2G at 532 nm and (d) its corresponding spectrum.

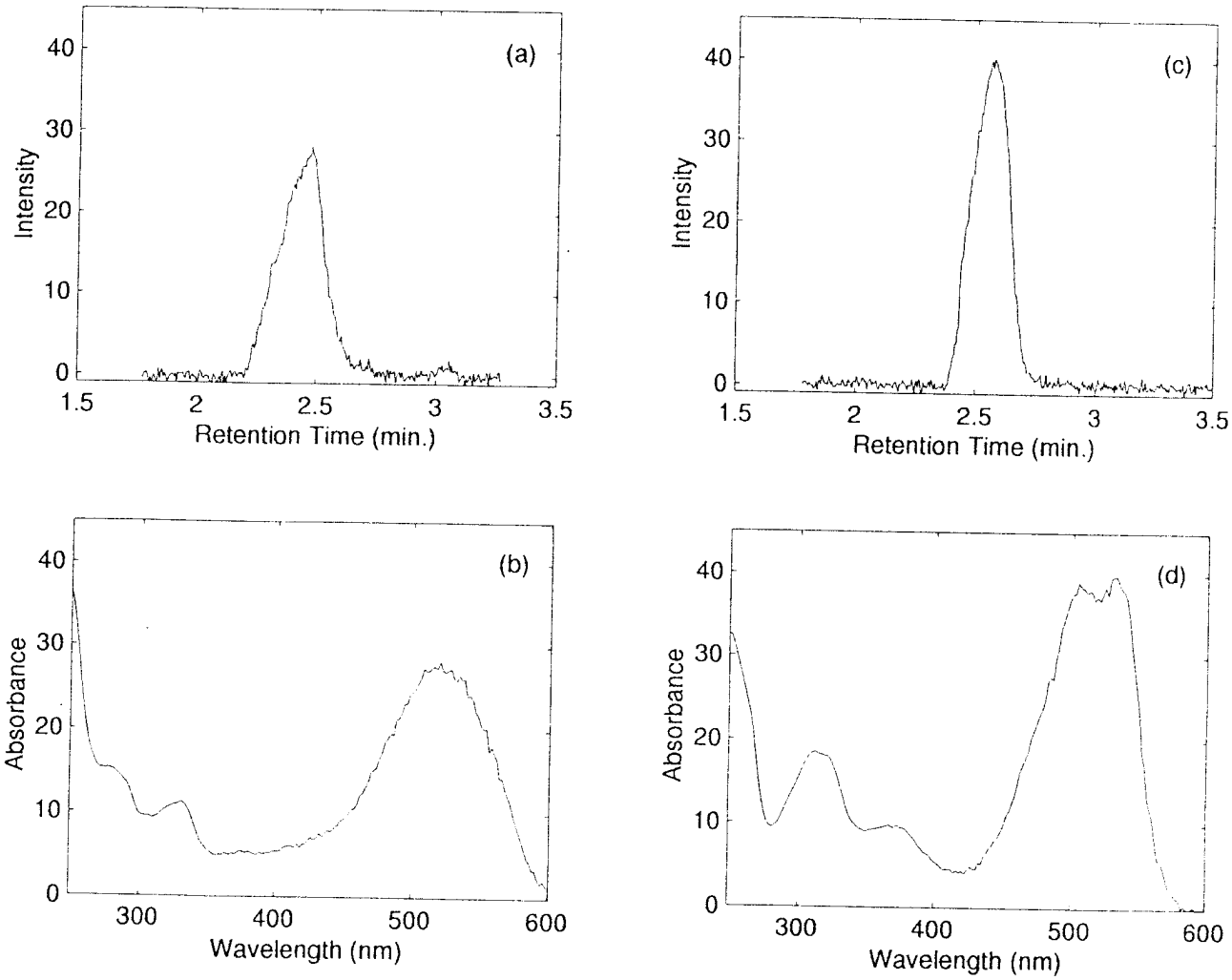


Figure 6.13 Resolved chromatograms and spectra for the sample solution of Red 2G and Amaranth as obtained with an injection volume of 5 μL through HELP analysis with window size 13 (a-b) and FWT-HELP analysis with window size 7 (c-d) respectively. The data set was processed with FWT treatment at resolution level 1.

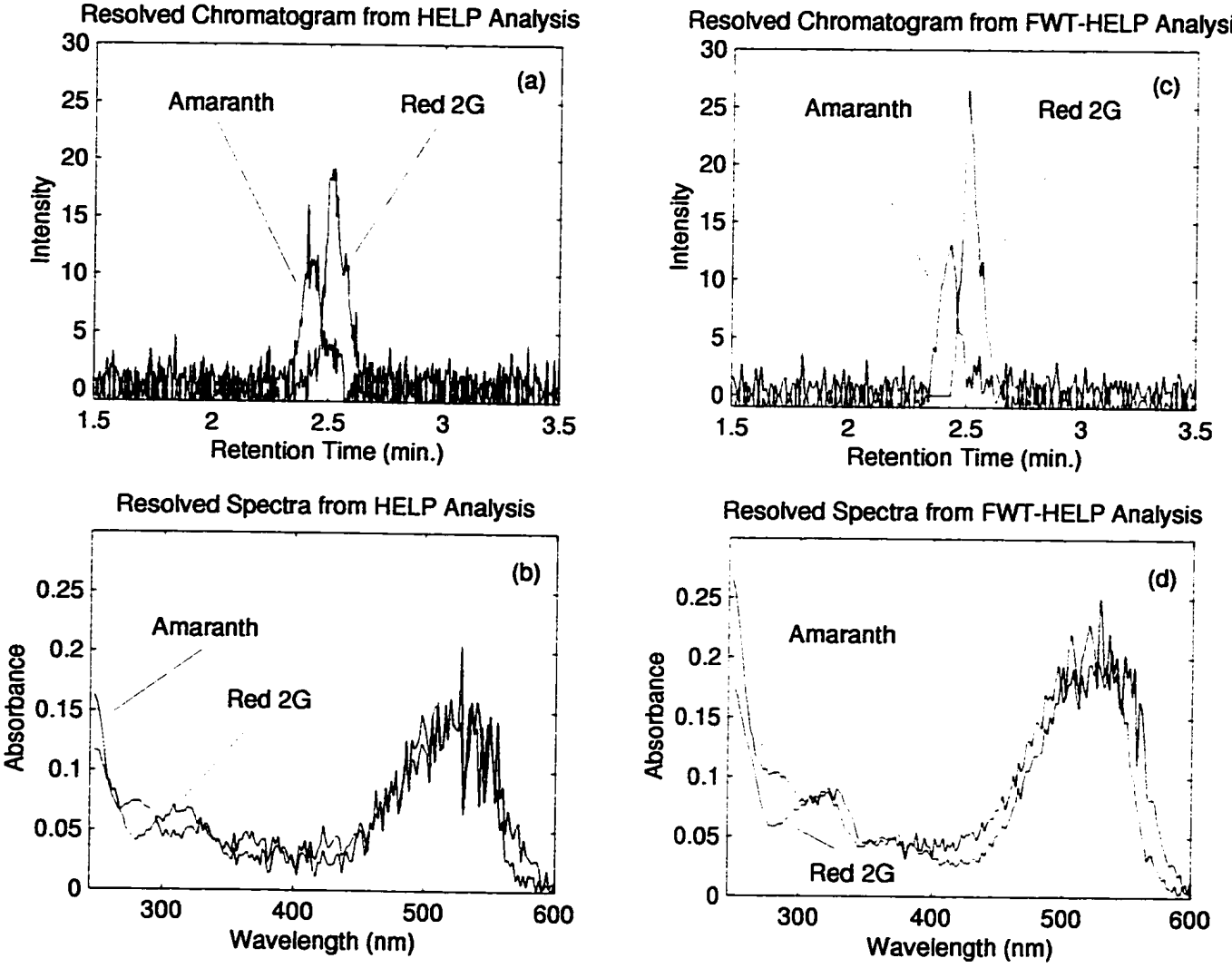
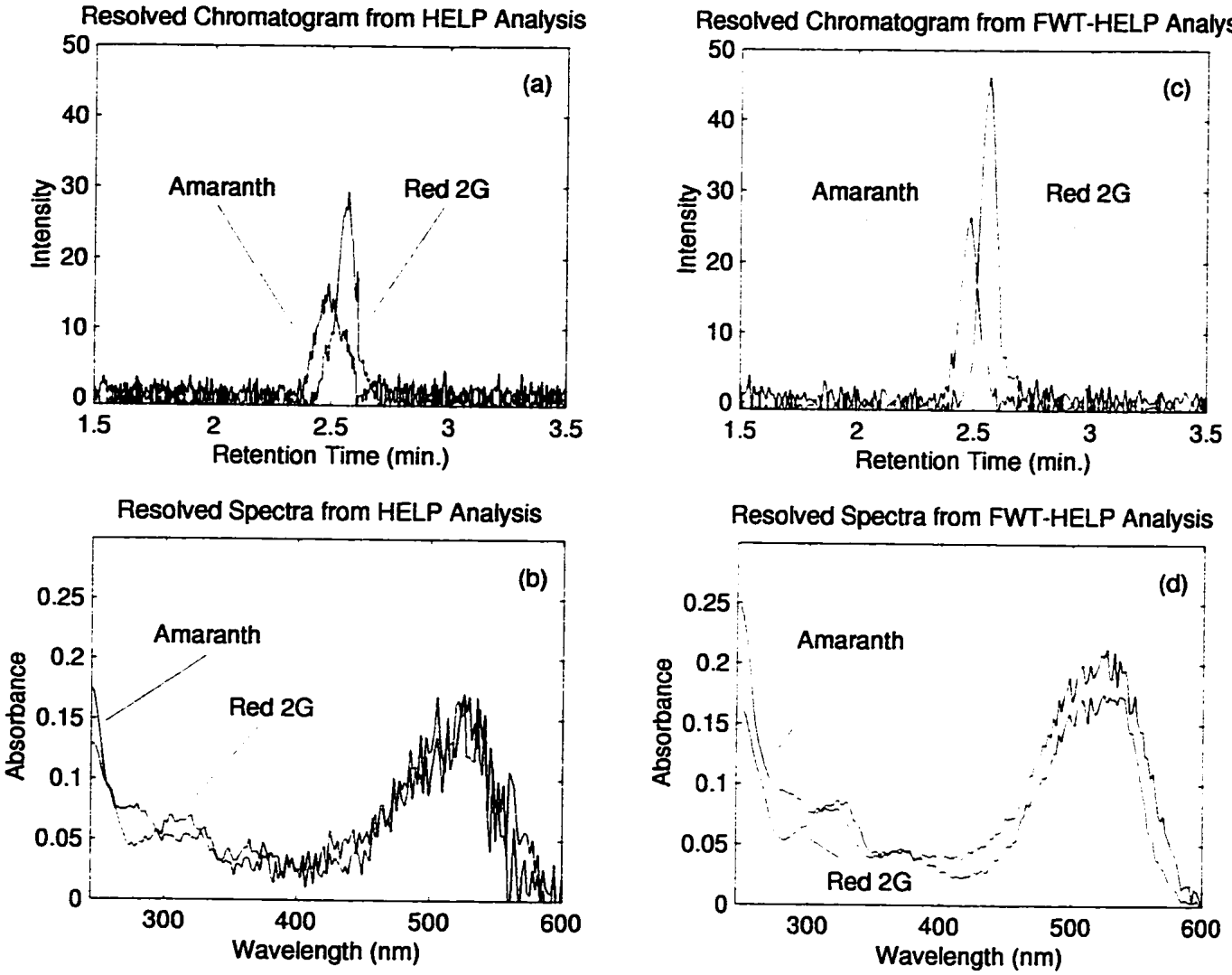


Figure 6.14 Resolved chromatograms and spectra for the sample solution of Red 2G and Amaranth as obtained with injection volume of 10 μL though HELP analysis with window size 19 (a-b) and FWT-HELP analysis with window size 7 (c-d) respectively. The data set was processed with FWT treatment at resolution level 1.



resolved chromatograms. Figure 6.15 give the resolved chromatograms and spectra of the sample that processed with FWT-HELP analysis at resolution level 2. A similar results as shown in Figures 6.13 and 6.14 were obtained.

Table 6.7 shows the results of peak center, peak height and peak area of the chromatographic peaks of the food dye mixture by different methods that includes HELP, FWT-HELP and direct estimation from Figure 6.7. Such parameters are usually required for both qualitative and quantitative analyses. In Figure 6.7, the peak area of individual peak was estimated by drawing a perpendicular line to bisect the overlapped chromatographic peak. This conventional method causes underestimation of the sample concentration. By utilizing the peak height or peak area from the resolved chromatogram with HELP or FWT-HELP treatment, a more accurate result can be obtained. In general speaking, the results from FWT-HELP treatment would be better than that from HELP treatment due to the denoising property of FWT.

6.5 Conclusion

In conclusion, FWT has been found to improve the HELP analysis by using a smaller of window size in the FSWM-EFA calculation for the HPLC-DAD data with low SNR value. Besides, it can also reduce the computational time in HELP computation by using the compressed data. HPLC-DAD study is a very complicated system. Parameters such as the degree of peak overlapping and SNR will affect the result from the HELP analysis. Yet, they vary with different experimental conditions.

Figure 6.15 Resolved chromatograms and spectra for the sample solution of Red 2G and Amaranth as obtained with injection volume of (a-b) 5 μL and (c-d) 10 μL though FWT-HELP analysis with window size 7 respectively. The data set was processed with FWT treatment at resolution level 2.

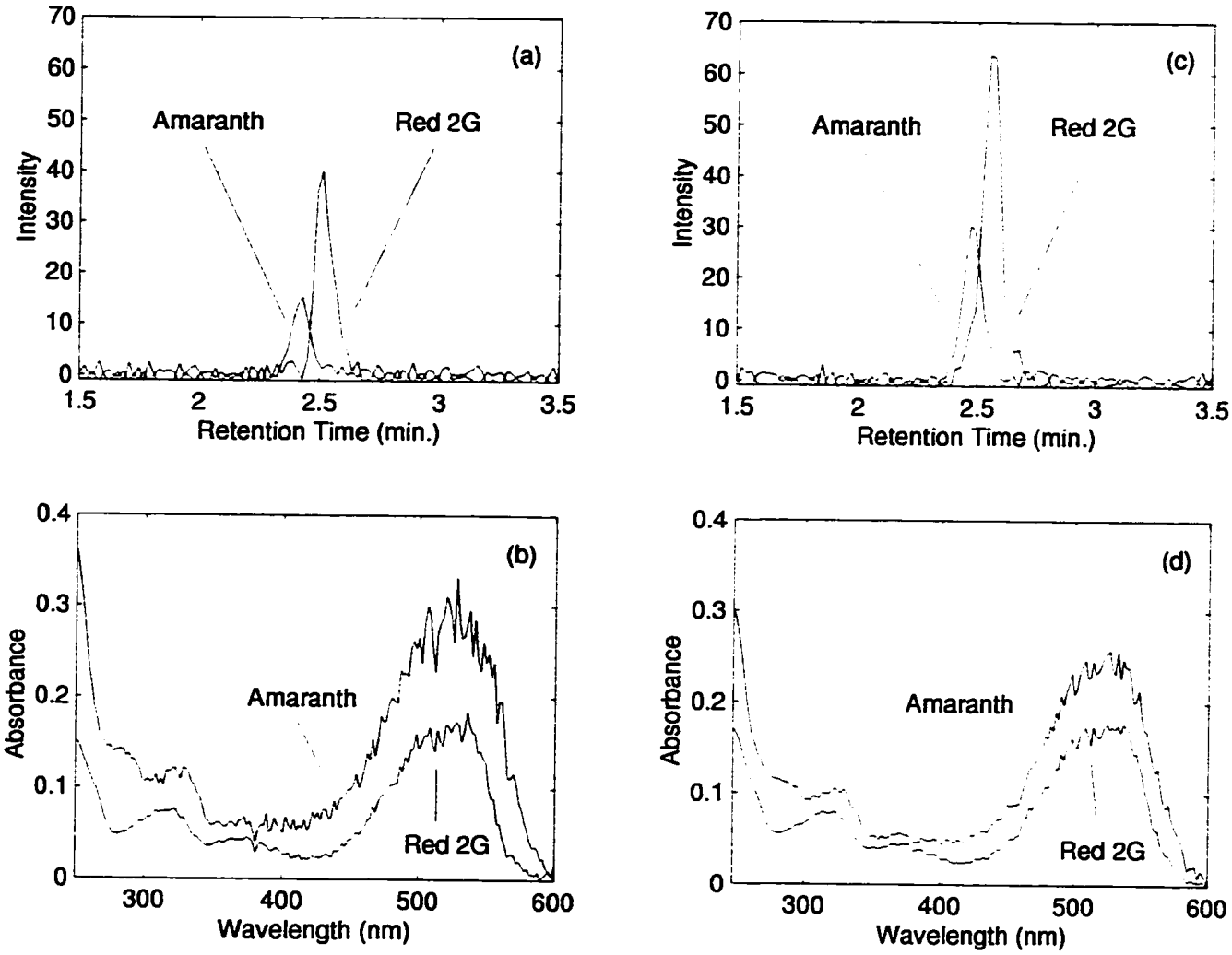


Table 6.7 Results of peak center, peak height and peak area of the chromatographic peaks of the food dye mixture with various analyzing method.

Analyzing Method	Injection volume / μL	Peak center / min.		Peak height / unit			Peak area / square unit		
		Amaranth	Red 2G	Amaranth	Red 2G	Ratio	Amaranth	Red 2G	Ratio
HELP	5	2.40	2.51	16.22	19.16	1:1.18	256.94	370.21	1:1.44
	10	2.48	2.57	16.52	29.54	1:1.79	403.52	554.77	1:1.37
FWT-HELP	5	2.43	2.51	13.08	26.77	1:2.05	116.13	217.56	1:1.87
with $j = 1$	10	2.48	2.56	26.25	46.22	1:1.76	255.53	395.71	1:1.55
FWT-HELP	5	2.41	2.50	15.45	40.09	1:2.59	69.15	172.78	1:2.50
with $j = 2$	10	2.47	2.55	30.13	63.32	1:2.10	153.09	320.47	1:2.09

(cont.)

(cont. Table 6.7)

Estimation									
Figure 6.7a	5	2.43	2.51	2.91	3.67	1:1.26	41.69	64.05	1:1.54
Figure 6.7b	5	2.45	2.51	2.25	3.24	1:1.44	30.44	46.86	1:1.54
Figure 6.7c	10	2.48	2.56	3.41	5.21	1:1.53	33.45	99.80	1:2.98
Figure 6.7d	10	2.49	2.56	3.50	5.40	1:1.54	33.72	99.17	1:2.94

So, it is not easy to generalize some conditions that the HELP algorithm might fail to work. This new method can be adopted for chemical analysis of complicated multiple component systems such as traditional Chinese medical herb using HPLC-DAD system. More information is available on the next chapter.

CHAPTER 7

Analysis of the Water Soluble Constituents of *Cordyceps Sinensis* with Fast Wavelet Transform Heuristic Evolving Latent Projections

7.1 Introduction

According to a recent study by the Massachusetts Institute of Technology (MIT) consultant group, Hong Kong is recommended to develop into an international manufacturing center of traditional Chinese medicine (TCM) in the future (Wang *et al.*, 1997). TCM refers to health care alternative or supplement which is produced on empirical experiences accumulated over a thousand years or more in China and other Asian countries. It is expected that there is a dramatic growth for TCM and its associated products in the world on the next century (Wong, 1997). However, the effectiveness of TCM products depends greatly on the quality of the natural materials used (Lee, 1997). Quality control of Chinese medicines has become a priority issue in Hong Kong (But and Kwok, 1996). Besides, safety on the TCM products is another issue to be concerned by many international organizations such as the Food and Drug Administration (FDA) in the United States. In Hong Kong, the practice of TCM and the related products are unregulated by the government. According to a recent report by The Chinese University of Hong Kong (Wong *et al.*, 1997), women, older residents, unemployed workers, low skill laborers, current smokers and citizens who dissatisfied with the quality of private sector clinics seek significantly to consult TCM practitioners and use TCM products. So, quality assurance based on both qualitative and quantitative analysis should be performed on raw materials and TCM products before they are sold to the market.

In recent year, various spectroscopic and chromatographic approaches have been used or are being explored to analyze TCM products (Che, 1997). These techniques

include thin layer chromatography (TLC), gas chromatography (GC), high performance liquid chromatography (HPLC), capillary electrophoresis (CE) and hyphenated instruments such as gas or liquid chromatography coupled with mass spectrometry (GC-MS or LC-MS). Several commonly used TCM products have been analyzed with the above mentioned techniques by the Laboratory for the Chemical Quality Control of Chinese Medicines and Health Foods in Hong Kong (But and Kwok, 1996; But *et al.*, 1996; But *et al.*, 1997).

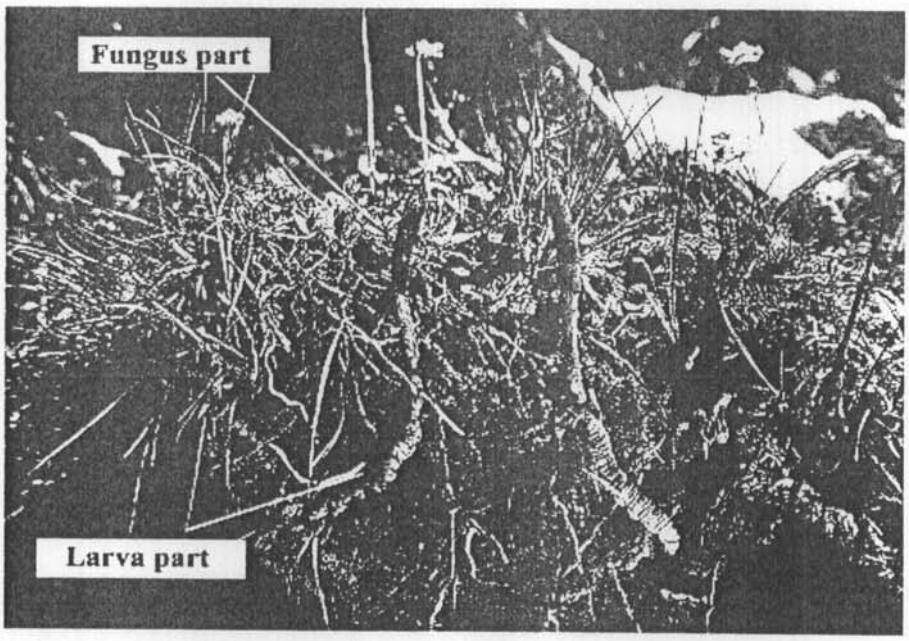
In terms of analytical chemistry, chemical constituents of Chinese herbs are a complex black system (Liang *et al.*, 1993a). There is no a priori information concerning the chemical composition of the samples. We need to develop a suitable method to analyze the major constituents. The use of computer and appropriate chemometrics techniques, such as fuzzy logic, factor analysis and image analysis is helpful in identifying the major and minor constituents from data generated from various chemical techniques (Xiong and Liu, 1998). In the previous chapter, the fast wavelet transform heuristic evolving latent projections (FWT-HELP) was developed to analyze data from high performance liquid chromatography coupled with a diode array detector (HPLC-DAD). In this chapter, the performance of the proposed FWT-HELP algorithm in the real situation was explored again. A traditional Chinese medicinal herb, *Cordyceps sinensis* (冬蟲夏草), was chosen as an example in this work and analyzed with HPLC-DAD.

Cordyceps sinensis is an expensive, traditional and precious dried Chinese medicinal herb belonging to the fungus category. The stroma are formed by *Cordyceps*

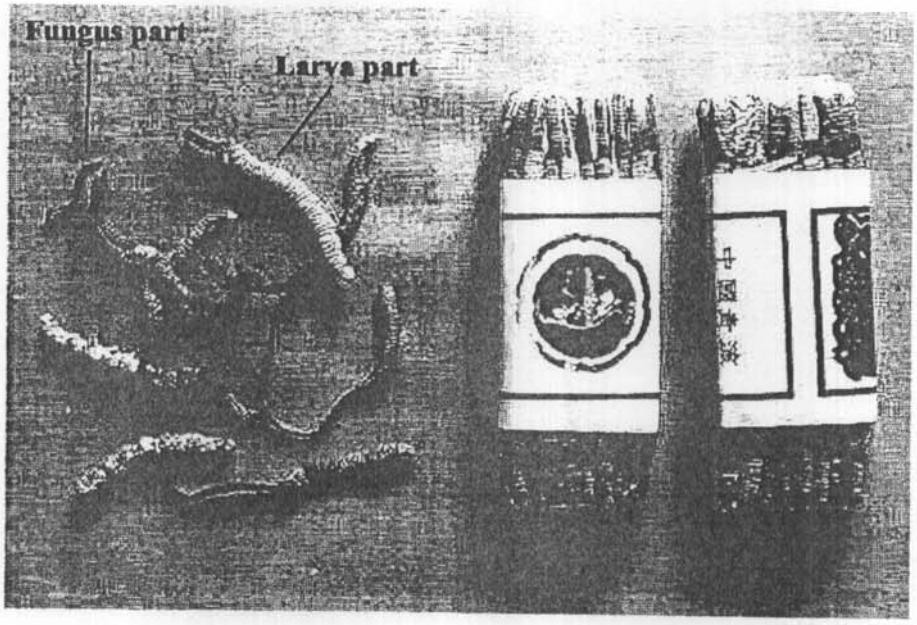
sinensis (Berk.) Sacc. parasitized on the larva of *Hepialus armoricanus* Oberthru (Ou, 1995) (Figure 7.1). In traditional Chinese medicine, *Cordyceps sinensis* is relatively nontoxic and has an antiasthmatic effect and is also an anticancer agent (Huang, 1993). It can be used to regulate and ensure the normal functioning of various parts of the body, strengthen the immune system and promotes overall vitality and longevity. Usually, it is utilized as a nutritional additive to be cooked with meat. Some literature reported that the fungus and larva part of the *Cordyceps sinensis* have different chemical constituents (Xu *et al.*, 1998). Such a difference may be originated from land of production and climate.

In *Cordyceps sinensis*, several types of chemicals such as nucleotide, ergosterol, mannite and amino acid (Xu *et al.*, 1988) have been identified successfully with modern analytical instruments. However, it is quite difficult to find out a simple method and an optimal condition to analyze all components at the same time. It is because the concentrations of the active ingredients in Chinese herbs are quite low. With the help of hyphenated instruments and chemometrics techniques, simultaneously qualitative and quantitative analysis of Chinese herbs becomes possible. In recent years, the development of hyphenated instruments is very fast. Different types of analytical devices can interface together to form a new type of instrument. High performance liquid chromatography coupled with diode array detector (HPLC-DAD) and gas chromatography interfaced to Fourier transform infrared spectrometer (GC-FTIR) or mass spectrometer (GC-MS) are examples of this kind. The advancement in hyphenated instruments leads to the development of new methods for data interpretation and evaluation (Toft, 1995). Commonly used methods for such purpose include evolving

Figure 7.1 Appearance of *Cordyceps sinensis* in (a) nature and (b) dried form.



(a)



(b)

Source: Pharmacopoeia Commission of the Ministry of Public Health, P.R. China
"A Coloured Atlas of The Chinese Materia Medica Specified in Pharmacopoeia of The People's Republic of China", 1995 Edition
Joint Publishing (H.K.) Co., Ltd. - Hong Kong, pp.127 (1996)

factor analysis (EFA) (Maeder and Zilian, 1988; Keller and Massart, 1992a), fixed size moving window evolving factor analysis (FSMW-EFA) (Keller and Massart, 1991) and heuristic evolving latent projections (HELP) (Kvalheim and Liang, 1992; Liang *et al.*, 1992).

Qualitative and quantitative analysis with HPLC-DAD system is much better than with conventional single-wavelength detector. The diode array detector (DAD) provides a way for multiwavelength detection and performs peak purity assessment on chromatographic peaks (Wickham, 1993). In practice, the capacity of DAD is not fully exploited. Several problems are encountered with the use of DAD for chemical analysis. For example, in many laboratories, DAD is only utilized as a single-wavelength detector. Chromatogram from a single wavelength cannot represent all components in a sample. Although some users try to monitor chromatogram with two or three wavelengths, only limited amount of information can be extracted. This may lead to wrong conclusion in the analysis. Wavelength selection is another problem to be found in HPLC-DAD study. During the operation, the optimal wavelength which gives the maximum absorbance for a particular component is usually selected for producing chromatograms of the sample. As we know that different substances have different responses on different wavelengths. Measurements from a few groups of wavelength cannot provide enough data for qualitative and/or quantitative analysis. As a result, concentration of a particular component in the sample will be either over-estimated or under-estimated. The third problem is arisen from the peak purity assessment. This function can be found in most of commercially available diode array software packages. The manufacturers use their own algorithm such as comparison of normalized spectrum

(Hewlett-Packard, 1994) to determine the purity of a particular chromatographic peak. However, only a simple value is provided from the calculation and it does not tell us the nature of the impurities in such peak. The last problem is the volume of experimental data. A typical run of experiment with HPLC-DAD can generate several mega-byte of data. We need an effective method to compress the data for storage or further processing.

In this work, the fast wavelet transform heuristic evolving latent projection (FWT-HELP) as described in the previous chapter was applied to analyze of water soluble constituents, nucleosides, in *Cordyceps sinensis* with HPLC-DAD. A comparison of the major chemical components in fungus and larva parts was performed. The results derived from FWT-HELP algorithm for the samples were compared with standards and further confirmed with GC-MS studies. It was found that our proposed method is a powerful method for quantitative and qualitative analysis of Chinese herbs.

7.2 Method of Investigation

FWT-HELP algorithm is a combination of fast wavelet transform and heuristic evolving latent projections. FWT is acting as a pre-processing step for the HELP treatment which involves data compression and denosing on experimental data. HELP is a type of factor analysis method that used widely in chemometrics. The method can be employed to determine the minimum number of components from HPLC-DAD data and to resolve chromatograms and spectra for individual components. With the help of

FWT, sensitivity of the HELP algorithm is improved especially at low sample concentration and low signal-to-noise ratio (SNR).

Heuristic evolving latent projections algorithm was first proposed by Kvalheim and Liang in 1992 for analyzing data from liquid chromatography with diode array detection (LC-DAD) (Kvalheim and Liang, 1992; Liang *et al.*, 1992). It is classified as a factor analysis or principal components analysis (PCA) techniques for multivariate data analysis. This algorithm was demonstrated to be better than the conventional methods such as EFA, ITTFA and AR (Toft and Kvalheim, 1994; Grung and Kvalheim, 1995). HELP has been applied successfully in an LC-UV study involving drug isomers (Liang *et al.*, 1992) and chlorophyll degradation experiments (Liang *et al.*, 1993b) and HPLC-DAD study of polycyclic aromatic hydrocarbons (PAHs) content in city air of Hong Kong (Shen *et al.*, 1997; 1998a-b).

In this work, the Daubechies wavelet, D_{16} (Daubechies 1992) which was shown to give optimal performance in Chapter 6 was selected in the wavelet computation. Besides, the translation-rotation transformation (TRT) method (Hayes *et al.*, 1973; Chau and Tam, 1994) and the coefficient position retaining (CPR) method as developed in Chapter 4 were also applied to solve the side-loop and number of data point problems. The former problem occurs when there is a sudden data change in the data during periodical extension at both end of the data. The later one originates when the length of a data set is not equal to 2^p with p for any positive integers. Chromatograms at each wavelength of the HPLC-DAD data were processed with FWT. The scale coefficients $\boldsymbol{C}^{(J)}$ at resolution level 1 for each wavelength from the FWT treatment are recombined

together to form a new data set. Then, such data were analyzed with the HELP algorithm to determine the minimum number of components present and to resolve pure chromatograms and spectra of the individual components.

7.3 Experimental Section

Experimental HPLC-DAD data of *Cordyceps sinensis* were provided by Prof. Y. Z. Liang from the College of Chemistry and Chemical Engineering, Hunan University, P. R. of China. Details of the sample preparation and experimental conditions are given as follows.

7.3.1 Samples Preparation

Samples of *Cordyceps sinensis* from Tibet Autonomous Region of P. R. China was brought from a Chinese herb store (Jiu Zhi Tang, Changsha, P. R. China). The water soluble constituents, nucleosides, were extracted by the following method. 0.5 g of fungus part of *Cordyceps sinensis* was added to 20 mL of distillated water at room temperature for the extraction of nucleosides. The sample-water mixture was placed into an ultrasonic bath for 2 hours (Guo *et al.*, 1998). Then, the sample-water mixture was filtered and the filtrate was vacuum-dried. Finally, the residue was dissolved in 10 mL of methanol. Similar extraction procedures were also performed on 0.5 g of the larva part.

7.3.2 Apparatus and Chemicals

HPLC-DAD data were acquired from a Shimadzu LC-4A HPLC system (Shimadzu Scientific Instruments Inc., MD, US) coupled with a Shimadzu SPD-M6A photodiode array detector (Shimadzu Scientific Instruments Inc., MD, US). All separations were performed on a 250 × 4.6 mm i.d. Zorbax-ODS column (DuPont, DE, US). GS-MS data were acquired on a Simadzu GC-17A GC system (Shimadzu Scientific Instruments Inc., MD, US) coupled with a Shimadzu QP-5000 MS system (Shimadzu Scientific Instruments Inc., MD, US). All separations were performed on a 30 m × 0.25 mm i.d. HP-50+ (OV-17) column (Hewlett-Packard, Germany). The sample-water mixture was extracted with an ultrasonic cleaner CQ250 (Ultrasonic Instruments Factory of Shenghai, P. R. China).

Standards of adenosine and uridine were acquired from Sigma (Sigma-Aldrich Corporation, MO, US). All solvents for HPLC measurements are of HPLC grade while the other solvents used in this work are of analytical grade.

7.3.3 Experimental Conditions

The mobile phase for HPLC measurement was a mixture of methanol-water (15:85 v/v). The flow rate was 1.0 mL/min. The column temperature was maintained at 25 °C. The spectra were collected in the wavelength range of 195 to 312 nm with a spectral resolution of 1 nm. The scanning rate was set to be 0.32 s. 5 μL of sample

solution from fungus and larva parts were injected, respectively, into the HPLC-DAD system.

In the GC-MS measurement, the carrier gas was high purity nitrogen gas. The flow rate and pressure was maintained at 0.5 mL/min and 10^{-7} mmHg respectively. The column temperature programming was set from room temperature to 285 °C at the rate 80 °C/min. 1 μL of sample solution from fungus and larva parts were injected, respectively, into the GC-MS system.

7.3.4 Data Analysis

The data obtained for the fungus part consists of 118 chromatograms and 3,748 spectra. The data for larva part consists of 118 chromatograms and 3,610 spectra. The HPLC-DAD data were first converted to the corresponding ASCII codes before processing. All computations were carried out using the Fast Wavelet Transform Heuristic Evolving Latent Projections (FWT-HELP) Toolbox Version 1.0 as developed in this work. It was coded in MATLAB® for UNIX® Version 5.1 (The MathWorks, Inc., 1996) under IRIX™ 6.2 environment on a SGI Indigo^2 Power Extreme workstation (Silicon Graphics, Inc., CA, US) with a 75MHz MIPS R8000™ processor. Details of the FWT-HELP toolbox can be found in Appendices 6.1 and 6.2 and Figure 6.3.

7.4 Results and Discussion

Figures 7.2 and 7.3 showed the chromatograms of the fungus and larva parts of *Cordyceps sinensis* at selected wavelengths. As we pointed out in the previous section, we cannot just use one wavelength for qualitative analysis. In Figures 7.2 and 7.3, one can find that the chromatographic peaks of some components appears at a particular range of wavelengths only. If a wrong wavelength is selected for qualitative analysis, the number of components present in the sample will be underestimated. With the use of our proposed FWT-HELP method, spectro-chromatographic data on a range of wavelengths can be employed for analysis to improve the accuracy of analysis. For the problem of peak purity, it can be solved by using both the eigenvalue plot and latent projective graph. To illustrate how FWT-HELP works in the real case, an example calculation on our sample will be shown on the next section.

7.4.1 Sample FWT-HELP Computation

Figures 7.4a and 7.4b show, respectively, the chromatograms from 11.0 to 14.0 minutes and the spectra from 195 to 312 nm of the fungus part of *Cordyceps sinensis*. Let $\boldsymbol{X}_{11.0-14.0}$ represents the HPLC-DAD data obtained in the above time interval. By using the conventional data analyzing method, it is quite difficult to identify the number of components from the chromatograms and spectra. With the use of the peak purity function as provided by the instrument such as HP ChemStation (Hewlett-Packard, Germany), it just gives an arbitrary number for the purity of a chromatographic peak. Such number only provides very limited amount of

Figure 7.2 Chromatograms of the fungus part of *Cordyceps sinensis* at wavelength (a) 234 nm, (b) 260 nm and (c) 294 nm.

