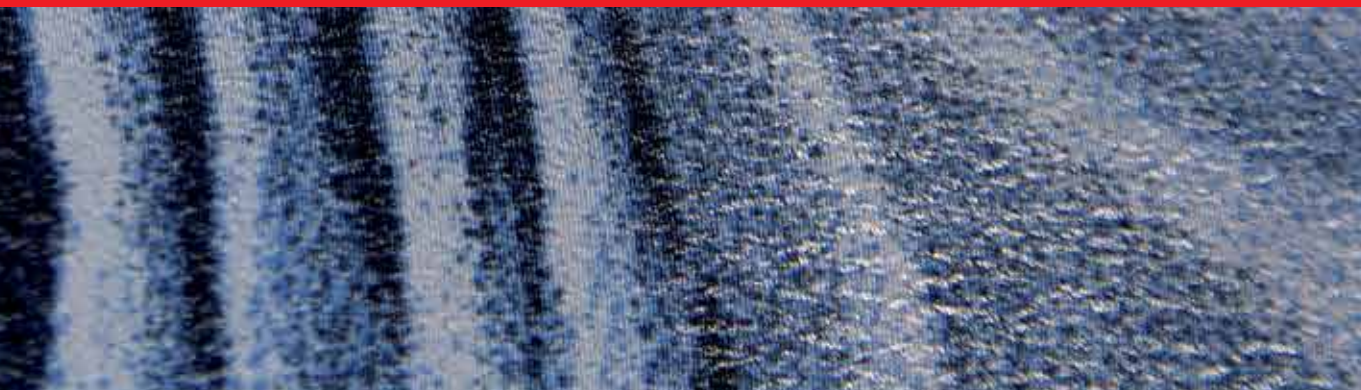
IntechOpen

Wavelet Theory and Its Applications

Edited by Sudhakar Radhakrishnan



WAVELET THEORY AND ITS APPLICATIONS

Edited by **Sudhakar Radhakrishnan**

Wavelet Theory and Its Applications
http://dx.doi.org/10.5772/intechopen.71240
Edited by Sudhakar Radhakrishnan

Contributors

Husam Alzaq, Michaela Kiermeier, Seedahmed Mahmoud, Jusak Jusak, Takaaki Sato, Riichi Kajiwara, Ichiro Takashima, Toshio Iijima, Joan Nix, Bruce McNevin, Perry Xiao, Shwan Dyllon, Talbi Mourad, Bouhlel Med Salim, Sabeur Sabeur Abid, Edwin Muchebve, Yoshiyuki Nakamura, Hiroshi Kamiya, Boris I. Lembrikov, Yossef Ben-Ezra, Moshe Schwartz, Segev Zarkovsky, Jesús Villa, Ismael De La Rosa, Gustavo Rodríguez, Jorge Luis Flores, Rumen Ivanov, Guillermo García-Torales, Daniel Alaniz, Efren Gonzalez, Gholamreza Naser, Peyman Yousefi, Hadi Mohammadi, Sudhakar Radhakrishnan

Notice

First published in London, United Kingdom, 2018 by IntechOpen
eBook (PDF) Published by IntechOpen, 2019
IntechOpen is the global imprint of INTECHOPEN LIMITED, registered in England and Wales, registration number: 11086078, The Shard, 25th floor, 32 London Bridge Street
London, SE19SG – United Kingdom
Printed in Croatia

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library

Additional hard and PDF copies can be obtained from orders@intechopen.com

Wavelet Theory and Its Applications
Edited by Sudhakar Radhakrishnan

p. cm.

Print ISBN 978-1-78923-432-9

Online ISBN 978-1-78923-433-6

eBook (PDF) ISBN 978-1-83881-536-3

Meet the editor

Dr. R. Sudhakar is currently a professor and the head of the Department of Electronics and Communication Engineering, Dr. Mahalingam College of Engineering and Technology, Pollachi, India. He is an editorial board member of three international journals, namely, *International Journal of Computer Theory and Engineering*, *International Journal of Computer and Electrical Engineering*, and *International Arab Journal of Information Technology*. He is a reviewer of 16 international journals, namely, *IEEE Transactions on Systems, Man, and Cybernetics: Systems* by IEEE; *International Arab Journal of Information Technology* coming from Zarqa University, Jordan; *International Journal of Computer and Electrical Engineering* published by International Association of Computer Science and Information Technology Press (IACSIT); *International Journal of Computer Theory and Engineering* published by International Association of Computer Science and Information Technology Press (IACSIT); *Journal of Electrical and Electronics Engineering Research*; *Iranian Journal of Electrical and Computer Engineering* published by Iranian Research Institute for Electrical Engineering, JD, Iran; *Journal of Optical Engineering* published by SPIE, USA; *Journal of Electronic Imaging* published by SPIE, USA; *The Imaging Science Journal* published by Taylor & Francis Group; *International Journal of Computational Science and Engineering (IJCSE)* published by Inderscience; *International Journal of Image Mining (IJIM)* published by Inderscience; *International Journal of Biomedical Engineering and Technology (IJBET)* published by Inderscience; journal for *Image Analysis & Stereology* from the International Society for Stereology; *ETRI Journal* from Korea; *AEU—International Journal of Electronics and Communications* by Elsevier; and *IET Image Processing*. He wrote two books titled *Research Issues in Image Compression Using Wavelet Variants* published by Lambert Academic Publishing (LAP), Germany (2010), and *Practicing Signals and Systems Laboratory Using MATLAB* published by Lambert Academic Publishing (LAP), Germany (2010), and two book chapters titled "Wavelet-Based Image Compression" in the book titled *Computational Intelligence Techniques in Handling Image Processing and Pattern Recognition* published by Lambert Academic Publishing (LAP), Germany (2010), and "Analysis of Hand Vein Images Using Hybrid Techniques" in *Hybrid Intelligence Techniques for Image Analysis and Understanding* published by Wiley. He edited three books titled *Effective Video Coding for Multimedia Applications*, *Applications of Digital Signal Processing through Practical Approach*, and *Recent Advances in Image and Video Coding* all published by InTech, Croatia (2011, 2015, and 2016). He has published 95 papers in international and national journals and conference proceedings. His areas of research include digital image processing, image analysis, wavelet transforms, and digital signal processing.

Contents

Preface

The concept of wavelets is quite new and different to the readers. Original research works were carried out in the field of mathematics, electrical engineering, physics, applied science, and engineering such as geophysics and finance.

A wavelet is a wavelike oscillation with amplitude that begins at zero, increases, and then decreases back to zero. It can typically be visualized as a "brief oscillation" like the one recorded by a heart monitoring system. Generally, wavelets have specific properties that make them useful for signal processing application.

Wavelet theory is applicable to several areas. Wavelet transforms normally come with time-frequency representation for analog signals. For any practical purpose, people use discrete wavelet transforms, which use discrete-time filter banks for implementation. These filter banks are called the wavelet and scaling coefficients in wavelet nomenclature. These filter banks may contain either finite impulse response (FIR) or infinite impulse response (IIR) filters. To understand these filters, signal processing concepts are very much essential.

Scope of the book: This book is the research outcome of various researchers and professors who have contributed a lot in applying wavelets in different application areas. This book suits researchers doing their research in the area of wavelets and its applications. The understanding of fundamentals of linear algebra and signal processing concepts is essential for the readers before reading this book. This book gives fundamental idea of wavelets and also how wavelets can be applied in different fields such as biomedical, prediction analysis, and finance. I do not have any doubts that this book will motivate researchers in different parts of the world to use the concepts, which are available in this book for their research.

Structure of the book: The book contains 12 chapters grouped into four sections. The reader of the book is expected to know the fundamentals of linear algebra and signal processing, which are available in the standard books. The book starts with an introductory chapter. Section 2 deals with the role of wavelets in biomedical application containing three chapters. Section 3 describes how wavelets can be applied in the field of finance containing two chapters. Section 4 concentrates on the usage of wavelets in contemporary applications containing seven chapters.

Acknowledgment

I thank the Almighty for showering blessings and giving me intelligence and energy to complete this work. I thank all the professors and scientists who have contributed to the different chapters. My sincere thanks go to the management of Dr. Mahalingam College of Engineering and Technology and Prof. C. Ramaswamy, Secretary, NIA Institutions, for their encouragement and patronage rendered to carry out this work. I am indebted to my wife

Mrs. Vinitha Mohan. Her support, encouragement, quiet patience, and unwavering love undeniably led me to the successful completion of the work. I have no words to thank my son S.V. Hemesh for bearing my preoccupation, understanding, and love he gave me. I am grateful to the IntechOpen publishing team specially Mr. Markus Mattila, Publishing Process Manager, who constantly helped me in bringing out this book.

Dr. R. Sudhakar
Professor, Department of Electronics and Communication Engineering
Dr. Mahalingam College of Engineering and Technology
India

Introduction

Chapter 1

Introductory Chapter: Understanding Wavelets

Sudhakar Radhakrishnan

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/intechopen.78388

1. Introduction

In this section, let us discuss some fundamentals which are required to understand wavelets. Signals which are coming from a source are normally in time domain. Examples are sinusoidal signal, bio-medical signal, etc. Anytime domain signal can be processed or transformed into frequency domain (spectral domain) using mathematical transformations. Fourier transform is one of the popular or famous transform that will convert a time domain signal into frequency domain signal without any loss of generality.

While plotting time domain signal, we use time in the x-axis and amplitude in the y-axis. The hidden information present in the signal cannot be revealed in the time domain hence a transform domain is required. The frequency content or spectrum of a signal is simply the frequency content (spectral components) of the signal. The frequency spectrum of a signal depicts what are all the frequencies exist in the signal. When plotting frequency domain, we use frequency in the x-axis and amplitude in the y-axis.

Normally for any signal, if the frequency content is not changing with respect to time is called as stationary signal. Example can be a sinusoidal signal where the frequency 'X' Hz is not changing irrespective of the cycle. Unfortunately, real time signals are nonstationary signal where the frequency content of the signal is keeping on changing. The best example is biological signals. Suppose when we are looking at an ECG (electrocardiograph) signal. The typical shape of a healthy ECG signal is well known to cardiologists. Any significant deviation from that shape is usually considered to be a symptom of a pathological condition. Doctors analyse these cases not only in time domain, they are using frequency domain also to confirm the pathological condition.

2. Fourier transform (FT)

To understand wavelets, let us go deep into the literature. The first and main contribution regarding the frequency content or spectrum came from the French mathematician John Baptise Fourier. He showed that any periodic function can be represented as an infinite sum of periodic complex exponential functions and named as Fourier transform (FT) [1].

Eqs. (1) and (2) represent the forward and inverse Fourier transform

$$X(j\Omega) = \int_{-\infty}^{\infty} x(t)\, e^{-j\Omega t}\, dt \tag{1}$$

$$x(t) = \int_{-\infty}^{\infty} X(j\Omega)\, e^{j\Omega t}\, d\Omega \tag{2}$$

where x(t) represents the time domain signal, X(jΩ) represents the frequency content of the signal and $\Omega = 2 \times \pi \times F$ and

$$e^{j\Omega t} = \cos(\Omega t) + j\sin(\Omega t) \tag{3}$$

Eq. (3) represents that any complex exponential is expressed as real part of cosine function and imaginary part of sine function with the corresponding frequency. As per Eq. (1), the input signal is multiplied with cosine function and sine function at all the time intervals and added (integrated) to yield the frequency content. The concept is best illustrated in **Figures 1** and **2**. **Figure 1** shows the implementation at analysis side. Here, each blue coloured square is a narrow band pass filter with the cut off frequency Ω_0, Ω_1, ….

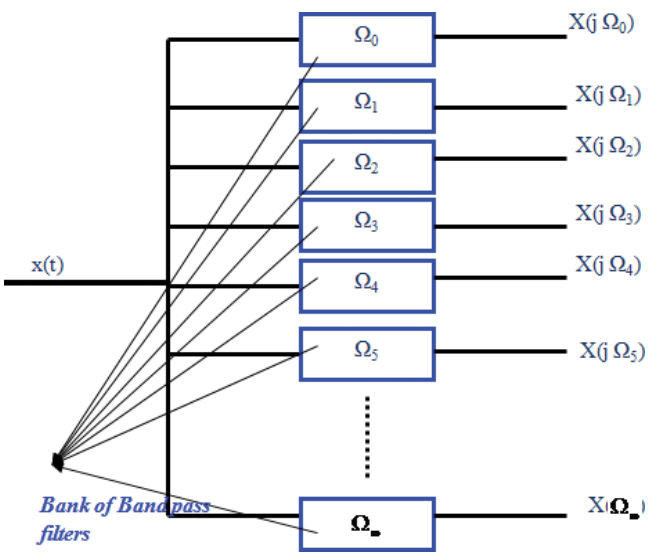


Figure 1. Fourier transform analysis side.