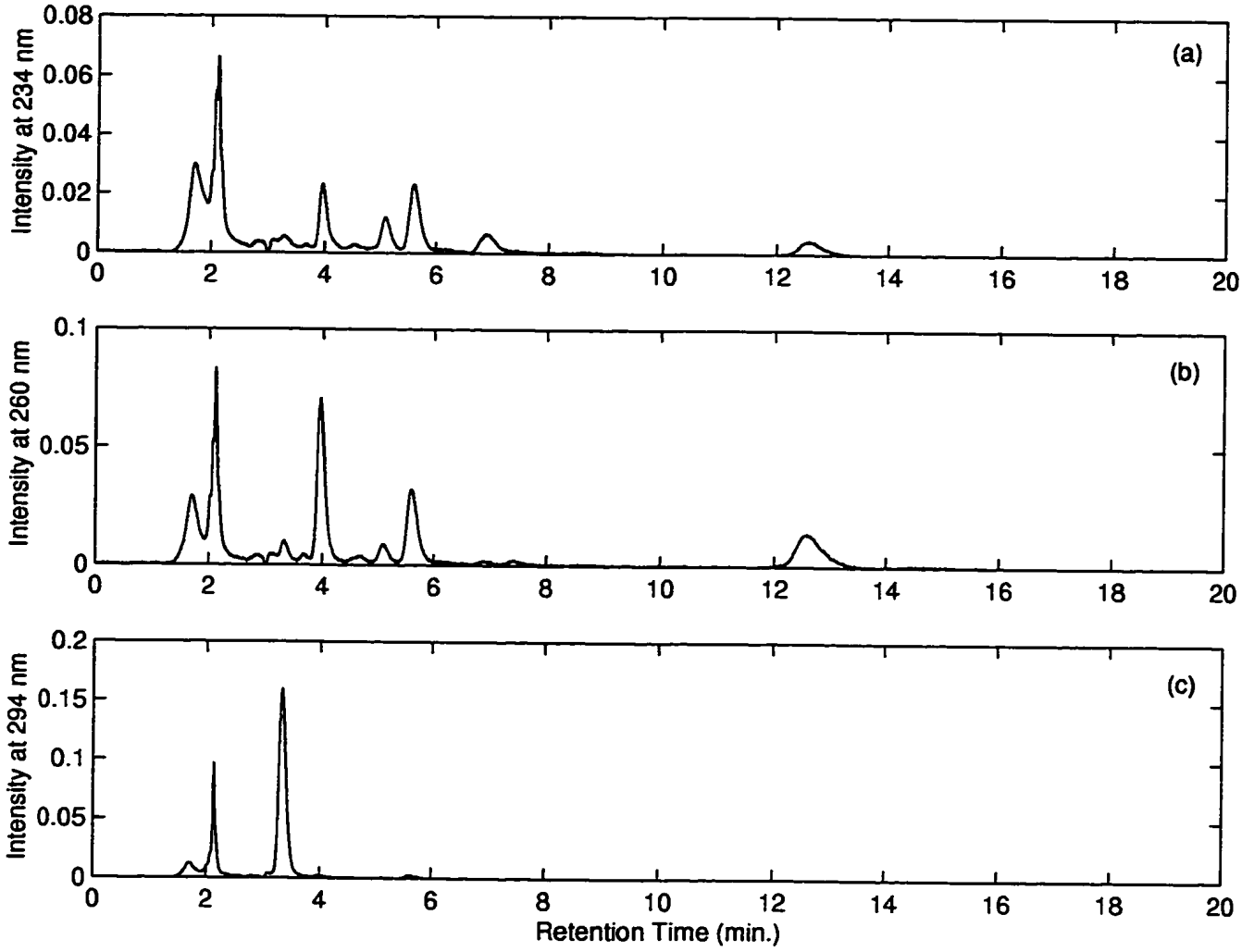


Figure 7.3 Chromatograms of the larva part of *Cordyceps sinensis* at wavelength (a) 234 nm, (b) 260 nm and (c) 294 nm.

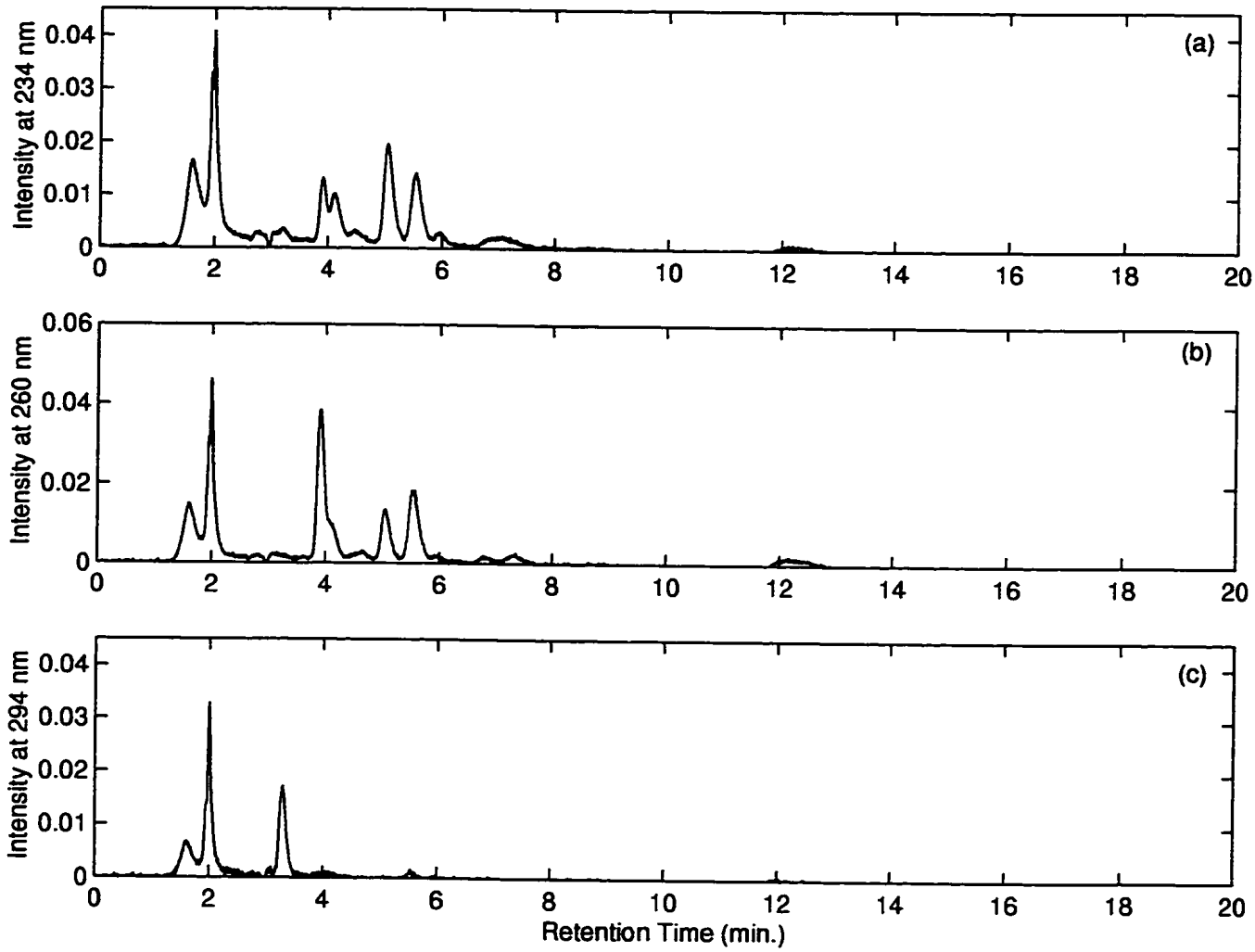
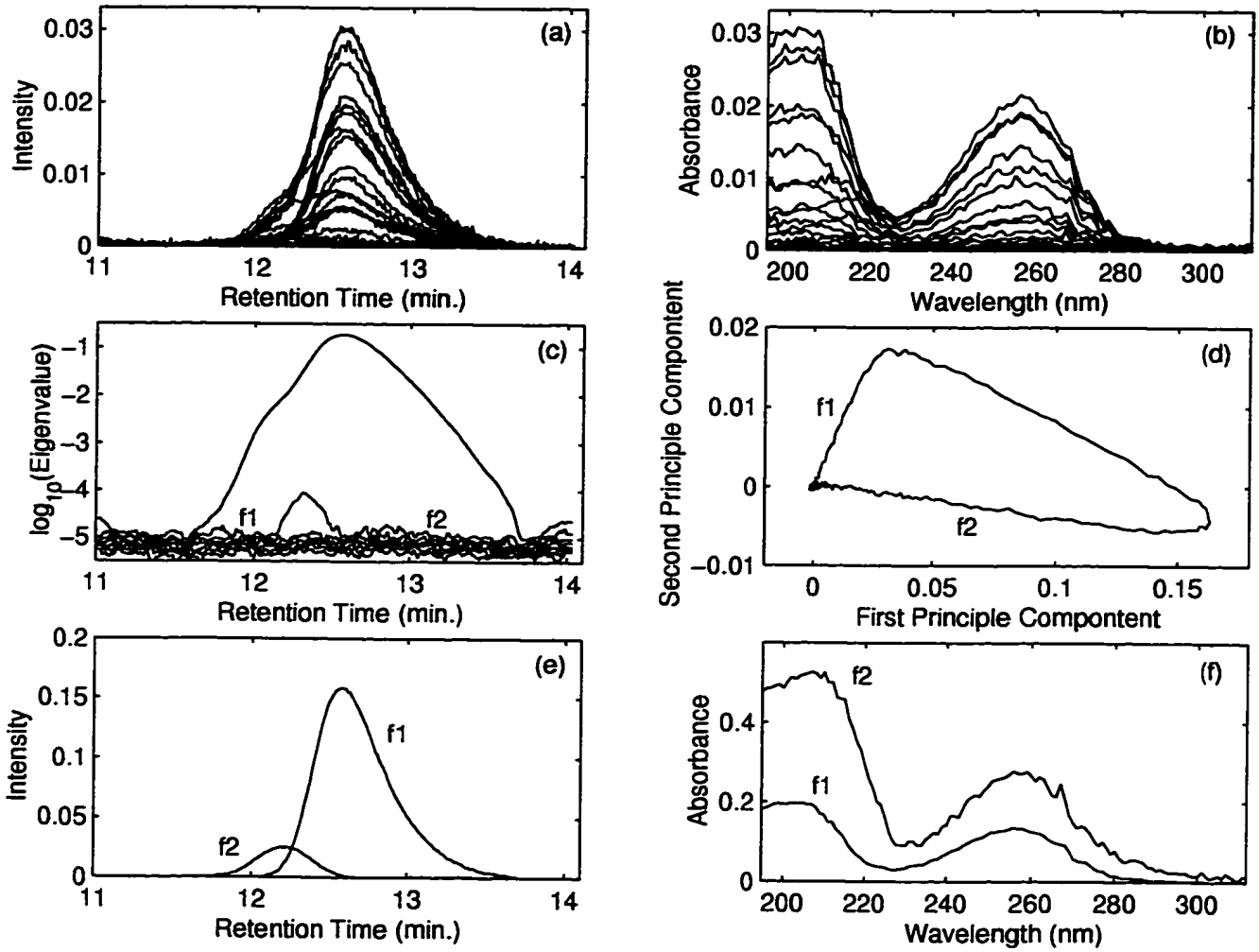


Figure 7.4 (a) The selective chromatograms of the fungus part from 11.0 to 14.0 minutes with an interval of 5 nm. (b) The spectra of the fungus part in the range of 195 to 312 nm with an interval of 0.1 minute. (c) The corresponding eigenvalue plot with a window size of 7 and (d) the latent project graph of $X^{WT}_{11.0-14.0}$ as well as (e) the resolved chromatograms and (f) spectra for components f1 and f2 in the fungus part which are obtained with FWT-HELP treatment.



information on peak purity and requires experience of the investigator to estimate the results.

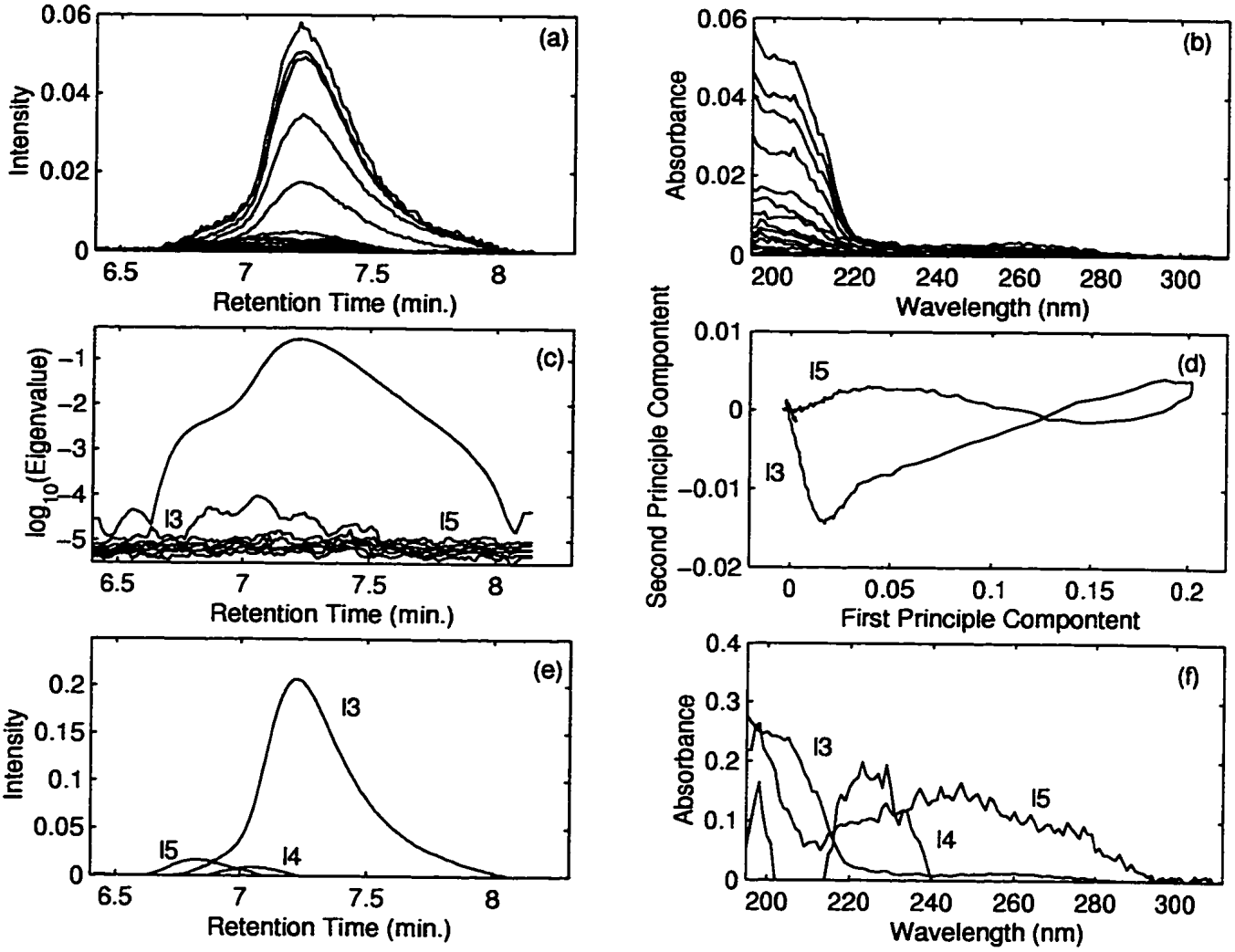
In our proposed method, $X_{11.0-14.0}$ was transformed by the Daubechies D_{16} wavelet function at the first resolution level on the chromatographic axes. The scale coefficients $C_{\lambda}^{(1)}$, with λ being the measured wavelengths, were combined to form a new data set $X_{11.0-14.0}^{WT}$ for HELP analysis. Then, an eigenvalue plot (Figure 7.4c) was obtained by performing the FSMW-EFA scheme on $X_{11.0-14.0}^{WT}$ with a windows size of 7. Two eigenvalue lines were found to be above the noise level. As a result, at least two components f1 and f2 could be identified from $X_{11.0-14.0}^{WT}$. This result can be further confirmed by the latent projective graph (Figure 7.4d). Two straight line segments which passed through the origin can be identified and these represents the two selected regions of the required components. The selective regions for components f1 and f2 are located at 11.8 to 12.1 and 12.8 to 13.5 minutes and represent by $X_{11.8-12.1}^{WT\_selective}$ and $X_{12.8-13.5}^{WT\_selective}$ respectively. From the results of both the eigenvalue plot and the latent projective graph, the system under study is a typical two- component case. The pure spectra of the corresponding components, S_{f1}^{pure} and S_{f2}^{pure} (Figure 7.4f), can be obtained by performing a singular value decomposition (SVD) on $X_{11.8-12.1}^{WT\_selective}$ and $X_{12.8-13.5}^{WT\_selective}$, and are represented by the first column of the loading matrix $P_{11.8-12.1}^{f1}$ and $P_{12.8-13.5}^{f2}$. Afterward, the pure chromatograms of the corresponding components, C_{f1}^{pure} and C_{f2}^{pure} (Figure 7.4e), are determined by the following equations:

$$C^{pure} = XS^{pure}\left(\left(S^{pure}\right)^{t} S^{pure}\right)^{-1} \tag{7.1}$$

where $C^{pure} = \left[C_{f1}^{pure}, C_{f2}^{pure}\right]$ and $S^{pure} = \left[S_{f1}^{pure}, S_{f2}^{pure}\right]$.

For a more complicated system, the component stripping approach should be used (Liang and Kvalheim, 1993). Resolution can only be performed by using a stepwise stripping procedure. To investigate the HELP analysis with more than two components, HPLC-DAD data of the larva part of *Cordyceps sinensis* was used. Let $X_{6.4-8.1}$ represents the HPLC-DAD data of the larva part with the chromatograms (Figure 7.5a) from 6.4 to 8.1 minutes and the spectra (Figure 7.5b) from 195 to 312 nm. Again, both the chromatograms and spectra cannot provide enough information to determine the peak purity. The data was transformed with wavelet transform under the same condition as that described previously and became $X_{6.4-8.1}^{WT}$. Figures 7.5c and 7.5d showed the eigenvalue plot and the latent projective graph of $X_{6.4-8.1}^{WT}$. The pattern of the latent projective graph especially the closed loop indicated that more than two components are identified. The selective region for components 13 and 15 were located at 6.7 to 6.8 and 7.3 to 7.4 minutes and the zero concentration region for components 13 and 15 were located at 7.3 to 8.0 and 6.6 to 6.8 minutes. With these information, one can determined the required rotation vectors r_{13} and r_{15} from the score matrix $T_{6.7-6.8,7.3-8.0}^{13}$ and $T_{7.3-7.4,6.6-6.8}^{15}$. The first part of the subscript represents the position of the selective region while the latter one represents the position of the zero concentration region of the component. The pure chromatograms of the related components, C_{13}^{pure} and C_{15}^{pure} (Figure 7.5e), can be calculated by the following equations:

Figure 7.5 (a) The selective chromatograms of the larva part from 6.4 to 8.1 minutes with an interval of 5 nm. (b) The spectra of the larva part in the range of 195 to 312 nm with an interval of 0.1 minute. (c) The corresponding eigenvalue plot with a window size of 7 and (d) the latent project graph of $X^{WT}_{6.4-8.1}$ as well as (e) the resolved chromatograms and (f) spectra for components 13, 14 and 15 in the larva part which are obtained with FWT-HELP treatment.



$$C_{13}^{pure} = T_{6.4-8.1} r_{13} \tag{7.2}$$

and

$$C_{15}^{pure} = T_{6.4-8.1} r_{15} \tag{7.3}$$

where $T_{6.4-8.1}$ represents the score matrix of $X_{6.4-8.1}^{WT}$. Then, the pure spectra of the corresponding components, S_{13}^{pure} and S_{15}^{pure} (Figure 7.5f), can be obtained from their relevant selective regions. After determining the pure chromatograms and spectra for components 13 and 15, the component stripping approach was applied on $X_{6.4-8.1}^{WT}$ to remove the contribution from components 13 and 15 as follows:

$$X_{6.4-8.1}^{WT\,new} = X_{6.4-8.1}^{WT} - C_{15}^{pure}\left(S_{15}^{pure}\right)^{t} - C_{15}^{pure}\left(S_{15}^{pure}\right)^{t}. \tag{7.4}$$

Finally, the HELP algorithm was applied to $X_{6.4-8.1}^{WT\,new}$ to resolve the pure chromatogram and spectrum of component 14.

In the above two examples, we have shown the procedures for resolving the pure chromatogram and spectrum for the individual components with FWT-HELP algorithm. By adopting similar analysis method, one can resolve chromatograms and spectra for the remaining peaks in the HPLC-DAD data of *Cordyceps sinensis*. Figures 7.6 and 7.7 show the overall resolved chromatograms of both fungus and larva parts of *Cordyceps sinensis* and Figures 7.8 and 7.9 give the spectra of the corresponding components in the fungus and larva parts. Usually, the numbers of chemical components are considered to be the same for the fungus and the larva part. Yet, 17 components were resolved in the fungus part. On the other hand, 19 components were resolved from the data of the larva part. The results as shown in Figures 7.6b and 7.7b are for reference only. The resolved chromatographic peaks for the components l13 to l19 and f11 to f17 may not represent a true chemical

Figure 7.6 (a) The overall resolved chromatograms of the fungus part with the FWT-HELP algorithm and (b) a magnified plot of (a).

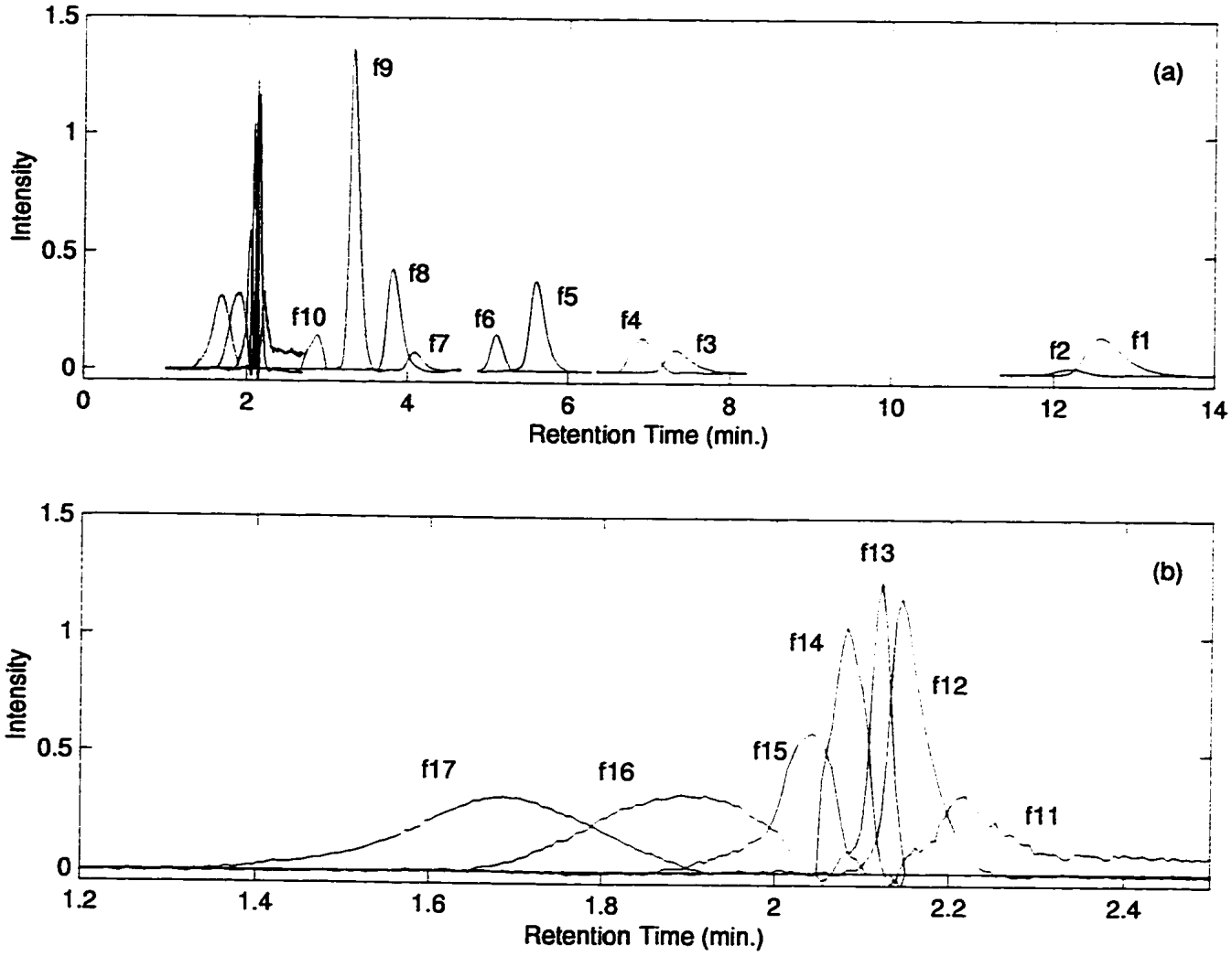


Figure 7.7 (a) The overall resolved chromatogram of the larva part with the FWT-HELP algorithm and (b) a magnified plot of (a).

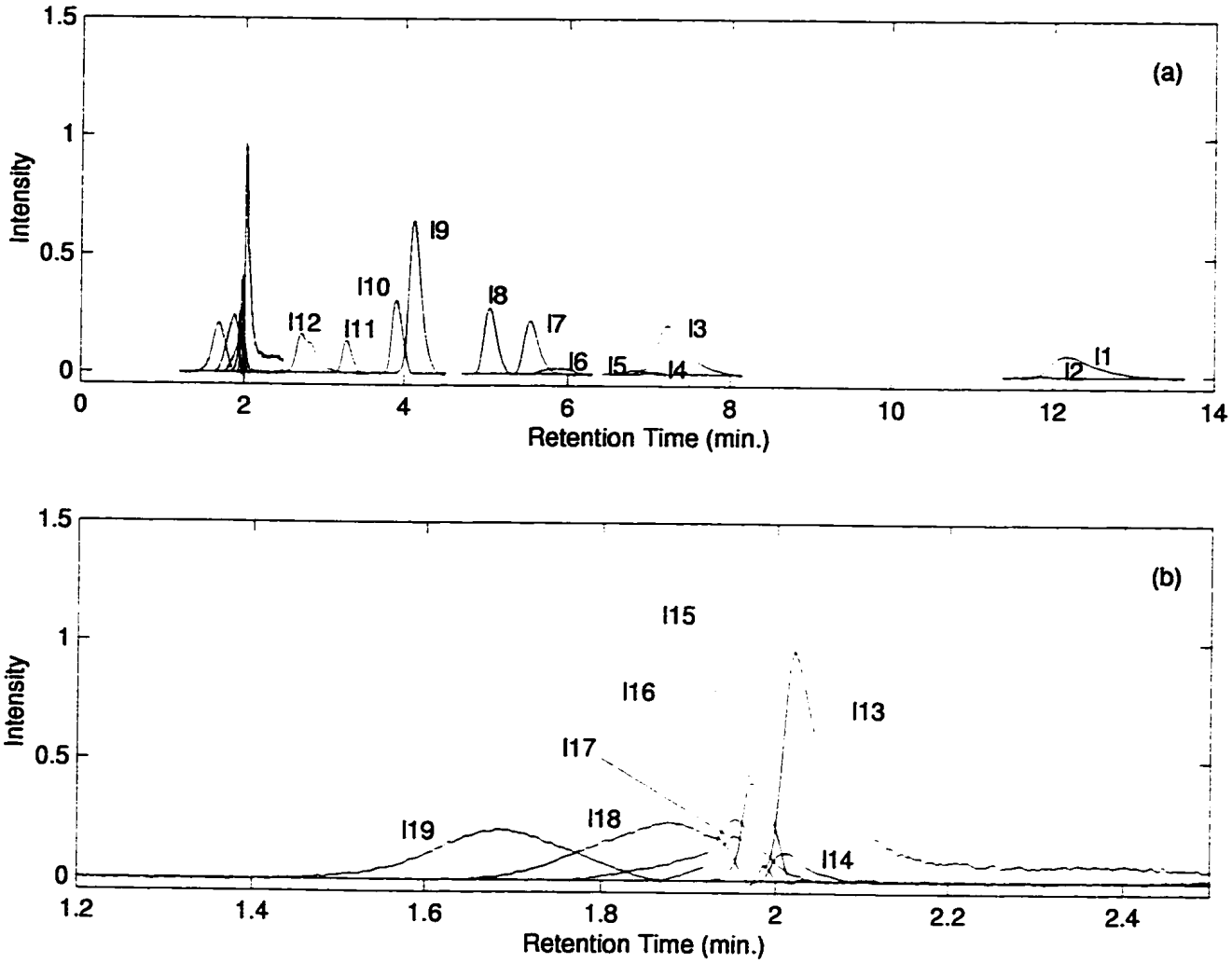
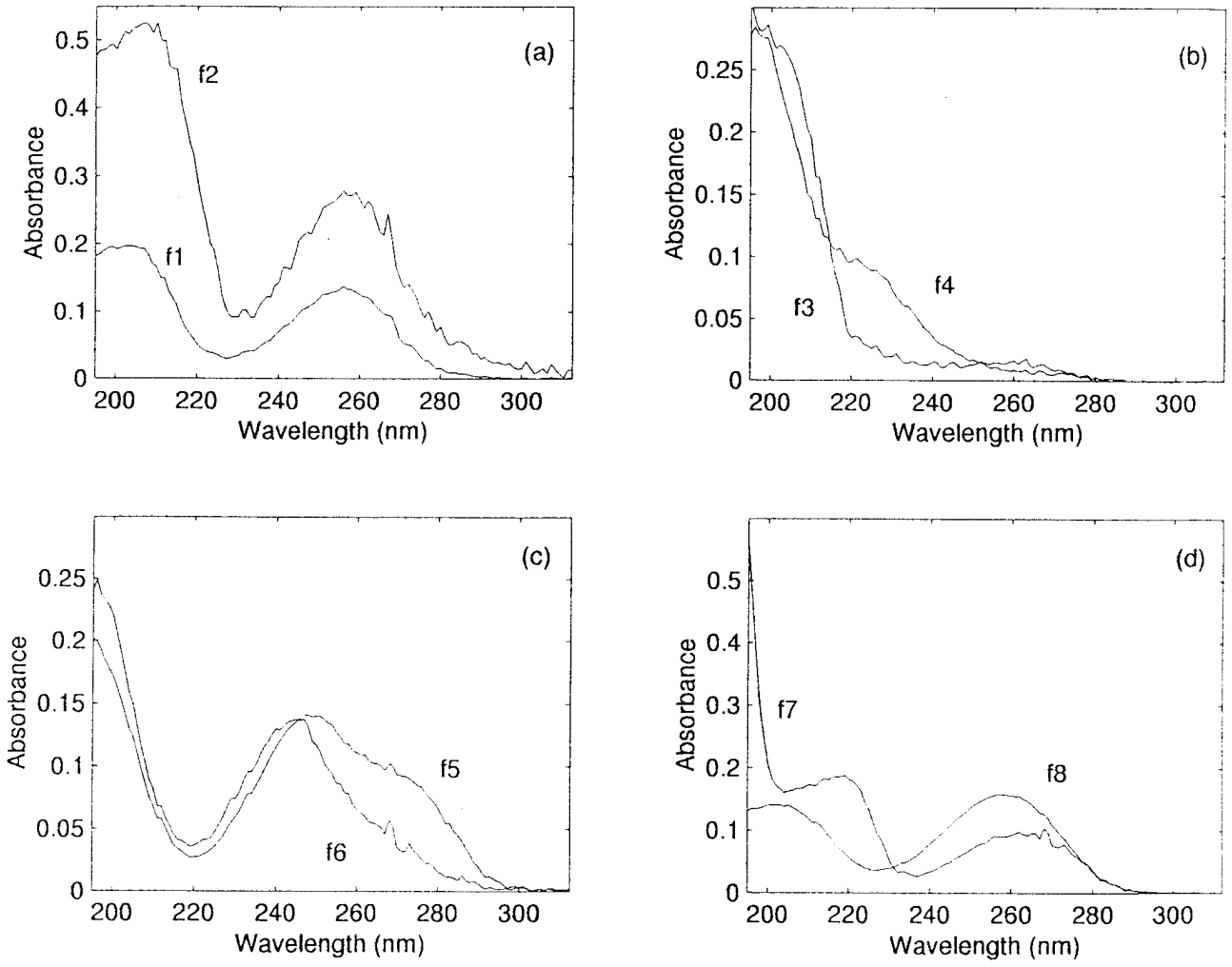


Figure 7.8 Resolved spectra of the 17 constituting components in the fungus part.



(cont.)