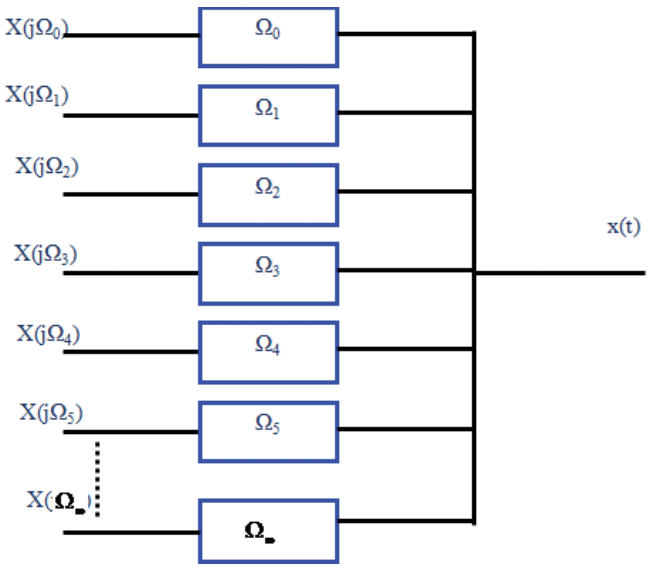


Figure 2. Fourier transform synthesis side.

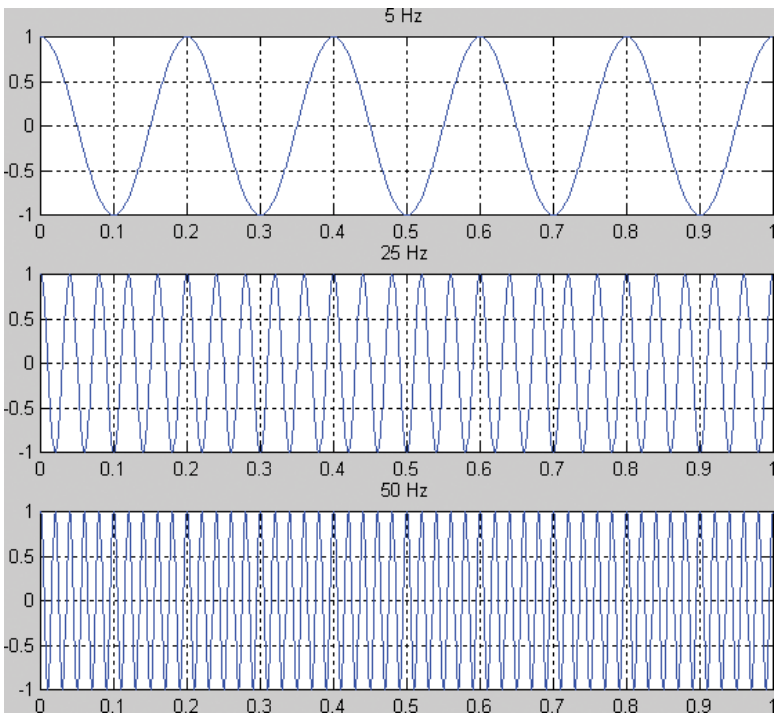


Figure 3. Cosine signal with three different frequencies 5, 25, and 50 Hz.

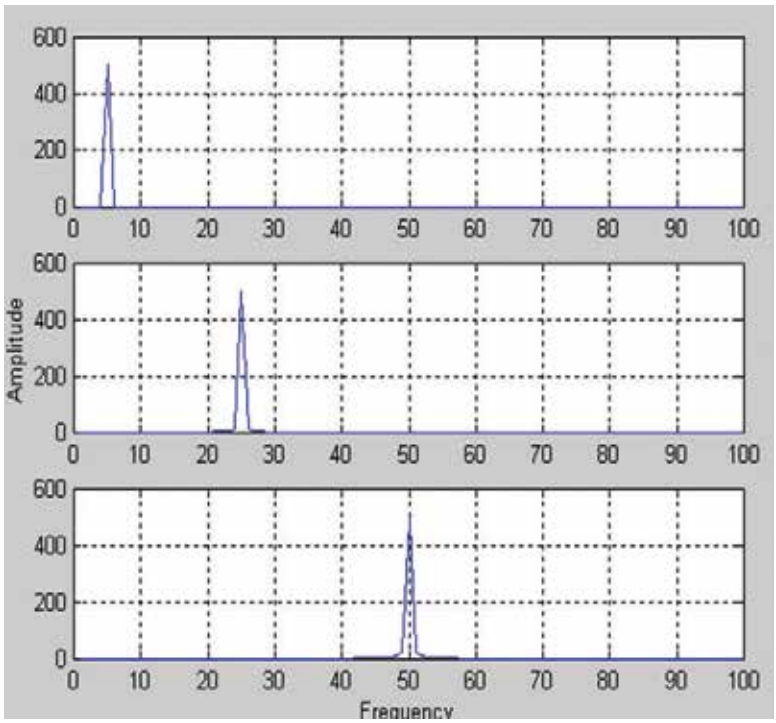


Figure 4. Spectrum cosine signal with three different frequencies 5, 25, and 50 Hz.

Figure 2 shows the implementation at synthesis side. Here also, each blue coloured square is a narrow band pass filter with the cut off frequency Ω_0, Ω_1, ….

In **Figure 1**, if the result of the particular filter is large then, we can understand that the particular frequency component is dominant. If there is no output in any filter means that the particular frequency component is zero, i.e., single signal is passed multiple filter and the outputs are analysed and the reverse is happening in **Figure 2** which is called as synthesis.

To understand the above one, see the following illustration shown in **Figure 3** where the individual frequencies are indicated as 5, 25, and 50 Hz. The corresponding spectrum is shown in **Figure 4**.

Figure 5 indicates a signal which contains all the three frequencies 5, 25, and 50 Hz mixed together and its corresponding spectrum.

Fourier transform is very much useful for a stationary signal. This means that Fourier transform clearly indicates what are all the frequency components exist in the given signal independent of time. Fourier transform completely fails for a nonstationary signal.

$$x_4(t) = \cos(2\pi \cdot 5 \cdot t) + \cos(2\pi \cdot 25 \cdot t) + \cos(2\pi \cdot 50 \cdot t)$$

$$x_4(t) \xleftrightarrow{\mathcal{F}} X_4(\omega)$$

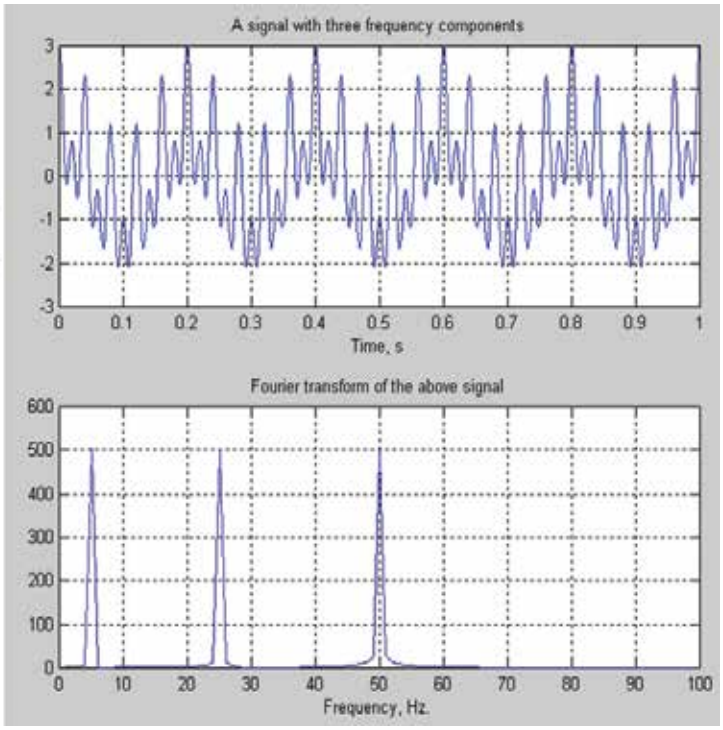


Figure 5. Cosine signal with three combined frequencies 5, 25, and 50 Hz and its corresponding spectrum.

3. Short-time Fourier transform (STFT)

As discussed previously that FT is not suitable for a nonstationary signal, a new set of transform is required which will provide timing and frequency information. The research moved forward and a conclusion was made that possibility of considering some portion of a nonstationary signal as stationary. This means that the long duration signal must be chopped for a short duration and possibility of finding the frequency components in that interval and this has to be completed for the entire signal to know the entire frequency components are present. The transform which provides this opportunity is short time Fourier transform (STFT). The STFT equation is given by

$$STFT_x^W(t, \Omega) = \int_{-\infty}^{\infty} [x(t) \times W(t - \tau)]\, e^{-j\Omega t}\, dt \tag{4}$$

The above equation indicates that the input signal x(t) is chopped by a window with a duration of 'τ' and Fourier transform is taken. In other words, the signal is assumed to be stationary for the interval 'τ'. This process is repeated for the entire duration of signal. Now, somewhat the

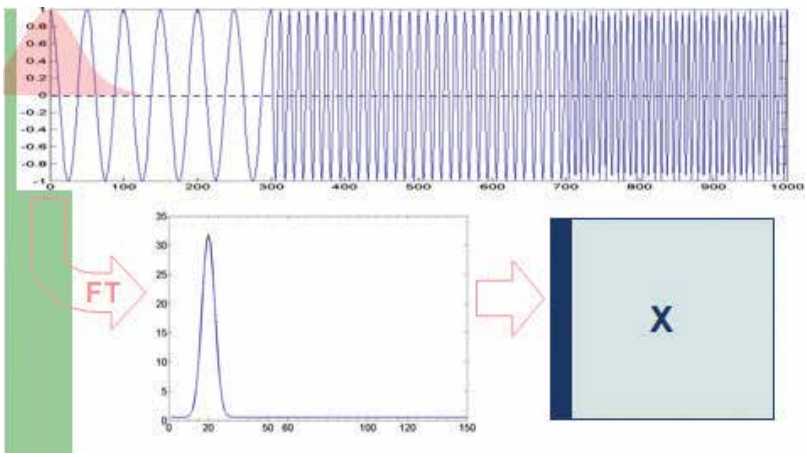


Figure 6. STFT of a nonstationary signal for one step.

problem faced by Fourier transform is solved and we get the frequency components with the window duration specified. **Figure 6** indicates how STFT is taken for a nonstationary signal for one step of the window.

Now, the selection of the window width plays a vital role in STFT. The narrow window we select leads to poor frequency resolution and good time resolution which is shown in **Figure 7**. The opposite effect will happen if we select a wider window which is illustrated in **Figure 8**.

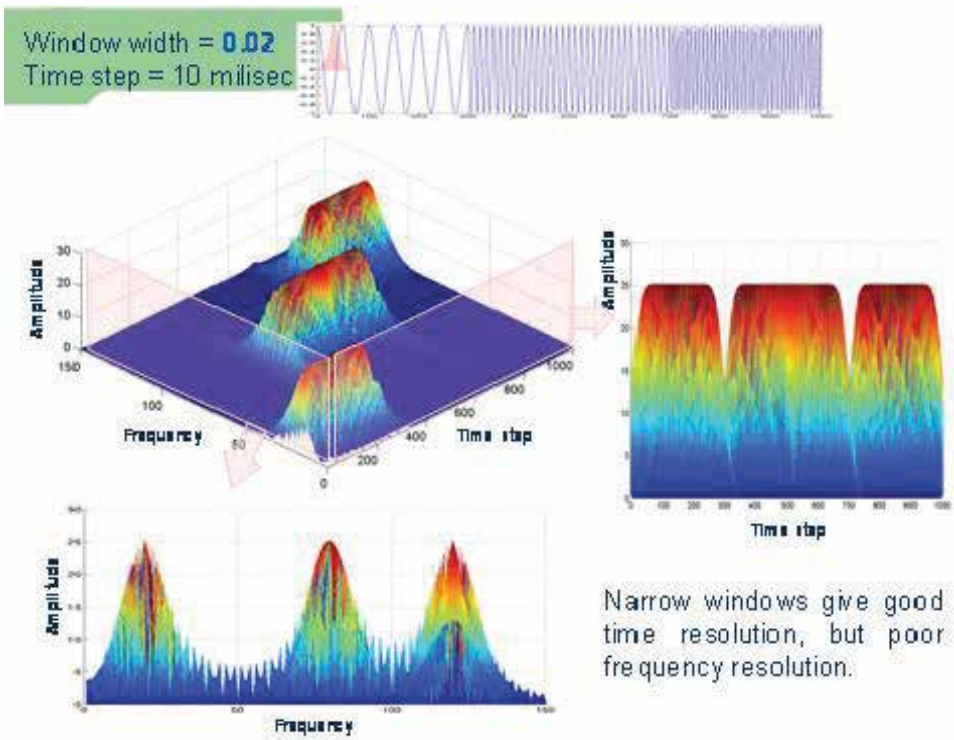


Figure 7. Resolution for a narrower window.

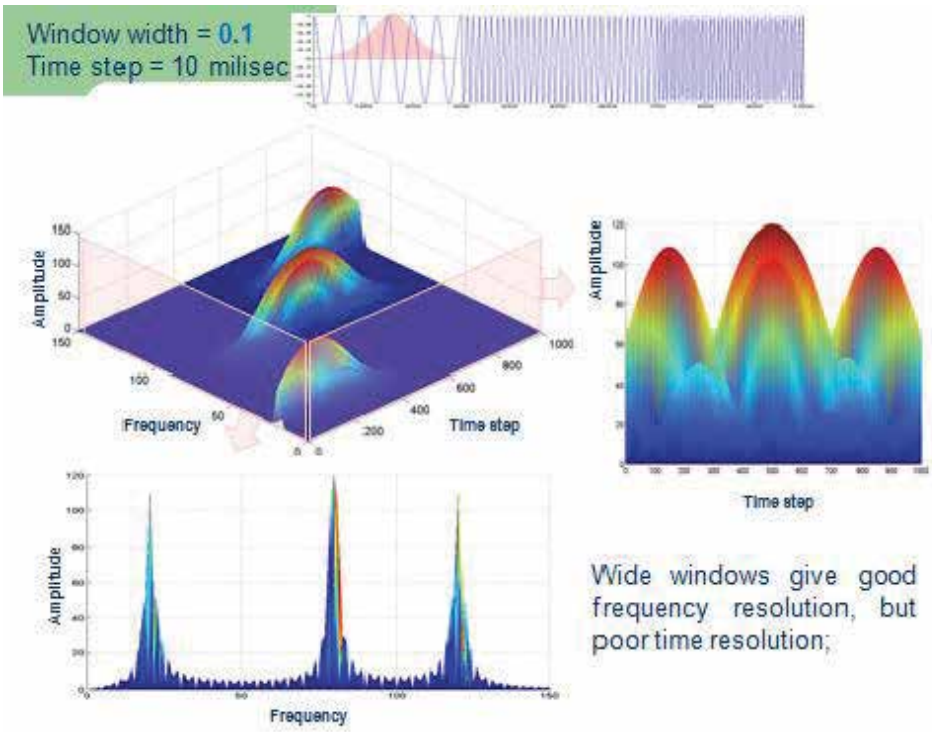


Figure 8. Resolution for a wider window.

Hence, the conclusion is STFT [1] that will provide a solution to the problem faced by Fourier transform but the drawback is there is a constant width of the window which is used and hence it provides only fixed resolution.

Normally, signal can carry both low frequency and high frequency components. To capture both, we need different widths of the window which is not provided by STFT. To understand more clearly, the concept of multiresolution is not there in STFT. Hence, we need a new transform which provides a solution to the above. The solution is wavelet transform.

4. Continuous wavelet transform (CWT)

The basic idea behind wavelet transform is, a new basis(window) function is introduced which can be enlarged or compressed to capture both low frequency and high frequency component of the signal (which relates to scale). The equation of wavelet transform [2, 3] is given in Eq. (5).

$$W(a,b) = \frac{1}{\sqrt{|a|}} \int_{-\infty}^{\infty} x(t)\, \psi\left(\frac{t-b}{a}\right) dt \tag{5}$$

where W(a,b) is called the wavelet coefficient, 'a' is called the scaling parameter and 'b' is the shifting or translational parameter. $\psi(t)$ is called the mother wavelet. Different dilations and translations lead to different daughter wavelets.

Any original data or signal can be represented in terms of a wavelet expansion. The best representation of a data using a wavelet depends on the best or close wavelet of what we are choosing. There are many numbers of wavelets available as per the literature. Some of the examples of the wavelets are Haar and Daubechies [3]; under Gaussian-based wavelets, we have Mexican hat wavelet and Morlet wavelet; under polynomial-based wavelets, we have Battle-lemarie, Coiflet and Spline-based wavelets; and under Sinc wavelets, we have Meyer wavelet and Shannon wavelet.

5. Discrete wavelet transform (DWT)

From the previous understanding, it is clear that CWT is a redundant transform, which means that the translation parameter 'b' and scaling parameter 'a' seem to be infinite making them difficult in terms of implementation. It is always seems to be CWT that is computable but not implementable. The solution for the implementation of wavelet transform arises from discrete wavelet transform (DWT). Sampling in the time-frequency plane on a dyadic (octave) grid is happening in DWT that makes them efficient in terms of implementation. The scaling parameter 'a' is replaced by 2^{-j} and 'b' is made proportional to 'a', i.e., $b = k\,2^{-j}$. Here 'j' is called as scaling parameter and 'k' is the proportionality constant taking the role of shifting parameter in DWT. Substituting $a = 2^{-j}; b = 2^{-j}k$ (j and k are integers) in Eq. (5), we get Eq. (6).

$$CWT_x^{\psi}(a,b) = \Psi_x^{\psi}(a,b) = \frac{1}{\sqrt{|a|}}\int_t x(t)\psi\left(\frac{t-b}{a}\right)dt$$
$$\psi_{j,k}(t) = \sqrt{2^j}\,\psi(2^j t - k)\quad j,k \in Z \tag{6}$$

In multiresolution analysis, the signal can be viewed as the sum of a smooth ("coarse") part—reflects main features of the signal (approximation signal) and a detailed ("fine") part—faster fluctuations represent the details of the signal [1]. The separation of the signal

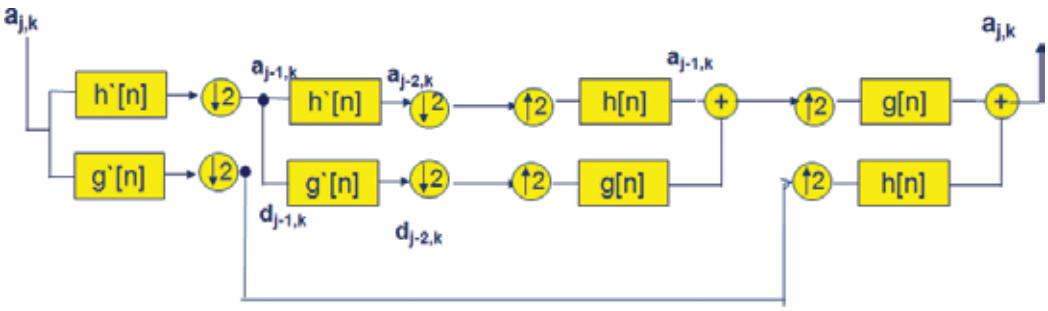


Figure 9. Filter bank implementation of DWT (courtesy by Robi polikar).

into two parts is determined by the resolution given. This discussion introduces scaling function '$\varphi_{j,k}(t)$' and wavelet function '$\psi_{j,k}(t)$' and the corresponding approximation coefficients resulting from scaling function is denoted as $a_{j,k}$ and the detailed coefficients resulting from wavelet function is denoted as $d_{j,k}$. In implementation perspective, any signal is represented in Eq. (7).

$$x(t) = \sum_{k=-\infty}^{\infty} a_{j_0,k}\,\varphi_{j_0,k}(t) + \sum_{j=-\infty}^{j_0}\sum_{k=-\infty}^{\infty} d_{j,k}\,\psi_{j,k}(t) \tag{7}$$

Using Filter bank [2, 4], the implementation of Eq. (7) results in **Figure 9**, where h′[n] and g′[n] are the low pass and high pass filters in the analysis side and h[n] and g[n] are low pass and high pass filters in the synthesis side using the corresponding wavelet.

6. Conclusion

I hope that this chapter gives a definite and thoughtful introduction to all the beginners who are new to wavelets. As there are different number of wavelets available with different signal processing properties like compact support, symmetry, regularity and vanishing moments make them suitable in the field of signal de-noising, detecting discontinuities and breakdown points in a signal, compressing images, identifying pure frequencies, seismic and geophysical signal processing, video compression, acoustic data analysis, nuclear engineering, neurophysiology, music, magnetic resonance imaging, speech discrimination, optics, fractals, turbulence, earthquake-prediction, radar, human vision, etc. Some of applications from the perspective scientists and researchers are discussed in the forthcoming chapters.

Author details

Sudhakar Radhakrishnan

Address all correspondence to: sudha_radha2000@yahoo.co.in

Department of Electronics and Communication Engineering, Dr. Mahalingam College of Engineering and Technology, Pollachi, India

References

[1] The wavelet tutorial by Robipolikar. https://cseweb.ucsd.edu/~baden/Doc/wavelets/polikar_wavelets.pdf

[2] Vetterli M, Herley C. Wavelets and filter banks: Theory and design. IEEE Transactions on Signal Processing. 1992;**40**:2207-2232

[3] Daubechies I. Orthonormal bases of compactly supported wavelets. Communications on Pure and Applied Mathematics. 1988;**41**:906-966

[4] Abate A et al. Wavelets and Subbands Fundamentals and Applications. Newyork: Birkhäuser c/o Springer Verlag; 2002

Section 2

Wavelets in Bio-medical Application

Chapter 2

Wavelet Correlation Analysis for Quantifying Similarities and Real-Time Estimates of Information Encoded or Decoded in Single-Trial Oscillatory Brain Waves

Takaaki Sato, Riichi Kajiwara, Ichiro Takashima and Toshio Iijima

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/intechopen.74810

Abstract

Even in cases when we recognize identical objects or when we behave similarly, the spatiotemporal activities in the brain are likely to fluctuate to various degrees. Temporally fluctuating responses easily decrease by averaging replicate measures. We previously developed a wavelet correlation analysis that tolerates the across-trial oscillatory phase variability observed in odor-induced cortical responses. The wavelet correlation analysis revealed a change in the neuronal information redundancy of transient and oscillatory brain waves from the dependencies on stimulus experience (high redundancy) to stimulus quality (low redundancy) between the input and output layers of the anterior piriform cortex in guinea pigs. We report on its application to estimate information in the fine temporal structures of single-trial brain waves. By using a set of standard brain waves for each information in a given category, the highest wavelet correlation coefficients provided the first candidate of estimated information with 75% accuracy. Moreover, the probability of including the correct information for the two upper candidates, regardless of information redundancy of the signal sources, was >92%. The wavelet correlation analysis is useful for similarity analyses and real-time estimates of in-brain information and for its application to brain-machine interfaces or medical/research tools.

Keywords: cross-correlation analysis, electroencephalography, information redundancy, odor representation, oscillatory local field potentials, real-time estimation, wavelet transformation

1. Introduction

In the sensory system, a stimulant likely activates stimulant-specific subsets of neurons with a stimulant-specific response profile through the sensory pathway from the sensory organ to the primary sensory cortex, resulting in identical sensory perception of the stimulant. At different stages of this neuronal information processing, the redundancy in sensory information changes by summing or subtracting overlapping signals from cognate and noncognate receptors for common and unique elements. The sensory systems generate oscillatory activities between related cortical regions and the thalamus, except in the olfactory system. The olfactory system generates oscillatory activities in the first and second olfactory centers, the olfactory bulb, and the anterior piriform cortex (aPC). It is significantly more difficult to quantify the degree of similarity or difference in these transient oscillatory responses compared to stationary oscillatory activities. We previously developed a wavelet correlation analysis that is phase-tolerant for transient oscillatory responses and demonstrated a stimulus dependency of the odor-evoked oscillatory brain waves (oscillatory local field potentials, osci-LFPs) in the aPC output layer and an experience dependency in the input layer [1]. These results suggest that the redundancy in the neural representation of olfactory information may change in the aPC.