(cont. Figure 7.8)

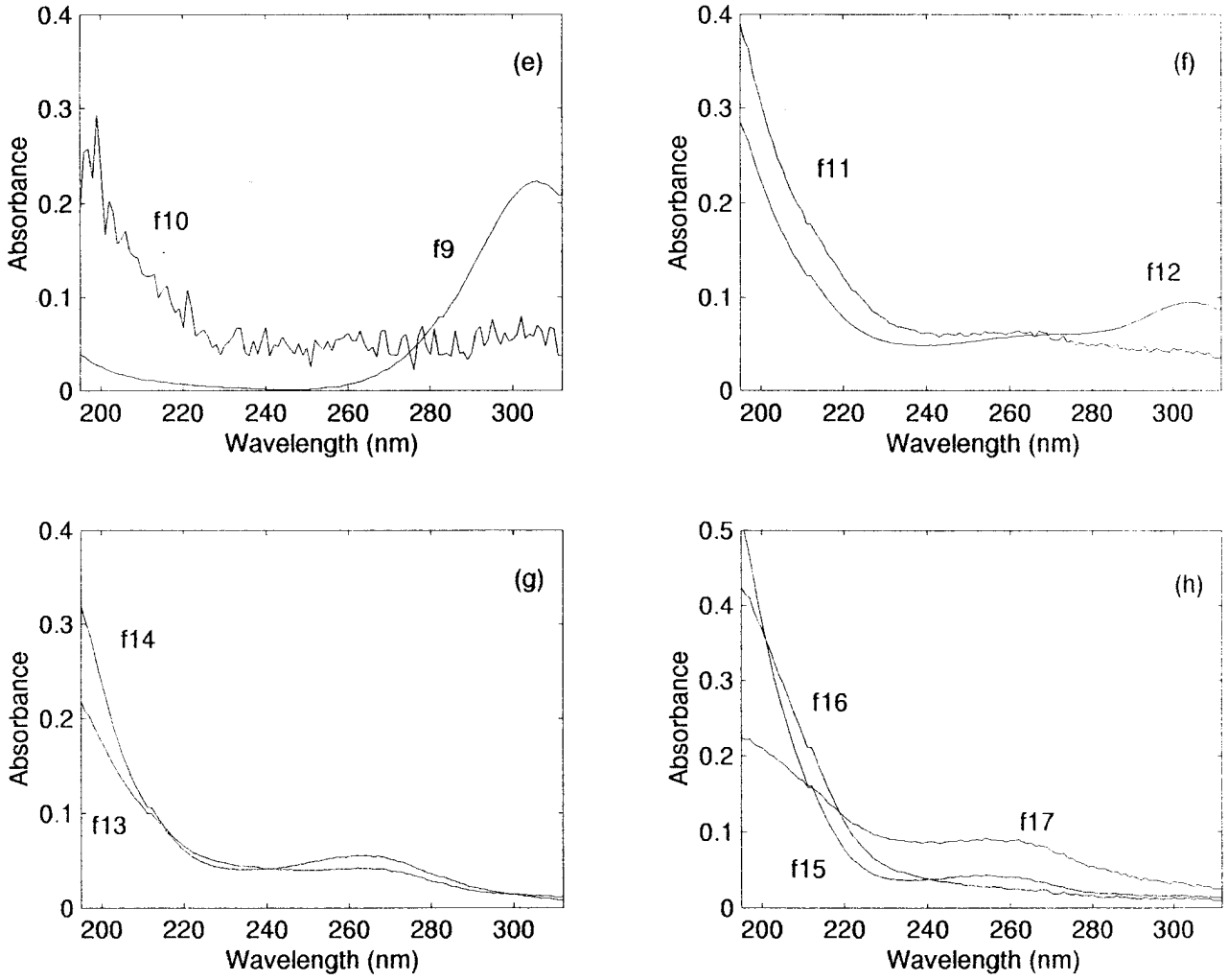
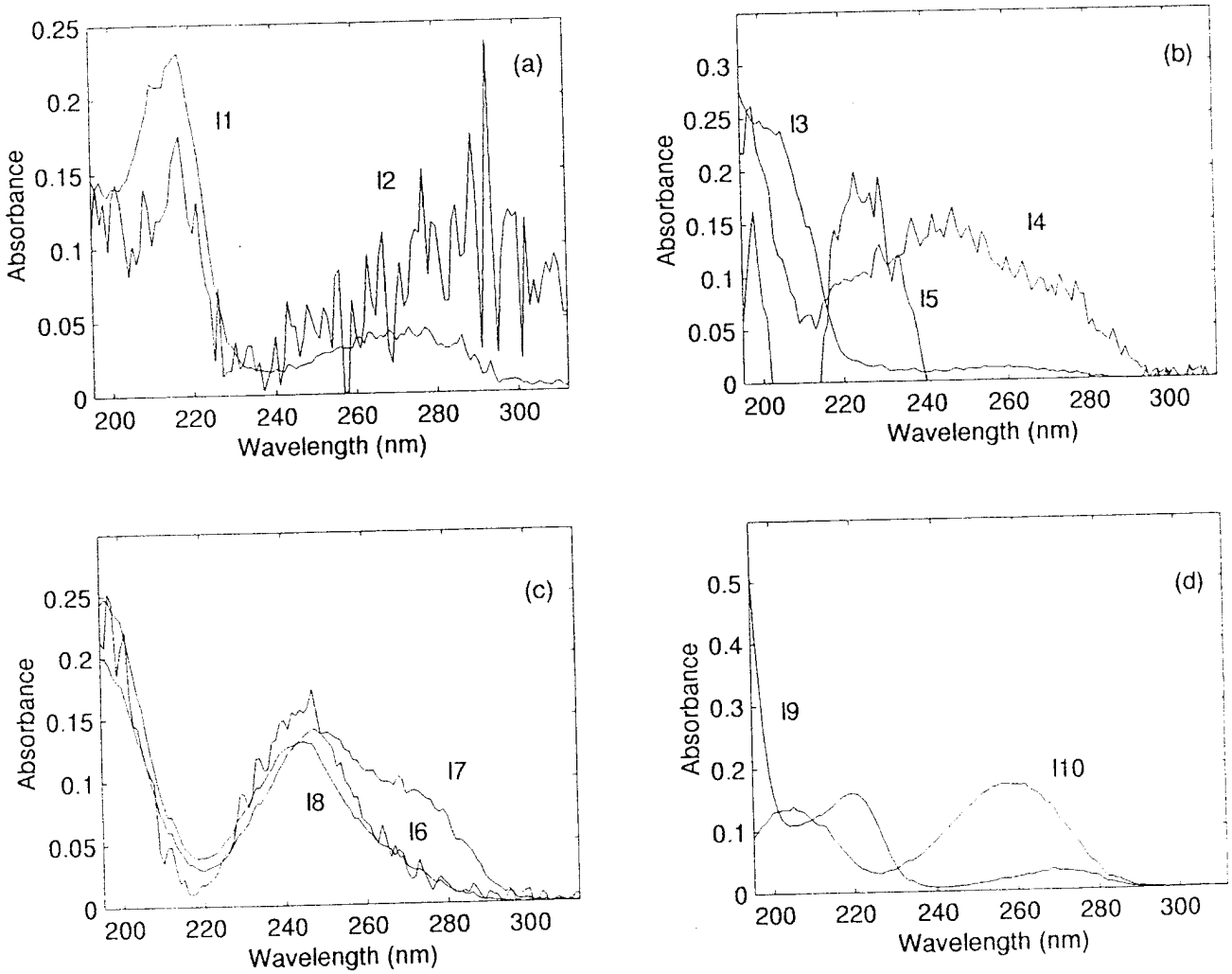
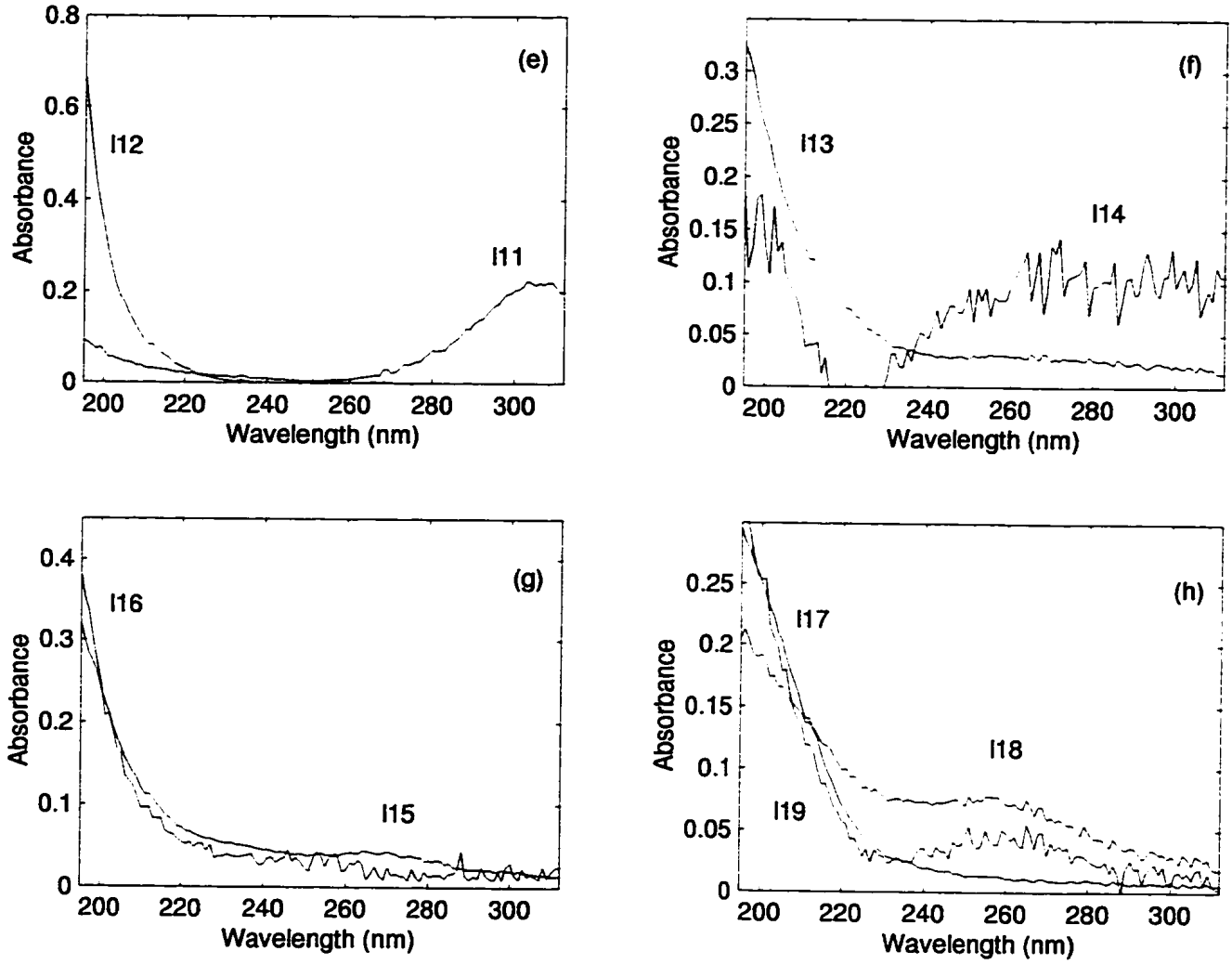


Figure 7.9 Resolved spectra of the 19 constituting components in the larva part.



(cont.)

(cont. Figure 7.9)



components. It could be reflected by the noisy resolved spectra in both Figures 7.8 and 7.9. It was because the chromatographic peaks at such region were highly overlapped. Another suitable mobile phase should be used to separate the chromatographic peaks at that regions. So, such regions will be ignored in the following discussion.

By observation, only 10 and 12 components could be identified roughly from the chromatograms of the fungus (Figure 7.2) and larva (Figure 7.3) parts. Although the results in Figures 7.6b and 7.7b are quite good, it only provide some information on the purity of the chromatographic peaks at that region. In order to have a better results, a suitable mobile phase should be adopted in another HPLC-DAD experiments to separate component in such region. Since the concentrations of some components such as f2, l2, l4 and l5 are very low, it is very difficult to identify them directly from the original chromatograms. By comparing the resolved UV spectra of all components in both fungus and larva parts, we found that there are some differences in their chemical compositions. All components except f1 in the fungus part could be found in the larva part.

7.4.2 Peak Purity Assessment

In terms of peak purity assessment, our proposed method is much better than that from some of the software of the HPLC-DAD instrument such as the HP ChemStation (Hewlett-Packard, Germany). In these kinds of softwares, one usually obtain an arbitrary number to represent the purity of a peak. A set point value should

be assigned prior to the peak purity assessment. A false result on peak purity will be provided for a bad value. Setting and suitable set point value also requires extensive knowledge and experience and it varies for different cases. However, the FWT-HELP method can provide an accurate peak purity test on the HPLC-DAD data and provide more information such as the minimum number of impurities present and their nature.

Several chemometrics techniques have been proposed to check the peak purity in HPLC-DAD system (Seaton *et al.*, 1986; Castledine *et al.*, 1992; Keller and Massart, 1992b; Gilliard and Ritter, 1997; Zissis *et al.*, 1997). Most of these methods are based on the change in the eigenvalue plot only. However, with the FWT-HELP algorithm, both eigenvalue plot and latent projective graph are utilized. Peak purity is mainly reflected from the pattern of the latent projective graph. If a peak consists only one component and without any background, a straight line which passes through the origin will be observed. For a peak consists with two components, two separated straight lines pass through the origin and a loop will be observed (Figure 7.4d). The loop segment is due to the change of the overall UV spectrum of the components under the chromatographic peak. For a peak consists of more than two components, a more complicated latent projective graph will be observed (Figure 7.5d). With the help of the component stripping approach, a new latent projective graph is obtained at each stage of computation so that one can resolve individual components in a step-by-step manner. In some cases, attention has to be paid on analyzing the latent projective graph. As stated in a previous study by Keller *et al.* (Keller *et al.*, 1992), the straight line for a component with low concentration is very

short. We need to enlarge the area near to the origin to identify the correct selective region of such component. Otherwise, wrong interpretation will result. Other artifacts in the latent projective graph will also affect the peak purity assessment. More information can be found in the work by Keller *et al.* (Keller *et al.*, 1992).

7.4.3 Roles of Fast Wavelet Transform in HELP Computation

Fast wavelet transform performs two major roles in the HELP computation. It compresses the HPLC-DAD data to a lower resolution level and denoises at the same time. FWT is a linear transformation (Daubechies, 1992) and it will cause only mild information loss after data compression. The HELP computation prefers to be carried out on a data set at a lower resolution or dimension because a lot of time is spent on the eigenvalue computation (Walczak and Massart, 1997c). A smaller data set can definitely reduce the computational time. As demonstrated in the previous section, the data dimension for the fungus and larva parts are 3,748 × 118 and 3,600 × 118 respectively. A typical chromatographic peak has 200 to 500 data points. By performing FWT on the data set at resolution level 1, the data size can be reduce by 50%. So, the computational time on eigenvalue calculation can be reduced by 50%.

It should be mentioned that over the denoising property associated with the FWT treatment do improve the sensitivity of the HELP operation. It was found that after denoising the data set with FWT, the eigenvalue plot can be obtained with a smaller window size. Figures 7.10 and 7.11 show a comparison of the eigenvalue plots and latent projective graphs of the data sets from the fungus and larva part. If

Figure 7.10 The eigenvalue plots for the fungus part from 11.0 to 14.0 minutes with a window size of (a) 7 and (b) 9 without FWT treatment. The eigenvalue plot (c) for the same data set which is processed with FWT and with a window size of 7. (d) and (e) The corresponding latent projective graphs as obtained without and with FWT treatment.

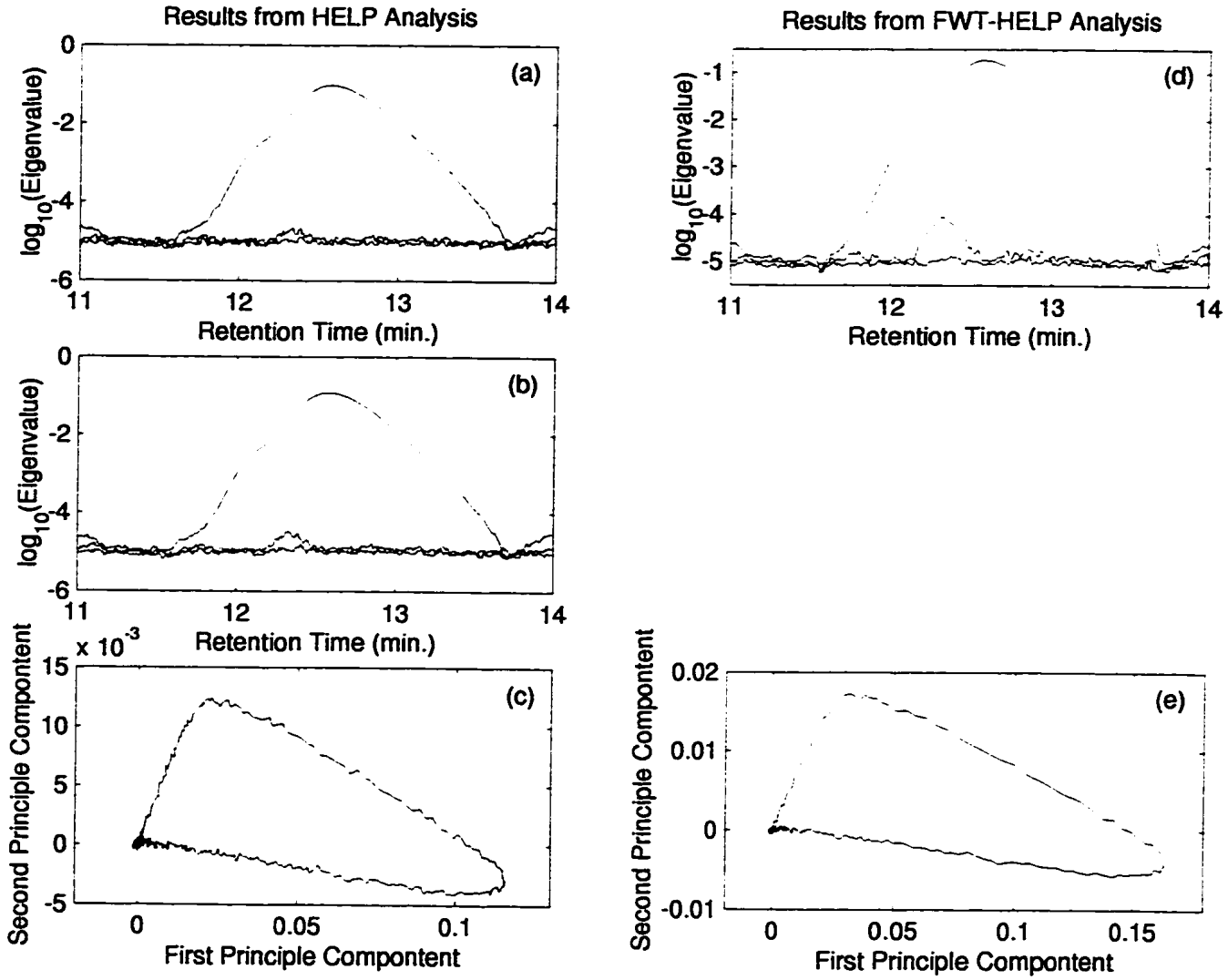
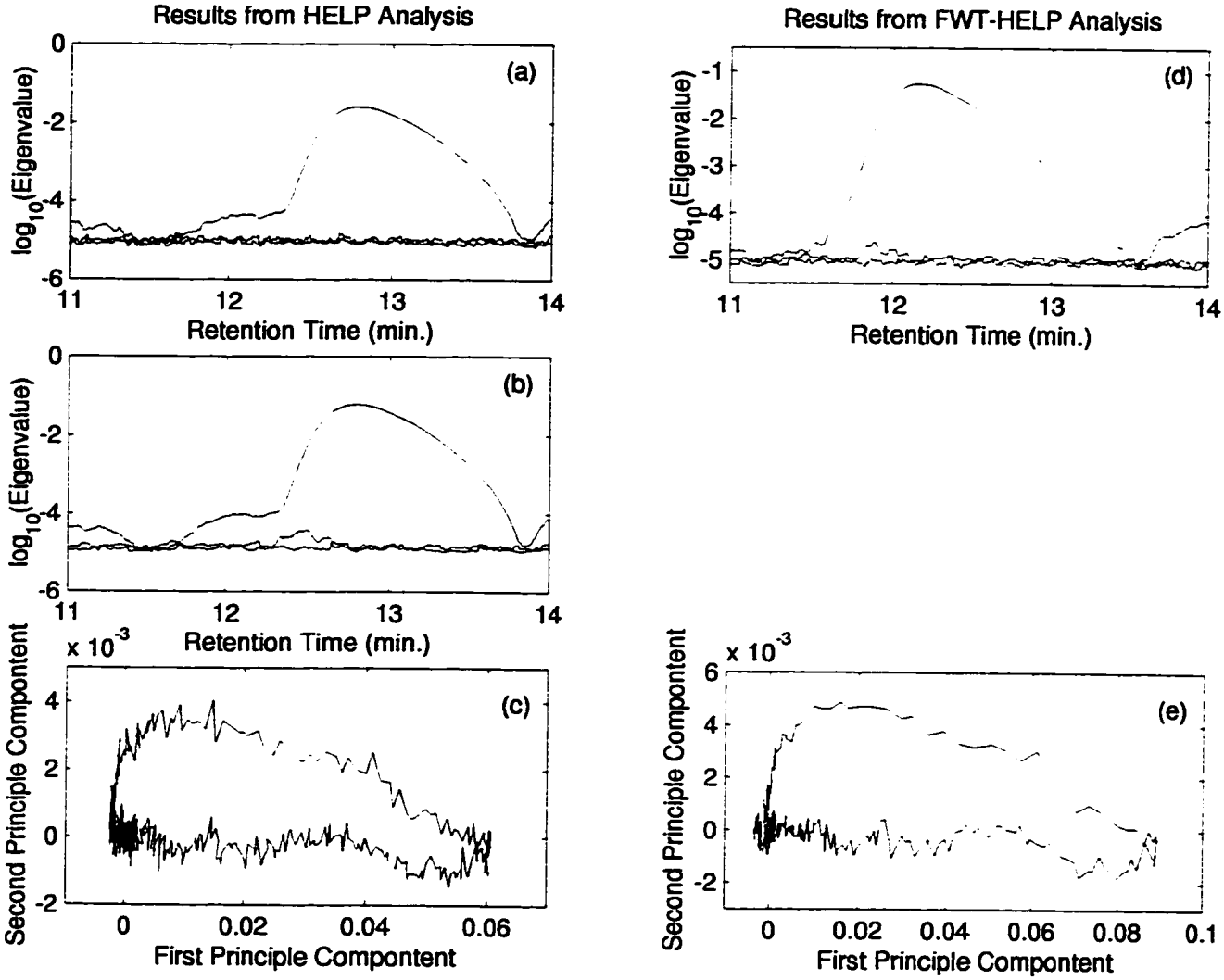


Figure 7.11 The eigenvalue plots for the larva part from 11.0 to 14.0 minutes with a window size of (a) 7 and (b) 17 without FWT treatment. The eigenvalue plot (c) for the same data which is processed with FWT and with a window size of 7. (d) and (e) The corresponding latent projective graphs as obtained without and with FWT treatment.



the relative concentration between components is very low, such as component l1 and l2 in Figure 7.7, it is quite difficult to detect them using the conventional HELP algorithm alone. Although a large window size can be employed to solve the problem, as described in Chapter 6, this would affect the selection of the selective region for a particular component and the quality of the resolved chromatograms and spectra. By employing FWT treatment, the relative eigenvalue ratio between the two eigenvalue lines can be improved. Such an improvement would benefit for the components at low concentration or signal-to-noise ratio. Besides, the latent project graph becomes less noisy to facilitate the determination of the selective regions.

7.4.4 Results from GC-MS Studies

In this work, GC-MS studies were also performed to confirm the major components in both fungus and larva parts of *Cordyceps sinensis* which have medicinal application. Table 7.1 shows the results from GC-MS studies. Six major components namely adenosine, hypoxanthine, inosine, thymine, thymidine and uridine were identified by the GC-MS method. In the previous section, we have shown that the two components, f1 and f2, can be resolved from $X_{11.0-14.0}$ by the WT-HELP algorithm. Similar results were also obtained from GC-MS study which indicates that f1 is adenosine and f2 is hypoxanthine. In order to attain the same result by using the conventional chromatographic methods, we need very long time to find out a suitable experimental condition. In this example, we have demonstrated that the FWT-HELP algorithm is a very powerful method on analyzing HPLC-DAD data of complex real system.

Table 7.1 Results of qualitative analysis of the fungus and larva parts of *Cordyceps sinensis* from GC-MS studies.

Compound Name	Peak number in fungus part	Peak number in larva part
Adenosine	f1	Nil
Hypoxanthine	f2	l1
Inosine	f4	l3
Thymine	f5	l7
Thymidine	f6	l8
Uridine	f8	l10

7.4.5 Results of Quantitative Analysis

Quantitative analysis for adenosine and uridine were also performed in this studies. By employing the two-way experimental data from HPLC-DAD, one can obtain a more accurate result in quantitative analysis because both the chromatographic and spectral information is available to assist the investigation. In the conventional method, a calibration curve is set up for quantitative analysis (Snyder and Kirkland, 1979). It is usually a plot of peak height or peak area against the concentration of sample injected. In general, the linear portion of the curve is chosen for quantitative analysis of the unknown sample. However, if the resolution between two chromatographic peaks is less than 1.5, it is quite difficult to determine their peak areas correctly. Commonly used methods to determine the peak area in such cases include drawing a perpendicular line or a tangent line to cut the peak (Snyder and Kirkland, 1979). But, these methods will introduce errors in the analysis and their performance depend on the relative concentration of components in the peak (Liang *et al.*, 1993c). In our work, peak volume of the resolved component instead of peak height or peak area was chosen for quantitative analysis. The peak volume is calculated by performing an integration of the chromatographic peak over a selective range of wavelength .

The calibration curves of adenosine and uridine were obtained by injection of various amount of standard adenosine and uridine into the HPLC-DAD instrument. We have found that adenosine and uridine show a linear relationship in the concentration range 7.425 to 297.0 μg/mL and 3.525 to 141.0 μg/mL respectively.

Let V and A represents the peak volume and the concentration of the sample respectively. The calibration curves of adenosine and uridine can be expressed as follows:

$$V^{adenosine} = -4.184A^{adenosine} + 3.098 \quad (\gamma = 1.000) \tag{7.5}$$

and $$V^{uridine} = -0.2698A^{uridine} + 2.220 \quad (\gamma = 0.9998) \tag{7.6}$$

where γ represents the regression coefficient. According to the results of the calibration curves, 0.15% of adenosine and 0.22% of uridine were found in the fungus part and 0.14% of uridine was found in the larva part. From the results of standard addition method, the recovery rate of adenosine and uridine in fungus part is 100.45% and 99.97% and that for uridine in larva part is 100.81%.

7.5 Conclusions

In conclusion, the FWT-HELP algorithm was applied successfully to analyze nucleosides in fungus and larva parts of *Cordyceps sinensis* with HPLC-DAD. FWT can be employed to compress and denoise the data set obtained from HPLC-DAD measurement and act as a pre-processing step of the HELP treatment. 17 and 19 components were resolved from the HPLC-DAD data of fungus and larva parts respectively with the FWT-HELP computation. 6 and 5 components were identified quantitatively from the fungus and larva parts, respectively, with GC-MS technique. Our study has demonstrated the advantages of the proposed method over the conventional ones in quantitative analysis of complex real systems.

CHAPTER 8

Summary of Results

Two classes of chemometrics techniques namely signal processing and factor analysis were applied successfully on chemical and biochemical studies. Wavelet transform (WT) which is a signal processing technique was employed in compression and spectral library search of infrared (IR) spectra and approximate derivative calculation of analytical signal. The heuristic evolving latent projections (HELP) algorithm which is a factor analysis technique was coupled with fast wavelet transform (FWT) to form a new chemometrics technique. This new technique was applied to analyze a Chinese herb medicine, *Cordyceps sinensis*. A brief summary of the findings in this work was given in the following paragraphs.

In Chapter 2, major signal techniques for spectral data compression in analytical chemistry were introduced. Five methods that include binary encoding, factor analysis, Fourier transform, spline and wavelet transform were discussed. Although Fourier transform was employed as the major technique in chemical studies for signal processing in the past, its status will be replaced by wavelet transform.

The development of WT in chemistry is less than 10 years. However, it has been applied successfully in various fields of chemistry since 1989. A literature survey was performed and 91 publications were found in the areas of analytical chemistry, quantum chemistry and chemical physics from 1989 to August 1998. WT was employed mainly for data compression and signal denoising of analytical signals. An introduction of wavelet transform and wavelet packet transform (WPT) as well as their applications in chemistry were descried in Chapter 3.

Fast wavelet transform and wavelet packet transform were proposed to compressed a small IR spectral library which consisted of twenty organic compounds. As shown in Chapter 4, the spectral library size can reduce 90% by compressing individual IR spectrum to the 4th resolution level with the use of the Daubechies wavelet function D_{16} and cutoff value ε = 0.20. A new algorithm called coefficient position retaining (CPR) method was also developed in this work to process spectral data with odd number data length, which cannot be handled by conventional FWT and WPT methods. Results found that the FWT or WPT compressed spectral library can minimize the search time by using the direct matching method. The overall performance of FWT and WPT methods in compression and library searching were much better than the fast Fourier transform method. A Spectral Library Compression and Search (SLCS) Toolbox Version 1.0, which was coded in MATLAB®, was developed in this study for the required computation.

A novel method as described in Chapter 5 was developed to perform approximate derivative calculation by using FWT. It was found that the difference between the scale coefficients as obtained from FWT treatment at resolution level 1 using the Daubechies wavelet functions D_8 and D_{18} was equal to the approximate first derivative of a particular signal. A higher order derivative could also be obtained by repeating the proposed algorithm on the scale coefficients. As compared with the conventional numerical derivative computation method, the FWT method has three major advantages. Firstly, it could enhance the signal-to-noise ratio (SNR) even at higher derivative order calculation without extra effort in data manipulation. Secondly, both the derivative and smoothing calculations could be combined in a single step. Thirdly, most of the

properties that found in conventional method were retained in the new method. The proposed method was test with both synthetic data and experimental data with satisfactory results. A MATLAB® program called WTDeriv was developed to perform the approximate derivative calculation.

From Chapter 6 onwards, the research works focused on the application of factor analysis together with FWT. A factor analysis called heuristic evolving latent projections was chosen in this study. A new algorithm was developed by coupling FWT to the HELP analysis for pre-processing of data from high performance liquid chromatography coupled with diode array detector (HPLC-DAD). With the denoising property of FWT, it could improve the HELP analysis by using a smaller of window size in the FSWM-EFA calculation for the HPLC-DAD data with low SNR value. On the other hand, the data compression property of FWT could reduce the computational time in HELP analysis. Individual chromatogram at each wavelength was compressed and denoised at the resolution level 1 with the Daubechies wavelet function D_{16}. Scale coefficients $C^{(1)}(\lambda)$ at each wavelength were recombined to form a new data set for HELP analysis. The proposed method was tested with both synthetic data and experimental data from food dye mixture with satisfactory results. FWT-HELP computation was performed by the Fast Wavelet Transform Heuristic Evolving Latent Projections (FWT-HELP) Toolbox Version 1.0.

Finally, in Chapter 7, the FWT-HELP algorithm was applied to analyze water soluble constituents, nucleosides, in fungus and larva parts of *Cordyceps sinensis* which is a traditional and precious dried Chinese medicinal herb with HPLC-DAD. 17 and 19

components were resolved from the HPLC-DAD data of fungus and larva parts respectively with the FWT-HELP computation. The results indicated that FWT can enhance the HELP analysis especially for components with low concentration. With the use of conventional HELP analysis, such low concentration components could not be identified correctly.

In conclusion, wavelet transform was applied into different fields of analytical chemistry for data compression, denoising and pre-processing. Novel methods were developed to enhance the performance of the existing methods. Wavelet transform will be adopted in the other fields of chemistry in the future.

CITED REFERENCES

[1] Adam, M.H. and Black, I.
"Error-free Storage Compression of Binary-coded Infrared Spectra"
Analytica Chimica Acta, Vol. 189, pp.353-363 (1986)

[2] Adam, M.J.
"Chemometrics in Analytical Spectroscopy"
The Royal Society of Chemistry - Cambridge, pp.54-62 (1995)

[3] Alsberg, B.K.
"Representation of Spectra by Continuous Function"
Journal of Chemometrics, Vol. 7, pp. 177-193 (1993)

[4] Alsberg, B.K. and Kvalheim, O.M.
"Compression of nth-order Data Arrays by B-splines Part 1: Theory"
Journal of Chemometrics, Vol. 7, pp.61-73 (1993)

[5] Alsberg, B.K., Nodland, E. and Kvalheim, O.M.
"Compression of nth-order Data Arrays by B-splines Part 2: Application to Second-order FT-IR Spectra"
Journal of Chemometrics, Vol. 8, pp.127-145 (1994)

[6] Alsberg, B.K., Woodward, A.M. and Kell, D.B.
"An Introduction to Wavelet Transforms for Chemometricians: A Time-Frequency Approach"
Chemometrics and Intelligent Laboratory Systems, Vol. 37, pp.215-239 (1997a)

[7] Alsberg, B.K., Woodward, A.M., Winson, M.K., Rowland, J. and Kell, D.B.
"Wavelet Denoising of Infrared Spectra"
Analyst, Vol. 122, pp.645-652 (1997b)

[8] Alsberg, B.K., Woodward, A.M., Winson, M.K., Rowland, J. and Kill, D.B.
"Variable Selection in Wavelet Regression Models"
Analytical Chimica Acta, Vol. 368, pp.29-44 (1998)

[9] Amsterdam, J.
"Data Compression with Huffman Coding"
Byte, Vol. 11(5), pp. 99-108 (1986)

[10] Antonoy, L. and Stoyanoy, S.
"Approach for Increased Information from the Second-derivative Spectra in UV-VIS Absorption Spectroscopy"
Applied Spectroscopy, Vol. 47, pp.172-1715 (1993)

[11] Antonoy, L. and Stoyanov, S.
"Step by Step Filter - An Approach for Noise Reduction in the Derivative UV-visible Spectra"
Analytica Chimica Acta, Vol. 324, pp.77-83 (1996)

[12] Aries, R.E., Lidiard, D.P. and Spragg, R.A.
"Principal Component Analysis"
Chemistry in Britain, Vol. 27, pp.821-824 (1991)

[13] Askar, A., Cetin, A.E. and Rabitz, H.
"Wavelet Transform for Analysis of Molecular Dynamics"
Journal of Physical Chemistry, Vol. 100, pp.19165-19173 (1996)

[14] Auf der Heyde, T.P.E.
"Analyzing Chemical Data in More Than Two Dimensions A Tutorial on Factor and Cluster Analysis"
Journal of Chemical Education, Vol. 67, pp.461-469 (1990)

[15] Azarraga, L.V., Williams, R.R. and de Hasete, J.A.
"Fourier Encoded Data Searching of Infrared Spectra (FEDS/IRS)"
Applied Spectroscopy, Vol. 35, pp.466-469 (1981)

[16] Bao, L.J. and Mo, J.Y.
"An Modify Fourier Transform Method for Processing High Noise Level Electrochemical Signal"
Chinese Science Bulletin (Chinese Edition), Vol. 43(1), pp.42-45 (1998) [In Chinese, without English abstracts]

[17] Bao, L.J., Mo, J.Y. and Tang, Z.Y.
"Comparative Study on Signal Processing in Analytical Chemistry by Fourier and Wavelet Transforms"
Acta Chimica Sinica, Vol. 55, pp.907-914 (1997a) [In Chinese, with English abstract]

[18] Bao, L.J., Mo, J.Y. and Tang, Z.Y.
"The Application in Processing Analytical Chemistry Signals of a Cardinal Spline Approach to Wavelets"
Analytical Chemistry, Vol. 69, pp.3053-3057 (1997b)

[19] Bao, L.J., Mo, J.Y. and Tang, Z.Y.
"The Application of Spline Wavelet and Fourier Transformation in Analytical Chemistry"
Chemical Journal of Chinese Universities, Vol. 19, pp.193-197 (1998) [In Chinese, with English abstract]

[20] Bao, L.J., Tang, Z.Y. and Mo, J.Y.
"The Application of Spline Wavelet and Fourier Transform in Analytical Chemistry" in New Trends in Chemometrics by Liang, Y.Z., Nortvedt, R., Kvalheim, O.M., Shen, H.L. and Xu, Q.S. (Eds.), First Internation Conference on Chemometrics in China held at Zhangjiajie, China, 17-22 October, 1997
Hunan University Press - Changsha, China, pp.197-198 (1997c)

[21] Barak, P.
"Smoothing and Differentiation by an Adaptive-degree Polynomial Filter"
Analytical Chemistry, Vol. 67, pp.2758-2762 (1995)

[22] Barclay, V.J., Bonner, R.F. and Hamilton, I.P.
"Application of Wavelet Transforms to Experimental Spectra: Smoothing, Denoising, and Data Set Compression"
Analytical Chemistry, Vol. 69, pp.78-90 (1997)

[23] Beylkin, G.
"On the Representation of Operators in Bases of Compactly Supported Wavelets"
SIAM Journal of Numerical Analysis, Vol. 6, pp.1716-1740 (1992)

[24] Binkley, D.P. and Dessy, R.E.
"Linear Parameter Estimation of Fused Peak Systems in the Spatial Frequency Domain"
Analytical Chemistry, Vol. 52, pp.1335-1344 (1980)

[25] Blaffert, T.
"Computer-assisted Multicomponent Spectral Analysis with Fuzzy Data Sets"
Analytica Chimica Acta, Vol. 161, pp.135-148 (1984)

[26] Borde, B.L.
"New Wavelet Class for Time Structure Identification" in Wavelet Application II by Szu, H.H., Proceedings of SPIE - The International Society for Optical Engineering, Vol. 2491, held at Orlando, U.S., 17-21 April 1995,
SPIE - The International Society for Optical Engineering - Washington, D.C., pp.1073-1085 (1995)

[27] Bos, M. and Hoogendam, E.
"Wavelet Transform for the Evaluation of Peak Intensities in Flow-injection Analysis"
Analytica Chimica Acta, Vol. 267, pp.73-80 (1992)

[28] Bos, M. and Vrielink, J.A.M.
"The Wavelete Transform for Pre-processing IR Spectra in the Identification of Mono- and Di-substituted Benzenes"
Chemometrics and Intelligent Laboratory Systems, Vol. 23, pp.115-122 (1994)

[29] Brereton, R.G.
"Chemometrics in Analytical Chemistry A Review"
Analyst, Vol. 112, pp.1635-1657 (1987)

[30] Brereton, R.G.
"Chemometrics Application of Mathematics and Statistics to Laboratory Systems"
Ellis Horwood - New York, pp.11-26 (1990)

[31] Brereton, R.G.
"Deconvolution of Mixtures by Factor Analysis"
Analyst, Vol. 120, pp.2313-2336 (1995)

[32] Brewster, M.E., Fann, G.I. and Yang, Z.Y.
"Wavelets for Electronic Structure Calculations"
Journal of Mathematical Chemistry, Vol. 22, pp.117-142 (1997)

[33] Brown, S., Blank, T.B., Sum, S.T. and Weyer, L.G.
"Chemometrics"
Analytical Chemistry, Vol. 66, pp.315R-359R (1994)

[34] Brown, S., Sum, S.T. and Despagne, F.
"Chemometrics"
Analytical Chemistry, Vol. 68, 21R-62R (1996)

[35] Brown, S.D.
"Signal Processing and Data Enhancement" in Practical Guide to Chemometrics by Haswell, S.J. (Ed.)
Marcel Dekker, Inc. - New York, pp.239-269 (1992)

[36] Brown, S.D.
"Has the Chemometrics Revolution Ended? Some Views on the Past, Present and Future of Chemometrics"
Chemometrics and Intelligent Laboratory Systems, Vol. 30, pp.49-58 (1995)

[37] Buckheit, J.B. and Donoho, D.L.
"WaveLab and Reproducible Research" in Wavelet and Statistics by Antoniadis, A. and Oppenheim, G. (Eds.)
Springer-Verlag New York Inc. - New York, pp.55-81 (1995)

[38] Bush, C.A.
"Fourier Method for Digital Data Smoothing in Circular Dichroism Spectrometry"
Analytical Chemistry, Vol. 46, pp.890-895 (1974)

[39] But, P.P.H. and Kwok, I.M.Y.
"Newsletter on Quality Control of Chinese Medicines", No. 1,
Laboratory for the Chemical Quality Control of Chinese Medicines and Health Foods,
The Chinese University of Hong Kong, Hong Kong (1996)