Sensory systems are incorporated in higher brain functions that synergistically control animal behaviors through multiple neural systems including sensory, memory, decision, motor, or other systems. Generally, all neural systems would maintain the reliability of signal processing in identical activities of identical subsets of neurons in identical time courses through neural pathways with acceptable across-trial variability. This suggests that brain waves in identical behaviors could be, to some extent, reproduced in each brain. Small fluctuations, however, sometimes change oscillatory phases across trials, as has been observed in odor-induced oscillatory brain waves [1]. The fine temporal structures of phase-fluctuated oscillatory activities responsible for informational differences are easily lost by averaging several brain waves, even for identical information in each brain. Associations of single-trial brain waves with in-brain information have been rarely studied. Regarding mental states, the most important individual-independent frequencies of electroencephalography (EEG) are 7–12 Hz at the P1 electrode and <5 Hz at Fz for attention, 10–20 Hz at F4 for fatigue, and 4–7 Hz at Fz and 10–20 Hz at Cz for frustration, with even greater variations in frequencies observed across individuals [2]. Alpha-band oscillations (8–13 Hz) exert top-down influences on the early visual processing for attention orienting [3] and are sensitive markers in the auditory memory loading process [4]. As a test case, we applied a wavelet correlation analysis to estimate odor information in the fine temporal structures of single-trial brain waves.

2. Wavelet correlation analysis

2.1. Characteristics of odor-evoked oscillatory brain waves in the aPC

Odor-evoked oscillatory brain waves in the aPC are not stationary over the time window of interest, even in an ex vivo isolated whole brain with attached nose preparation under the

http://dx.doi.org/10.5772/intechopen.74810

condition of no inputs from the nonolfactory sensory systems (**Figure 1**) [1, 5]. Oscillatory brain waves initiate during the 1-s odor presentation before the peak of the receptor potential, the electro-olfactogram (EOG) (the lowest trace in **Figure 1**) [1]. A pair of quite different odors, lavender essential oil (Lav), and a mixture of three fatty acids—mc4 + mc6 + mc8 (mc468)—were selected as plant- and animal-related odors, respectively. Linalool (Lina) and *n*-butanoic acid (mc4) were selected as the single-compound odors of Lav and mc468, respectively, with partial overlaps of the activated olfactory receptors and their respective signal pathways with their original mixtures as well as 0.1 Lav (10-fold diluted Lav). As expected, oscillatory brain waves of a pair of quite different Lav and mc468 odors look dissimilar in the initial phase but are partially similar in the late phase.

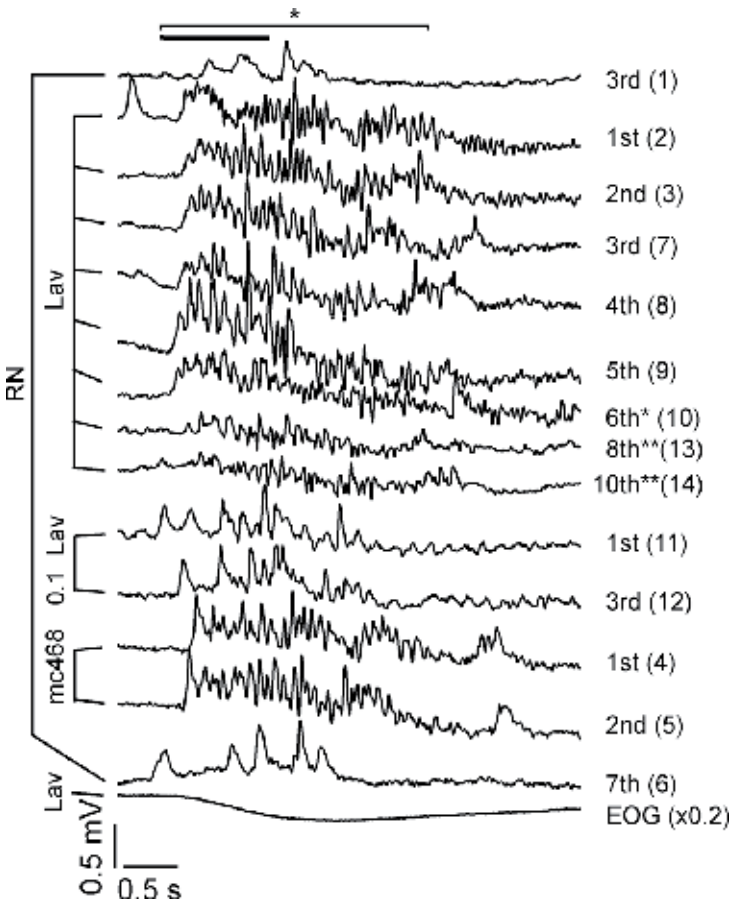


Figure 1. Odor-evoked oscillatory brain waves in layer I of the anterior piriform cortex (aPC) [1]. Time courses of low-pass-filtered (0–45 Hz) oscillatory brain waves and the receptor potential (electro-olfactogram, EOG) at the centromedial or caudocentral** site of the aPC in the isolated whole brain are shown for three odors (Lav, lavender essential oil as an odor from a plant; 0.1 Lav (10-fold diluted Lav); and mc468, a mixture of three fatty acids as an imitated odor from animals). Ringer solution (RN) was used as a control. The odor or RN was presented to the nose of the isolated brain for 1 or 4 s* (only for the sixth Lav), as indicated by the horizontal bar in the in-presentation order for each odor (entire presentation order). The responses in the 2.5-s time window* of interest were analyzed.

The correlations of the temporal profiles of oscillatory brain waves in the aPC for a 2.5-s time window, which comprised the 1-s odor presentation and the following 1.5 s, were not homogeneously high between identical odors (**Figure 2A**) [1]. Only a few identical odor pairs for Lav or 0.1 Lav demonstrated relatively high correlations (0.7–0.74), whereas the remaining pairs demonstrated intermediate (0.47–0.69) or low (0.29) correlations. These low correlations are attributable to the independent fluctuations in the oscillatory phase angles and powers including a few synchronous cycles (indicated by the daggers), in the fast Fourier transform (FFT) components even between identical odors, indicating that oscillatory responses are not strictly phase-locked to the stimulus onset (**Figure 3**) [1]. The spurious high correlations of the 0–45 Hz components are attributable to the similarities in the temporal profiles of the 0–2 Hz components [1]. The 0–2 Hz component resulted in high correlations (>0.77) for all the Lav and 0.1-Lav pairs (**Figure 2B**), whereas the 2–45 Hz components resulted in low correlations (<0.4) for all pairs (**Figure 2C**). To address these weaknesses of the conventional analyses, we tested a novel correlation analysis of wavelet profiles.

2.2. Wavelet correlation analysis procedure for oscillatory brain waves in the time window of interest

Figure 4 shows the procedure for the wavelet transformation and its conversion to a data array for the wavelet correlation analysis [1]. The wavelet time-frequency power profiles enable us to quantify the similarity of the odor-evoked oscillatory brain waves. The wavelet transform is like a running, windowed Fourier transform; it uses a certain window size and slides it along in time, computing the FFT at each time using only the data within the window. The original wavelet software libraries were provided by Torrence and Compo [6] and modified with respect to the following points. Because of the spurious high correlations in the low-frequency band, all 0–2 Hz components were removed prior to the phase-tolerant analysis of the 2–45 Hz components of the oscillatory brain waves. The 2–45 Hz bandpass-filtered brain waves (**Figure 4A**) were subjected to a Morlet wavelet analysis by using the following equations:

$$W_n(s) = \sum_{n'=0}^{N-1} x_{n'} \Psi^* \left[\frac{(n' - n)\delta t}{s} \right] \tag{1}$$

$$\Psi_0(\eta) = \pi^{-1/4} e^{i\omega_0 \eta} e^{-\eta^2/2} \tag{2}$$

$$\omega_j = \frac{\omega_0 + \sqrt{2 + \omega_0{}^2}}{4\pi s_j} \tag{3}$$

$$s_j = s_o 2^{j\delta j} \ (j = 0, 1, \ldots \mathrm{J}) \tag{4}$$

$$\mathrm{J} = \delta j^{-1} ln \frac{\mathrm{N}\delta t}{s_0} \tag{5}$$

where (*) indicates the complex conjugate, $\omega_0 = 6$, N = 2048, $\delta t = 0.001$, $s_0 = 2\delta t$, and $\delta j = 0.1$. The wavelet power spectrum, $|W_n(s)|2$, was plotted in the 1.89–42.78 Hz frequency (ω_j) range

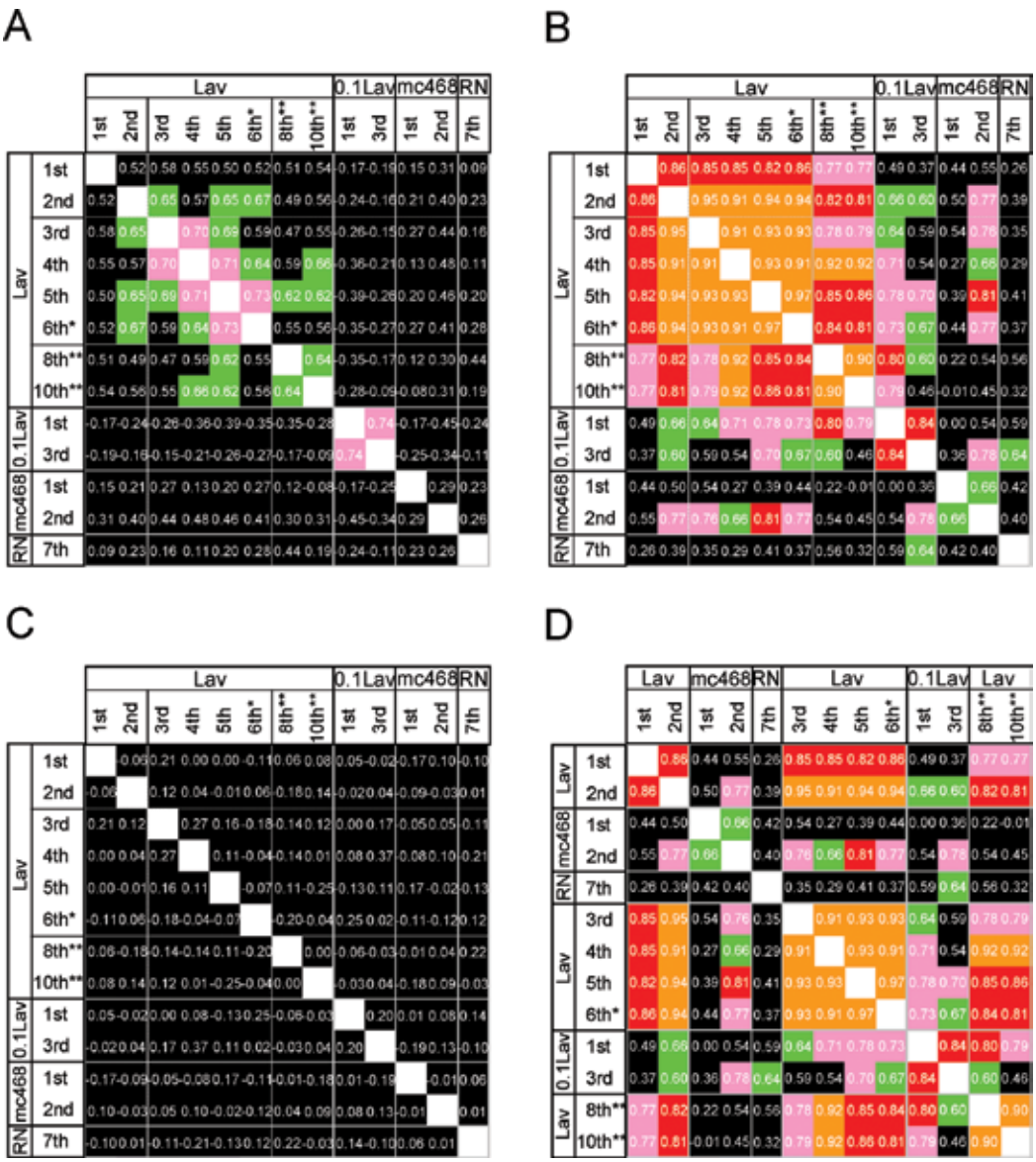


Figure 2. Correlation matrices among odor-evoked oscillatory brain waves in layer I of the aPC [1]. (A) Matrix of cross-/autocorrelations of the 0–45 Hz components of the odor-evoked oscillatory brain waves in the 2.5-s time window* of interest (shown in **Figure 1**). Some of the identical odor pairs produced high correlations >0.7. Identical odors are grouped in the order of stimulus presentation. (B) Cross-/autocorrelation matrix of the 0–2 Hz components of the odor-evoked oscillatory brain waves. (C) Cross/autocorrelation matrix of the 2–45 Hz components of the odor-evoked oscillatory brain waves. By omitting the 0–2 Hz component, all correlations were reduced to <0.4. (D) the matrix in B rearranged in the entire presentation order did not demonstrate an approach of the high correlations of the 0–2 Hz components to the diagonal line (between the dashed lines). The color represents the respective amplitude range of the cross-correlations: black, <0.60; green, 0.60–0.69; pink, 0.70–0.79; red, 0.80–0.89; orange, 0.90–0.99; and white, 1.00.

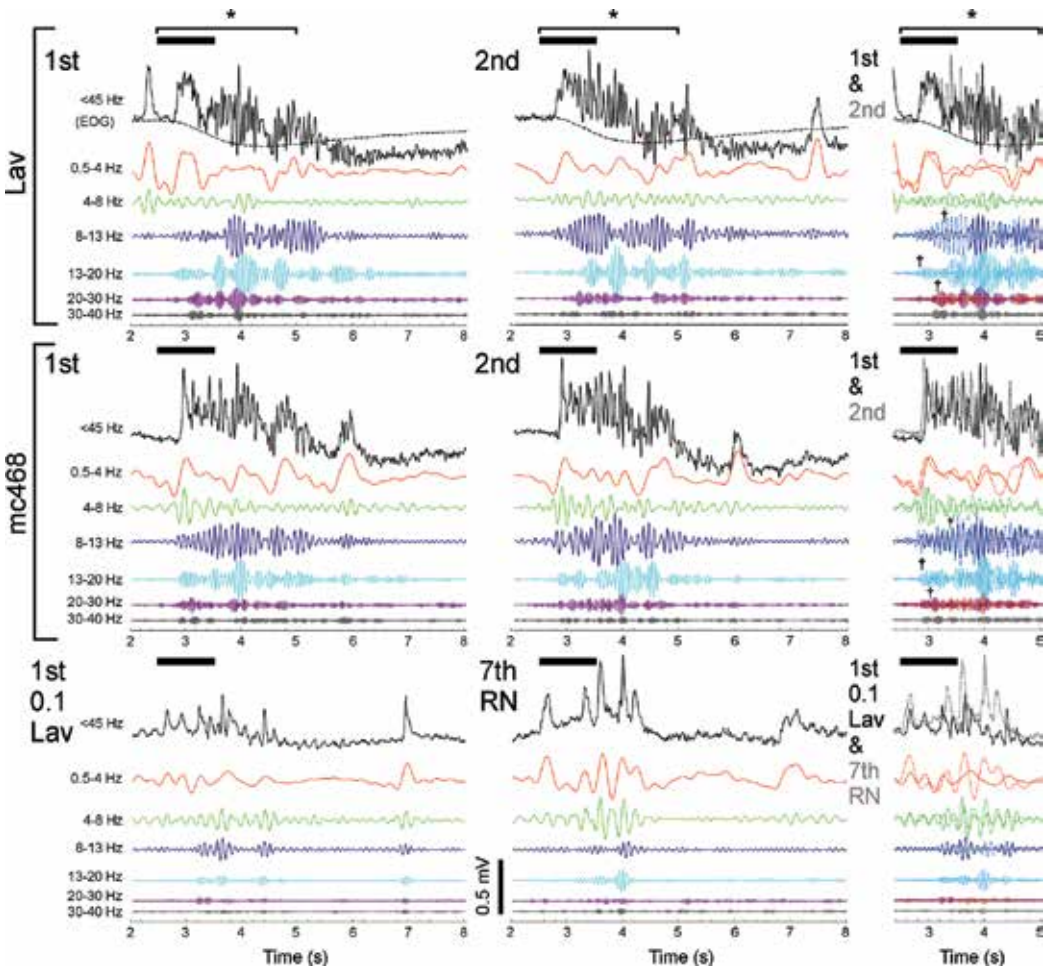


Figure 3. The oscillatory phases of the odor-evoked oscillatory brain waves differed between identical stimuli [1]. The 0–45 Hz and six frequency band components of the odor-evoked oscillatory brain waves were obtained by using an FFT bandpass filter. The two responses in the left and middle columns were superimposed on the respective frequency bands in the right column, indicating the trial-by-trial oscillatory phase differences and their fluctuations. The phase-matching points are indicated by the daggers.

(**Figure 4B**) [1]. To avoid the frequency-dependent errors that increase at the edges of epochs, the 8192 data points (2^{13} sequential points at the 1000 Hz sampling rate) were divided into seven epochs of 2048 (2^{11}) data points (2048 ms, centered every 1024 data points to the 7336th data point) with a 50% overlap and subjected to wavelet transformations (**Figure 4B**) [1]. Around the edge of each epoch, the time series was padded with the actual data ($s \geq 0$) or zeros ($s < 0$). To reconstruct a continuous wavelet transform from 0 to 8191 ms, the middle two quarters of each epoch of seven wavelets were combined (**Figure 4B**) [1]. Compared to the average wavelet power of the pre-stimulus period (10–2057 ms, marked with double asterisks in **Figure 4A**), the wavelet power in the regions within the black lines was highly significant

http://dx.doi.org/10.5772/intechopen.74810

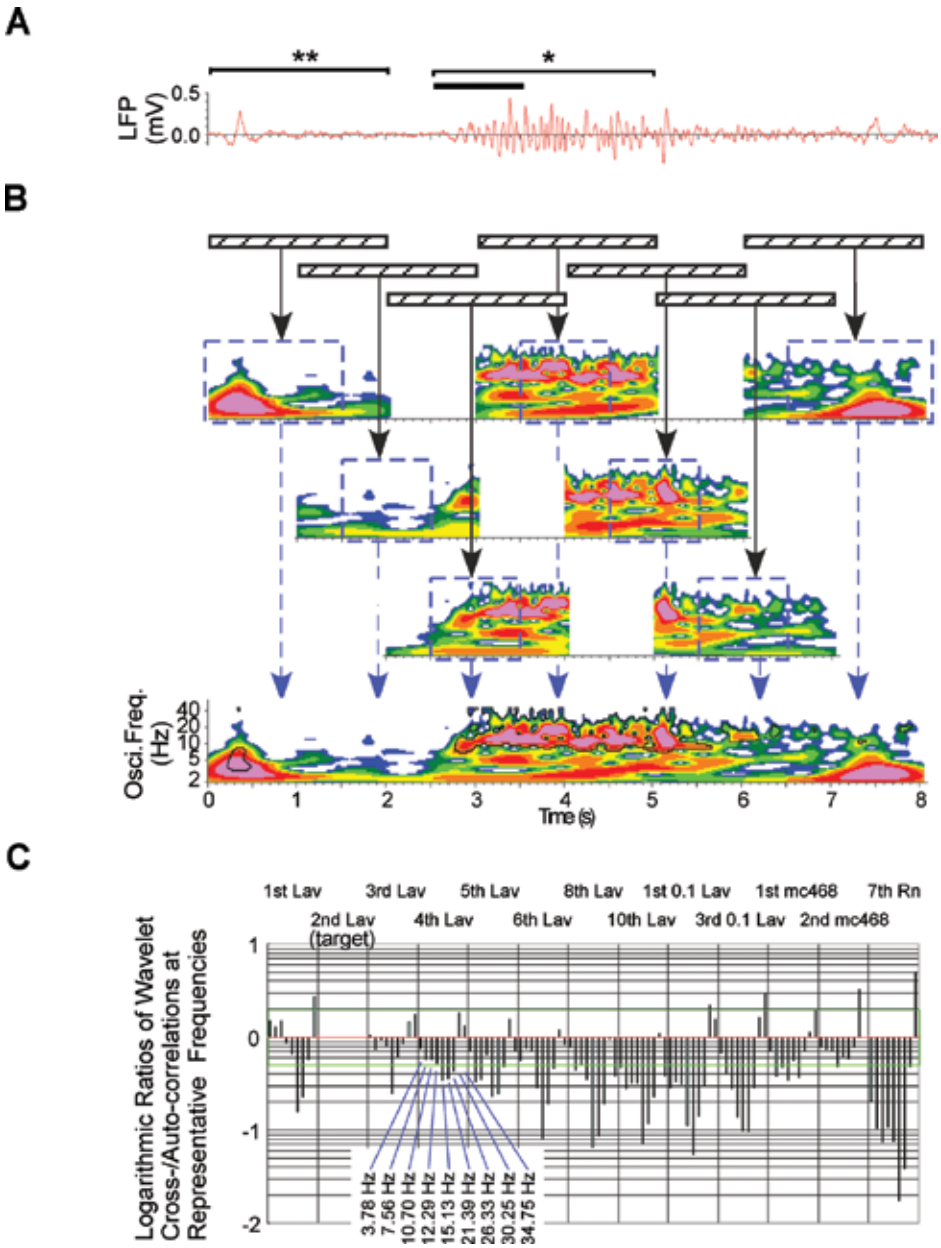


Figure 4. Wavelet transformation and wavelet cross-correlation profile of an oscillatory response [1]. (**A**) The 2–45 Hz component of a single-trial 1-s odor-evoked oscillatory brain wave (oscillatory local field potentials, osci-LFPs) in the anterior piriform cortex in an isolated guinea-pig whole brain (second presentation of lavender odor, indicated by the bold bar). (**B**) A Morlet wavelet time-frequency power spectrum of the second Lav-evoked oscillatory brain wave. Subsequently, seven sets of 2048-point wavelet transformations of the oscillatory brain waves were computed. (**C**) A columnar array of wavelet cross-/autocorrelations of the second Lav-evoked response. One of the responses for the 2.5-s time window at nine representative frequencies and sets of logarithmic ratios of the cross-correlation to the autocorrelation between wavelet pairs of the second Lav-evoked response (target) were serially concatenated into a data array, in which the wavelet correlations were calculated as correlation coefficients.

($P < 0.0001$, chi-squared test, **Figures 4B** and **5**) across all recordings from the same preparation at each frequency [1].

We calculated correlation coefficients between logarithmic ratio arrays of the cross-correlations to the autocorrelations of the wavelet power profile for the time window of interest at the following nine representative frequencies (selected from the calculated wavelet frequencies) to quantify the similarities of the wavelet time-frequency power profiles between identical and different odors:

Delta (2–4 Hz): 3.78 Hz.

Theta (4–8 Hz): 7.56 Hz.

Alpha (8–13 Hz): 10.7 Hz for the dominant oscillation and 12.29 Hz.

Low beta (13–20 Hz): 15.13 Hz.

High beta (20–30 Hz): 21.39 and 26.33 Hz.

Gamma (30–45 Hz): 30.25 and 34.75 Hz.

The cross-correlation was calculated as the sum of the products of the wavelet power for a pair comprising the target response ($|W_t(s,f_i)|$) and one of the other responses ($|W_n(s,f_i)|$) at the representative frequencies (f_i) for T1 [ms] ≤ s ≤ T2 [ms]. In a similar manner, the nine sums of the squared wavelet power for the target response were used to calculate the autocorrelation. Moreover, the logarithms of the ratios [$R_n(f_i)$] of the cross-correlations to the autocorrelations at the representative frequencies (f_i) were used to equalize the contributions of the increases and decreases in the response amplitude to the correlation analysis:

$$R_n(f_i) = \frac{\sum_{s=\mathrm{T1}}^{\mathrm{T2}} |W_n(s,f_i)||W_t(s,f_i)|}{\sum_{s=\mathrm{T1}}^{\mathrm{T2}} |W_t(s,f_i)||W_t(s,f_i)|} \tag{6}$$

A serially concatenated columnar array of all sets of the nine logarithmic ratios of the cross-correlations to the autocorrelations of the target response in the identical order of responses is a form of a wavelet cross-correlation profile (**Figure 4C**) [1]. The wavelet correlations were calculated as the correlation coefficients between these columnar arrays and employed to quantify the similarities of the odor-evoked oscillatory brain waves in the aPC.