[40] But, P.P.H., Kwok, I.M.Y. and Leung, C.H.
"Newsletter on Quality Control of Chinese Medicines", No. 2,
Laboratory for the Chemical Quality Control of Chinese Medicines and Health Foods,
The Chinese University of Hong Kong, Hong Kong (1996)

[41] But, P.P.H., Kwok, I.M.Y. and Leung, C.H.
"Newsletter on Quality Control of Chinese Medicines", No. 3,
Laboratory for the Chemical Quality Control of Chinese Medicines and Health Foods,
The Chinese University of Hong Kong, Hong Kong (1997)

[42] Byrnes, J.S., Byrnes, J.L., Hargreaves, K.A. and Berry, K.
"Wavelet and Their Application"
Kluwer Academic Publishers - Boston (1994)

[43] Calais, J.L.
"Wavelets — Something for Quantum Chemistry?"
International Journal of Quantum Chemistry, Vol. 58, pp.541-548 (1996)

[44] Castledine, J.B., Fell, A.F., Modin, R. and Sellberg, B.
"Novel Techniques for Peak Purity/Homogeneity Assessment of Sulfasalazine by Liquid Chromatography with Diode Array Detection"
Analytical Proceedings, Vol. 29, pp.100-104 (1992)

[45] Chau, F.T.
"Introducing Gran Plots to Undergraduates"
Education in Chemistry, Vol. 27, pp.109-110 (1990)

[46] Chau, F.T. and Chung, W.H.
"Using MATLAB to Assist Undergraduates in Learning Chemometrics"
Journal of Chemical Education, Vol. 72, pp.A84-A85 (1995)

[47] Chau, F.T., Gao, J.B., Shih, T.M. and Wang, J.
"Compression of Infrared Spectral Data Using the Fast Wavelet Transform Method"
Applied Spectroscopy, Vol. 51, pp.649-659 (1997)

[48] Chau, F.T., Shih, T.M., Gao, J.B. and Chan, C.K.
"Application of the Fast Wavelet Transform Method to Compress Ultraviolet-visible Spectra"
Applied Spectroscopy, Vol. 50, pp.339-349 (1996)

[49] Chau, F.T. and Tam, K.Y.
"Application of the Fast Fourier Transform Method for Compression of Spectral Data Obtained From a Photodiode Array Spectrophotometer"
Computers and Chemistry, Vol. 18, pp.13-20 (1994)

[50] Chau, F.T., Tse, H.K., and Cheng, F.L.
"Modified Gran Plots of Very Weak Acids on a Spreadsheet"
Journal of Chemical Education, Vol. 67, pp.A8 (1990)

[51] Che, C.T.
"Standardization of Chinese Medicines"
Symposium on "Advancing The TCM Industry into the Next Century: Current Challenges and Issues" held at Hong Kong, 27-29 May, 1997, pp.S2.4 (1997)

[52] Chen, J., Shan, J. and Zhang, M.S.
"Wavelet Transform of Infrared Spectra of Poly(Acrylamide-Sodium Acrylate)Hydrogel Copolymers"
Chinese Journal of Analytical Chemistry, Vol. 25, pp.172-174 (1997) [In Chinese, with English abstract]

[53] Chen, J., Zhong, H.B., Pan, Z.X. and Zhang, M.S.
"Application of Wavelet Transform in Differential Pulse Voltammetric Data Processing"
Chinese Journal of Analytical Chemistry, Vol. 24, pp.1002-1006 (1996) [In Chinese, with English abstract]

[54] Cho, K., Arias, T.A., Joannopoulos, J.D. and Lam, P.K.
"Wavelets in Electronic Structure Calculations"
Physical Review Letters, Vol. 71, pp.1808-1811 (1993)

[55] Chui, C.K.
"An Introduction to Wavelets"
Academic Press - Boston (1992a)

[56] Chui, C.K.
"Wavelets: A Tutorial in Theory and Application"
Academic Press - Boston (1992b)

[57] Chui, C.K.
"Wavelets: A Mathematical Tool for Signal Processing"
SIAM Press - Philadelphia (1997)

[58] Chui, C.K., Montefusco, L. and Puccio, L.
"Wavelets: Theory, Algorithms and Applications"
Academic Press - San Diego (1994)

[59] Cody, M.A.
"The Fast Wavelet Transform Beyond Fourier Transforms"
Dr. Dobb's Journal, Vol. 17(4), pp.16-28 and 100-101 (1992)

[60] Cody, M.A.
"The Wavelet Packet Transform"
Dr. Dobb's Journal, Vol. 19(4), pp.44-54 and 100 (1994)

[61] Coifman, R.R. and Wickerhauser, M.V.
"Entropy-based Algorithms for Best Basis Selection"
IEEE Transactions on Information Theory, Vol. 38, pp.713-719 (1992)

[62] Coifman, R.R., Meyer, Y., Quake, S. and Wickerhauser, M.V.
"Signal Processing and Compression with Wavelet Packet" in Wavelet and Their Application by Byrnes, J.S., Byrnes, J.L., Hargreaves, K.A. and Berry, K. (Eds.)
Kluwer Academic Publishers - Boston , pp.363-379 (1994)

[63] Collantes, E.R., Duta, R., Welsh, W.J., Zielinski, W.L. and Brower, J.
"Preprocessing of HPLC Trace Impurity Patterns by Wavelet Packets for Pharmaceutical Fingerprinting Using Artifical Neural Networks"
Analytical Chemistry, Vol. 69, pp.1392-1397 (1997)

[64] Crawford, E.F. and Larsen, R.D.
"Clipped Free Induction Decay Signal Analysis"
Analytical Chemistry, Vol. 49, pp.508-510 (1977)

[65] Dai, X.D., Joseph, B. and Motard, R.L.
"Introduction to Wavelet Transform and Time-frequency Analysis" in Wavelet Application in Chemical Engineering by Motard, R.L. and Joseph, B. (Eds.)
Kluwer Academic Publishers - Massachusetts, pp.1-32 (1994)

[66] Daiguji, M. Kudo, O. and Wada, T.
"Application of Wavelet Analysis to Fault-detection in Oil Refinery"
Computers & Chemical Engineering, Vol. 21, pp.S1117-S1122 (1997)

[67] Daubechies, I.
"Orthonormal Bases of Compactly Supported Wavelets"
Communications on Pure and Applied Mathematics, Vol. 41, pp.909-996 (1988)

[68] Daubechies, I.
"The Wavelet Transform, Time-frequency Localization and Signal Analysis"
IEEE Transactions on Information Theory, Vol. 36, pp.961-1005 (1990)

[69] Daubechies, I.
"Ten Lectures on Wavelets"
SIAM Press - Philadelphia (1992).

[70] Daubechies, I.
"Two Recent Results on Wavelets: Wavelet Bases for the Interval, and Biorthogonal Wavelets Diagonalizing the Derivative Operator" in Recent Advances in Wavelet Analysis by Schumaker, L.L. and Webb, G. (Eds.)
Academic Press, Inc. - Boston, pp.237-257 (1994)

[71] de Cerqueira, E.O. and Poppi, R.J.
"Using Dynamic Data Exchange to Exchange Information Between Visual Basic and Matlab"
Trends in Analytical Chemistry, Vol. 15, pp.500-503 (1996)

[72] de Haseth, J.A. and Azarraga, L.V.
"Interferogram-based Infrared Search System"
Analytical Chemistry, Vol. 53, pp.2292-2296 (1981)

[73] Depczynski, U., Jetter, K., Molt, K. and Niemöller, A.
"The Fast Wavelet Transform on Compact Intervals as a Tool in Chemometics. I. Mathematical Background"
Chemometrics and Intelligent Laboratory Systems, Vol. 39, pp.19-27 (1997)

[74] Dohan, K. and Whitfield, P.H.
"Identification and Characterization of Water Quality Transients using Wavelet Analysis. I. Wavelet Analysis Methodology"
Water and Science Technology, Vol. 36, pp.325-335 (1997)

[75] Donoho, D.L.
"De-noising by Soft-thresholding"
IEEE Transitions on Information Theory, Vol. 41, pp.613-627 (1995)

[76] Ehrentreich, F., Nikolov, S.G., Wolkenstein, M. and Hutter, H.
"The Wavelet Transform: A New Preprocessing Method for Peak Recognition of Infrared Spectra"
Mikrochimica Acta, Vol. 128, pp.241-250 (1998)

[77] Fang, H. and Chen, H.Y.
"Wavelet Analyses of Electroanalytical Chemistry Responses and an Adaptive Wavelet Filter"
Analytica Chimica Acta, Vol. 346, pp.319-325 (1997)

[78] Fell, A.F.
"Present and Future Prespectives in Derivative Spectroscopy"
UV Spectrometry Group Bulletin, No. 8, pp.5-31 (1980)

[79] Fenyö, D., Zhang, W.Z., Chait, B.T. and Beavis, R.C.
"Internet-based Analytical Chemistry Resources: A Model Project"
Analytical Chemistry, Vol. 68, pp.721A-726A (1996)

[80] Fischer, P. and Defranceschi, M.
"Looking at Atomic Orbitals throught Fourier and Wavelet Transforms"
International Journal of Quantum Chemistry, Vol. 45, pp.619-636 (1993)

[81] Fischer, P. and Defranceschi, M.
"The Wavelet Transform A New Mathematical Tool for Quantum Chemistry" in Conceptual Trends in Quantum Chemistry by Kryachko, E.S. and Calais, J.L. (Eds.)
Kluwer Academic Publishers - Boston, 227-247 (1994a)

[82] Fischer, P. and Defranceschi, M.
"Iterative Process for Solving Hartree-Fock Equations by Means of a Wavelet Transform"
Applied and Computational Harmonic Analysis, Vol. 1, pp.232-241 (1994b)

[83] Fischer, P. and Defranceschi, M.
"Representation of the Atomic Hartree-Fock Equations in a Wavelet Basis by Means of the BCR Algorithm" in Wavelets: Theory, Algorithms and Applications by in Chui, C.K., Montefusco, L. and Puccio, L. (Eds.)
Academic Press, Inc. - San Diego, pp.495-506 (1994c)

[84] Fischer, P. and Defranceschi, M.
"Numerical Solution of the Schrödinger Equation in a Wavelet Basis for Hydrogen-like Atoms"
SIAM Journal of Numerical Analysis, Vol. 35, pp.1-12 (1998)

[85] Gao, J.B., Chau, F.T. and Shih, T.M.
"Wavelet Transform Method for De-noising Spectral Data from UV-VIS Spectrophotometer"
Southeast Asian Bulletin of Mathematics, Vol. 20, pp.85-90 (1996)

[86] Gemperline, P.J.
"A Priori Estimates of the Elution Profiles of the Pure Components in Overlapped Liquid Chromatography Peaks Using Target Factor Analysis"
Journal of Chemical Information and Computer Science, Vol. 24, pp. 206-212 (1984)

[87] Gemperline, P.J.
"Target Transformation Factor Analysis with Linear Inequality Constrants Applied to Spectroscopic-chromatographic Data"
Analytical Chemistry, Vol. 58, pp.2656-2663 (1986)

[88] Gilliard, J.A. and Ritter, C.
"Use of Simulated Liquid Chromatography-diode Array Detection Data for the Definition of a Guide Curve in Peak Purity Assessment by Spectral Comparison"
Journal of Chromatography A, Vol. 786, pp.1-11 (1997)

[89] Glasser, L.
"Fourier Transforms for Chemists Part I. Introduction to the Fourier Transform"
Journal of Chemical Education, Vol. 64, pp.A228-A233 (1987a)

[90] Glasser, L.
"Fourier Transforms for Chemists Part II. Fourier Transforms in Chemistry and Spectroscopy"
Journal of Chemical Education, Vol. 64, pp.A260-266 (1987b)

[91] Glasser, L.
"Fourier Transforms for Chemists Part III. Fourier Transforms in Data Treatment"
Journal of Chemical Education, Vol. 64, pp.A306-313 (1987c)

[92] Gorry, P.A.
"General Least-squares Smoothing and Differentiation by the Convolution (Savitzky-Golay) Method"
Analytical Chemistry, Vol. 62, 570-573 (1990)

[93] Gran, G.
"Determination of the Equivalence Point in Potentiometric Titrations Part II"
Analyst, Vol. 77, pp.661-671 (1952)

[94] Grossman, A. and Morlet, J.
"Decomposition of Hardy Functions into Square Integrable Wavelets of Constrant Shape"
SIAM Journal on Mathematical Analysis, Vol. 15, pp.723-736 (1984)

[95] Grung, B. and Kvalheim, O.M.
"Resolution of Multicomponent Profiles with Partial Selectivity. A Comparison of Direct Methods"
Chemometrics and Intelligent Laboratory Systems, Vol. 29, pp.75-87 (1995)

[96] Guillemain, P., Kronland-Martinet, R. and Martens, B.
"Estimation of Spectral Lines with the Help of the Wavelet Transform: Applications in NMR Spectroscopy" in Wavelets and Applications by Meyer, Y. (Ed.)
Proceedings of the Second International Conference on Wavelets and Their Applications held in Marseille, France, May, 1989
Springer Verlag - Pairs, pp.38-60 (1992)

[97] Guo, C., Zhu, J., Zhang, C. and Zhang, J.L.
"Determination of Adenosine and 3'-Deoxyadenosine in Cordyceps militaris (L.) Link. By HPLC"
China Journal of Chinese Materia Medica, Vol. 23, pp.236-237 and 256 (1998) [In Chinese with English abstract]

[98] Hämäläinen, M.D., Liang, Y.Z., Kvalheim, O.M. and Andersson, R.
"Deconvolution in One-dimensional Chromatography by Heuristic Evolving Latent Projections of Whole Profiles Retention Time Shifted by Simplex Optimization of Cross-correlation Between Target Peaks"
Analytica Chimica Acta, Vol. 271, pp.101-114 (1993)

[99] Hangac, G., Wieboldt, R.C., Lam, R.B. and Isenhour, T.L.
"Compression of an Infrared Spectral Library by Karhunen-Loeve Transformation"
Applied Spectroscopy, Vol. 36, pp.40-47 (1982)

[100] Hargis, L.G., Howell, J.A. and Sutton, R.E.
"Ultraviolet and Light Absorption Spectrometry"
Analytical Chemistry, Vol. 68, pp. 169R-183R (1996)

[101] Harrington, P.B. and Isenhour, T.L.
"Closure Effects on Infrared Spectral Library Search Performance"
Applied Spectroscopy, Vol. 41, pp.1298-1302 (1987)

[102] Haswell, S.J.
"Introduction to Chemometrics" in Practical Guide to Chemometrics by Haswell, S.J. (Ed.)
Marcel Dekker, Inc. - New York, pp.1-3 (1992)

[103] Hayes, J.W., Glover, D.E., Smith, D.E. and Overton, M.W.
"Some Observations on Digital Smoothing of Electroanalytical Data Based on the Fourier Transformation"
Analytical Chemistry, Vol. 45, pp.277-284 (1973)

[104] Hecht, H.G.
"Mathematics in Chemistry An Introduction to Modern Methods"
Prentice Hall - New Jersey, 1990, pp.142-173 (1990)

[105] Heite, F.H., Dupuis, P.F., Van 'T Klooster, H.A. and Dijkstra, A.
"Numerical Taxonomy and Information Theory Applied to Feature Selection from Filed Infrared Spectra for Automated Interpretation"
Analytica Chimica Acta, Vol. 103, pp.313-321 (1978)

[106] Hernández, E. and Weiss, G.
"A First Course on Wavelets"
CRC Press - Boca Raton, (1996)

[107] "HPLC^{3D} ChemStation: Installing and Understanding Your Spectra Module"
Hewlett-Packard Company - Germany, pp.25-54 (1994)

[108] Hilton, M.L.
"Wavelet and Wavelet Packet Compression of Electrocardiograms"
IEEE Transactions on Biomedical Engineering, Vol. 44, pp.394-402 (1997)

[109] Hobert, H.
"Library Search - Principles and Applications" in Chemometrics in Environmental Chemistry - Applications by Einax, J. (Ed.)
Springer-Verlag Berlin Heidelberg - Germany, pp.1-23 (1995)

[110] Hoch, J.C. and Stern, A.S.
"NMR Data Processing"
Wiley-Liss, Inc. - New York, pp.144-151 (1996)

[111] Hopke, P.K.
"Target Transform Factor Analysis"
Chemometrics and Intelligent Laboratory Systems, Vol. 6, pp.7-19 (1989)

[112] Horlick, G.
"Digital Data Handling of Spectra Utilizing Fourier Transforms"
Analytical Chemistry, Vol. 44, pp.943-947 (1972)

[113] Howell, J.A. and Hargis, L.G.
"Ultraviolet and Light Absorption Spectrometry"
Analytical Chemistry, Vol. 66, pp. 445R-461R (1994)

[114] Huang, K.C.
"The Pharmacology of Chinese Herbs"
CRC Press - Boca Raton, pp.207-208 (1993)

[115] Hubbard, B.B.
"The World According to Wavelets : The Story of a Mathematical Technique in the Making"
A. K. Peters - Massachusetts (1996)

[116] Hutter, H., Brunner, C., Nikolov, S., Mittermayr, C.R. and Grasserbauer, M.
"Imaging Surface Spectroscopy for Two- and Three-Dimensional Characterization of Materials"
Fresenius' Journal of Analytical Chemistry, Vol. 355, pp.585-590 (1996)

[117] Idris, F. and Panchanathan, S.
"Storage and Retrieval of Compressed Images Using Wavelet Vector Quantization"
Journal of Visual Languages and Computing, Vol. 8, pp.289-301 (1997)

[118] Jeffery, G.H., Bassett, J., Mendham, J. and Denney, R.C.
"Vogel's Textbook of Quantitative Chemical Analysis", 5th ed.,
Longman Scientific & Technical - England, pp.574-578 (1989)

[119] Jouan-Rimbaud, D., Walczak, B., Poppi, R.J., de Noord, O.E. and Massart, D.L.
"Application of Wavelet Transform to Extract the Relevant Component from Spectral Data for Multivariate Calibration"
Analytical Chemistry, Vol. 69, 4317-4323 (1997)

[120] Kaiser, G.
"A Friendly Guide to Wavelets"
Birkhäuser - Boston, pp.ix-xiv (1994)

[121] Karjalainen, E.J.
"The Spectrum Reconstruction Problem Use of Alternating Regression for Unexpected Spectral Components in Two-dimensional Spectroscopies"
Chemometrics and Intelligent Laboratory Systems, Vol. 7, pp.31-38 (1989)

[122] Karjalainen, E.J. and Karjalainen, U.P.
"Data Analysis for Hyphenated Techniques"
Elsevier Science B.V. - Amsterdam, pp.127-210 (1996)

[123] Keller, H.R. and Massart, D.L.
"Peak Purity Control in Liquid Chromatography with Photodiode-array Detection by a Fixed Size Moving Window Evolving Factor Analysis"
Analytica Chimica Acta, Vol. 246, pp.379-390 (1991)

[124] Keller, H.R. and Massart, D.L.
"Peak Purity Control in Liquid Chromatography with Photodiode-array Detection by A Fixed Size Moving Window Evolving Factor Analysis"
Analytica Chimica Acta, Vol. 246, pp.379-390 (1991)

[125] Keller, H.R. and Massart, D.L.
"Artefacts in Evolving Factor Analysis-based Methods for Peak purity Control in Liquid Chromatography with Diode-array Detection"
Analytica Chimica Acta, Vol. 263, pp.21-28 (1992b)

[126] Keller, H.R. and Massart, D.L.
"Evolving Factor Analysis"
Chemometrics and Intelligent Laboratory Systems, Vol. 12, pp.209-224 (1992a)

[127] Keller, H.R., Massart, D.L., Liang, Y.Z. and Kvalheim, O.M.
"A Comparison of the Heuristic Evolving Latent Projections and Evolving Factor Analysis Methods for Peak Purity Control in Liquid Chromatography with Photodiode Array Detection"
Analytica Chimica Acta, Vol. 267, pp.63-71 (1992)

[128] Kemp, W.
"NMR in Chemistry A Multinuclear Introduction"
Macmillan - London, pp.158-168 (1986)

[129] Kirmse, D.W. and Westerberg, A.W.
"Resolution Enhancement of Chromatography Peaks"
Analytical Chemistry, Vol. 43, pp.1035-1039 (1971)

[130] Klawun, C. and Wilkins, C.L.
"Neural Network Assisted Rapid Screening of Large Infrared Spectral Databases"
Analytical Chemistry, Vol. 67, pp.374-378 (1995)

[131] Krauss, T.P., Shure, L. and Little, J.N.
"Signal Processing Toolbox User's Guide"
The MathWorks, Inc. - Massachusetts, pp.2-111-2-114 (1994)

[132] Kuo, J.E., Wang, H. and Pickup, S.
"Multidimensional Least-squares Smoothing Using Orthogonal Polynomials"
Analytical Chemistry, Vol. 63, pp.630-635 (1991)

[133] Kuroda, R., Wada, T., Kokubo, Y. and Oguma, K.
"Ion Interaction Chromatography of Nitrilatriacetatocomplexes of the Rare Earth Elements with Post-column Reaction Detection"
Talana, Vol. 40, 237-241 (1993)

[134] Kvalheim, O.M. and Liang, Y.Z.
"Heuristic Evolving Latent Projections: Resolving Two-way Multicomponent Data. 1. Selectivity, Latent-Projective Graph, Datascope, Local Rank, and Unique Resolution"
Analytical Chemistry, Vol. 64, pp.936-946 (1992)

[135] Lam, R.B., Foulk, S.J. and Isenhour, T.L.
"Clipped Fourier Transform Mass Spectral Compression Algorithm for Microcomputer-compatible Search Systems"
Analytical Chemistry, Vol. 53, pp.1679-1684 (1981)

[136] Lee, D.
"The Quest for Quality"
Hong Kong Industrialist, November, pp.50-54 (1997)

[137] Lee, O. and Wade, A.P.
"Generalized Fourier Smoothing of Flow Injection Analysis Data"
Analytical Chemistry, Vol. 66, pp.4507-4513 (1994)

[138] Leung, K.M. and Chau, F.T.
"A Review on Signal Compression of Spectroscopic Data in Analytical Chemistry"
Acta Physico-Chimica Sinica, Vol. 13, pp.857-864 (1997)

[139] Li, J.W.
"Development and Evaluation of Flexible Empirical Peak Functions for Processing Chromatographic Peaks"
Analytical Chemistry, Vol. 69, pp.4452-4462 (1997)

[140] Li, Z.M., Borrmann, A. and Martens, C.C.
"Wavelet Analysis of Condensed Phase Molecular Dynamics"
Chemical Physics Letters, Vol. 214, pp.362-366 (1993)

[141] Liang, Y.Z., Brereton, R.G., Kvalheim, O.M. and Rahmani, A.
"Use of Chemometrics Factor Analysis for Chromatographic Integration: Application to Diode-array High-performance Liquid Chromatography of Mixtures of Chlorophyll a Degradation Products"
Analyst, Vol. 118, pp.779-790 (1993c)

[142] Liang, Y.Z., Hämäläinen, M.D., Kvalheim, O.M. and Andersson, R.
"Assessment of Peak Origin and Purity in One-dimensional Chromatography by Experimental-design and Heuristic Evolving Latent Projections"
Journal of Chromatography A, Vol. 662, pp.113-122 (1994)

[143] Liang, Y.Z. and Kvalheim, O.M.
"Resolution of Strongly Overlapping Two-way Multicomponent Data by Means of Heuristic Evolving Latent Projections"
Journal of Chemometrics, Vol. 7, pp.15-43 (1993a)

[144] Liang, Y.Z. and Kvalheim, O.M.
"Unique Resolution of Hidden Minor Peaks in Multidetection Chromatography by First-order Differentiation and Orthogonal Projections"
Analytica Chimica Acta, Vol. 276, pp.425-440 (1993b)

[145] Liang, Y.Z. and Kvalheim, O.M.
"Heuristic Evolving Latent Projections: Resolving Hyphenated Chromatographic Profiles by Component Stripping"
Chemometrics and Intelligent Laboratory Systems, Vol. 20, pp.115-125 (1993c)

[146] Liang, Y.Z., Kvalheim, O.M., Keller, H.R., Massart, D.L., Kiechle, P. and Erni, F.
"Heuristic Evolving Latent Projections: Resolving Two-way Multicomponent Data. 2. Detection and Resolution of Minor Constituents"
Analytical Chemistry, Vol. 64, pp.947-953 (1992)

[147] Liang, Y.Z., Kvalheim, O.M. and Manne, R.
"White, Grey and Black Multicomponent Systems A Classification of Mixture Problems and Methods for their Quantitative Analysis"
Chemometrics and Intelligent Laboratory Systems, Vol. 18, pp.235-250 (1993)

[148] Liang, Y.Z., Kvalheim, O.M., Rahmani, A. and Brereton, R.G.
"A Two-way Procedure for Background Correction of Chromatographic/ Spectroscopic Data by Congruence Analysis and Least-squares Fit of the Zero-component Regions: Comparison with Double-centering"
Chemometrics and Intelligent Laboratory Systems, Vol. 28, pp.265-279 (1993a)

[149] Liang, Y.Z., Kvalheim, O.M., Rahmani, A. and Brereton, R.G.
"Resolution of Strongly Overlapping Two-way Multicomponent Data by Means of Heuristic Evolving Latent Projections"
Journal of Chemometrics, Vol. 7, pp.15-43 (1993b)

[150] Liu, W., Li, J.P., Xiong, J.H., Pan, Z.X. and Zhang, M.S.
"The Compress of IR Spectra by Using Wavelet Neural Network"
Chinese Science Bulletin (Chinese Edition), Vol. 42(8), pp.824-827 (1997a) [In Chinese, without English abstract]

[151] Liu, W., Li, J.P., Xiong, J.H., Pan, Z.X. and Zhang, M.S.
"The Compress of IR Spectra by Using Wavelet Neural Network"
Chinese Science Bulletin (English Edition), Vol. 42(10), pp.822-825 (1997b)

[152] Liu, W., Wang, Y.M., Pan. Z.X., Zhou, W. and Zhang, M.S.
"Simultaneous Determination of Molybdenum and Tungsten using A Wavelet Neural Network"
Chinese Journal of Analytical Chemistry, Vol. 25, pp.1189-1191 (1997d) [In Chinese, with English abstract]

[153] Liu, W., Xiong, J.H., Wang, H. and Wang, Y.M.
"The Recognition of UV Spectra by Using Wavelet Neural Network"
Chemical Journal of Chinese Universities, Vol. 18, pp.860-863 (1997c) [In Chinese, with English abstract]

[154] Lo, S.C. and Brown, C.W.
"Infared Spectral Search for Mixtures in Large-size Libraries"
Applied Spectroscopy, Vol. 45, pp.1628-1632 (1991)

[155] Losev, A.
"Smoothing by Spline Functions: Applications in Electron Spectroscopy"
Journal of Electron Spectroscopy and Related Phenomena, Vol. 50, pp.c19-c23 (1990)

[156] Lowry, S.R., Huppler, D.A. and Anderson, C.R.
"Data Base Development and Search Algorithms for Automated Infrared Spectral Identification"
Journal of Chemical Information and Computer Science, Vol. 25, pp.235-241 (1985)

[157] Lu, X.Q. and Mo, J.Y.
"Spline Wavelet Multi-resolution Analysis for High-noise Digital Signal Processing in Ultraviolet-visible Spectrophotometry"
Analyst, Vol. 121, pp.1019-1024 (1996)

[158] Lu, X.Q. and Mo, J.Y.
"Spline Wavelet Multifrequency Channel Filters for High Noise Digital Signal Processing in Voltammetry"
Acta Scientiarum Naturalium Universitatis Sunyatseni, Vol. 36, pp.129-130 (1997)

[159] Lu, X.Q., Mo, J.Y., Kang, J.W. and Gao, J.Z.
"Method of Processing Discrete Data for Deconvolution Voltammetry II. Spline Wavelet Transformation"
Analytical Letters, Vol. 31, pp.529-540 (1998)

[160] Lupu, M. and Todor, D.
"A Singular Value Decomposition Based Algorithm for Multicomponent Exponential Fitting of NMR Relaxation Signals"
Chemometrics and Intelligent Laboratory Systems, Vol. 29, pp.11-17 (1995)

[161] Madden, H.H.
"Comments on the Savitzky-Golay Convolution Method for Least-squares Fit Smoothing and Differentiation of Digital Data"
Analytical Chemistry, Vol. 50, 1383-1386 (1978)

[162] Maeder, M.
"Evolving Factor Analysis for the Resolution of Overlapping Chromatographic Peaks"
Analytical Chemistry, Vol. 59, pp.527-530 (1987)

[163] Maeder, M. and Zilian, A.
"Evolving Factor Analysis: A New Multivariate Technique in Chromatography"
Chemometrics and Intelligent Laboratory Systems, Vol. 3, pp.205-213 (1988)

[164] Malinowski, E.R.
"Factor Analysis in Chemistry", 2nd ed.,
John Wiley & Sons, Inc. - New York (1991)

[165] Mailnowski, E.R.
"Window Factor Analysis: Theoretical Derivation and Application to Flow Injection Analysis Data"
Journal of Chemometrics, Vol. 6, pp.29-40 (1992)

[166] Maeder, M. and Zuberbuehler, A.D.
"The Resoultion of Overlapping Chromatographic Peaks by Evolving Factor Analysis"
Analytica Chimica Acta, Vol. 181, pp.287-291 (1986)

[167] Mailnowski, E.R.
"Window Factor Analysis: Theoretical Derivation and Application to Flow Injection Analysis Data"
Journal of Chemometrics, Vol. 6, pp.29-40 (1992)

[168] Mallat, S.G.
"Multifrequency Channel Decompositions of Images and Wavelet Model"
IEEE Transactions on Acoustics, Speech, and Signal Processing, Vol. 37, pp.2091-2110 (1989a)

[169] Mallat, S.G.
"Multiresolution Approximation and Wavelets"
Transactions of American Mathematical Society, Vol. 315, pp.69-88 (1989b)

[170] Mallat, S.G. and Hwang, W.L.
"Singularity Detection and Processing with Wavelets"
IEEE Transactions on Information Theory, Vol. 38, pp.617-643 (1992)

[171] Mallet, Y. Coomans, D. and de Vel, O.
"Recent Developments in Discriminant Analysis on High-dimensional Spectral Data"
Chemometrics and Intelligent Laboratory Systems, Vol. 35, pp.157-173 (1996)

[172] Mao, J.J., Su, Q.D. and Zhnag, M.S.
"Wavelet Analysis Applied in Photoacoustic Spectroscopy"
New Trends in Chemometrics by Liang, Y.Z., Nortvedt, R., Kvalheim, O.M., Shen, H.L. and Xu, Q.S. (Eds.), First Internation Conference on Chemometrics in China held at Zhangjiajie, China, 17-22 October, 1997
Hunan University Press - Changsha, China, pp.201-202 (1997)

[173] Marcus, M.
"Matrices and MATLAB® A Tutorial"
Prentice-Hall International, Inc. - New Jersey (1992)

[174] Massart, D.L., Vandeginste, B.G.M., Deming, S.N., Michotte, Y. and Kaufman, L.
"Chemometrics: A Textbook"
Elsevier Science Publishers B.V. - Amsterdam, pp.5-12 (1988)

[175] Meglen, R.R.
"Chemometrics: Its Role in Chemistry and Measurement Sciences"
Chemometrics and Intelligent Laboratory Systems, Vol. 3, pp.17-29 (1988)

[176] Meloun, M., Militký, J. and Forina, M.
"Chemometrics for Analytical Chemistry Volume 2: PC-Aided Regression and Related Methods"
Ellis Horwood - New York, pp.376-380 (1994)

[177] Meyer, V.R.
"Generation and Investigation of Chromatographic Peaks Using a Spreadsheet Program"
LC-GC International, Vol. 7, pp.590-596 (1994)

[178] Misiti, M., Misiti, Y., Oppenheim, G. and Poggi, J.M.
"MICRONDE: A Matlab Wavelet Toolbox for Signals and Images" in Wavelet and Statistics by Antoniadis, A. and Oppenheim, G. (Eds.)
Springer-Verlag New York Inc. - New York, pp.239-259 (1995)

[179] Misiti, M., Misiti, Y., Oppenheim, G. and Poggi, J.M.
"Wavelet Toolbox User's Guide", Version 1
The MathWorks - Massachusetts (1996)

[180] Mitra, S. and Bose, T.
"Processing of Analytical Signals using MATLAB"
Chemometrics and Intelligent Laboratory Systems, Vol. 22, pp.3-16 (1994)

[181] Mittermayr, C.R., Frischenschlager, H., Rosenberg, E. and Grasserbauer, M.
"Filtering and Integration of Chromatographic Data: A Tool to Improve Calibration?"
Fresenius' Journal of Analytical Chemistry, Vol. 358, pp.456-464 (1997b)

[182] Mittermayr, C.R., Nikolov, S.G., Hutter, H. and Grasserbauer, M.
"Wavelet Denoising of Gaussian Peaks: A Comparative Study"
Chemometrics and Intelligent Laboratory Systems, Vol. 34, pp.187-202 (1996)

[183] Mittermayr, C.R., Rosenberg, E. and Grasserbauer, M.
"Detection and Estimation of Heteroscedastic Noise by Means of the Wavelet Transform"
Analytical Communications, Vol. 34, pp.73-75 (1997a)

[184] Modisette, J.P., Nordlander, P., Kinsey, J.L and Johnson, B.R.
"Wavelet Bases in Eigenvalue Problems in Quantum Mechanics"
Chemical Physics Letters, Vol. 250, pp.485-494 (1996)

[185] Morrison, N.
"Introduction to Fourier Analysis"
John Wiley & Sons Inc. - New York, pp.388 (1994)

[186] Neue, G.
"Simplification of Dynamic NMR Spectroscopy by Wavelet Transforms"
Solid State Nuclear Magnetic Resonance, Vol. 5, pp.305-314 (1996)

[187] Nikolov, S.G., Hutter, H. and Grasserbauer, M.
"De-noising of SIMS Images via Wavelet Shrinkage"
Chemometrics and Intelligent Laboratory Systems, Vol. 34, pp.263-273 (1996)

[188] O'Haver, T.C.
"Teaching and Learning Chemometrics with MATLAB"
Chemometrics and Intelligent Laboratory Systems, Vol. 6, pp.95-103 (1989)

[189] O'Haver, T.C.
"An Introduction to Signal Processing in Chemical Measurement"
Journal of Chemical Education, Vol. 68, pp.A147-A150 (1991)

[190] O'Haver, T.C. and Begley, T.
"Signal-to-noise Ratio in Higher Order Derivative Spectrometry"
Analytical Chemistry, Vol. 53, pp.1876-1878 (1981)

[191] Oppenheim, A.V. and Schafer, R.W.
"Discrete-time Signal Processing"
Prentice-Hall - New Jersey, pp.514-661 (1989)

[192] Otto, M. and Bandemer, H.
"A Fuzzy Method for Component Identification and Mixture Evaluation in the Ultraviolet Spectral Range"
Analytica Chimica Acta, Vol. 191, pp.193-204 (1986)

[193] Ou, M.
"Chinese-English Manual of Common-Used in Traditional Chinese Medicine"
Joint Publishing (H.K.) Co., Ltd. - Hong Kong, pp.158-159 (1995)

[194] Owens, P.M. and Isenhour, T.L.
"Infrared Spectral Compression Procedure for Resolution Independent Search Systems"
Analytical Chemistry, Vol. 55, pp.1548-1553 (1983)

[195] Palavajjhala, S., Motard, R.L. and Joseph, B.
"Computational Aspects of Wavelets and Wavelet Transforms" in Wavelet Application in Chemical Engineering by Motard, R.L. and Joseph, B. (Eds.)
Kluwer Academic Publishers - Massachusetts. pp.33-83 (1994)

[196] Pan, Z.X., Shao, X.G., Zheng, H.B., Liu, W., Wang, H. and Zhang, M.S.
"Correction of Baseline Drift in High-performance Liquid Chromatography by Wavelet Transform"
Chinese Journal of Analytical Chemistry, Vol. 24, pp.149-153 (1996) [In Chinese, with English abstract]

[197] Penchev, P.N., Sohou, A.N. and Andreev, G.N.
"Description and Performance Analysis of an Infrared Library Search System"
Spectroscopy Letters, Vol. 29, pp.1513-1522 (1996)

[198] Penski, W.C., Padowski, D.A. and Bouck, J.B.
"Computer Storage and Search System for Infrared Spectra Including Peak Width and Intensity"
Analytical Chemistry, Vol. 46, pp.955-957 (1974)

[199] Perkampus, H.H.
"UV-VIS Spectroscopy and Its Applications"
Springer-Verlag - Berlin, pp.26 (1992)

[200] Permann, D. and Hamilton, I.
"Wavelet Analysis of Time Series for the Duffing Oscillator The Detection of Order within Chaos"
Physical Review Letters, Vol. 69, pp.2607-2610 (1992)

[201] Permann, D. and Hamilton, I.
"Wavelet Analysis of Time Series for the Weakly Forced and Weakly Damped Morse Oscillator"
Journal of Chemical Physics, Vol. 100, pp.379-386 (1994)

[202] Permann, D.N.S. and Teitelbaum, H.
"Wavelet Fast Fourier Transform (WFFT) Analysis of a Millivolt Signal for a Transient Oscillating Chemical Reaction"
Journal of Physical Chemistry, Vol. 97, 12670-12673 (1993)

[203] Pharmacopoeia Commission of the Ministry of Public Health, P.R. China
"A Coloured Atlas of The Chinese Materia Medica Specified in Pharmacopoeia of The People's Republic of China", 1995 Edition
Joint Publishing (H.K.) Co., Ltd. - Hong Kong, pp.127 (1996)

[204] Phillips, M.L. and White, R.L.
"Dependence of Chromatogram Peak Areas Obtained by Curve-fitting on the Choice of Peak Shape Function"
Journal of Chromatographic Science, Vol. 35, pp.75-81 (1997)

[205] Press, W.H., Teukalsky, S.A., Vetterling, W.T. and Flannery, B.P.
"Numerical Recipes in Fortran 90 The Art of Parallel Scientific Computing", 2nd ed., Cambridge University Press - Cambridge, pp.584-594 (1996)

[206] Provaznik, I. and Kozumplik, J.
"Wavelet Transform in Electrocardiography - Data Compression"
International Journal of Medical Informatics, Vol. 45, pp.111-128 (1997)

[207] Qian, S.N. and Sun, H.
"Data Compression Method Based on Wavelet Transform for Spectral Information"
Spectroscopy and Spectral Analysis (Beijing), Vol. 16, pp.1-8 (1996) [In Chinese, with English abstract]

[208] Rann, C.S.
"Automatic Sorting of Infrared Spectra"
Analytical Chemistry, Vol. 44, pp.1669-1672 (1972)

[209] Rasmussen, G.T. and Isenhour, T.L.
"Library Retrieval of Infrared Spectra Based on Detailed Intensity Information"
Applied Spectroscopy, Vol. 33, pp.371-376 (1979)

[210] Ratton, L., Kunt, T., McAvoy, T., Fuja, T., Cavicchi, R. and Semancik, S.
"A Comparative Study of Signal Processing Techniques for Clustering Microsensor Data (A First Step Towards an Artifical Nose)"
Sensors and Actuators B: Chemical, Vol. 41, pp.105-120 (1997)

[211] Ratzlaff, K.L. and Ratzlaff, E.H.
"The Role of the Microcomputer" in Practical Guide to Chemometrics by Haswell, S.J. (Ed.)
Marcel Dekker, Inc. - New York, pp.271-308 (1992)

[212] Ren, K. and Ren-Kurc, A.
"A New Numerical Method of Finding Potentiometric Titration End-points by Use of Rational Spline Functions"
Talanta, Vol. 33, pp.641-647 (1986)

[213] Rojas, F.S., Ojeda, C.B. and Cano Pavon, J.M.
"Derivative Ultraviolet-visible Region Absorption Spectrophotometry and Its Analytical Applications"
Talanta, Vol. 35, pp.753-761 (1988)

[214] Rzhevskii, A.M. and Mardilovich, P.P.
"Generalized Gans-Gill Method for Smoothind and Differentiation of Composite Profiles in Practice"
Applied Spectroscopy, Vol. 48, pp.13-20 (1994)

[215] Saito, N. and Coifman, R.R.
"Local Discrimiant Bases" in Wavelet Applications in Signal and Image Processing II by Laine, A.F. and Unser, M.A., Proceedings of SPIE - The International Society for Optical Engineering, Vol. 2303, held at San Diego, U.S., 27-29 July, 1994
SPIE - The International Society for Optical Engineering - Washington, D.C., pp.2-14 (1994)

[216] Saji, R. and Konno, H.
"The Uncertainty Relation of One-Soliton Solution with Nonlinear Schrödinger Equation and Its Performance as An Analyzing Wavelet"
Journal of the Physical Society of Japan, Vol. 67, pp.361-364 (1998)

[217] Savitzky, A. and Golay, M.J.E.
"Smoothing and Differentiation of Data by Simplified Least Square Procedures"
Analytical Chemistry, Vol. 36, pp.1627-1639 (1964)

[218] Schostack, K.J. and Malinowski, E.R.
"Investigation of Window Factor Analysis and Matrix Regression Analysis in Chromatography"
Chemometrics and Intelligent Laboratory Systems, Vol. 20, pp.173-182 (1993)

[219] Scott, D.R.
"Effects of Binary Encoding on Pattern Recognition and Library Matching of Spectral Data"
Chemometrics and Intelligent Laboratory Systems, Vol. 4, pp.47-63 (1988)

[220] Seaton, G.G.R., Marr, J.G.D., Clark, B.J. and Fell, A.F.
"Chemometric Methods for the Validation of Peak Homogeneity in HPLC"
Analytical Proceedings, Vol. 23, pp.424-426 (1986)

[221] Shao, X.G. and Cai, W.S.
"Resolution of Multicomponent Chromatograms by Window Factor Analysis with Wavelet Transform Preprocessing"
Journal of Chemometrics, Vol. 12, pp.85-93 (1998)

[222] Shao, X.G., Cai, W.S., Pan, Z.X. and Zhang, M.S.
"Wavelet Transorm and Its Applications in HPLC Analysis" in New Trends in Chemometrics by Liang, Y.Z., Nortvedt, R., Kvalheim, O.M., Shen, H.L. and Xu, Q.S. (Eds.), First Internation Conference on Chemometrics in China held at Zhangjiajie, China, 17-22 October, 1997
Hunan University Press - Changsha, China, pp.205-206 (1997a)

[223] Shao, X.G., Cai, W.S., Sun, P.Y., Zhang, M.S. and Zhao, G.W.
"Quantitative Determination of the Components in Overlapping Chromatographic Peaks using Wavelet Transform"
Analytical Chemistry, Vol. 69, pp.1722-1725 (1997b)

[224] Shao, X.G., Sun, P.Y., Cai, W.S. and Zhang, M.S.
"Wavelet Analysis and Its Application to the Resolution of the Overlapping Chromatograms"
Chemistry (Huaxue Tongbao), No. 8, pp.59-62 (1997c) [In Chinese, without English abstract] [a]

[225] Shao, X.G., Sun, P.Y., Cai, W.S. and Zhang, M.S.
"Resolution of Overlapping Chromatograms by Wavelet Transform"
Chinese Journal of Analytical Chemistry, Vol. 25, pp.671-674 (1997d) [In Chinese, with English abstract]

[226] Shen, H.L., Cui, H., Liang, Y.Z. and Lee, F.
"Qualitative and Quantitative Analysis of Polycyclic Aromatic Hydrocarbons in Environmental Samples by Heuristic Evolving Latent Projections"
Acta Chimica Sinica, Vol. 56, pp.378-384 (1998b) [In Chinese with English abstract]

[227] Shen, H.L., Liang, Y.Z., Yu, R.Q., Li, X.C. and Sun, X.X.
"Analysis of PAHs in Air-borne Particulates in Hong Kong City by Heuristic Evolving Latent Projections"
Science in China (Series B) (Chinese Edition), Vol. 27, pp.556-563 (1997) [In Chinese without English abstract]

[228] Shen, H.L., Liang, Y.Z., Yu, R.Q., Li, X.C. and Sun, X.X.
"Analysis of PAHs in Air-borne Particulates in Hong Kong City by Heuristic Evolving Latent Projections"
Science in China (Series B) (English Edition), Vol. 41, pp.21-29 (1998a)

[229] Shen, H.L., Wang, J.H., Liang, Y.Z. and Chen, W.C.
"Multiresolution Analysis of Hyphenated Chromatographic Data"
Chemical Journal of Chinese Universities, Vol. 18, pp.530-534 (1997a) [In Chinese, with English abstract]

[230] Shen, H.L., Wang, J.H., Liang, Y.Z., Pettersson, K., Josefson, M., Gottfries, J. and Lee, F.
"Chemical Rank Estimation by Multiresolution Analysis for Two-way Data in the Presence of Background"
Chemometrics and Intelligent Laboratory Systems, Vol. 37, 261-269 (1997b)

[231] Shew, S.L.
"Method and Apparatus for Determining Relative Ion Abundances in Mass Spectrometry Utilizing Wavelet Transforms" in U.S. Patent 5,436,447, 25 July, 1995, U.S. Patent and Trademark Office - Washington, D.C. (1995)

[232] Small, G.W., Rasmussen, G.T. and Isenhour, T.L.
"An Infrared Search System Based on Direct Comparsion of Interferograms"
Applied Spectroscopy, Vol. 33, pp.444-450 (1979)

[233] Smith, G.L.
"Principal Component Analysis: An Introduction"
Analytical Proceedings, Vol. 28, pp.150-151 (1991)

[234] Snyder, L.R. and Kirkland, J.J.
"Introduction to Modern Liquid Chromatography", 2nd Ed.,
A Wiley-Interscience Publication - New York, pp.15-82 and 541-574 (1979)

[235] Stark, P.B., Herron, M.M. and Matteson, A.
"Empirically Minimax Affine Mineralogy Estimates from Fourier Transform Infrared Spectrometry using a Decimated Wavelet Basis"
Applied Spectroscopy, Vol. 47, pp.1820-1829 (1993)

[236] Steinier, J., Termonia, Y. and Deltour, J.
"Comments on Smoothing and Differentiation of Data by Simplified Least Square Procedure"
Analytical Chemistry, Vol. 44, pp.1906-1909 (1972)

[237] Strang, G. and Nguyen, T.
"Wavelets and Filter Banks"
Wellesley-Cambridge Press - Massachusetts (1996)

[238] Strobel, H.A. and Heineman, W.R.
"Chemical Instrumentation A Systematic Approach", 3rd Ed.
John Wiley & Sons, Inc. - New York, pp. 824-829, 875 and 1037-1039 (1989)

[239] Suzuki, H., Kinjo, T., Hayashi, Y., Takemoto, M., and Ono, K.
"Wavelet Transform of Acoustic Emission Signals"
Journal of Accoustic Emission, Vol. 14, pp.69-84 (1996)

[240] Talsky, G.
"Derivative Spectrophotometry Low and High Order"
VCH - Weinheim (1994)

[241] Taswell, C.
"WavBox 4: A Software Toolbox for Wavelet Transform and Adaptive Wavelet Packet Decompositions" in Wavelet and Statistics by Antoniadis, A. and Oppenheim, G. (Eds.)
Springer-Verlag New York Inc. - New York, pp.361-375 (1995)

[242] "Using MATLAB", Version 5
The MathWorks, Inc. - Massachusetts (1996)

[243] Toft, J.
"Evolutionary Rank Analysis Applied to Multidetectional Chromatographic Structure"
Chemometrics and Intelligent Laboratory Systems, Vol. 29, pp.189-212 (1995)

[244] Toft, J. and Kvalheim, O.M.
"Multi-array Resolution Parameter for Multidetection Chromatography. Performance of Alternating Regression, Iterative Target Transformation Factor Analysis and Heuristic Evolving Latent Projections"
Chemometrics and Intelligent Laboratory Systems, Vol. 25, pp.61-75 (1994)

[245] Tymczak, C.J.
"Orthonormal Wavelet Bases for Quantum Molecular Dynamics"
Physical Review Letters, Vol. 78, pp.3654-3657 (1997)

[246] Toft, J. and Kvalheim, O.M.
"Multi-array Resolution Parameter for Multidetection Chromatography. Performance of Alternating Regression, Iterative Target Transformation Factor Analysis and Heuristic Evolving Latent Projections"
Chemometrics and Intelligent Laboratory Systems, Vol. 25, pp.61-75 (1994)

[247] Vaidyanathan, P.P.
"Orthonormal and Biorthonormal Filter Banks as Convolvers, and Convolutional Coding Gain"
IEEE Transactions on Signal Processing, Vol. 41, pp.2110-2130 (1993)

[248] Vandeginste, B.G., Derks, W. and Kateman, G.
"Multicomponent Self-modelling Curve Resolution in High-performance Liquid Chromatography by Iterative Target Transform Analysis"
Analytica Chimica Acta, Vol. 173, pp.253-264 (1985)

[249] Vankeerberghen, P., den Bogaert, B.V. and Massart, D.L.
"Dynamic Link Libraries I. Introduction"
Trends in Analytical Chemistry, Vol. 15, pp.206-208 (1996a)

[250] Vankeerberghen, P., den Bogaert, B.V. and Massart, D.L.
"Dynamic Link Libraries II. Development"
Trends in Analytical Chemistry, Vol. 15, pp.209-214 (1996b)

[251] Verdun, F.R., Giancaspro, C. and Marshall, A.G.
"Effects of Noise, Time-domain Damping, Zero-filling and the FFT Algorithm on the "Exact" Interpolation of Fast Fourier Transform Spectra"
Applied Spectroscopy, Vol. 42, pp.715-721 (1988)

[252] von Kitzing, E. and Schmitt, E.
"Configurational Space of Biological Macromolecules as seen by Semi-empirical Force Fields: Inherent Problems for Molecular Design and Strategies to Solve them by means of Hierarchical Force Fields"
Journal of Molecular Structure (Theochem), Vol. 336, pp.245-259 (1995)

[253] Walczak, B., Bogert, B. and Massart, D.L.
"Application of Wavelet Packet Transform in Pattern Recognition of Near-IR Data"
Analytical Chemistry, Vol. 68, pp.1742-1747 (1996)

[254] Walczak, B., Bouveresse, E. and Massart, D.L.
"Standardization of Near-infrared Spectra in the Wavelet Domain"
Chemometrics and Intelligent Laboratory Systems, Vol. 36, pp.41-51 (1997)

[255] Walczak, B. and Massart, D.L.
"Noise Suppression and Signal Compression using the Wavelet Packet Transform"
Chemometrics and Intelligent Laboratory Systems, Vol. 36, pp.81-94 (1997a)

[256] Walczak, B. and Massart, D.L.
"Wavelet Packet Transform Applied to A Set of Signals: A New Approach to the Best-basis Selection"
Chemometrics and Intelligent Laboratory Systems, Vol. 38, pp.39-50 (1997c)

[257] Walczak, B. and Massart, D.L.
"Wavelets - Something for Analytical Chemistry?"
Trends in Analytical Chemistry, Vol. 16, pp.451-463 (1997b)

[258] Wang, C.P. and Isenhour, T.L.
"Infrared Library Search on Principal Component Analyzed Fourier Transformed Absorption Spectra"
Applied Spectroscopy, Vol. 41, pp.185-194 (1987)

[259] Wang, C.P. and Isenhour, T.L.
"Infrared Library Search on Principal Component analyzed Fourier Transformed Absorption Spectra"
Applied Spectroscopy, Vol. 41, pp.185-195 (1987)

[260] Wang, D.I.C., Leung, J.C., Wu, I.C. and Gao, N.
"Biotechnology and Hong Kong" in Made by Hong Kong by Berger, S. and Lester, R.K. (Eds.)
Oxford University Press (China) Ltd. - Hong Kong, pp.249-292 (1997)

[261] Wang, H., Si, S.Z., Xiong, J.H. and Zhang, M.S.
"The Compression and Reconstruction of IR Spectra by Wavelet Transform"
Chemical Journal of Chinese Universities, Vol. 17, pp.1537-1540 (1996) [In Chinese, with English abstract]

[262] Wang, H., Pan, Z.X., Liu, W., Zhang, M.S., Si, S.Z. and Wang, L.P.
"The Determination of Potentiometric Titration End-Points by Using Wavelet Transform"
Chemical Journal of Chinese Universities, Vol. 18, pp.1286-1290 (1997b) [In Chinese, with English abstract]

[263] Wang, H., Xiao, J.H., Pan, Z.X. and Zhang, M.S.
"Wavelet Transform and Its Application in Chemistry"
Chemistry (Huaxue Tongbao), No.6, pp.20-23 (1997a) [In Chinese, without English abstract] [a]

[264] Warr, W.A.
"Spectral Database"
Chemometrics and Intelligent Laboratory Systems, Vol. 10, pp.279-292 (1991)

[265] Warr, W.A.
"Computer-assisted Structure Elucidation - Library Search and Spectral Data Collections"
Analytical Chemistry, Vol. 65, pp.1045A-1050A (1993a)

[266] Warr, W.A.
Computed-assisted Structure Elucidation - Indirect Database Approaches and Established Systems"
Analytical Chemistry, Vol. 65, pp.1087A-1095A (1993b)

[267] Wei, S.Q. and Chou, M.Y.
"Wavelets in Self-consistent Electronic Structure Calculations"
Physical Review Letters, Vol. 76, pp.2650-2653 (1996)

[268] Welsh, W.J., Lin, W.K., Tersigni, S.H., Collantes, E., Duta, R., Carey, M.S., Zielinski, W.L., Brower, J., Spencer, J.A. and Layloff, T.P.
"Pharmaceutical Fingerprinting: Evaluation of Neural Networks and Chemometrics Techniques for Distinguishing among Same-Product Manufacturers"
Analytical Chemistry, Vol. 68, 3473-3482 (1996)

[269] Whitfield, P.H. and Dohan, K.
"Identification and Characterization of Water Quality Transients using Wavelet Analysis. II. Application to Electronic Water Quality Data"
Water Science Technology, Vol. 36, pp.337-348 (1997)

[270] Whitfield, P.H. and Wade, N.L.
"Monitoring Transient Water Quality Events Electronically"
Water Resources Bulletin, Vol. 28, pp.703-711 (1992)

[271] Wickerhauser, M.V.
"Adapted Wavelet Analysis from Theory to Software"
A. K. Peter - Massachusetts, (1994a)

[272] Wickerhauser, M.V.
"Large-rank Approximate Principal Component Analysis with Wavelets for Signal Feature Discrimination and the Inversion of Complicated Maps"
Journal of Chemical Information and Computer Science, Vol. 34, pp.1036-1046 (1994b)

[273] Wickerhauser, M.V.
"Time Localization Techniques for Wavelet Transforms"
Croatica Chemica Acta, Vol. 68, pp.1-27 (1995)

[274] Wickham, D.
"Photodiode Array Absorbance Detection" in A Practical Guide to HPLC Detection by Parriott, D. (Ed.)
Academic Press, Inc. - San Diego, pp.67-109. (1993)

[275] Willson, P.D. and Edwards, T.H.
"Sampling and Smoothing of Spectra"
Applied Spectroscopy Review, Vol. 12, pp.1-81 (1976)

[276] Wolkenstein, M., Hutter, H. and Grasserbauer, M.
"Wavelet Filtering for Analytical Data"
Fresenius' Journal of Analytical Chemistry, Vol. 358, pp.165-169 (1997a)

[277] Wolkenstein, M., Hutter, H., Nikolov, S.G. and Grasserbauer, M.
"Improvement of SIMS Image Classigication by Means of Wavelet De-noising"
Fresenius' Journal of Analytical Chemistry, Vol. 357, pp.783-788 (1997b)

[278] Wolkenstein, M., Hutter, H., Nikolov, S.G., Schmitz, I. And Grasserbauer, M.
"Comparison of Wavelet Filtering with Established Techniques for De-noising of EPMA Images"
Journal of Trace and Microprobe Techniques, Vol. 15, pp.33-49 (1997c)

[279] Wong, A.
"The Boom to Come"
Hong Kong Industrialist, December, pp.30-35 (1997)

[280] Wong, T.W., Yu, T.S., Liu, J.L.Y., Lee, N.L. and Lloyd, O.L.
"Factors Associated with the Utilization of Traditional Chinese Medicine in a Small Town in Hong Kong"
American Journal of Chinese Medicine, Vol. XXV, pp.367-373 (1997)

[281] Xie, Q.T., Liu, W., Pan, Z.X. and Zhang, M.S.
"Wavelet Neural Network and Its Application in Chemical Spectra" in New Trends in Chemometrics by Liang, Y.Z., Nortvedt, R., Kvalheim, O.M., Shen, H.L. and Xu, Q.S. (Eds.), First Internation Conference on Chemometrics in China held at Zhangjiajie, China, 17-22 October, 1997
Hunan University Press - Changsha, China, pp.203-204 (1997)

[282] Xiong, W.G. and Liu, Y.W.
"Application of Computer in Chinese Medicine Analysis"
China Journal of Chinese Materia Medica, Vol. 23, pp.124-126 (1998) [In Chinese]

[283] Xu, W.H., Xue, Z. and Ma, J.W.
"Studies on the Water Soluble Constituents of *Cordyceps Sinensis* (Berk.) Sacc. - The Nucleosides"
Bulletin of Chinese Materia Medica, Vol. 13, pp.226-229 (1988) [In Chinese]

[284] Yan, L. and Mo, J.Y.
"Study on New Real-time Digitial Wavelet Filters to Electroanalytical Signals"
Chinese Science Bulletin (Chinese Edition), Vol. 40(17), pp.1567-1570 (1995) [In Chinese, without English abstract]

[285] Yu, K.Q., Li, H., Liu, J.F. and Song, J.F.
"Chronopotentiometric Stripping Voltammetry with Wavelet Analysis" in New Trends in Chemometrics by Liang, Y.Z., Nortvedt, R., Kvalheim, O.M., Shen, H.L. and Xu, Q.S. (Eds.), First Internation Conference on Chemometrics in China held at Zhangjiajie, China, 17-22 October, 1997
Hunan University Press - Changsha, China, pp.195 (1997)

[286] Zheng, X.P. and Mo, J.Y.
"The Couple Application of B-spline Wavelet and RLT Filtration in Staircase Voltammetry" in New Trends in Chemometrics by Liang, Y.Z., Nortvedt, R., Kvalheim, O.M., Shen, H.L. and Xu, Q.S. (Eds.), First Internation Conference on Chemometrics in China held at Zhangjiajie, China, 17-22 October, 1997
Hunan University Press - Changsha, China, pp.199-200 (1997)

[287] Zheng, X.P., Mo, J.Y. and Cai, P.X.
"Simultaneous Application of Spline Wavelet and Riemann-Liouville Transform Filtration in Electroanalytical Chemistry"
Analytical Communications, Vol. 35, pp.57-59 (1998)

[288] Zhong, H.B., Zheng, J.B., Pan, Z.X., Zhang, M.S. and Gao, H.
"Investigation on Application of Wavelet Transform in Recovering Useful Information from Oscillographic Signal"
Chemical Journal of Chinese Universities, Vol. 19, pp.574-549 (1998) [In Chinese, with English abstract]

[289] Zhu, Q.W. and Stillman, M.J.
"Expert Systems and Analytical Chemistry: Recent Progress in the Acexpert Project"
Journal of Chemical Information and Computer Science, Vol. 36, pp.497-509 (1996)

[290] Zissis, K.D., Brereton, R.G. and Escott, R.
"Monitoring of Impurities using High-performance Liquid Chromatography with Diode-array Detection: Eigenvalue Plots of Partially Overlapping Tailing Peaks"
Analyst, Vol. 122, pp.1007-1013 (1997)

[291] Zou, X.Y. and Mo, J.Y.
"Spline Wavelets Analysis of Step Voltammetry Signals"
Chemical Journal of Chinese Universities, Vol. 17, pp.1522-1527 (1996) [In Chinese, with English abstract]

[292] Zou, X.Y. and Mo, J.Y.
"Spline Wavelet Analysis for Voltammetric Signals"
Analytica Chimica Acta, Vol. 340, pp.115-121 (1997a)

[293] Zou, X.Y. and Mo, J.Y.
"Spline Wavelet Multifiltering Analysis"
Chinese Sciene Bulletin (Chinese Edition), Vol. 42(4), pp.382-385 (1997b) [In Chinese, without English abstract]

[294] Zou, X.Y. and Mo, J.Y.
"Spline Wavelet Multifiltering Analysis"
Chinese Science Bulletin (English Edition), Vol. 42(8), pp.640-644 (1997c)

[295] Zupan, J.
"Algorithms for Chemists"
John Wiley & Sons, Inc. - Chichester, pp.102-114 (1989)

[a] The English title was translated by the authors of this review.

APPENDICES

Appendix 4.1 Program description of the Spectral Library Compression and Search (SLCS) Toolbox Version 1.0

Program Name	Functions
Main Program	
LIBINIT.M	To initiate various program of the toolbox.
Spectral Library Set Up	
LIBSETUP.M	To set up a spectral library from individual spectrum data file.
LIBSETAD.M	To add new spectral data file into the spectral library.
LIBSETCD.M	To change data such as spectral data, compound name or delete spectral data record from the spectral library.
FWT Computation	
LIBFWTC.M	To compress spectral library by TRT-FWT algorithm.
LIBFWTR.M	To regenerate spectral library by IFWT-ITRT algorithm.
FWTCOMP.M	To perform TRT-FWT computation on a spectrum to determine the required parameters such as resolution level *J*, Daubechies coefficients and cutoff value.
IFWTCOMP.M	To perform IFWT-IFWT computation on a WT processed spectrum and compare the difference between regenerated and original spectrum.
WPT Computation	
LIBWPDC.M	To compress spectral library by TRT-WPT algorithm.
LIBWPDR.M	To regenerate spectral library by IWPT-ITRT algorithm.

(cont.)

(cont. Appendix 4.1)

WPDCOMP.M	To perform TRT-WPT computation on a spectrum to determine the required parameters such as resolution level, Daubechies coefficients and cutoff value.
IWPDCOMP.M	To perform IWPD-ITRT computation on a WPT processed spectrum and compare the difference between regenerated and original spectrum.

FFT Computation

LIBFFTC.M	To compress spectral library by TRT-FFT algorithm.
LIBFFTR.M	To regenerate spectral library by IFFT-ITRT algorithm.

Spectral Library Search

LIBSEFFT.M	To perform spectral library search with FFT compressed spectral library.
LIBSEFWT.M	To perform spectral library search with FWT compressed spectral library.
LIBSEWPD.M	To perform spectral library search with WPT compressed spectral library.
LIBHITX.M	To display a hit index table for a particular library search result.
LIBSORT.M	To sort the search results from various methods that include correlation coefficient, absolute derivative, absolute difference, square derivative and square difference.

Wavelet Computation

ENTROPY.M	To perform the Shannon-Weaver entropy calculation.
FWT.M	To perform FWT calculation at a particular resolution level J and Daubechies wavelet filters. (Command version)
IFWT.M	To perform IFWT calculation at a particular resolution level J and Daubechies wavelet filters. (Command version)
WPD.M	To perform WPT calculation at a particular resolution level J and Daubechies wavelet filters. (Command version)

(cont.)

(cont. Appendix 4.1)

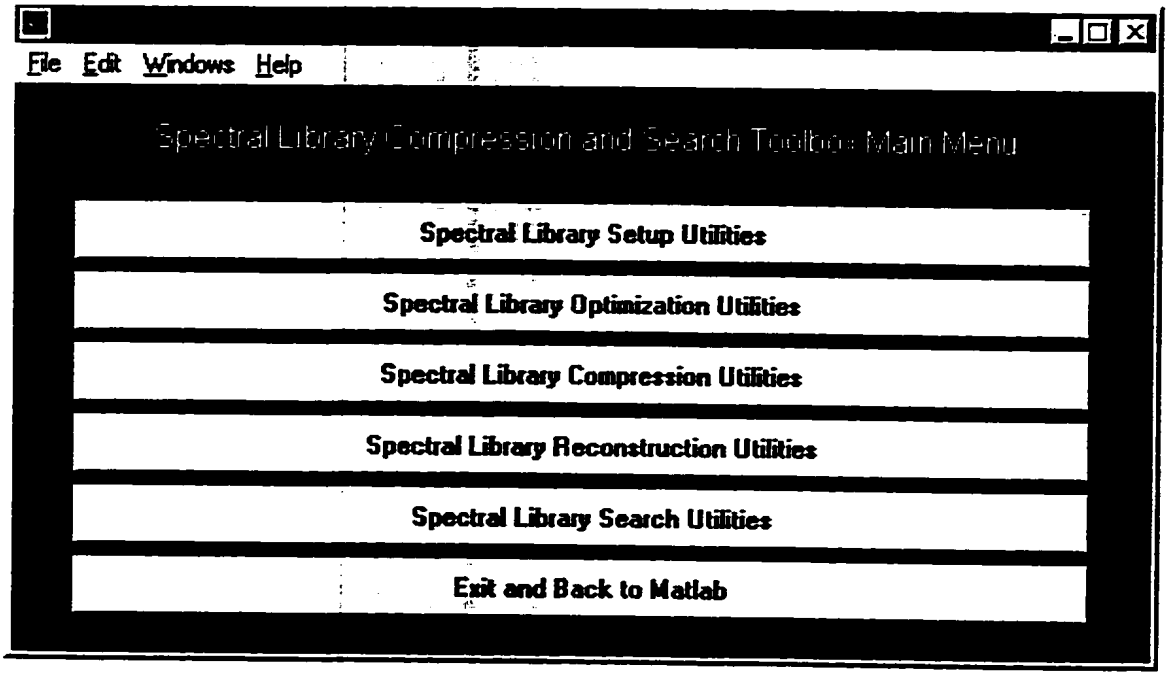
IWPD.M	To perform IWPT calculation at a particular resolution level *J* and Daubechies wavelet filters.
WAVECOEF.M	To calculate the Daubechies's wavelet coefficients.

Supporting Files

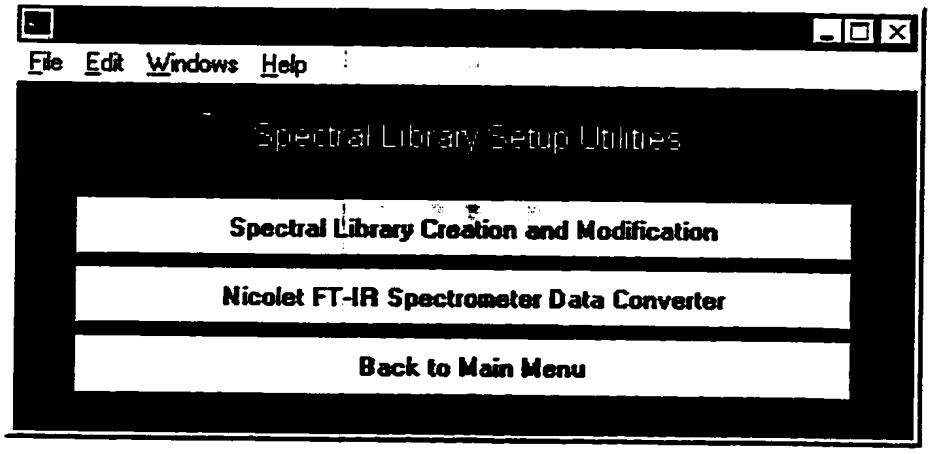
FILE2VAR.M	To assign data from a file to a particular variable name.
NAMECONV.M	To generate different file extension for a particular file.
TRT.M	To perform the TRT and ITRT computation on the spectral data.

Appendix 4.2 Screen capture from the Spectral Library Compression and Search (SLCS) Toolbox Version 1.0

Screen 4.2.1 Main menu of the Spectral Library Compression and Search Toolbox



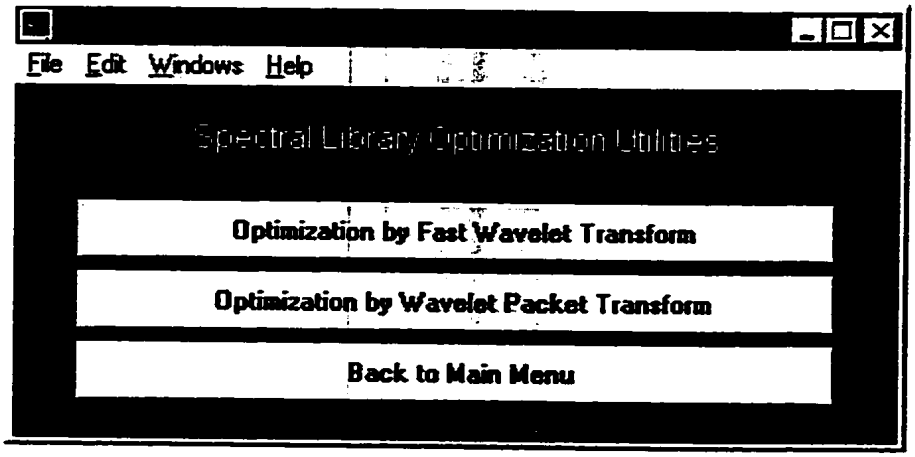
Screen 4.2.2 Menu for spectral library setup utilities.



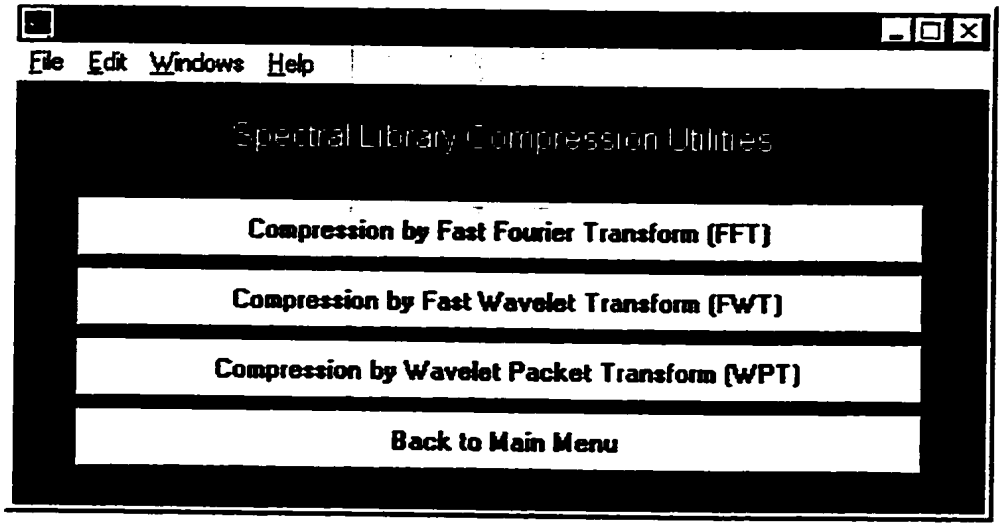
(cont.)

(cont. Appendix 4.2)

Screen 4.2.3 Menu for spectral library optimization.



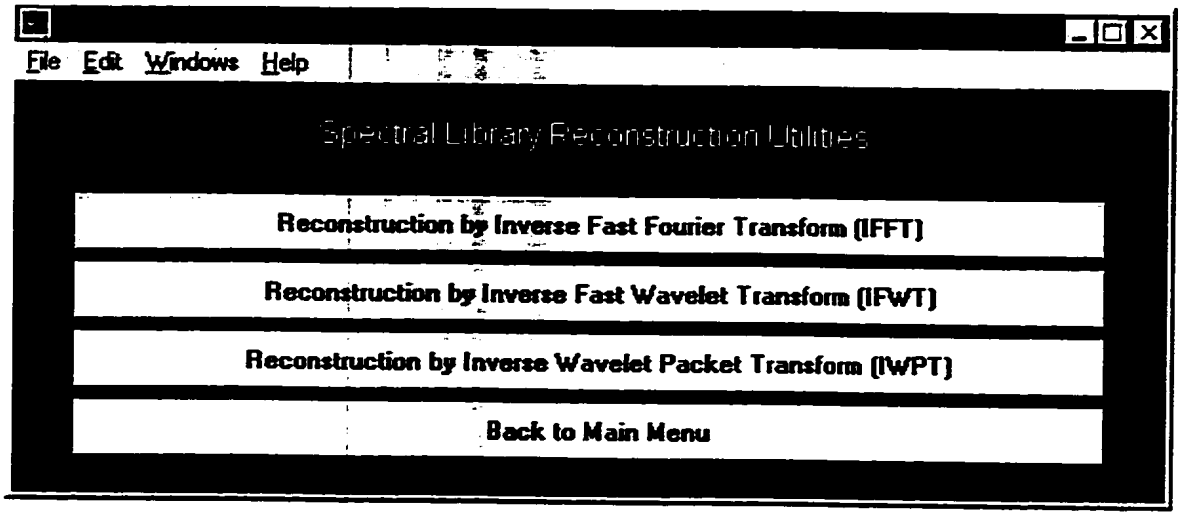
Screen 4.2.4 Menu for spectral library compression by using FFT, FWT and WPT.