Other mother wavelets such as Meyer and Mexican hat were considered to be inadequate for application to the odor-evoked oscillatory brain waves because their shapes appeared more dissimilar to any FFT components of the oscillatory brain waves than that of the Morlet (**Figure 3**). To date, except for one case [1], there are no published results of quantifying the similarities between oscillatory brain waves. Regarding the time-frequency power profiles, three reports were found. In one study, a discrete wavelet transform was used to identify and compare the timings of spike trains in an insect antennal lobe (corresponding to the mammal olfactory bulb) [7]. In another study, the Morlet wavelet transform was used to identify dominant oscillatory frequency bands and the synchrony between the oscillatory brain waves

http://dx.doi.org/10.5772/intechopen.74810

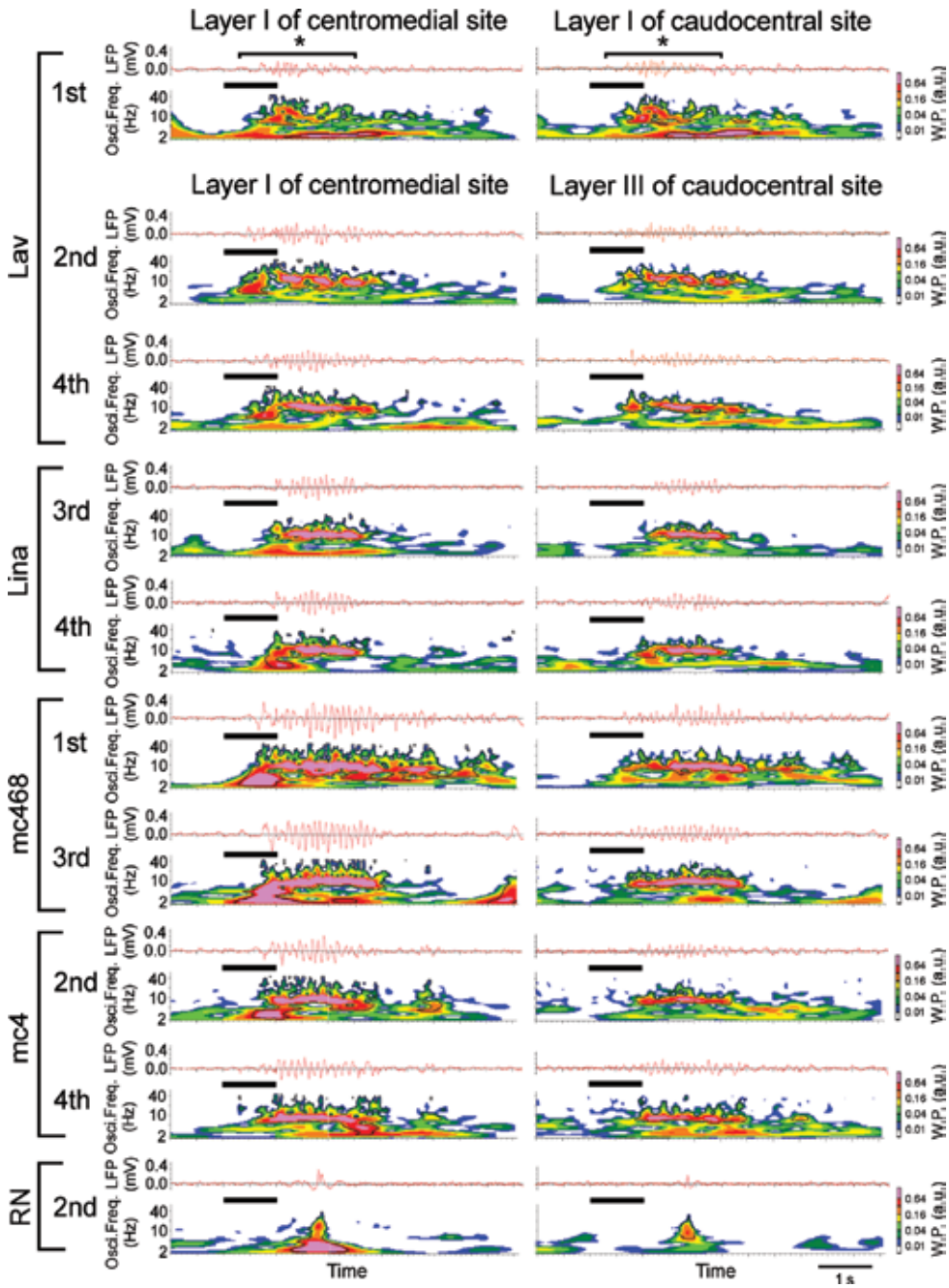


Figure 5. The wavelet profiles of odor-evoked oscillatory brain waves differed between the input and output layers of the aPC [1]. Of the 21 pairs of 1-s odor-evoked oscillatory brain waves (upper traces) that were simultaneously recorded in layers I (input) or III (output) of the aPC, 10 pairs are represented. In the wavelet time-frequency power profiles (lower traces) for the 2.2-s time window (marked by the asterisk), the ~10 Hz components remained prominent in layer III, whereas the <8 Hz components became less prominent compared to those in layer I. The in-stimulant presentation order is indicated. Statistically significant oscillatory powers were located within the black lines compared to those before presentation of odors ($P < 0.0001$, chi-squared test).

in different olfactory regions [8]. In the third study, the Hilbert transform was used to identify the dominant oscillations of the odor-evoked responses in the theta band in the posterior piriform cortex with phase-locked activities in the hippocampus in humans [9]. The Hilbert transform produced similar oscillation powers in a wide frequency range of 60–140 Hz, which is inconsistent with the decreased powers of the Morlet wavelet. Considering these results, we did not intend to analyze the odor-evoked oscillatory brain waves with the Meyer or Mexican hat mother wavelets or the Hilbert transform.

2.3. Wavelet correlation analysis of the time-frequency power profiles for revealing the stimulus dependency of odor-evoked oscillatory brain waves

The wavelet correlation analysis revealed that the olfactory information redundancy of a neural representation changes from experience (high redundancy) to a stimulus dependency (low redundancy) in the aPC [1]. The origins of the activities in layer I of the aPC are mainly the afferent fibers (input), association fibers, and postsynaptic inhibitory feedback input, whereas the activities in layer III primarily originate from the responses (output) of pyramidal cells, which are the principal neurons in the aPC and receive signals from multiple ORs. The wavelet profiles of identical odors resembled each other more than they resembled those of different odors in layers I (input signals) and III (output signals) of the aPC (**Figure 5**) [1]. In addition, the wavelet transformation visualized moderately clustered spot-like transient reductions in oscillatory power at frequencies just above 10 Hz in the odor-evoked oscillatory brain waves in layer I of the aPC (**Figure 5**). The most characteristic odor-dependent differences appeared in the initial phase of the wavelets for odor-evoked oscillatory brain waves in layer I of aPC. The mc468-evoked oscillatory brain wave was markedly greater especially at low frequencies in the initial phase than that of the Lav-evoked response [1].

The array data of the logarithmic ratios of the wavelet cross-/autocorrelations between 21 odor-evoked oscillatory brain waves differed slightly between layers I and III of the aPC (**Figure 6**) [1]. The lengths of the bars reflect the differences between a pair of oscillatory brain waves in such a way that the values of +1, 0, and −1 represent cross-correlations that are 10-fold, equal to, and one-tenth of the autocorrelation at the respective frequencies.

In layer III, the Lav odor pairs (broken yellow square in **Figure 7C**) showed homogeneously high correlations, except for the ninth Lav, whereas the identical Lav pairs in layer I resulted in more heterogeneous correlations (**Figure 7A**) [1]. In addition, the correlations between different single-component odors (Lina and mc4, in the broken blue squares in **Figure 7C**) decreased to <0.6 in layer III, whereas the corresponding correlations in layer I were mostly greater than 0.6 (**Figure 7A**) [1]. Notably, the heterogeneous correlations changed into an experience-dependent response similarity, which was observed for some of the odors in layer I of the aPC (a cluster of high correlations between the dashed lines in **Figure 7B** vs. **7A**) but was not clearly observed in layer III (**Figure 7D** vs. **7A**) as well as the 0–2 Hz components in layer I (**Figure 2D**) [1]. In layer III, the <8 Hz components decreased relative to those in layer I, with the prominent ~10 Hz oscillation remaining [1]. These results indicate a change in the neuronal information redundancy of transient and

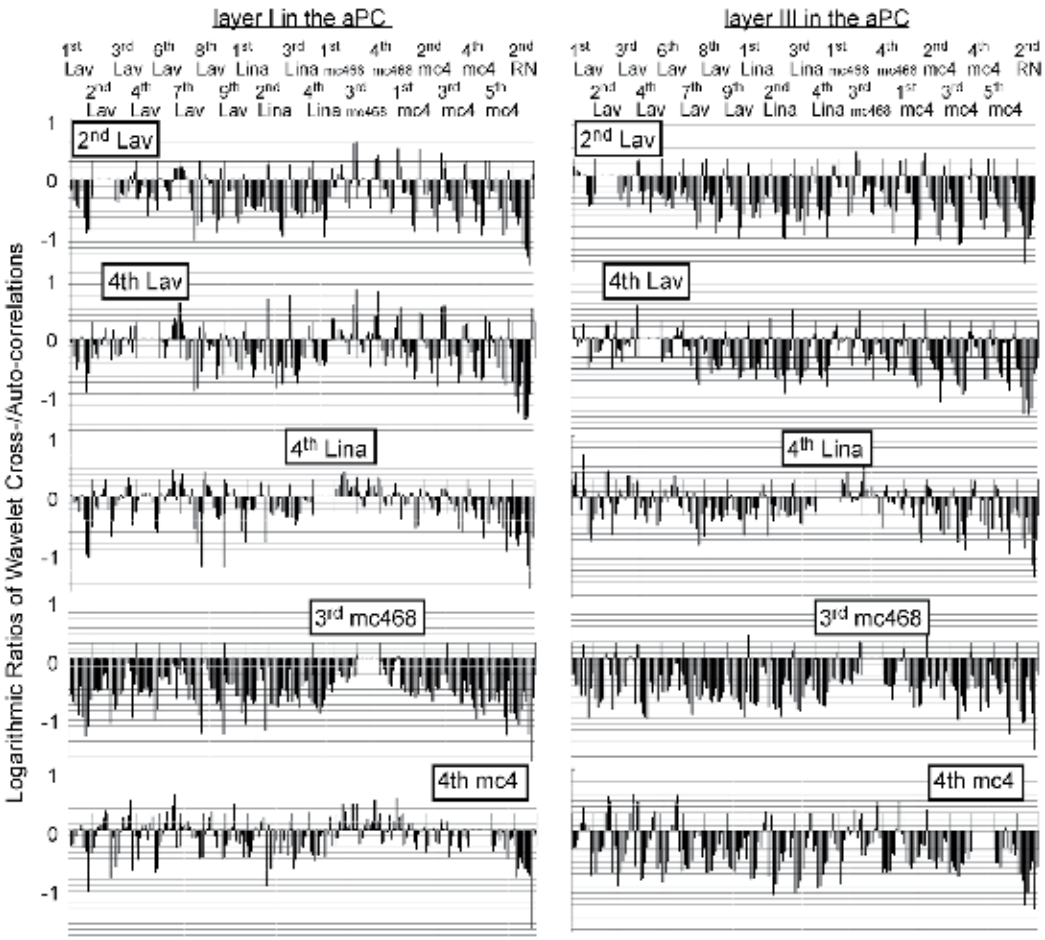


Figure 6. The wavelet cross-correlation profiles of odor-evoked oscillatory brain waves slightly differed between the input and output layers of the aPC [1]. The five pairs of logarithmic ratio arrays of the wavelet cross-/autocorrelations are exemplified. These ratio arrays suggest that the mc468-evoked responses markedly differed from those of Lav or Lina in each layer of the aPC and that they slightly differed between the input and output layers.

oscillatory brain waves from the dependencies on stimulus experience (high redundancy) to stimulus quality (low redundancy) between the input and output layers of the aPC. Recently, in the olfactory bulb that is upstream of the aPC in the olfactory pathway, stimulus history-dependent odor processing was observed [10]. This means that the wavelet correlation analysis had revealed a consistent experience dependency in input signals in the aPC from the olfactory bulb.