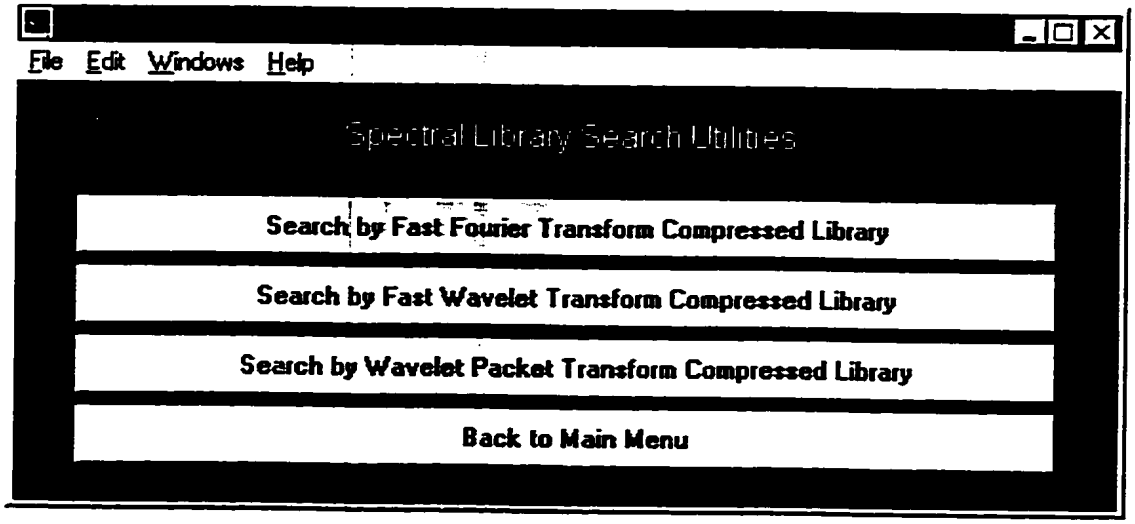
(cont.)

(cont. Appendix 4.2)

Screen 4.2.5 Menu for spectral library reconstruction by using IFFT, IFWT and IWPT



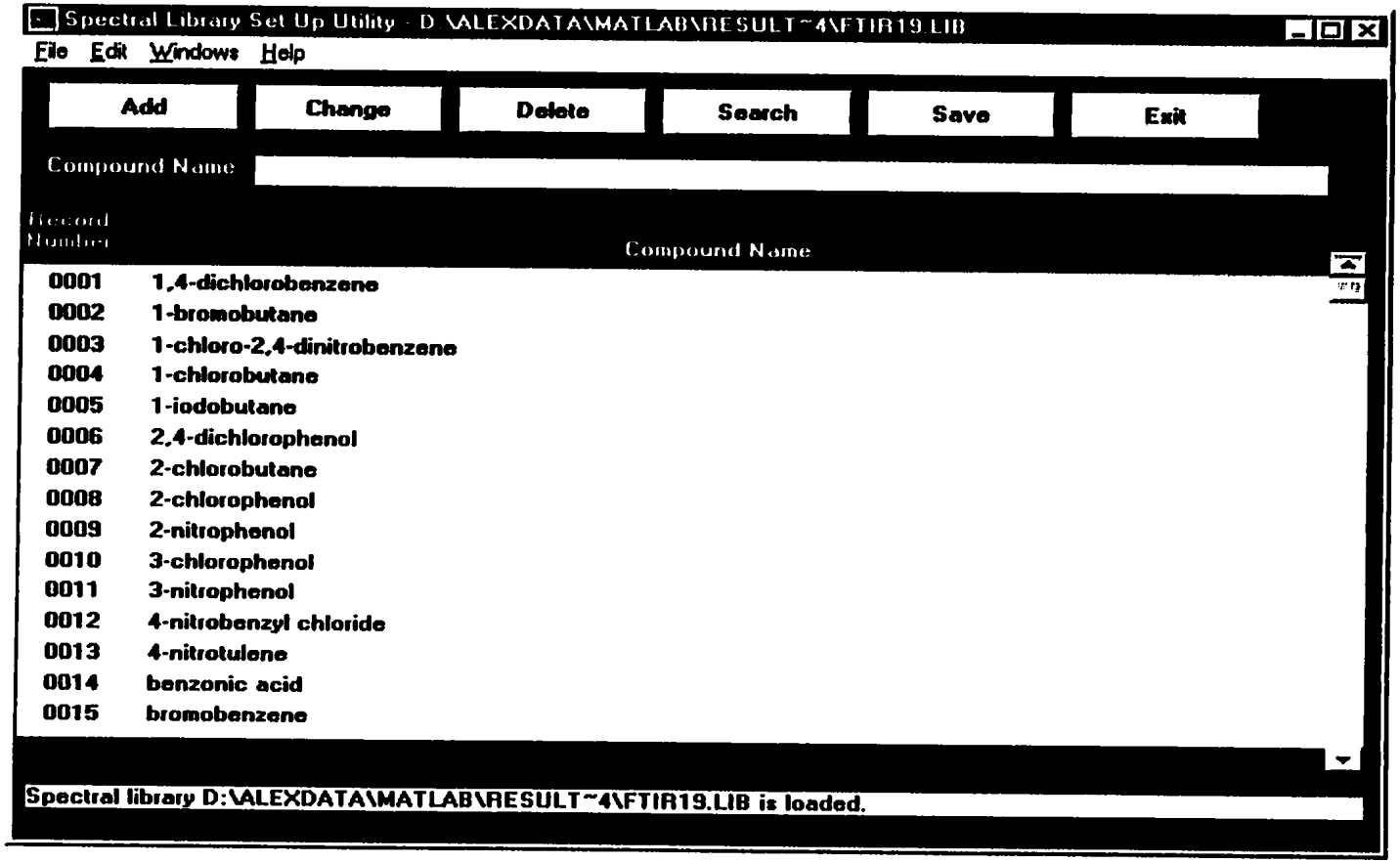
Screen 4.2.6 Menu for spectral library search by using FFT, FWT or WPT compressed library.



(cont.)

(cont. Appendix 4.2)

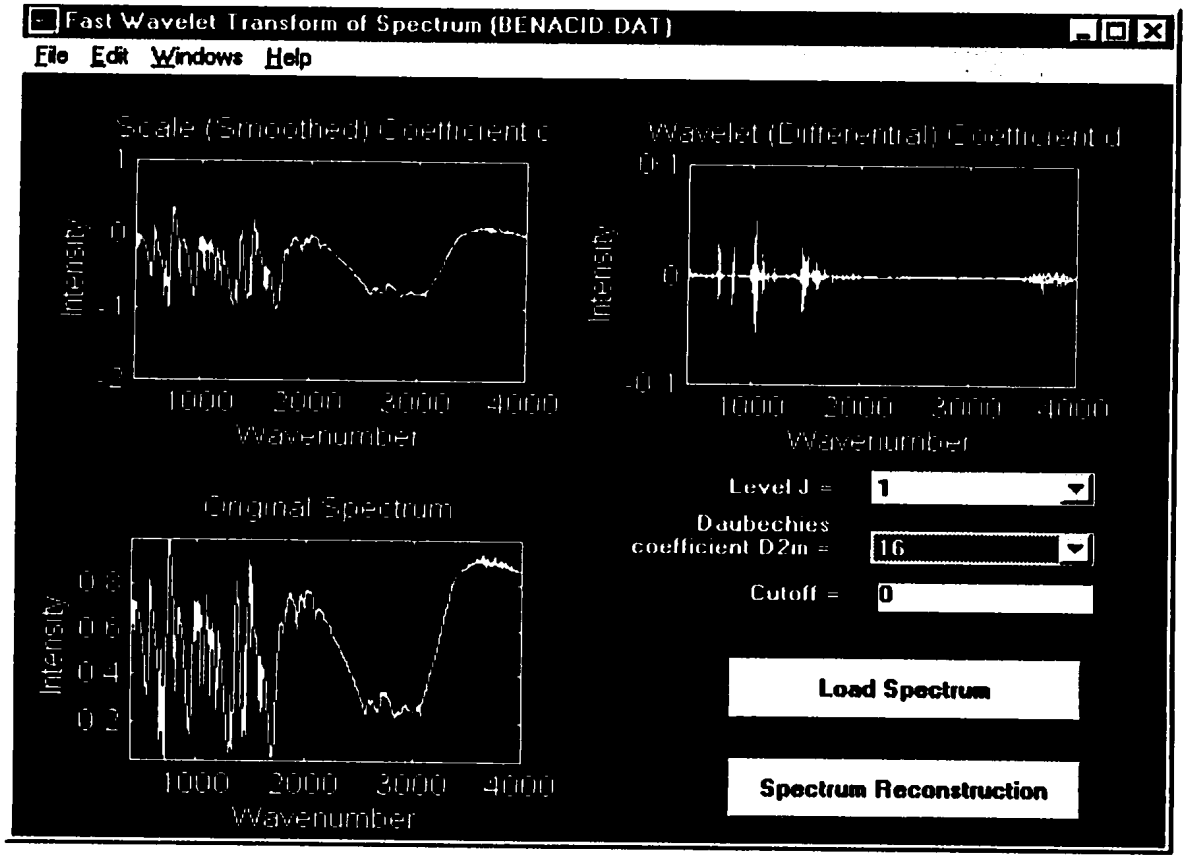
Screen 4.2.7 Program for modifying the spectral library.



(cont.)

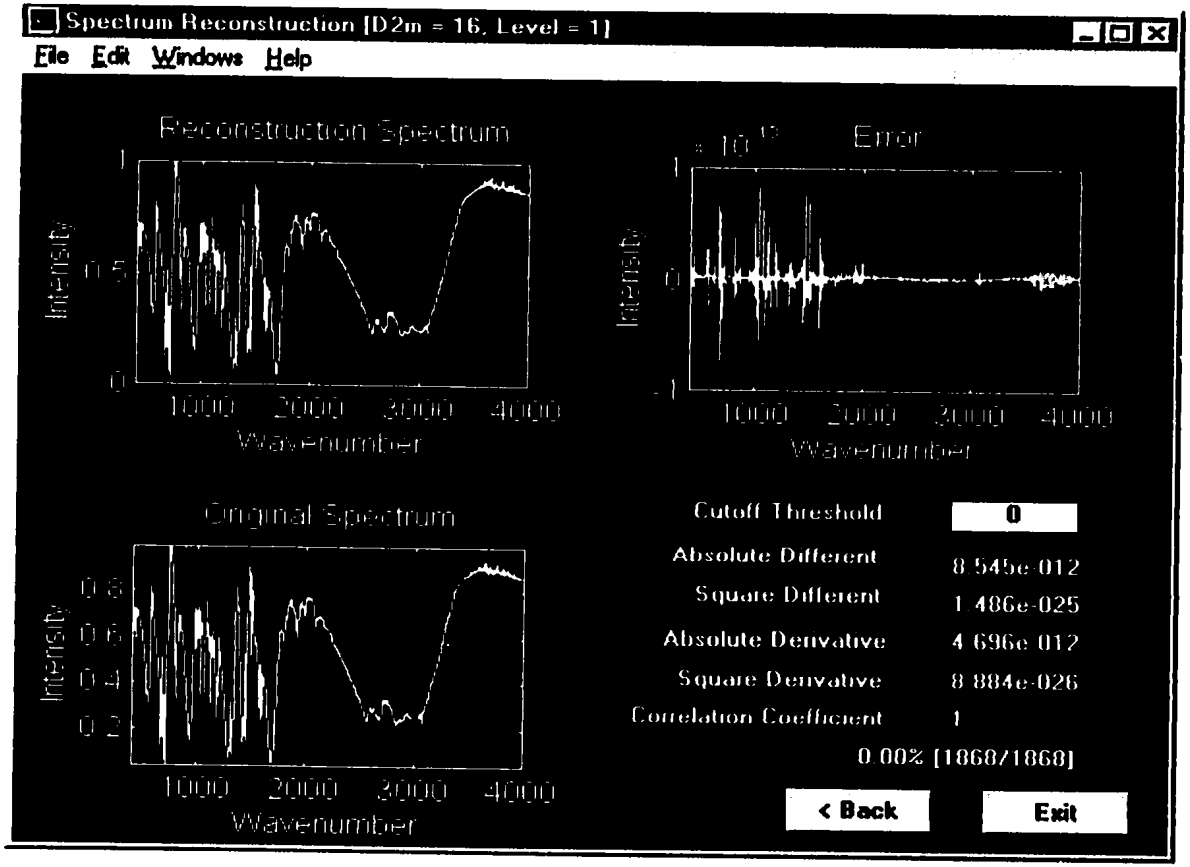
(cont. Appendix 4.2)

Screen 4.2.8 Program for optimization of parameters for fast wavelet transform computation.

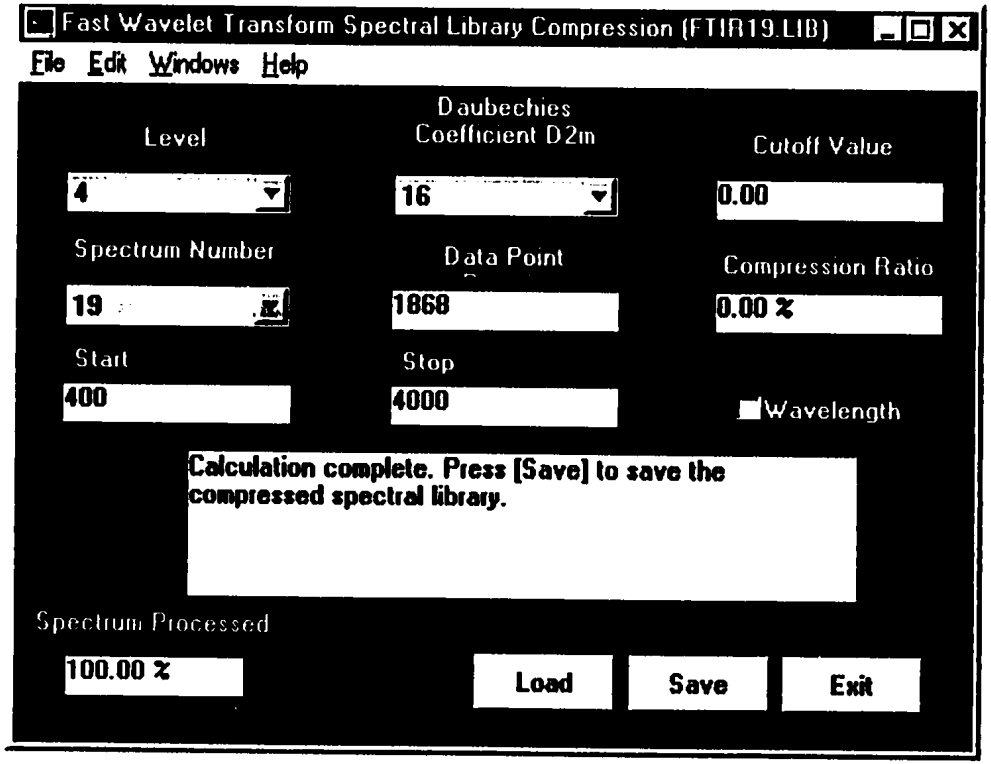


(cont.)

Screen 4.2.9 Results from spectrum reconstruction.

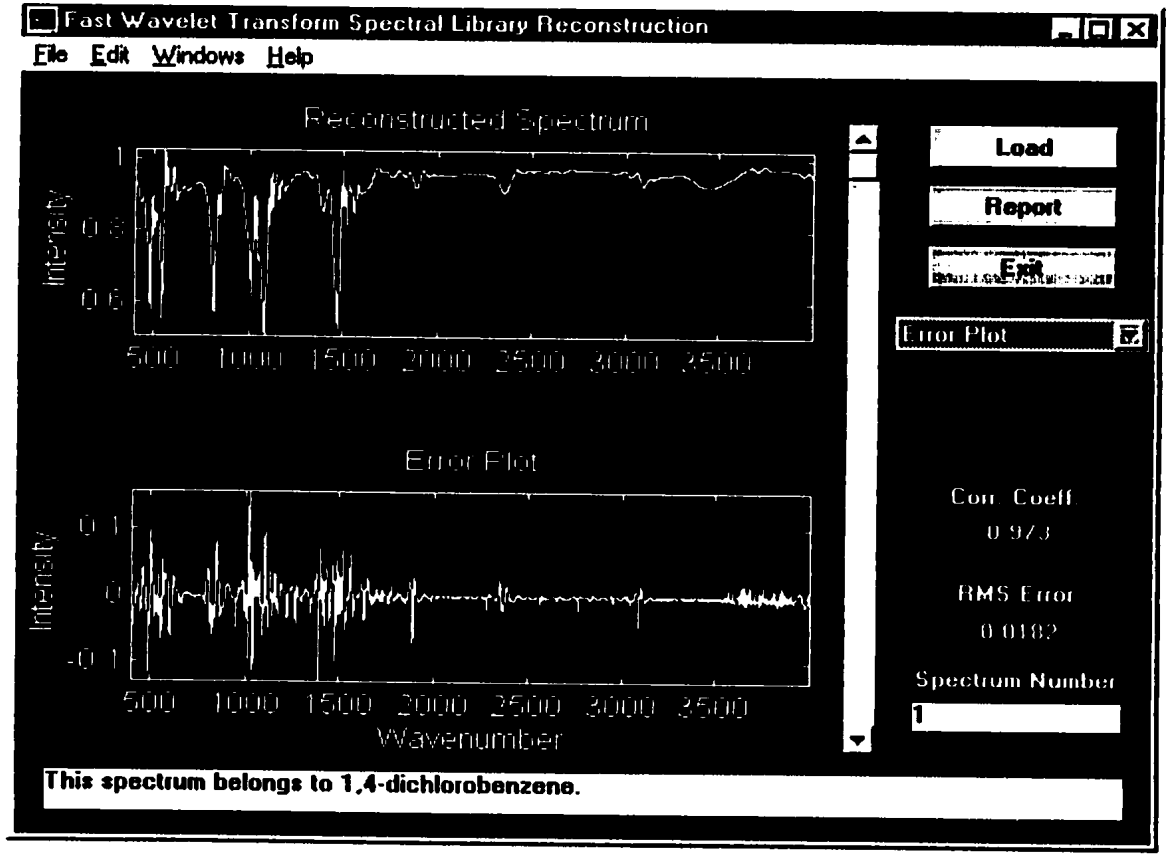


Screen 4.2.10 Program for spectral library compression with FWT.



(cont. Appendix 4.2)

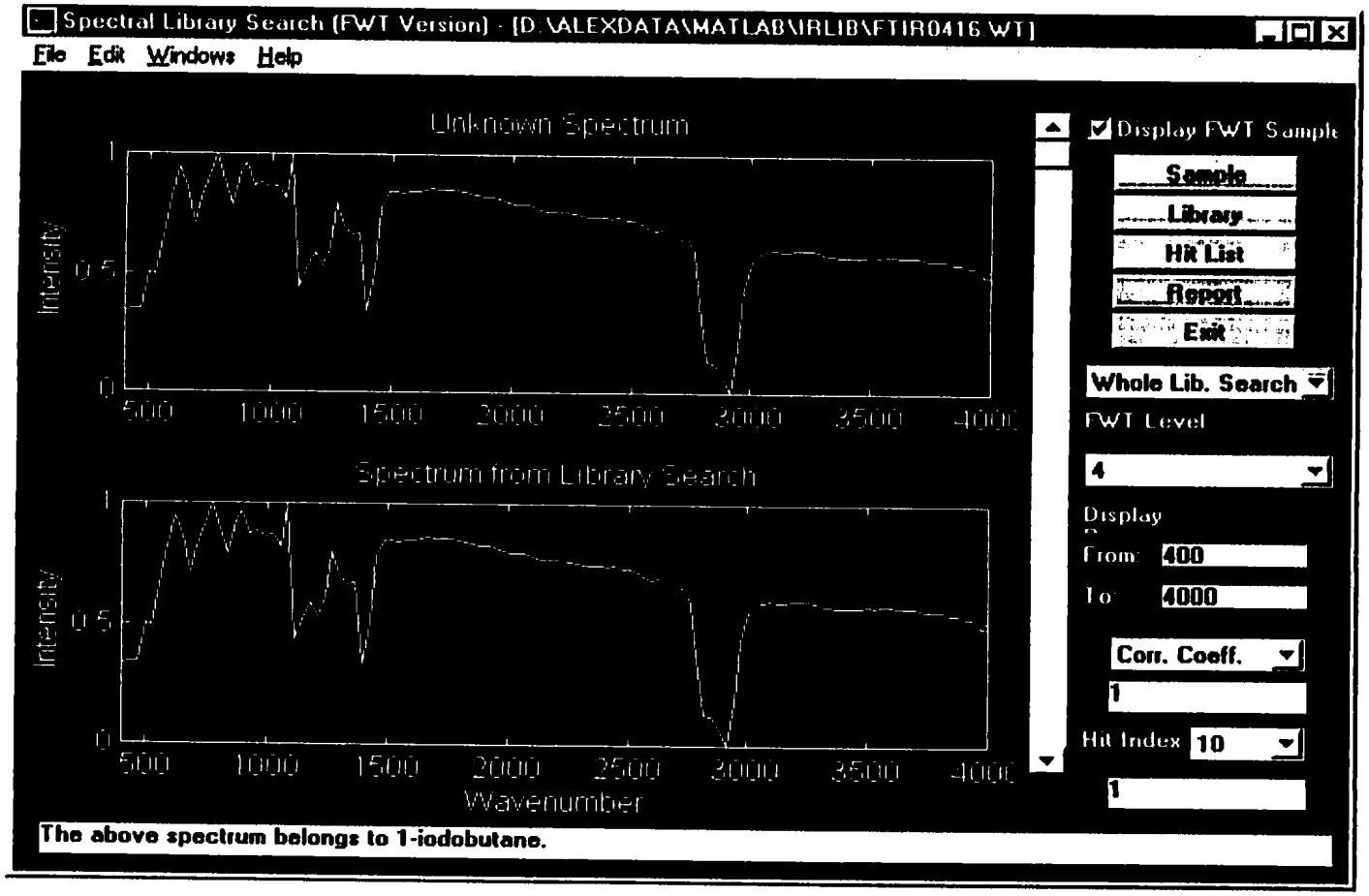
Screen 4.2.11 Program for spectral library reconstruction with IFWT.



(cont.)

(cont. Appendix 4.2)

Screen 4.2.12 Program for spectral library search by using FWT compressed library.



Appendix 4.3 File format of the individual IR spectrum data file and the spectral library that generated from FWT, WPT and FFT compression.

A.4.3.1 Spectral Library Format for FWT Treatment

Row number [a]	Column number	Description
1	1	Total number of data points per spectrum (N)
	2	Total number of spectra in the spectral library (M)
	3	Resolution level (J)
	4	Type of Daubechies wavelet filter D_{2m} (m)
	5	Cutoff value used in the compression process
	6	Location of the position information for each spectrum (*WritePos*)
	7	Starting wavenumber or wavelength of the IR spectrum
	8	Ending wavenumber or wavelength of the IR spectrum
	9	Indication for the unit of position information (wavenumber = 0; wavelength = 1)
	10	Location of the data for ITRT treatment
	11	Location of the compound name
	12-14	Spectral library file identification number
2 to M	1 to *WritePos*	Compressed IR spectral data
	WritePos+1 to 2×*WritePos*	Corresponding position of the IR spectral data in the wavelet domain
	2×*WritePos*+1	First data point $c_1^{(0)}$ in the original IR spectrum
	2×*WritePos*+2	Last data point $c_N^{(0)}$ in the original IR spectrum
	2×*WritePos*+3 to End of column	Compound name (Converted to ASCII code)

(cont.)

(cont. Appendix 4.3)

A.4.3.2 Spectral Library Format for WPT Treatment

Row number [a]	Column number	Description
1	1	Total number of data points per spectrum (N)
	2	Total number of spectra in the spectral library (M)
	3	Resolution level (J)
	4	Type of Daubechies wavelet filter D_{2m} (m)
	5	Cutoff value used in the compression process
	6	Location of the position information for each spectrum (*WritePos*)
	7	Starting wavenumber or wavelength of the IR spectrum
	8	Ending wavenumber or wavelength of the IR spectrum
	9	Indication for the unit of position information (wavenumber = 0; wavelength = 1)
	10	Location of the data for ITRT treatment
	11	Location of the compound name
	12-14	Spectral library file identification number
2 to M	1 to *WritePos*	Compressed IR spectral data
	WritePos+1 to 2×*WritePos*	Corresponding position of the IR spectral data in the wavelet domain
	2×*WritePos*+1 to 2×*WritePos*+2^J-1	Wavelet packet basis information
	2×*WritePos*+2^J	First data point $c_1^{(0)}$ in the original IR spectrum
	2×*WritePos*+2^J+1	Last data point $c_N^{(0)}$ in the original IR spectrum
	2×*WritePos*+2^J+2 to End of column	Compound name (Converted to ASCII code)

(cont.)

(cont. Appendix 4.3)

A.4.3.3 Spectral Library Format for FFT Treatment

Row number [a]	Column number	Description
1	1	Total number of data points per spectrum (N)
	2	Total number of spectra in the spectral library (M)
	3	Resolution level (J)
	4	Type of Daubechies wavelet filter D_{2m} (m)
	5	Cutoff value used in the compression process
	6	Location of the position information for each spectrum (*WritePos*)
	7	Starting wavenumber or wavelength of the IR spectrum
	8	Ending wavenumber or wavelength of the IR spectrum
	9	Indication for the unit of position information (wavenumber = 0; wavelength = 1)
	10	Location of the data for ITRT treatment
	11	Location of the compound name
	12-14	Spectral library file identification number
2 to M	1 to *WritePos*/2	Real part of the compress IR spectral data
	WritePos/2+1 to *WritePos*	Imaginary part of the compress IR spectral data
	WritePos+1 to 1.5×*WritePos*	Corresponding position of the IR spectral data in the wavelet domain
	1.5×*WritePos*+1	First data point $c_1^{(0)}$ in the original IR spectrum
	1.5×*WritePos*+2	Last data point $c_N^{(0)}$ in the original IR spectrum
	1.5×*WritePos*+3 to End of column	Compound name (Converted to ASCII code)

(cont.)

(cont. Appendix 4.3)

A.4.3.4 Spectral Library Format for data without Compression

Row number [a]	Column number	Description
1	1	Total number of data points per spectrum (*N*)
	2	Total number of spectra in the spectral library (*M*)
	3	Maximum length of the compound name (*NameSize*)
	4	Starting wavenumber or wavelength of the IR spectrum
	5	Ending wavenumber or wavelength of the IR spectrum
2 to *M*	1 to N	Raw IR spectral data
	N+1 to End of column	Compound name (Converted to ASCII code)

(cont.)

(cont. Appendix 4.3)

A.4.3.5 Raw IR Spectrum Data File [a]

Column number	Description
1 to *N*	Raw IR spectral data
N+1	Starting wavelength or wavelength of the IR spectrum
N+2	Ending wavelength or wavelength of the IR spectrum

[a] Row 1 is used to store general information of the spectral library. Rows 2 to *M* are used to store individual IR spectral data.

Appendix 4.4 File extension of files in the Spectral Library Compression and Search (SLCS) Toolbox Version 1.0

File extension	Description
CSV	Raw spectrum data file from Nicolet FT-IR spectrometer in comma-separated value format
DAT	Spectrum data file in tab-delimited format
LIB	Spectral library file without compression
FFT	Spectral library file with FFT compression
WT	Spectral library file with FWT compression
WP	Spectral library file with WPT compression
LOF	Log file of the spectral library with FFT compression
LOT	Log file of the spectral library with FWT compression
LOP	Log file of the spectral library with WPT compression
REF	Report of spectral library with FFT compression
RET	Report of spectral library with FWT compression
REP	Report of spectral library with WPT compression
SHF	Report on spectral library search with the use of FFT compressed library
SHP	Report on spectral library search with the use of FWT compressed library
SHT	Report on spectral library search with the use of WPT compressed library

Appendix 5.1 Macro for conversion of experimental data of HP 1050 HPLC-DAD system to ASCII text file.

```
!Macro name:  3DATA.MAC (Version 1.1)
!Author:      Mr. Alexander K.M. Leung
!             Dept. of Appl. Bio. & Chem. Tech., The Hong Kong Polytechnic Univ.
!Date:        22 May, 1997
!Modified at: 9 Jan, 1998
!             1. Include wavelength selection
!
!Usage:       To convert Hewlett Packard's data file, whose extension name is .d,
!             into ASCII format.
!
!Procedure:   1. Type DataMenu and press Enter in the command prompt to add a Data Menu.
!             2. Load the signal file in HP ChemStation by choosing Load Signal in File menu.
!             3. Choose Export from the Data Menu.
!             4. Modify the path and filename as required.
!             5. Choose required retention time and wavelength
!
!Output:      *.dat - Raw data file
!             *.tim - Retention time information file
!             *.wal - Wavelength information file
!             *.inf - General information file
!Variable List
!DataFile$     - Filename of the data file
!FileStr$      - Filename of the data to be saved
!InfoFile$     - Filename of the sample information
!OutStr$       - String of output data
!PathStr$      - Path name of the data to be saved
!TimeFile$     - Filename of the retention time data
!WaveFile$     - Filename of the wavelength data
!
!ColsNum       - Number of columns
```

(cont.)

(cont. Appendix 5.1)

```
!EndTime        - Ending retension time
!EndWave        - Ending wavelength
!RowsNum        - Number of rows
!RtnCode        - Button return code
!SpecData       - HPLC-DAD data
!StartTime      - Starting retension time
!StartWave      - Starting wavelength
!Start Main Program

Name StartConv
On Error ErrorHand
Local FileStr$, PathStr$, i, j
Flag=0
While Flag=0
  GetName
EndWhile
DelReg SpecData
DelReg ChroData
Flag=0
While Flag=0
  TimeRange
  WaveRange
EndWhile

DelReg SpecData
LoadSpectra , StartTime:EndTime, StartWave:EndWave, SpecData
RowsNum=DataRows(SpecData[1])
ColsNum=DataCols(SpecData[1])

!Output HPLC-DAD Information
Local TempStr1$, TempStr2$, TempStr3$, TempStr4$
```

(cont.)

(cont. Appendix 5.1)

```
BeginDialog "InfoDlg", 20, 40, 300, 100, "HPLC-DAD Data Information"
   StaticText 10, 10, 80, 10, "Input Filename:"
   TempStr1$=ObjHdrText$(_Config[1],DADataPath)+ObjHdrText$(_Config[1],DADataFile)+"\"
   StaticText 90, 10, 200, 10, TempStr1$
   StaticText 10, 20, 80, 10, "Output Filename:"
   StaticText 90, 20, 200, 10, DataFile$
   StaticText 10, 30, 80, 10, "Wavelength (nm):"
   TempStr2$=Val$(Data(SpecData[1], 0, 1))+" to "+Val$(Data(SpecData[1], 0, ColsNum))
   StaticText 90, 30, 200, 10, TempStr2$
   StaticText 10, 40, 80, 10, "Retention Time (min):"
   TempStr3$=Val$(Data(SpecData[1], -1, 1))+" to "+Val$(Data(SpecData[1], -1, RowsNum))
   StaticText 90, 40, 200, 10, TempStr3$
   StaticText 10, 50, 80, 10, "Data Dimension"
   TempStr4$=Val$(RowsNum)+" (C) x "+Val$(ColsNum)+ " (S)"
   StaticText 90, 50, 200, 10, TempStr4$
   OkButton 75, 70, 50, 15
   CancelButton 175, 70, 50, 15
EndDialog
RtnCode=ShowDialog("InfoDlg")
RemoveDialog("InfoDlg")
If RtnCode=0
   ErrorHand
EndIf
Print "Conversion in process, please wait ..."

!Output Sample Information
Print "Output general information file"
Open InfoFile$ For Output As #10
Print #10, "Sample Information"
Print #10, "=================="
Print #10, " "
```

(cont.)

(cont. Appendix 5.1)

```
TempStr$=ObjHdrText$(_Config[1],DADataPath)+ObjHdrText$(_Config[1],DADataFile)+"\"
Print #10, "Original Data Filename:    ", TempStr$
Print #10, "Operator                   ", ObjHdrText$(ChromRes[1], AcqOp)
Print #10, "Sample Name:               ", ObjHdrText$(ChromRes[1], SampleName)
Print #10, "Instrument Name:           ", ObjHdrText$(ChromRes[1], AcqInstName)
Print #10, "Injection Data & Time:     ", ObjHdrText$(ChromRes[1], InjDateTime)
Print #10, "Version of HPChem Station: ", ObjHdrText$(ChromRes[1], Version)
Print #10, " "
Print #10, "Wavelength Range (nm):     ", Data(SpecData[1], 0, 1), "to", Data(SpecData[1], 0, ColsNum)
Print #10, "Retention Time (min):        ", Data(SpecData[1], -1, 1), "to", Data(SpecData[1], -1,
RowsNum)
Print #10, "Data Dimension:            ", RowsNum, "x", ColsNum
Close #10

!Output Retention Time Data
Print "Output retention time information file, please wait ..."
Open TimeFile$ For Output As #10
For i=1 to RowsNum
    OutVal=Data(SpecData[1], -1, i)
    Print #10, OutVal
Next i
Close #10

!Output Wavelength Data
Print "Output wavelength information file, please wait ..."
Open WaveFile$ For Output As #10
For i=1 to ColsNum
    OutVal=Data(SpecData[1], 0, i)
    Print #10, OutVal
Next i
Close #10
```

(cont.)

(cont. Appendix 5.1)

```
!Output 3D Data to an ASCII File
Print "Output HPLC-DAD raw data file, please wait ..."
Open DataFile$ For Output As #10
For i=1 to RowsNum
    OutStr$=""
    For j=1 to ColsNum
        Print "No. of Spectra processed = ", i, "/", RowsNum , "(", trunc(i/RowsNum*100), "%)", "Point
=", j
        OutStr$=OutStr$+Val$(Data(SpecData[1], i, j))+" "
    Next j
    Print #10, OutStr$
Next i
Close #10
Print "Convertion complete."
EndMacro

!Get Output File Name
Name GetName
BeginDialog "OutputName", 20, 40, 300, 80, "Output Filename for HPLC-DAD"
   PathStr$=ObjHdrText$(_Config[1],DADataPath)+ObjHdrText$(_Config[1],DADataFile)+"\"
   FileStr$="SPECHRO"
   EditBox 60, 10, 200, 10, PathStr$, "SLAS", 14
   EditBox 60, 30, 200, 10, FileStr$, "SLAS", 14
   OKButton 10, 60, 50, 15
   CancelButton 70, 60, 50, 15
   PushButton 130, 60, 50, 15, "Default", 3
   StaticText 10, 10, 50, 10, "Path: "
   StaticText 10, 30, 50, 10, "Filename: "
   StaticText 10, 45, 100, 10, "No file extension is required."
EndDialog
RtnCode=ShowDialog("OutputName")
```

(cont.)

(cont. Appendix 5.1)

```
While RtnCode=3
   PathStr$=ObjHdrText$(_Config[1],DADataPath)+ObjHdrText$(_Config[1],DADataFile)+"\"
   FileStr$="SPECHRO"
   RtnCode=ShowDialog("OutputName")
EndWhile
RemoveDialog("OutputName")
If RtnCode=1
   If PathStr$[Len(PathStr$)]<>"\"
      PathStr$=PathStr$+"\"
   EndIf
Else
   ErrorHand
EndIf
If FileStat(Mode, PathStr$[1:Len(PathStr$)-1])=-1
   BeginDialog "ErrorDlg", 20, 40, 130, 55, "Error"
      OKButton 10, 30, 50, 15, "Yes"
      CancelButton 70, 30, 50, 15, "No"
      StaticText 10, 10, 100, 10, "Directory not found. Creat it?"
   EndDialog
   RtnCode=ShowDialog("ErrorDlg")
   RemoveDialog("ErrorDlg")
   If RtnCode=1
      Mkdir PathStr$[1:Len(PathStr$)-1]
   Else
      ErrorHand
   EndIf
EndIf
DataFile$=PathStr$+FileStr$+".DAT"
If FileStat(Mode, DataFile$)=1
   BeginDialog "FileDlg", 20, 40, 130, 55, "File"
      OKButton 10, 30, 50, 15, "Yes"
```

(cont.)

(cont. Appendix 5.1)

```
      CancelButton 70, 30, 50, 15, "No"
      StaticText 10, 10, 100, 10, "File is exist. Overwrite it?"
   EndDialog
   RtnCode=ShowDialog("FileDlg")
   RemoveDialog("FileDlg")
   If RtnCode=1
      Flag=1
   Else
      Flag=0
   EndIf
Else
   Flag=1
EndIf
InfoFile$=PathStr$+FileStr$+".INF"
TimeFile$=PathStr$+FileStr$+".TIM"
WaveFile$=PathStr$+FileStr$+".WAL"
Return
EndMacro

!Get Time Range
Name TimeRange
print "Please wait (time information extraction)..."

!Modify the following line if the file is too large.
!eg LoadChrom ,10:20,,, ChroData      to load data from 10 to 20 minutes.
LoadChrom ,,,, ChroData


RowsNum=DataRows(ChroData[1])
ColsNum=DataCols(ChroData[1])
StartTime=Data(ChroData[1],0,1)
```

(cont.)

(cont. Appendix 5.1)

```
EndTime=Data(ChroData[1],0,ColsNum)
StartValC$=Val$(StartTime)
EndValC$=Val$(EndTime)

BeginDialog "TimeRange", 20 ,40, 180, 100, "Select HPLC-DAD Data Time Range"
   StaticText 25, 10, 150, 10, "Select Time Range for Data Conversion"
   StaticText 25, 25, 80, 10, "Starting Time (min.): "
   StaticText 25, 45, 80, 10, "Ending Time (min.): "
   EditBox 95, 25, 50, 10, StartValC$, "SLAS", 14
   EditBox 95, 45, 50, 10, EndValC$, "SLAS", 14
   OkButton 25, 70, 50, 15
   CancelButton 100, 70, 50, 15
EndDialog
RtnCode=ShowDialog("TimeRange")
RemoveDialog("TimeRange")
If RtnCode=0
   ErrorHand
EndIf
If Val(StartValC$)>=Val(EndValC$)
   Flag=0
EndIf
If Val(StartValC$)<StartTime
   Flag=0
Else
   Flag=1
EndIf
If Val(EndValC$)>EndTime
   Flag=0
Else
   Flag=1
EndIf
```

(cont.)