2.4. Effects of changes in oscillatory components on the wavelet correlation analysis

We evaluated the ability of the wavelet correlation analysis to detect changes in oscillatory powers at specific frequencies by 0.2-fold step modified wavelet powers at 1–8 frequency

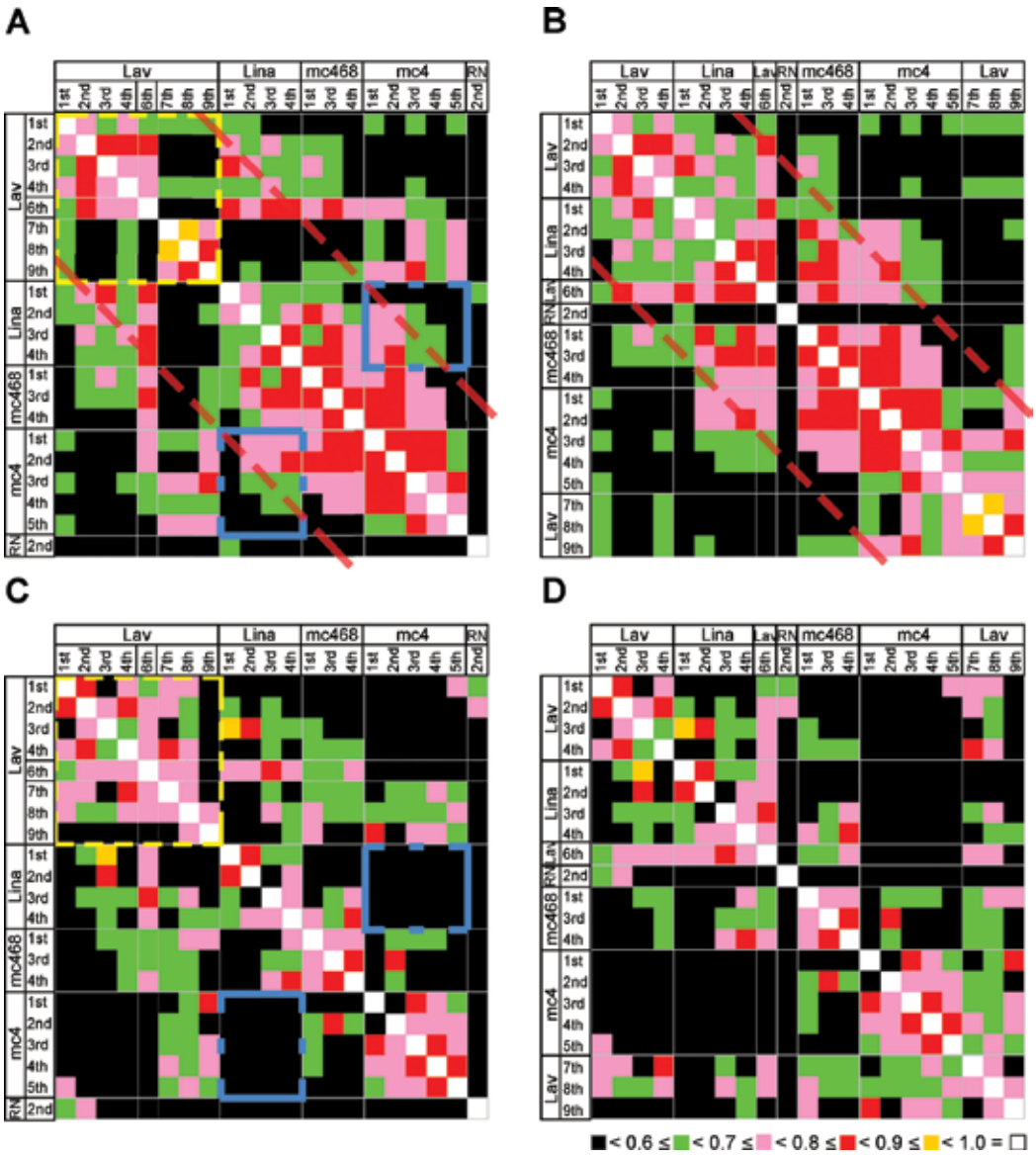


Figure 7. The wavelet correlation matrices of oscillatory brain waves differed between the input and output signals in the aPC [1]. (A) The wavelet correlation matrix of oscillatory brain waves in layer I (input) of the aPC. (B) The matrix in A rearranged in the entire presentation order. High correlations approached the diagonal line. (C) The wavelet correlation matrix of osci-LFPs in layer III (output) of the aPC. (D) The matrix in **C** rearranged in the entire presentation order. The colors representing power magnitudes are the same as in **Figure 2**.

bands (**Figure 8**) [1]. Greater decreases in correlations (0.4–0.7) were observed as a result of the 0.2-fold power modification at only 1–2 frequencies than those of eight frequencies (number/9 given in parentheses on the Y-axis). For 0.2-fold power amplification, the largest and smallest decreases were observed at 8–13 and 4–8 Hz, respectively. This analysis revealed that in the

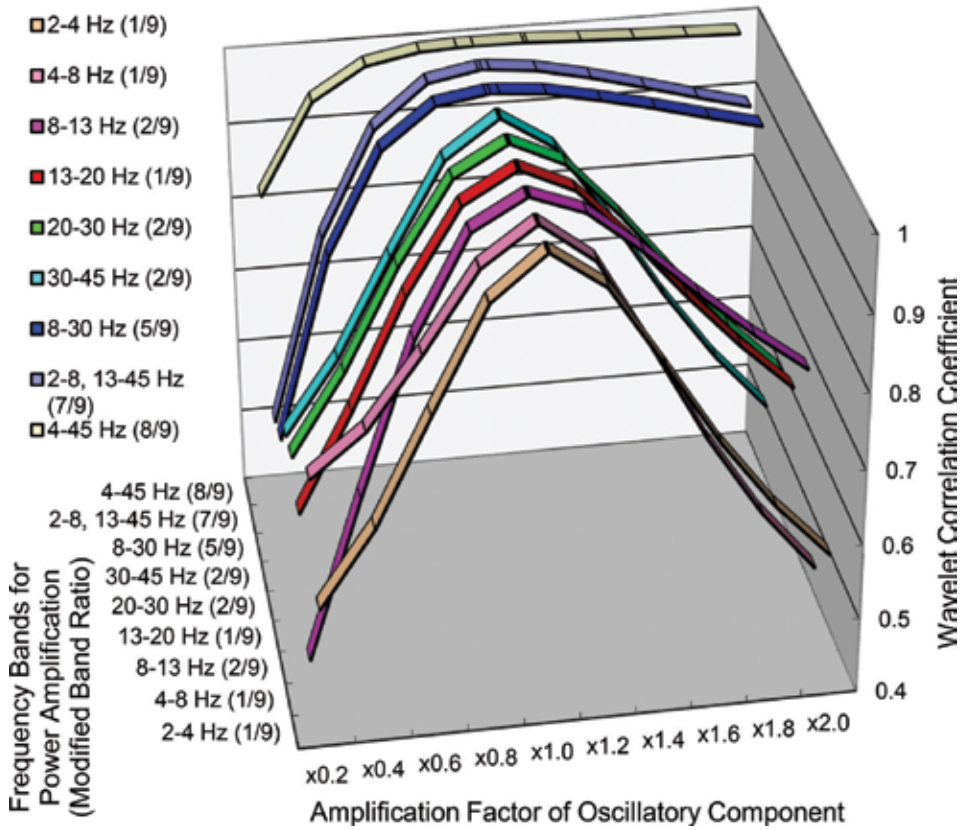


Figure 8. Sensitivity of the wavelet correlation analysis to changes in the oscillatory components [1]. A 0.2-fold power amplification resulted in the largest and smallest decreases in the wavelet correlations for 8–13 and 4–8 Hz, respectively. As the number of power-modified frequencies increased to more than four, changes in the wavelet correlations were reduced.

aPC, the 8–13 Hz component of the oscillatory brain waves contributes to the correlation coefficients more than the 4–8 Hz component. The wavelet correlation analysis enables the estimation of the relative contributions of oscillatory components to the similarities and differences between oscillatory brain waves.

3. Method for estimating in-brain information

3.1. Ranking of the correlation coefficients of several brain waves for identical information

Here, the odor-evoked brain waves were the same as those used in the previous section. To estimate the in-brain information, two standard brain waves, covering a wide range of variations for identical information, were selected. The criteria for selecting the two standard brain waves were as follows: (i) a brain wave with the highest pairwise correlation coefficient and a

high average of pairwise correlation coefficients in the given information for each individual and (ii) a brain wave with the second highest pairwise correlation coefficient and a differently ranked average of pairwise correlation coefficients in the given information for the same individual.

To select standard brain waves for the four odors, the correlation coefficients in the 2.2-s time window of interest were ranked between single-trial brain waves for all possible pairs of identical odors. Among the 28 pairs of brain waves for Lav, the highest correlation was obtained for the second Lav and fourth Lav pair that provided the fourth (median) and second highest averages of pairwise correlation coefficients, respectively (**Table 1**). The second highest correlation coefficient was obtained for the third and fifth Lav brain wave pair that provided the seventh and third highest averages of pairwise correlation coefficients, respectively. On the basis of the criteria, the fourth and third Lav brain waves were selected as the two standard brain waves for Lav information.

With regard to the pairwise correlation coefficients, their values for Lav pairs tended to be greater than those for mc4 pairs, and the values for Lina pairs tended to be greater than those for mc468 pairs. The lower correlation coefficients between identical odors suggest a greater across-trial variability in the time-frequency power profiles of single-trial brain waves, despite the tolerance of oscillatory phase differences. Similarly, the first and third Lina brain waves (**Table 2**), the fourth and first mc4 brain waves (**Table 3**), and the third and first mc468 brain waves (**Table 4**) were selected as standard brain waves for the respective information. These eight standard brain waves, as well as a control brain wave evoked by an odorless Ringer solution (second RN), were used as Set 1 of standard brain waves.

3.2. Estimates of the most probable information for single-trial brain waves using a pair of standard brain waves for each item of information

Using the wavelet correlation analysis, all possible pairwise correlation coefficients between a given single-trial brain wave and each standard brain wave (Set 1) were calculated. The first candidate was selected as the standard brain wave with the highest correlation coefficient to a target single-trial brain wave. The wavelet correlation analysis provided the first candidates for 12 single-trial brain waves with an accuracy of 75% (**Table 5**). An accuracy of 100% was achieved for Lina (2/2) and mc468 (1/1), whereas an accuracy of 67% was achieved for Lav (4/6) and mc4 (2/3). Notably, the single-trial brain waves tested were not any of the Set 1 standard brain waves. The accuracy of the first candidates was more than threefold higher than chance in five cases (20%). The probability of including the correct information for the two upper candidates was 92% (**Table 5**). However, the third candidates did not improve the probability of including the correct information for the three upper candidates (92%). In the estimates of information, candidates with correlation coefficients <0.6 were disregarded as nonspecific ones.

To compare the ideal set of standard brain waves (Set 1) with different sets of standard brain waves (standard Set 1-m) in terms of their accuracies for estimating information, wavelet correlation analyses were performed with partial replacements of standard brain waves. When one or three of the nine Set 1 standard brain waves were replaced with brain waves that did

	Ranking of wavelet correlations											Standard set									
Lav	**First Lav**	**Second Lav**	**Third Lav**	**Fourth Lav**	**Sixth Lav**	**Seventh Lav**	**Eighth Lav**	**Ninth Lav**	**Corr. coeff. Rank**	**Ave. corr. Coeff.**	**Ave. rank**	**Memo.**	**1**	**1-m1**	**1-m1p1**	**1-mp**	**2**	**2-m2p**	**s1**	**s1 m1**	**s2**
First Lav	1.00	0.63995	0.26	0.59	0.44	0.59	0.60	0.31	6	0.55	6										
Second Lav	0.64	1.00	0.60	0.73	0.60	0.52	0.47	0.11	1	0.59	4	median					○	○			
Third Lav	0.26	0.60	1.00	0.47	0.683	0.28	0.54	0.38	2	0.53	7	△	○	○	○						○
Fourth Lav	0.59	0.73	0.47	1.00	0.68	0.64	0.50	0.21	1	0.60	2	◎	○	○	○	○	○	○	○	○	
Sixth Lav	0.44	0.60	0.683	0.682	1.00	0.55	0.51	0.30	2	0.60	3										
Seventh Lav	0.59	0.52	0.28	0.63998	0.55	1.00	0.59	0.41	5	0.57	5										
Eighth Lav	0.60	0.47	0.54	0.50	0.51	0.59	1.00	0.66	4	0.61	1					○					
Ninth Lav	0.31	0.11	0.38	0.21	0.30	0.41	0.66	1.00	4	0.42	8										

Table 1. Pairwise wavelet correlations of single-trial brain waves for Lav in layer III of the aPC, their ranking, and various sets of standard brain waves.

http://dx.doi.org/10.5772/intechopen.74810

	Ranking of wavelet correlations								Standard set									
Lina	**First Lina**	**Second Lina**	**Third Lina**	**Fourth Lina**	**Corr. coeff. Rank**	**Ave. corr. Coeff.**	**Ave. rank**	**Memo.**	**1**	**1-m1**	**1-m1p1**	**1-mp**	**2**	**2-m2p**	**s1**	**s1 m1**	**s2**
First Lina	1.00	0.49	0.22	0.04	1	0.44	1	◎	○	○		○	○		○		
Second Lina	0.49	1.00	−0.13	0.18	1	0.38	3	Median			○	○	○	○			
Third Lina	0.22	−0.13	1.00	0.34	2	0.36	4	△	○								
fourth Lina	0.04	0.18	0.34	1.00	2	0.39	2			○	○			○		○	○

Table 2. Pairwise wavelet correlations of single-trial brain waves for Lina in layer III of the aPC, their ranking, and various sets of standard brain waves.