(cont. Appendix 5.1)

```
If Flag=1
   StartTime=Val(StartValC$)
   EndTime=Val(EndValC$)
EndIf
Endmacro

!Get Wavelength Range
Name WaveRange
print "Please wait (wavelength information extraction)..."
LoadSpectra ,StartTime:EndTime,, SpecData
RowsNum=DataRows(SpecData[1])
ColsNum=DataCols(SpecData[1])
StartWave=Data(SpecData[1],0,1)
EndWave=Data(SpecData[1],0,ColsNum)
StartValS$=Val$(StartWave)
EndValS$=Val$(EndWave)
BeginDialog "WaveRange", 20 ,40, 200, 100, "Select HPLC-DAD Data Wavelength Range"
   StaticText 25, 10, 160, 10, "Select Wavelength Range for Data Conversion"
   StaticText 25, 25, 120, 10, "Starting Wavelength (nm): "
   StaticText 25, 45, 120, 10, "Ending Wavelength (nm): "
   EditBox 130, 25, 50, 10, StartValS$, "SLAS", 14
   EditBox 130, 45, 50, 10, EndValS$, "SLAS", 14
   OkButton 25, 70, 50, 15
   CancelButton 125, 70, 50, 15
EndDialog
RtnCode=ShowDialog("WaveRange")
RemoveDialog("WaveRange")
If RtnCode=0
   ErrorHand
EndIf
If Val(StartValS$)>=Val(EndValS$)
```

(cont.)

(cont. Appendix 5.1)

```
   Flag=0
EndIf
If Val(StartValS$)<StartWave
   Flag=0
Else
   Flag=2
EndIf
If Val(EndValS$)>EndWave
   Flag=0
Else
   Flag=2
EndIf
If Flag=2
   StartWave=Val(StartValS$)
   EndWave=Val(EndValS$)
EndIf
Endmacro

!Add Menu
Name DataMenu
MenuAdd "&Data", "Export", StartConv
EndMacro

!Error Handling
Name ErrorHand
!Print "Error: Data convertion terminated."
Stop
EndMacro
```

Appendix 6.1 Program description of the Fast Wavelet Transform Heuristic Evolving Latent Projections (FWT-HELP) Toolbox Version 1.0

Program Name	Functions
Main Program	
WTHELP.M	To initiate various program of the toolbox.
Supporting Program	
BGREMOVE.M	To remove background in the data set.
CALEIGEN.M	To calculate the overall eigenvalues and determine the minimium number of factors in the data.
ETA.M	To perform eigenstructure tracking analysis.
HELPANA1.M	To analysis the number of components from the eigenvalue plot by the user.
HELPANA2.M	To determine the zero concentration and selective region from the eigenvalue plot by the user.
HELPANA3.M	To select suitable resolve chromatograms and spectra for data subtraction in component stripping step.
HELPANA4.M	To display and save overall result.
LATPLOT.M	To plot the latent projective plot.
RANGECAL.M	To calculate the actual time and wavelength range.
ROTMAT.M	To calculate the rotational matrix r in HELP analysis.
SCALE.M	To calculate the scaling factor for the resolved chromatogram and spectrum.
SELRANGE.M	To select time and wavelength ranges of data for HELP analysis.
SELWAVE.M	To select parameter for different wavelet functions.
WTANAL.M	To perform wavelet transform on HPLC-DAD data.

(cont.)

(cont. Appendix 6.1)

HPLC-DAD Data Simulation Program

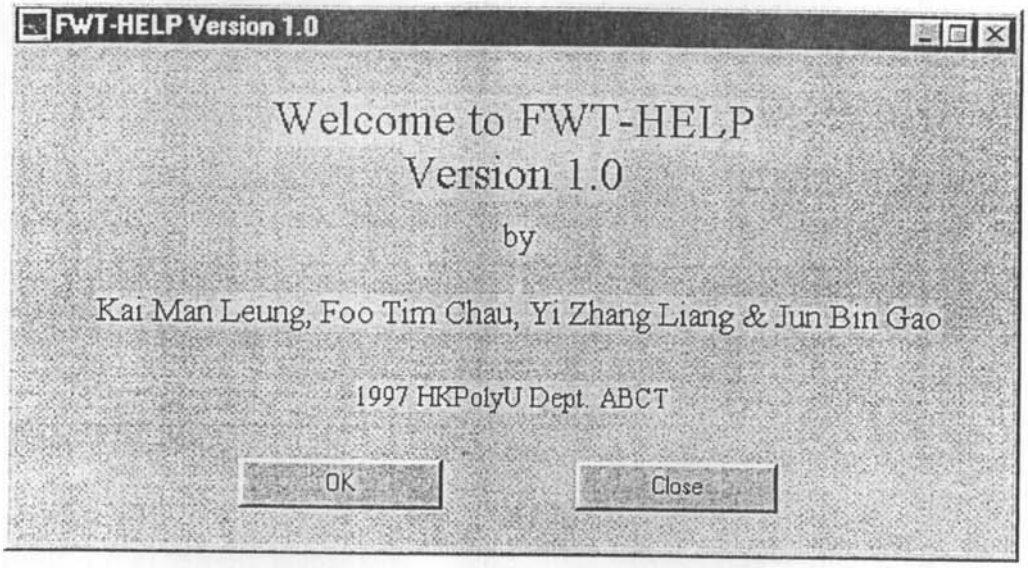
DATASIM.M	To simulate a HPLC-DAD data.
GENEMG.M	To generate a data set by an exponentially modified gaussian (EMG) function.
GENGAUS.M	To generate a data set by a Gaussian function.
GENLOREN.M	To generate a data set by a Lorentzian function.

Wavelet Transform Related Program

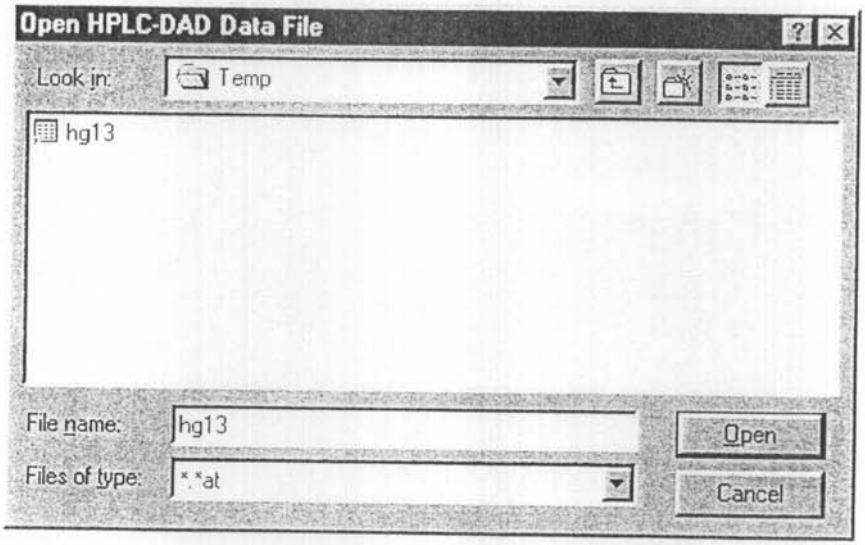
DAUB.M	To generate Daubechies wavelet filters.
WT1D.M	To perform fast wavelet transform calculation.

Appendix 6.2 Screen captures of the of the Fast Wavelet Transform Heuristic Evolving Latent Projections (FWT-HELP) Toolbox Version 1.0

Screen 6.2.1 Welcome screen of the FWT-HELP Toolbox.



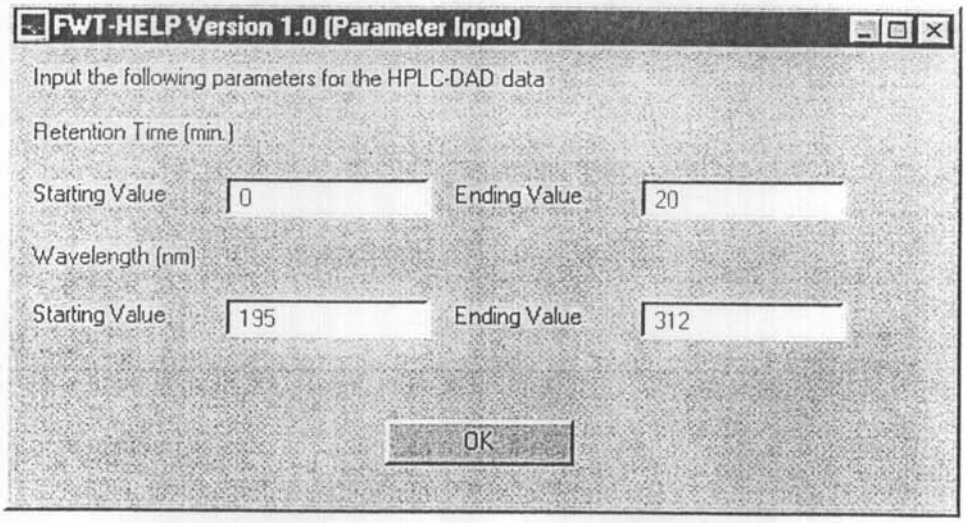
Screen 6.2.2 Loading of HPLC-DAD data file.



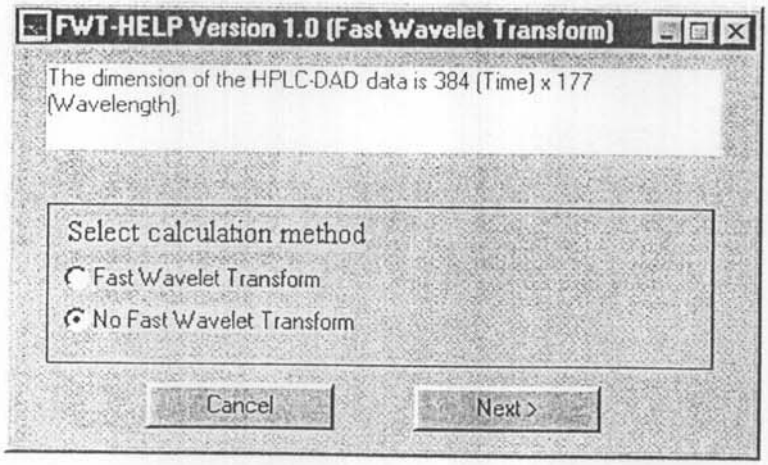
(cont.)

(cont. Appendix 6.2)

Screen 6.2.3 Input of actual retention time and wavelength of the HPLC-DAD data.

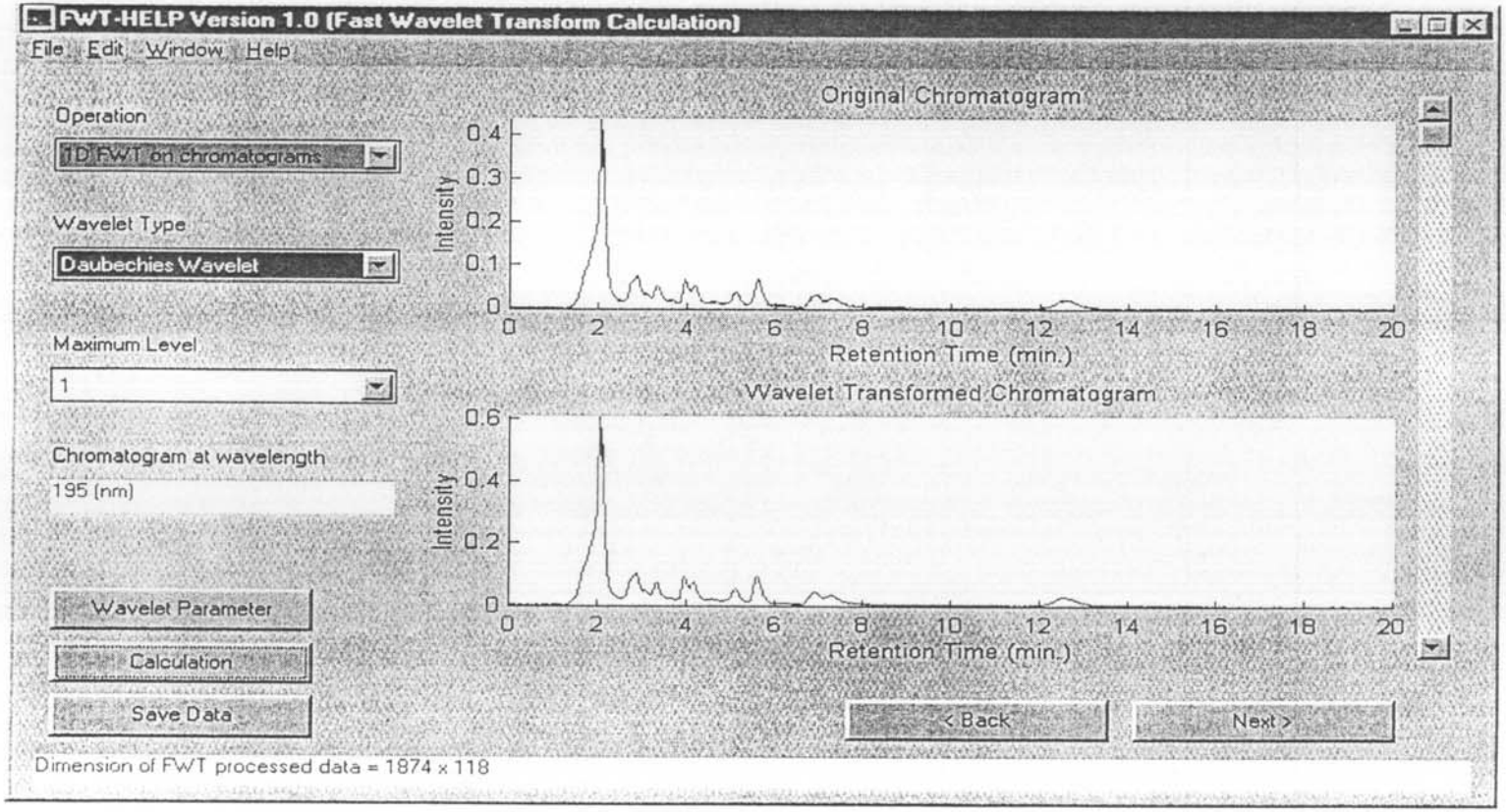


Screen 6.2.4 Selection of preprocessing method.



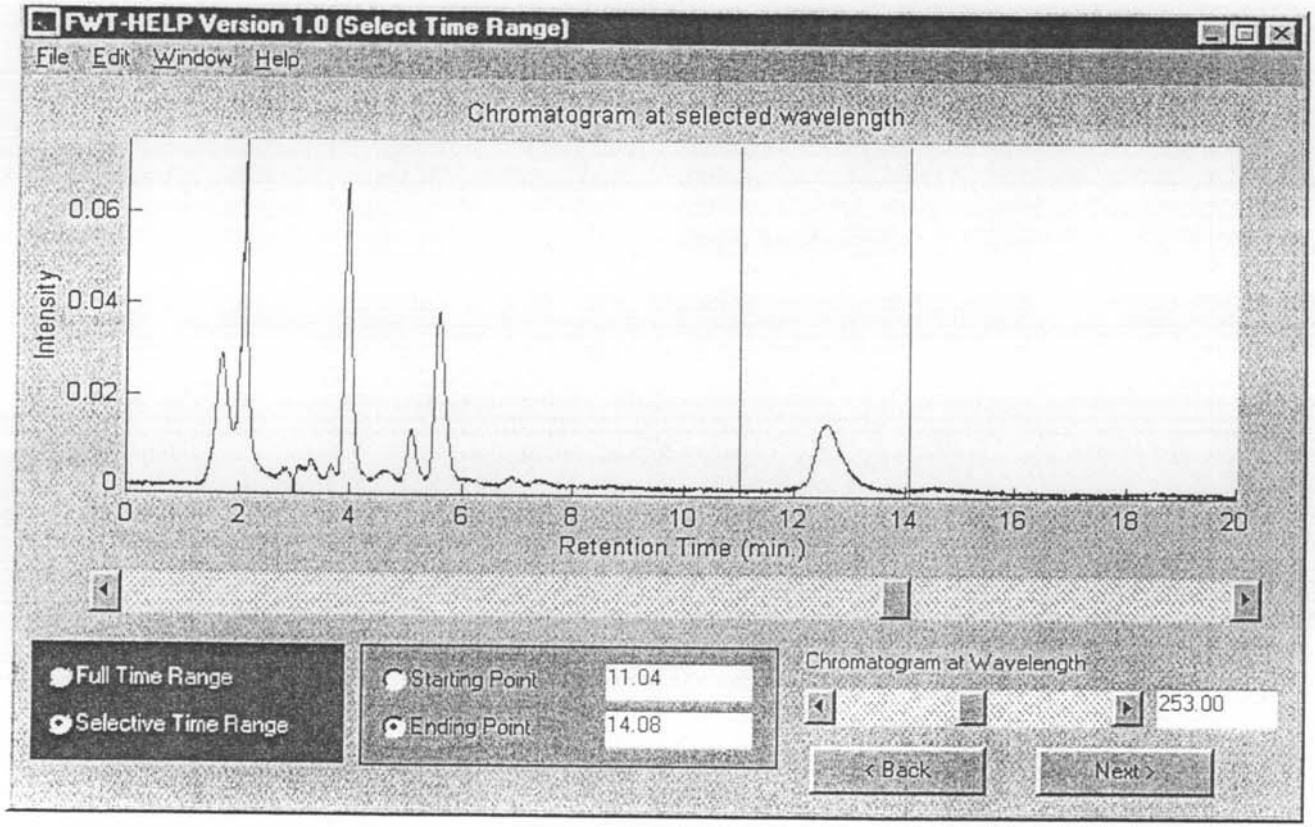
(cont.)

Screen 6.2.5 Program for FWT computation.



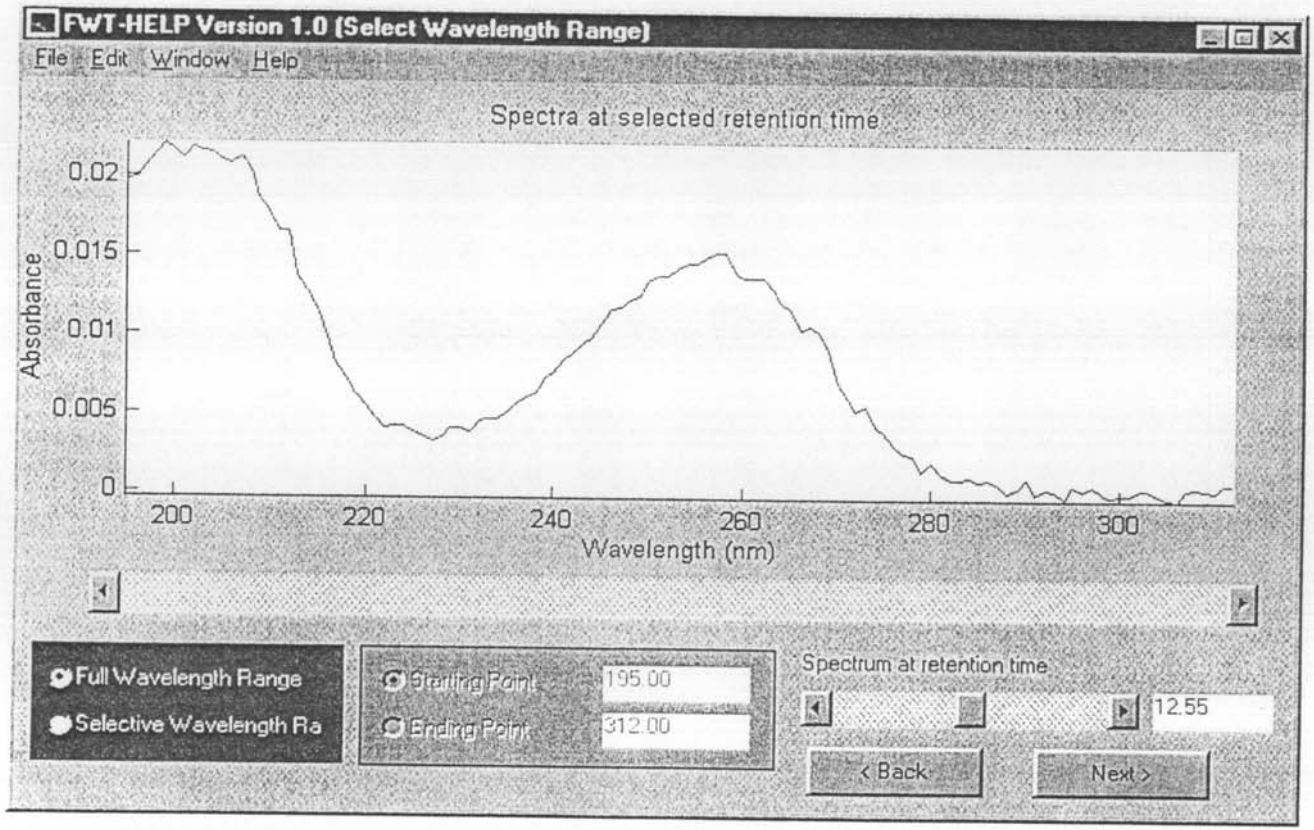
(cont. Appendix 6.2)

Screen 6.2.6 Program for selection of chromatographic peak for HELP analysis.



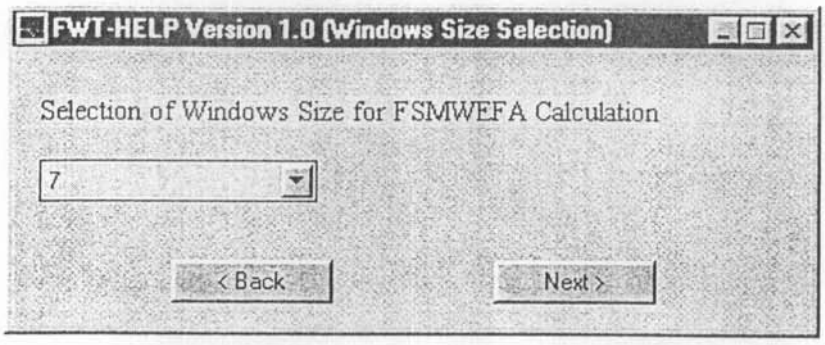
(cont.)

Screen 6.2.7 Program for selection of wavelength range for HELP analysis.



(cont. Appendix 6.2)

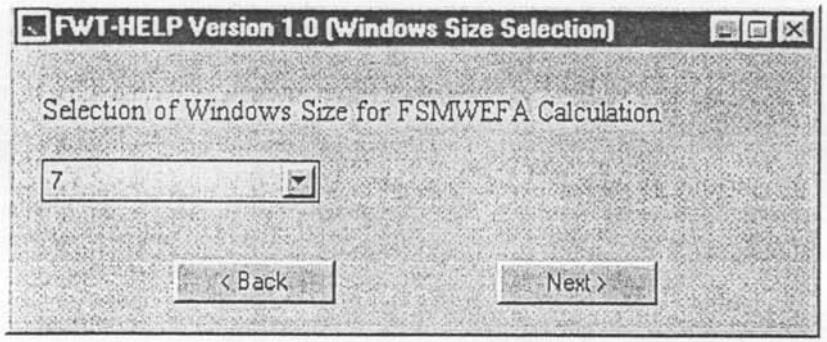
Screen 6.2.8 Program for selection of window size for FSMWEFA calculation.



(cont.)

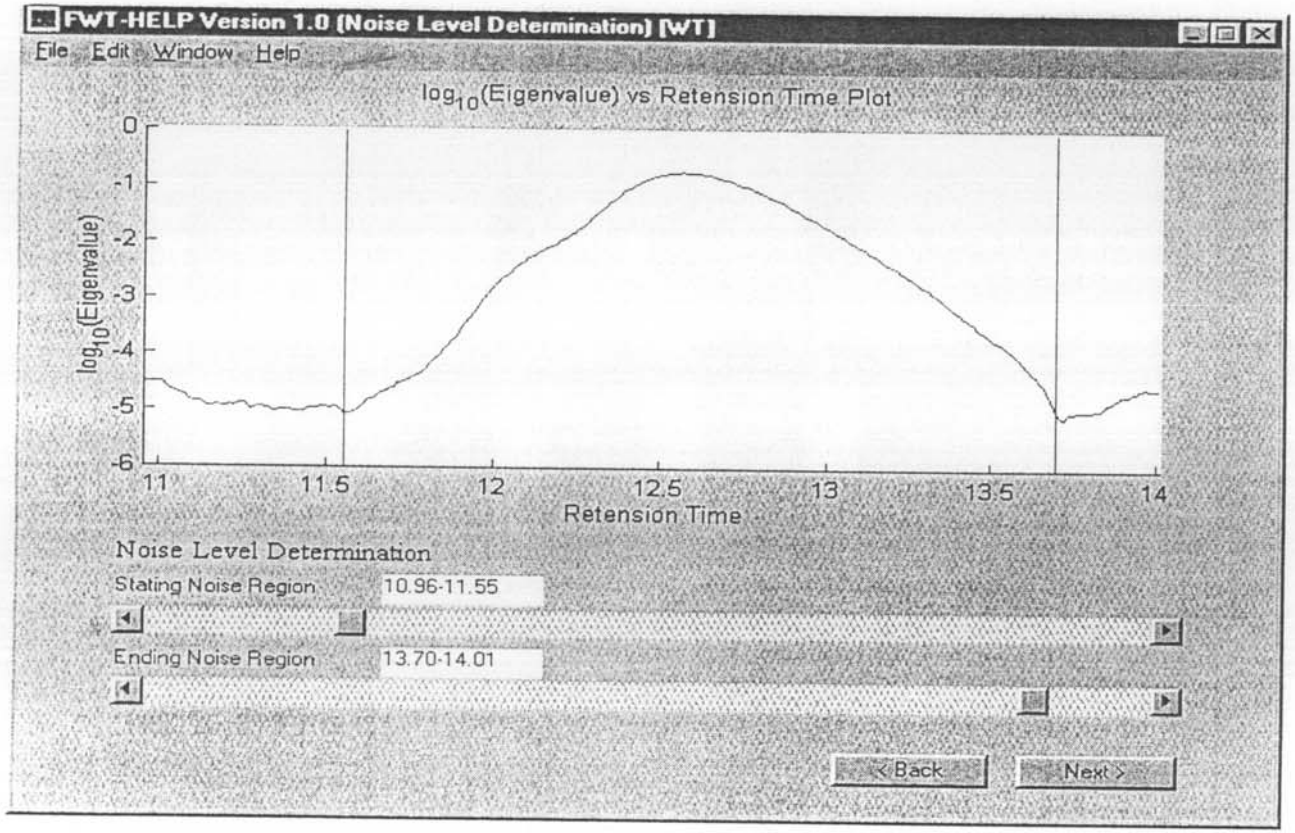
(cont. Appendix 6.2)

Screen 6.2.8 Program for selection of window size for FSMWEFA calculation.



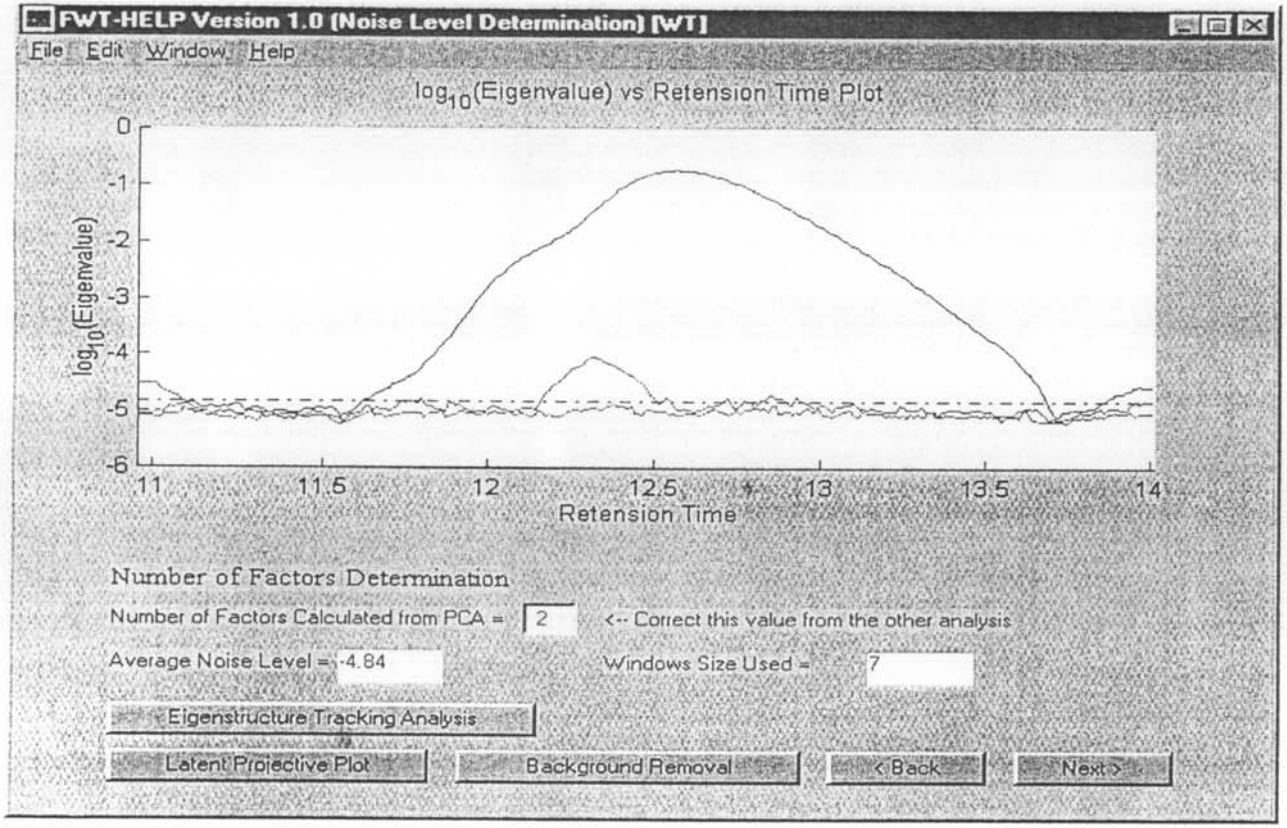
(cont.)

Screen 6.2.9 Program for selection of noise level in the eigenvalue plot.



(cont. Appendix 6.2)

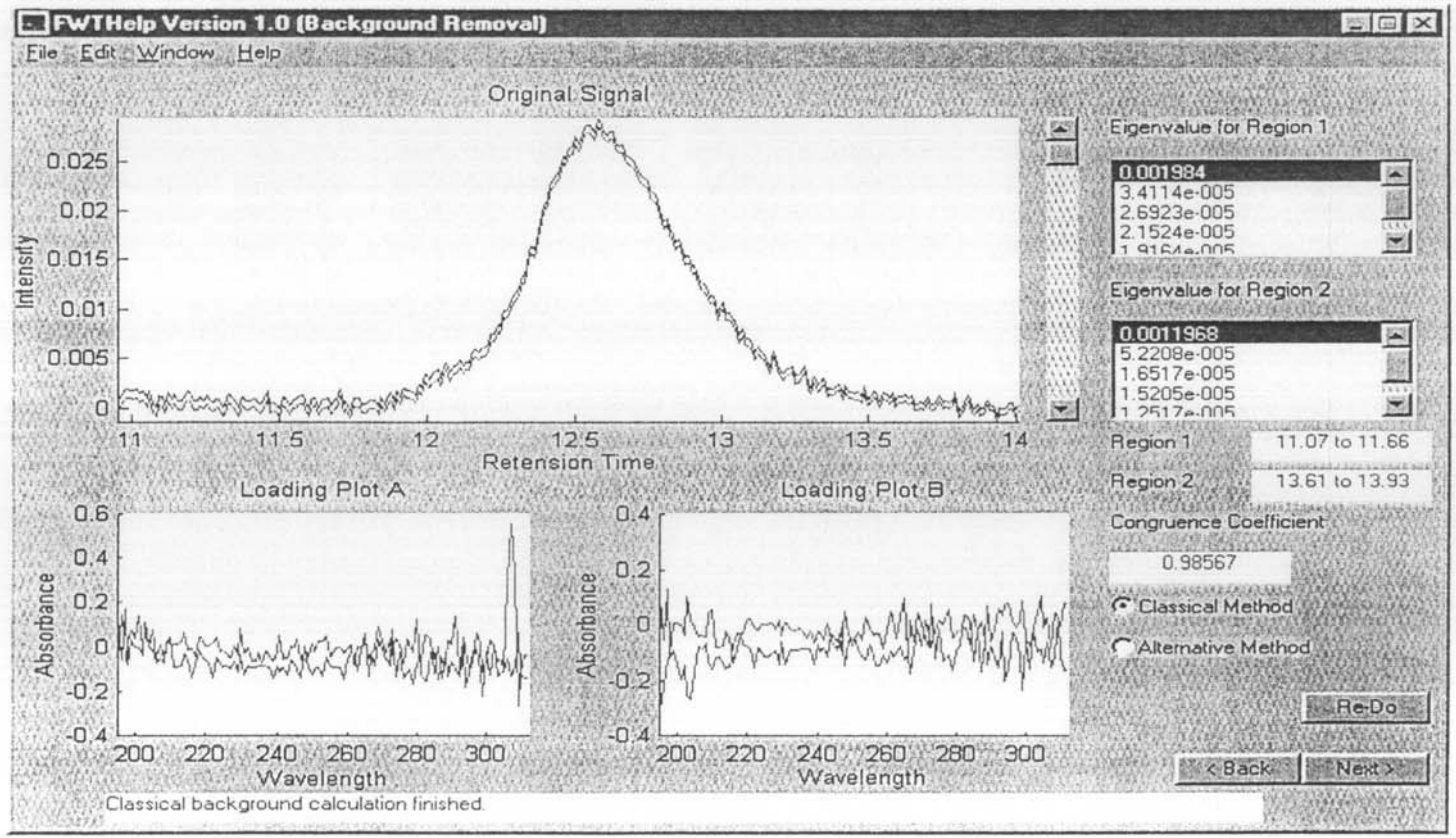
Screen 6.2.10 Program for calculation of eigenvalue plot.



(cont.)

(cont. Appendix 6.2)

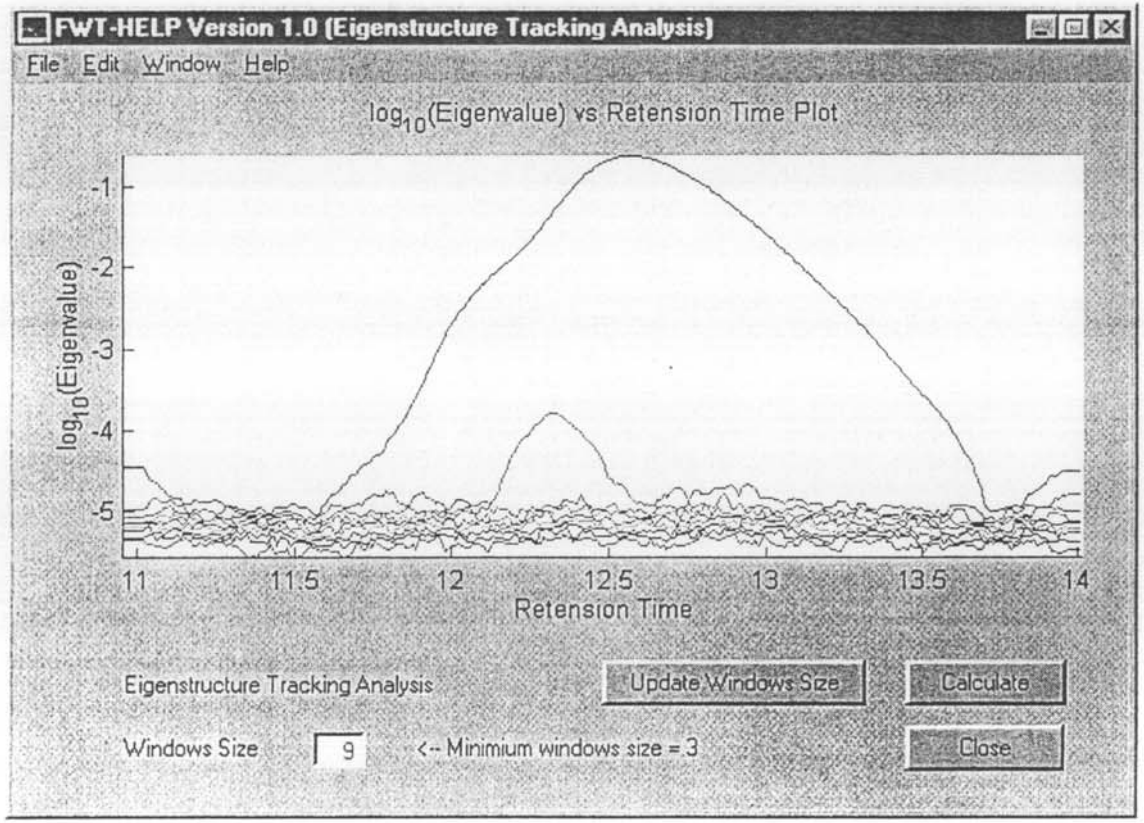
Screen 6.2.11 Program for background removal.



(cont.)

(cont. Appendix 6.2)

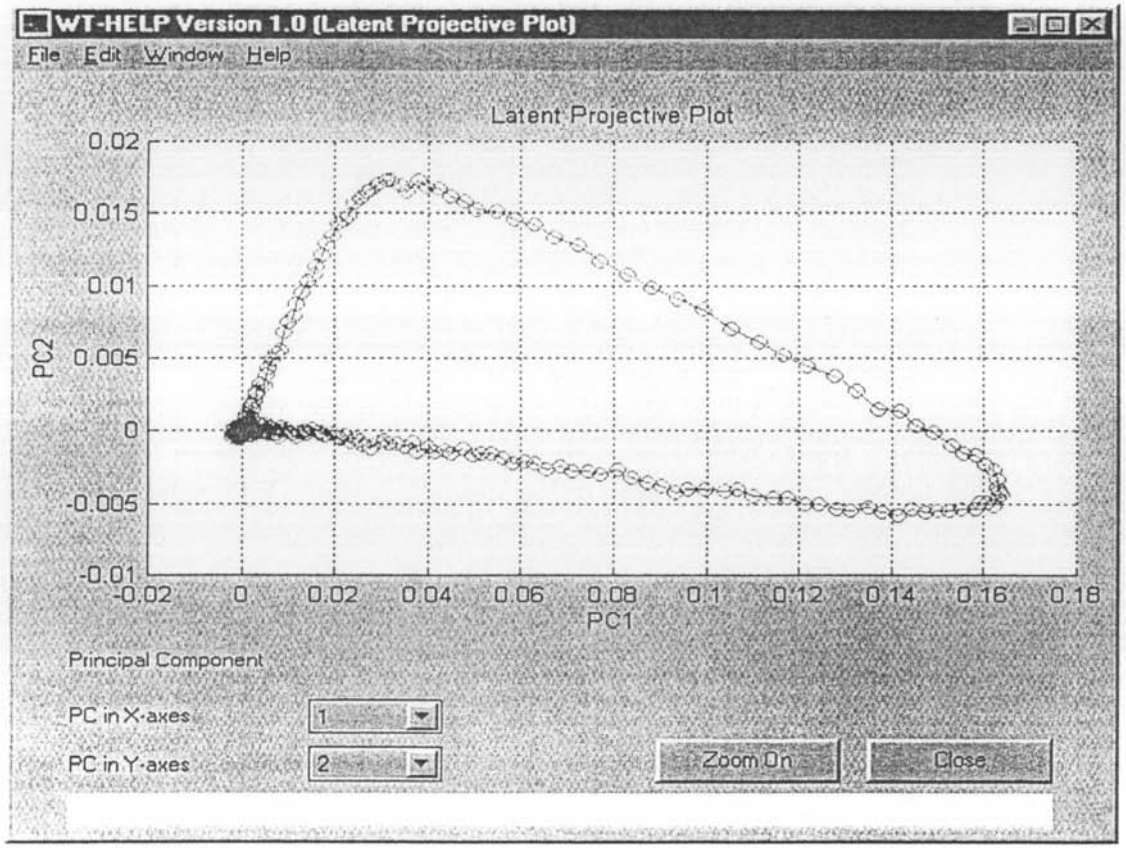
Screen 6.2.12 Program for eigenvalue tracking analysis (for selection of new window size in FSMWEFA calculation).



(cont.)

(cont. Appendix 6.2)

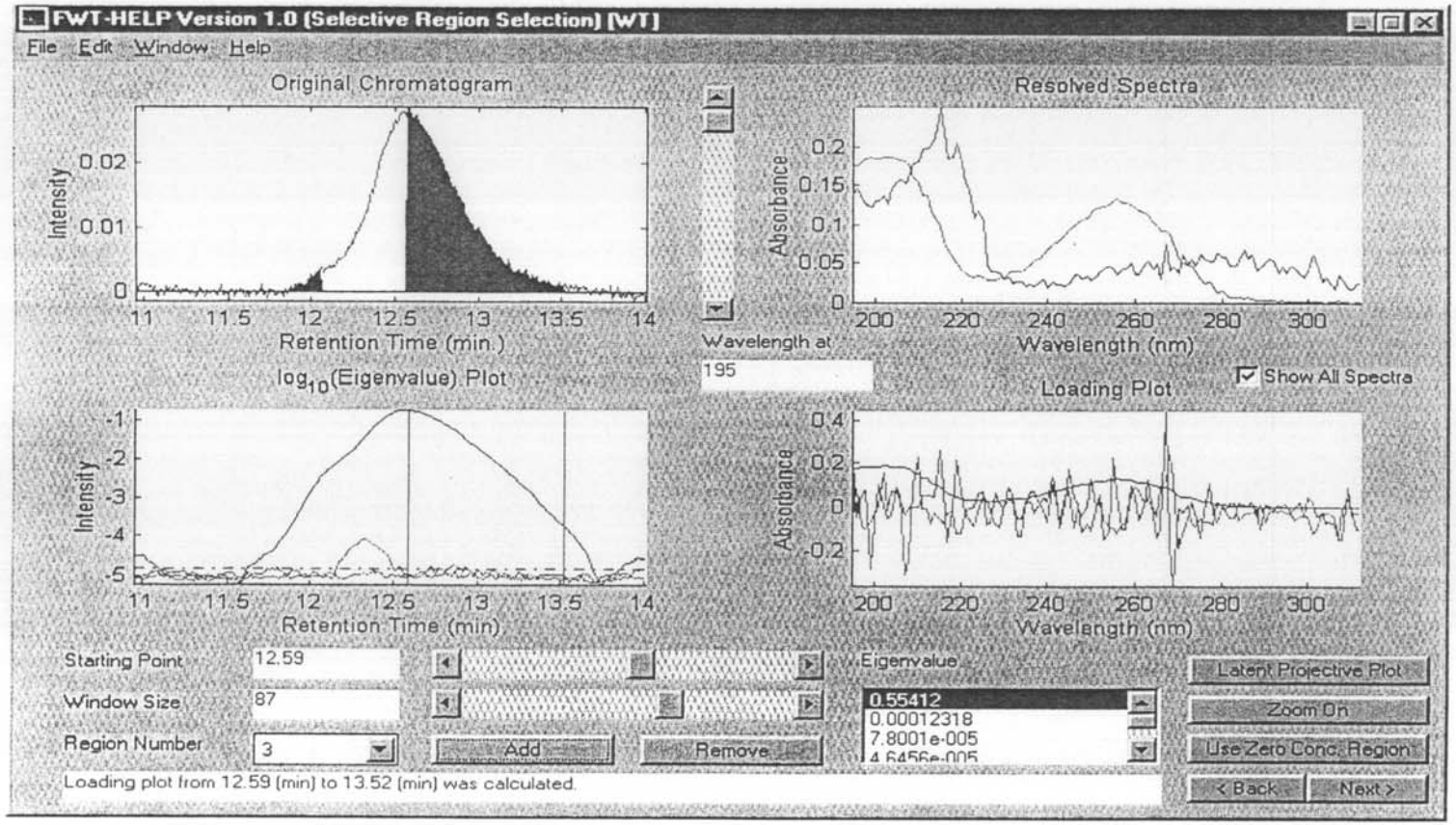
Screen 6.2.13 Program for generation of the latent projective graph.



(cont.)

(cont. Appendix 6.2)

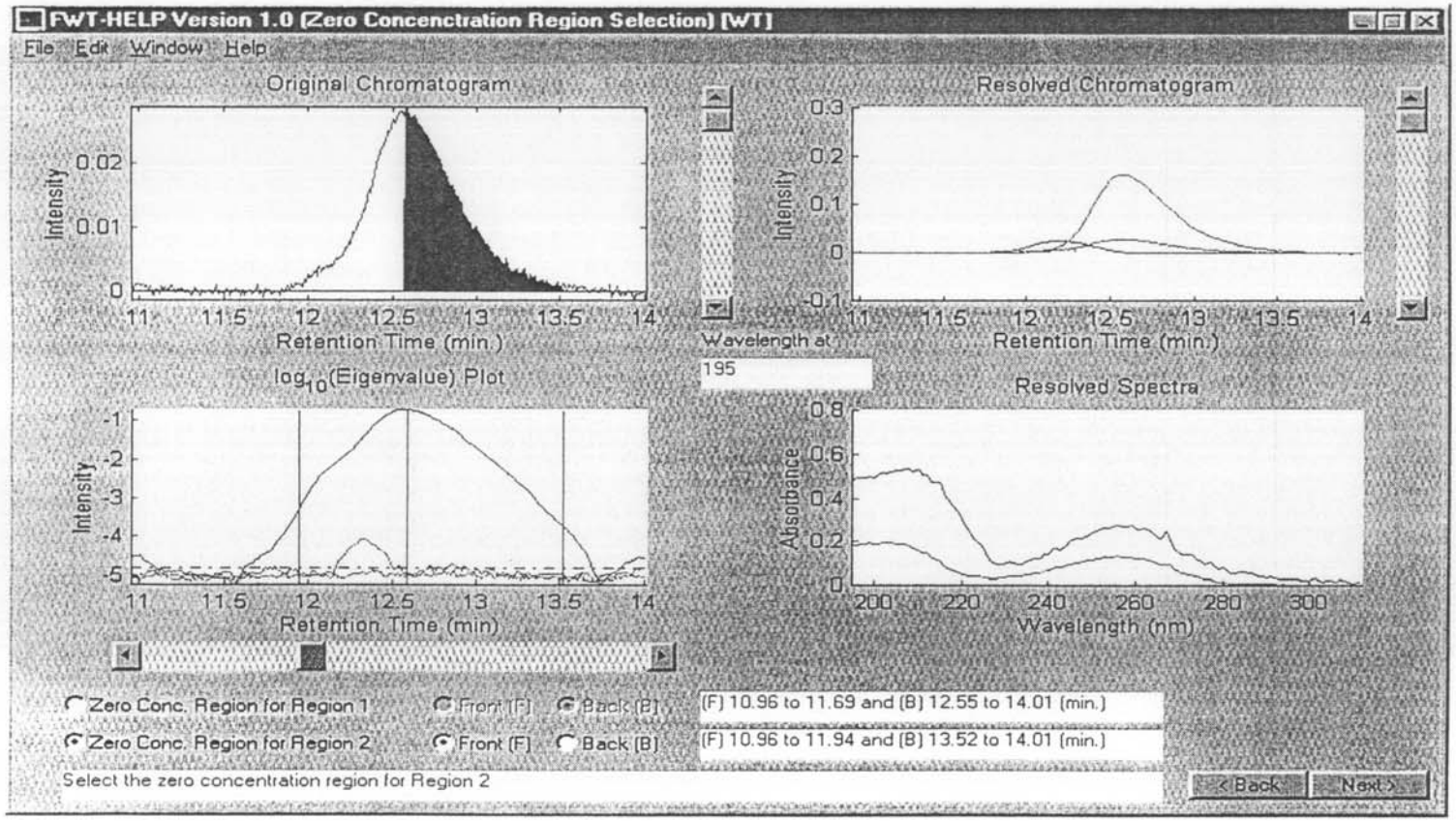
Screen 6.2.14 Program for choosing selective regions.



(cont.)

(cont. Appendix 6.2)

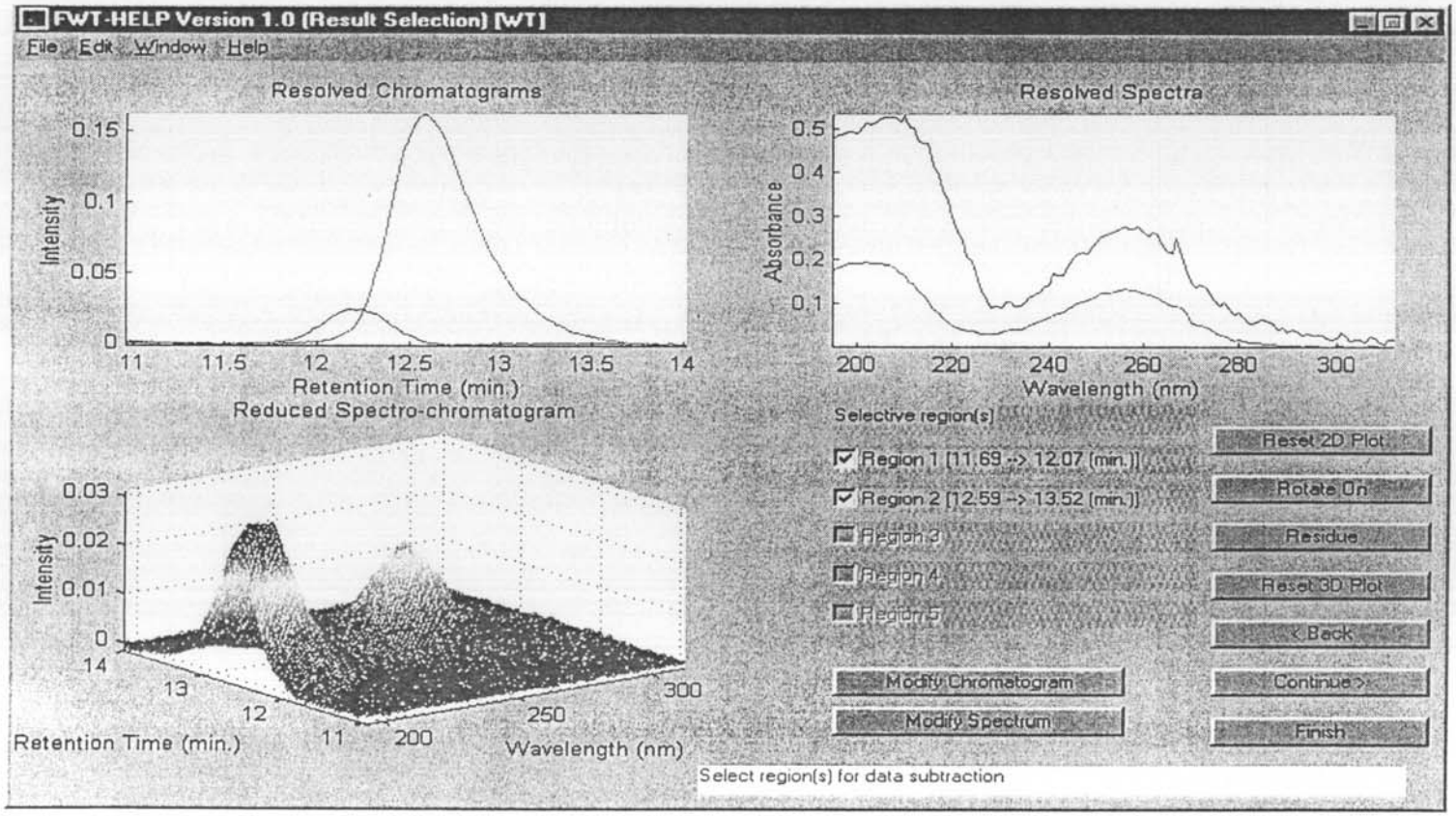
Screen 6.2.15 Program for selection of the zero concentration regions.



(cont.)

(cont. Appendix 6.2)

Screen 6.2.16 Program for removal of unwanted part from the resolved chromatograms and spectra.



(cont.)

PUBLICATIONS AND PRESENTATIONS

Papers

[1] Leung, K.M, Chau, F.T., Kwok, P.H. and Lau, W.T.
"CHEMISTOOLS for Windows: A Computer Program for Chemical Information System"
Computers & Chemistry, Vol. 21, pp.161-166 (1997)

[2] Leung, K.M. and Chau, F.T.
"A Review on Signal Compression of Spectroscopic Data in Analytical Chemistry"
Acta Physico-Chimica Sinica (物理化學學報), Vol. 13, pp.857-864 (1997)

[3] Wang, D.C., Chau, F.T., Lee, E.P.F., Leung, A.K.M. and Dyke, J.M.
"The $\tilde{X}^2\Pi$ and $\tilde{A}^2\Sigma$ States of FCN^+ and $ClCN^+$: *ab Initio* Calculations and Simulation of the HeI Photoelectron Spectra of FCN and ClCN"
Journal of Molecular Physics, Vol. 93, pp.995-1005 (1998)

[4] Leung, A.K.M., Chau, F.T., Gao, J.B. and Shih, T.M.
"Application of Wavelet Transform in Infrared Spectrometry: Spectral Compression and Library Search"
Chemometrics and Intelligent Laboratory Systems, in press (1998)

[5] Leung, A.K.M., Chau, F.T. and Gao, J.B.
"A Review on Application of Wavelet Transform Techniques in Chemical Analysis: 1989-1997"
Chemometrics and Intelligent Laboratory Systems, in press (1998)

[6] Leung, A.K.M., Chau, F.T. and Gao, J.B.
"Wavelet Transform: A Novel Method for Derivative Calculation in Analytical Chemistry"
Analytical Chemistry, in press (1998)

[7] Gong, F., Peng, Y.G., Cui, H., Liang, Y.Z., Leung, A.K.M. and Chau, F.T.
"Hyphenated Chromatography Applied to Multip-component Determination of *Cordyceps Sinensis*"
Acta Pharmaceutica Sinica (藥學學報), accepted for publication (1998) [In Chinese with English Abstract]

[8] Chan, A.C.M., Lo, S.C.L., Chau, F.T. and Leung, A.K.M.
"Application of Image Analysis Techniques to Crossed Two-dimensional Immunoelectrophoresis Study"
Computer and Chemistry, submitted for publication (1998)

[9] Chau, F.T., Leung, A.K.M., Liang, Y.Z. and Chan, B.T.P.
"Sensitivity Enhancement in Factor Analysis of Two-way Hyphenated Data with Fast Wavelet Transform"
Analytical Chemistry, submitted for publication (1998)

[10] Gao, J.B., Chau, F.T., Leung, A.K.M. and Shih, T.M.
"Spectral Noise Removal by Wavelet Low-pass Filters"
Analytica Chimica Acta, submitted for publication (1998)

[11] Gong, F., Peng, Y.G., Cui, H., Liang, Y.Z., Leung, A.K.M. and Chau, F.T.
"HELP Applied to Traditional Chinese Medicine Analysis - Determination of Components of *Cordyceps Sinensis*"
Chemistry Journal of Chinese Universities (高等學校化學學報), submitted for publication (1998) [In Chinese with English Abstract]

[12] Szeto Y.B., Chau F.T. and Leung A.K.M.
"DacStat: A Microsoft® Excel® Visual Basic Program for Analytical Chemistry and Food Science"
Journal of Chemical Education, submitted for publication (1998)

[13] Gao, J.B., Leung, A.K.M., Chau, F.T. and Shih, T.M.
"Compression of Infrared Spectrum using Wavelet Packet-Optimal Bit Allocation-Hauffman Coding Method"
Preparing for publication (1998)

[14] Lau, C.K., Leung, A.K.M., Lo, S.C.L., Chau, F.T. and Kwok, F.
"A model for the ATP-binding Domain of Pyridoxal Kinase with Experimentally Identified Domain Boundary"
Preparing for publication (1998)

[15] Leung, A.K.M., Chau, F.T., Gong, F., Liang, Y.Z. and Gao, J.B.
"Analysis of the Water Soluble Constituents of *Cordyceps Sinensis* with Wavelet Transform Heuristic Evolving Latent Projections"
Preparing for publication (1998)

[16] Liang, Y.Z., Leung, A.K.M. and Chau, F.T.
"A Roughness Penalty Approach and Its Application for Noisy Hyphenated Chromatographic Two-way Data"
Preparing for publication (1998)

Conference and Symposium Presentations

[1] Leung, K.M. and Chau, F.T.
"Application of Chemometrics Techniques in Chemical and Biochemical Studies"
A poster presented at The Third Symposium on Chemical Postgraduate Research in Hong Kong, Hong Kong, 24 February, 1996, pp.P-56 (1996)

[2] Gao, J.B., Chau, F.T., Shih, T.M. and Leung, K.M.
"Compression of Infrared Spectral Data using Wavelet Transform Techniques"
A poster presented at The First International Symposium on Inorganic Chemistry for Chinese Scientists (ISICS 96), Hong Kong, 14-19 July, 1996, pp.88 (1996)

[3] Leung, K.M., Chau, F.T. and Gao, J.B.
"Wavelet Application in Spectral Compression and Library Search in Analytical Chemistry"
A poster presented at The 213th National Meeting of the American Chemical Society (ACS), San Francisco, CA, US, 13-17 April, 1997, pp.183-COMP (1997)

[4] Chan, B.T.P., Leung, A.K.M., Chau, F.T., Chan, W.W.L., Wu, J.Y., Kwok, I.M.Y. and But, P.P.H.
"Application of Image Analysis and Chemometrics Techniques to Chemical Analysis of Ginseng"
A poster presented at The 215th National Meeting of the American Chemical Society (ACS), Dallas, TX, US, 29 March - 2 April, 1998, pp.73-ANAL (1997)

[5] Leung, K.M., Chau, F.T., Liang, Y.Z. and Gao, J.B.
"A Modified Chemometrics Method for Two-way Data Analysis by Using Fast Wavelet Transform"
An oral presentation presented at The Fifth Symposium on Chemistry Postgraduate Research in Hong Kong in conjunction with The First Annual Congress of The Hong Kong Chemical Society, Hong Kong, 25 April, 1998, pp.Oral-4 (1998)

[6] Chau, F.T., Leung, K.M., Liang, Y.Z., Gao, J.B. and Chan, T.P.
"Sensitivity Enhancement in Factor Analysis of Two-way Hyphenated Data using Fast Wavelet Transform"
A poster presented at The 216th National Meeting of American Chemical Society (ACS), Boston, MA, US, 23-27 August, 1998, pp. 30-ANAL (1998)

[7] Liang, Y.Z., Leung, K.M. and Chau, F.T.
"A Roughness Penalty Approach and Its Application for Noisy Hyphenated Chromatographic Two-way Data"
Will be presented at The Third International Symposium of Worldwide Chinese Scholar on Analytical Chemistry (ISWCSAC98), Hong Kong, 16-19 December, 1998.

Books/Chapters in Books

[1] Chau, F.T. and Leung, A.K.M.
"Application of Wavelet Transform in Spectroscopic Studies" in Wavelets in Analytical Chemistry by Walczak B. (Ed.)
Elsevier, to be published (2000)

[2] Chau, F.T. and Leung, A.K.M.
"Application of Wavelet Transform in Processing Chromatographic Data" in Wavelets in Analytical Chemistry by Walczak B. (Ed.)
Elsevier, to be published (2000)

[3] Chau, F.T. and Leung, A.K.M.
"Application of Wavelet Transform in Electrochemical Studies" in Wavelets in Analytical Chemistry by Walczak B. (Ed.)
Elsevier, to be published (2000)

Internal Publications

[1] Leung, A.K.M.
SGI Indigo^2 Workstation Root Administration Guide (1997)

[2] Leung, A.K.M.
SGI O_2 Workstation Root Administration Guide (1997)

[3] Leung, A.K.M.
SGI Octane Workstation Root Administration Guide (1998)

[4] Leung, A.K.M.
Webpages for Chemical Physics and Chemometrics Research Group, Department of Applied Biology and Chemical Technology, The Hong Kong Polytechnic University (1997)
URL address: http://fg702-6.abct.polyu.edu.hk

Abstract.1 The Third Symposium on Chemical Postgraduate Research in Hong Kong (1996)

P-56

Application of Chemometrics Techniques in Chemical and Biochemical Studies

K. M. Leung and F. T. Chau

Department of Applied Biology and Chemical Technology

The Hong Kong Polytechnic University, Hung Hom, Hong Kong

Abstract

Chemometrics is a special area that makes use of statistical, mathematical and other logic-based methods in the field of chemistry and in particular in analytical chemistry[1]. By applying chemometric techniques, useful information can be extracted from simpler and less informative instruments as completely as possible. Besides, it can be employed to avoid expensive, difficult and time consuming analyses. In this work, different chemometrics techniques such as principal component analysis, signal compression and multivariate calibration were utilized to process spectroscopic and chromatgraphic data[2-4] for qualitative and quantitative analyses of trace components in some chemical and biological systems.

References

1. Haswell S. J., *Practical Guide to Chemometrics*, Marcel Dekker, Inc., New York, 1992.
2. Malinowski E. R., *Factor Analysis in Chemistry*, 2nd. Ed., John Wiley & Sons, Inc., New York, 1991.
3. Harrington P. B. and Isenhour T. L., *Anal. Chem.*, 1988, 60, 2687-2692.
4. E. V. Thomas, *Anal. Chem.*, 1994, 66, 795A-804A.

Abstract.2 The First International Symposium on Inorganic Chemistry for Chinese Scientists (ISICS 96) (1996)

COMPRESSION OF INFRARED SPECTRAL DATA USING WAVELET TRANSFORM TECHNIQUES

Junbin Gao[a], Foo-tim Chau[b], Tsi-man Shih[c], and Kai-man Leung[b]
[a]Department of Mathematics, Huazhong University of Science and Technology, Wuhan 430072, China
[b]Department of Applied Biology and Chemical Technology, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong
[c]Department of Applied Mathematics, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong

With the rapid expansion of the chemical database, there is a great demand for developing efficient spectral compression techniques. One method often employed involves encoding only peak locations into library data files. This approach allows efficient data compression. Yet much of the spectral intensity information is removed. To incorporate this kind of information while implementing data compression, it requires sophisticated mathematical techniques. In one case, the factor analysis procedure was used to give a significant reduction in infrared (IR) spectral library size [1]. In this way, a five-fold reduction in the file size was achieved. Adams et al. [2] proposed an error-free data compression scheme based on the Huffman shift to code IR spectra for the compression purpose with saving storage up to 70% of the digital memory space. Also, FFT methods were utilized in compressing IR spectra [3 -- 5] with good compression ratios. In this work, we present IR spectral compression algorithms utilizing the newly developed mathematical techniques. Wavelet Transform (WT) and Wavelet Packet Transform (WPT) [6 -- 8] which have been shown to play an important role in signal processing. In our research, three different compression procedures have been developed, i.e., the wavelet-based thresholding algorithm (WT-Th), which has been demonstrated to be better for UV-VIS spectral compression [9], wavelet-based optimal bit allocation with Huffman coder (WT-OBA-H) and wavelet-packet-based optimal bit allocation algorithm (WPT-OBA). In order to compare the performance of the proposed algorithms, we apply these algorithms including the FFT compression algorithm to IR spectrum of Caffeine. The compression ratios obtained are, respectively, 20% for FFT, 38% for WT-Th, 85% for WT-OBA-H and 92% for WPT-OBA with Daubechies wavelet function D_{16} and resolution level J=6 in the wavelet case. In addition, the relationship between the Daubechies wavelet series and the root-mean-square-errors (RMSE) or reduction rates as deduced from algorithms on different IR spectra was detected. We prefer to use WPT-OBA treatment by which a higher compression ratio with lower RMSE compared to the other algorithms can be achieved. In this work, different Daubechies wavelets and the bit rate levels (corresponding to the expected compression ratio) were attempted to study the performance of the WPT-OBA algorithm. It was found that the combination of the bit rate level with a value of 12 and the D_{16} wavelet is good and efficient in compressing IR spectral data with significant reduction in the storage space and diminishing RMSE's.

(cont.)

(cont. Abstract.2)

References
1. G. Hangac, R.C. Wieboldt, R.B. Lam and T.L. Isenhour, Applied Spectrosc., 36 (1982) 40-47.
2 M.J. Adams and I. Black, Anal. Chem. Acta, 189 (1986) 353-363.
3. R.B. Lam, S.J. Foulk and T.L. Isenhour, Anal. Chem. 53 (1981) 1679-1684.
4. L.V. Azarraga, R.R. Williams and J.A. Haseth, Appl. Spectrosc., 35 (1981) 466-469.
5. P.M. Owens and T.L. Isenhour, Anal. Chem., 55 (1983) 1548-1553.
6. S. Mallat, IEEE Trans. on Acoustic Speech and Signal Processing, 37 (1989) 2091-2110.
7. I. Daubechies, "Ten Lectures on Wavelets", SIAM Press, Philadelphia, 1992.
8. R. Coifman and M. Wickerhauser, IEEE Trans. Information Theory, 38 (1992) 713-719.
9. F.T. Chau, T.M. Shih, J.B. Gao and C.K. Chan, Applied Spectrosc., 50 (1996) 399-410.

Abstract.3 The 213th National Meeting of the American Chemical Society (ACS) (1997)

COMP

183.

WAVELET APPLICATION IN SPECTRAL COMPRESSION AND LIBRARY SEARCH IN ANALYTICAL CHEMISTRY Kai-man Leung, Department of Applied Biology & Chemical Technology, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong.

Within the past several years, a new mathematical technique called wavelet transform (WT) has been proposed for signal processing in chemical studies owing to its efficiency, availability of large numbers of basis functions and speed in data treatment. In this work, two different types of wavelet transform were applied to compress a set of Fourier transform infrared (FT-IR) spectra. They are fast wavelet transform (FWT) and its derivative wavelet packet decomposition (WPD). The average compression ratios obtained are 92.7% for FWT and 92.8% for WPD with Daubechies function and resolution level J=4. The wavelet coefficients were used to set up a FT-IR spectral library for spectral library search. By utilizing the wavelet coefficients at the 4th resolution level for spectral library searching, the searching time could be increased up to 30% when compared with the traditional method by using fast Fourier transform.

Abstract.4 The 215th National Meeting of the American Chemical Society (ACS) (1998)

ANYL

073.

APPLICATION OF IMAGE ANALYSIS AND CHEMOMETRICS TECHNIQUES TO CHEMICAL ANALYSIS OF GINSENG.

Benny Tsz-pun Chan[a], Alexander Kai-man Leung[a], Foo-tim Chau[a], Willy Wing-lai Chan[a], Jian-yong Wu[a], Iris Man-yan Kwok[b] and Paul Pui-hay But[b], [a]Department of Applied Biology and Chemical Technology, The Hong Kong Polytechnic University, Hunghom, Kowloon, Hong Kong and [b]Department of Biology and Chinese Medicinal Material Research Centre, The Chinese University of Hong Kong, Shatin, N.T., Hong Kong.

Chinese medicines have about four thousand year of therapeutic history. Ginseng is one of the most valued herbs in Chinese medicines and is widely used all over the world. Therefore, the quality control of ginseng is very important. Nowadays, qualitative and quantitative analyses different types of ginseng are mainly carried out by using thin layer chromatography (TLC) and high-performance liquid chromatography (HPLC). In this investigation, a HPLC instrument coupled with a diode array detector system (HPLC-DAD) and the TLC separation technique were used for studying ginseng. Image analysis and chemometrics techniques were applied to analyze data of ginseng as obtained from TLC and HPLC-DAD studies. Firstly, a cost effective image analysis system and a software package TLCQA were implemented in this work for studying TLC patterns. With the aids of a flatbed scanner and an IBM PC compatible of image capture, a new quantitative method was proposed for improving qualitative and semi-quantitative analyses of thin layer chromatograms of ginseng samples. Besides, the heuristic evolving latent projections method (HELP) coupled with wavelet transform (WT) were also developed to resolve overlapped chromatograms from HPLC-DAD investigation. The software package WTHELP was coded for HELP and WT calculations. Results of investigations on *Panax ginseng* and *P. quinquefolium* by using the proposed methods were compared with those from conventional techniques.

Abstract.5 The Fifth Symposium on Chemistry Postgraduate Research in Hong Kong in conjunction with The First Annual Congress of The Hong Kong Chemical Society (1998)

Oral - 4

A Modified Chemometrics Method for Two-Way Data Analysis by Using Fast Wavelet Transform

Kai-man Leung[a], Foo-tim Chau[a], Yi-zeng Liang[b] and Jun-bin Gao[c]

[a] *Department of Applied Biology and Chemical Technology, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong*

[b] *College of Chemistry and Chemical Engineering, Hunan University, Changsha, 410082, Hunan, P. R. China.*

[c] *Department of Mathematics, Huazhong University of Science and Technology, Wuhan, 4300?4 Hubei, P. R. China.*

Abstract

In recent year, the development of hyphenated instruments such as HPLC-DAD and GC-MS has been very fast in analytical chemistry. With the use of chemometrics techniques such as principle component analysis, more valuable information can be extracted from data acquired from these instruments as completely as possible [1]. In this work, the heuristic evolving latent projections (HELP) method [2] was chosen for analyzing experimental data from HPLC-DAD. It was found that the HELP method failed to analysis samples with strongly overlapped chromatographic peaks at low concentration level. Fast wavelet transform (FWT) [3-5] has been proposed to solve this problem by enhancing the sensitivity of the HELP method. The HPLC-DAD data was pre-processed with FWT by using Daubechies wavelet filter and then analyzed with the HELP method. The result indicated that the sensitivity of the HELP method can be improved and the approach facilitated for the identification of the selective region in the data. This new chemometrics method can be adopted for chemical analysis of complicated multiple component system such as Chinese herb using HPLC-DAD

References

[1] E.J. Karjalainen and U.P. Karjalainen (Eds.), Data Analysis for Hyphenated Techniques, Elsevier, The Netherlands, 1996.

[2] O.M. Kvalheim and Y.Z. Liang, Anal. Chem., 64 (1992) 936-946.

[3] F.T. Chau, T.M. Shih, J.B. Gao and C.K. Chan, Appl. Spectrosc., 50 (1996) 339-349.

[4] J.B. Gao, F.T. Chau, T.M. Shih, SEA Bull. Math., 20 (1996) 85-90.

[5] F.T. Chau, J.B. Gao, T.M. Shih and J. Wang, Appl. Spectrosc., 51 (1997) 649-659

Abstract.6 The 216th National Meeting of the American Chemical Society (ACS) (1998)

ANYL

030.

Sensitivity Enhancement in Factor Analysis of Two-Way Hyphenated Data Using Fast Wavelet Transform, Foo-tim Chau[a]. Kai-man Leung[a]. Yi-zeng Liang[b]. Jun-bin Bao[c] and Tsz-pun Chan[a]. [a]Department of Applied Biology and Chemical Technology, The Hong Kong Polytechnic University, Hong Kong. [b]College of Chemistry and Chemical Engineering, Hunan University, Changsha, P.R. China. [c]Department of Mathematics, Huazhong University of Science and Technology, Wuhan, P.R. China.

Fast wavelet transform (FWT) was proposed as a preprocessing step of the heuristic evolving latent projections (HELP) analysis. Data from spectro-chromatograms of a HPLC-DAD system were treated by FWT with the use of the Daubechies wavelet D_{16} function at resolution level 1 for each wavelength. The scale coefficients at this resolution level and individual wavelengths were recombined to form a new data set for HELP analysis. It was found that the FWT method can reduce the window size and enhance the sensitivity of the HELP analysis by means of signal denoising. Besides, the HPLC-DAD data in the compressed form can also reduce the computational time of the HELP analysis. Both simulated and experimental HPLC-DAD data were employed to test our proposed method.

Abstract.7 The Third International Symposium of Worldwide Chinese Scholar on Analytical Chemistry (ISWCSAC98) (1998)

A ROUGHNESS PENALTY APPROACH AND ITS APPLICATION FOR NOISY HYPHENATED CHROMATOGRAPHIC TWO-WAY DATA

Yi Zeng LIANG[a], Kai Man LEUNG[b], Foo Tim CHAU[b]

[a] College of Chemistry and Chemical Engineering, Hunan University, Changsha 410082.
[b] Department of Applied Biology and Chemical Technology, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong

ABSTRACT
In order to improve signal detection and resolution of chemical components with very low concentrations in the hyphenated chromatographic two-way data, the effectiveness of measurement errors from the instruments on these two aspects is first investigated in the present paper. A new smoothing technique called roughness penalty approach is introduced to reduce the influence of the measurement errors. The results showed that the proposed method can enhance the detection ability significantly. The resolved spectra after such a smooth treatment using the roughness penalty approach can also be significantly improved. The performance of the method are assessed by the analysis of real hyphenated two-way data as well as simulated data.

REFERENCES
1. Adam, Chemometrics in Analytical Spectroscopy, The Royal Society of Chemistry, Cambridge, 1995.
2. S.E. Leurgans, R.A. Moyeed and B.W. Silverman, J.R. Statist. Soc. B55 (1993) 725.
3. P.J. Green and B.W. Silverman, Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach, Chapman and Hall, London, 1994.