	Ranking of wavelet correlations									Standard set									
mc4	**First mc4**	**Second mc4**	**Third mc4**	**Fourth mc4**	**Fifth mc4**	**Corr. coeff. Rank**	**Ave. corr. Coeff.**	**Ave. rank**	**Memo.**	**1**	**1-m1**	**1-m1p1**	**1-mp**	**2**	**2-m2p**	**s1**	**s1 m1**	**s2**
First mc4	1.00	0.04	0.467	0.35	0.15	2	0.40	5	△	○		○			○			
Second mc4	0.04	1.00	0.25	0.366	0.368	4	0.40	4										
Third mc4	0.467	0.25	1.00	0.46	0.18	2	0.47	2					○					
Fourth mc4	0.35	0.37	0.46	1.00	0.472	1	0.53	1	◎	○	○	○	○	○	○	○	○	○
Fifth mc4	0.15	0.37	0.18	0.472	1.00	1	0.43	3	Median		○			○				

Table 3. Pairwise wavelet correlations of single-trial brain waves for mc4 in layer III of the aPC, their ranking, and various sets of standard brain waves.

	Ranking of wavelet correlations							Standard set									
mc468	**First mc468**	**Third mc468**	**Fourth mc468**	**Corr. coeff. Rank**	**Ave. corr. Coeff.**	**Ave. rank**	**Memo.**	**1**	**1-m1**	**1-m1p1**	**1-mp**	**2**	**2-m2p**	**s1**	**s1 m1**	**s2**
First mc468	1.00	0.14	0.05	2	0.39	3	△	○	○							
Third mc468	0.14	1.00	0.23	1	0.46	1	◎	○	○	○	○	○	○	○	○	○
Fourth mc468	0.05	0.23	1.00	1	0.43	2	Median			○	○	○	○			

Table 4. Pairwise wavelet correlations of single-trial brain waves for mc468 in layer III of the aPC and various sets of standard brain waves.

Standard brain waves	Third Lav	Fourth Lav	First Lina	Third Lina	First mc468	Third mc468	First mc4	Fourth mc4	Second RN	Highest corr.		
Third Lav	1.00	0.67	0.85	0.66	0.63	0.50	0.44	0.36	0.45	Lina		
Fourth Lav	0.67	1.00	0.58	0.61	0.70	0.54	0.28	0.54	0.41	mc468		
First Lina	0.85	0.58	1.00	0.60	0.58	0.40	0.41	0.30	0.46	Lav		
Third Lina	0.66	0.61	0.60	1.00	0.63	0.72	0.54	0.58	0.41	mc468		
First mc468	0.63	0.70	0.58	0.63	1.00	0.728	0.63	0.723	0.40	mc468		
Third mc468	0.50	0.54	0.40	0.72	0.73	1.00	0.50	0.63	0.52	mc468		
First mc4	0.44	0.28	0.41	0.54	0.63	0.50	1.00	0.73	0.35	mc4		
Fourth mc4	0.36	0.54	0.30	0.58	0.72	0.63	0.73	1.00	0.37	mc4		
Second RN	0.45	0.41	0.46	0.41	0.40	0.52	0.35	0.37	1.00	—		
Single-trial brain waves										**Estimated information**	**Second candidate (>0.6)**	**Third candidate (>0.6)**
First Lav	0.56	0.77	0.47	0.48	0.53	0.47	0.39	0.59	0.49	Lav	—	—
Second Lav	0.69	0.82	0.62	0.56	0.51	0.43	0.26	0.41	0.58	Lav	Lav	—
Sixth Lav	0.774	0.78	0.766	0.79	0.69	0.51	0.40	0.46	0.39	Lina	Lav	Lav
Seventh Lav	0.53	0.79	0.42	0.65	0.71	0.63	0.50	0.75	0.51	Lav	mc4	mc468
Eighth Lav	0.641	0.693	0.52	0.689	0.68	0.56	0.63	0.63	0.33	Lav	Lina	mc468
Ninth Lav	0.51	0.43	0.46	0.61	0.73	0.51	0.74	0.72	0.20	mc4	mc468	mc4
Second Lina	0.71	0.44	0.79	0.53	0.56	0.29	0.48	0.29	0.28	Lina	Lav	—
Fourth Lina	0.652	0.56	0.654	0.79	0.71	0.63	0.57	0.47	0.24	Lina	mc468	Lina
Fourth mc468	0.58	0.56	0.54	0.775	0.777	0.86	0.60	0.63	0.33	mc468	mc468	Lina
Second mc4	0.36	0.44	0.23	0.60	0.68	0.84	0.58	0.80	0.51	mc468	mc4	mc468
Third mc4	0.35	0.35	0.25	0.55	0.66	0.61	0.81	0.85	0.34	mc4	mc4	mc468
Fifth mc4	0.36	0.45	0.25	0.57	0.54	0.55	0.68	0.81	0.41	mc4	mc4	—
								Correct rate		75%	92%	92%

Table 5. Estimated information of single-trial brain waves in layer III of the aPC by ranking of wavelet correlations using two standard brain waves (set 1).

not meet the criteria, there were no changes in the 75% accuracy for the first candidates, and a 92% probability of including the correct information for the two upper candidates was observed. Nevertheless, there were some exchanges between correct and incorrect estimates for identical information (data not shown).

In contrast, by using the pair of brain waves with the highest pairwise correlation coefficients as the two standard brain waves for each odor (standard Set 2), the accuracies of estimation were reduced by 100% for Lina (2/2 → 0/2) and 34% for Lav (4/6 → 2/6), but no change occurred for mc468 (1/1) and mc4 (2/3) (**Table 6**). This standard Set 2 provided a total accuracy of 42% (33% reduction) and a 75% probability (17% reduction) of including the correct information for the two upper candidates (**Figure 9**). By replacing two of the nine Set 2 standard brain waves with one that did not meet the criteria, the accuracy for the first candidates increased by 25% and the 92% probability of including the correct information for the two upper candidates was recovered (**Figure 9**). Therefore, the proposed criteria of selecting standard brain waves with a wide variation are likely appropriate and achieve better estimation than the selection of those with a narrow range (the most similar brain wave pairs).

3.3. Estimates of the most probable information for single-trial brain waves with a standard brain wave for each item of information

By using a set of single standard brain waves for four odors that met only the first criterion (standard Set s1), a similar accuracy of estimated information and probability of including the correct information for the two upper candidates was obtained for the 12 target brain waves (data not shown). The Set s1 standard brain waves were composed of the fourth Lav, first Lina, third mc468, fourth mc4, and second RN. Among the 16 target brain waves, the accuracy and probability slightly decreased by 6 and 4%, respectively, compared to those of the 12 target brain waves (data not shown). When one or two of the five Set-s1 standard brain waves were replaced with those that did not meet the criteria, the accuracy was reduced to 67 or 42%, respectively (data not shown). The probability of including the correct information for the two upper candidates was also reduced by 9 and 25%, respectively. For the 16 target brain waves, the accuracy and probability showed almost no changes when one of the five Set s1 standard brain waves was replaced, whereas the accuracy and probability for the estimated information were reduced by 13% when two of the Set s1 standard brain waves were replaced (data not shown).

3.4. Single-trial brain waves composed of redundant signals in the olfactory pathway exhibiting a similar accuracy and probability for estimated information

It is interesting to examine the accuracy of the wavelet correlation analysis for predicting the in-brain information of single-trial brain waves comprising redundant signals in layer I of the aPC. By using a set of standard brain waves that meet the proposed criteria for the redundant brain waves recorded in layer I (standard Set 1r), the wavelet correlation analysis provided a similar accuracy (75%) of estimated information and probability (100%) of including the correct information for the two upper candidates (**Table 7**) compared to the results observed for the brain waves recorded in layer III (**Table 5**). In contrast, by using the pairs of brain waves

Standard brain waves	Second Lav	Fourth Lav	First Lina	Second Lina	Third mc468	Fourth mc468	Fourth mc4	Fifth mc4	Second RN	Highest corr.		
Second Lav	1.00	0.80	0.57	0.47	0.40	0.37	0.35	0.39	0.55	Lav		
Fourth Lav	0.80	1.00	0.47	0.37	0.50	0.54	0.50	0.46	0.35	Lav		
First Lina	0.57	0.47	1.00	0.80	0.38	0.56	0.28	0.24	0.42	Lina		
Second Lina	0.47	0.37	0.80	1.00	0.28	0.49	0.29	0.29	0.27	Lina		
Third mc468	0.40	0.50	0.38	0.28	1.00	0.85	0.54	0.55	0.48	mc468		
Fourth mc468	0.37	0.54	0.56	0.49	0.85	1.00	0.58	0.53	0.29	mc468		
Fourth mc4	0.35	0.50	0.28	0.29	0.54	0.58	1.00	0.83	0.27	mc4		
Fifth mc4	0.39	0.46	0.24	0.29	0.55	0.53	0.83	1.00	0.38	mc4		
Second RN	0.55	0.35	0.42	0.27	0.48	0.29	0.27	0.38	1.00	—		
Single-trial brain waves										**Estimated information**	**Second candidate (>0.6)**	**Third candidate (>0.6)**
First Lav	0.78	0.76	0.42	0.36	0.44	0.41	0.58	0.70	0.43	Lav	Lav	mc4
Third Lav	0.61	0.58	0.87	0.76	0.49	0.62	0.35	0.36	0.41	Lina	Lina	mc468
Sixth Lav	0.65	0.73	0.70	0.74	0.56	0.71	0.41	0.52	0.32	Lina	Lav	mc468
Seventh Lav	0.62	0.74	0.36	0.40	0.57	0.63	0.72	0.65	0.41	Lav	mc4	mc4
Eighth Lav	0.61	0.68	0.47	0.52	0.57	0.63	0.65	0.72	0.33	mc4	Lav	mc4
Ninth Lav	0.27	0.40	0.42	0.42	0.47	0.60	0.75	0.70	0.18	mc4	mc4	mc468
Third Lina	0.52	0.57	0.55	0.54	0.69	0.77	0.54	0.60	0.40	mc468	mc468	mc4
Fourth Lina	0.34	0.50	0.62	0.70	0.60	0.84	0.44	0.40	0.18	mc468	Lina	Lina
First mc468	0.50	0.66	0.56	0.55	0.71	0.79	0.66	0.55	0.35	mc468	mc468	mc4
First mc4	0.17	0.22	0.37	0.45	0.45	0.57	0.75	0.68	0.27	mc4	mc4	—
Second mc4	0.26	0.42	0.21	0.15	0.83	0.70	0.71	0.72	0.48	mc468	mc4	mc4
Third mc4	0.19	0.32	0.24	0.28	0.60	0.63	0.85	0.77	0.28	mc4	mc4	mc468
								Correct rate		42%	75%	75%

Table 6. Estimated information of single-trial brain waves in layer III of the aPC by ranking of wavelet correlations using two standard brain waves with the highest pairwise correlation coefficients (set 2).

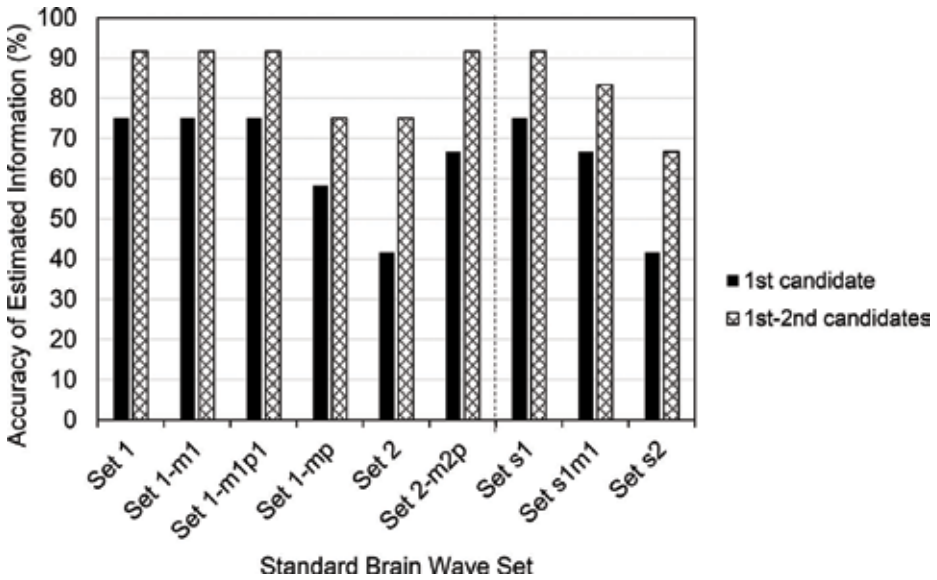


Figure 9. Variation-dependent changes in the accuracy of estimated information of single-trial brain waves in layer III of the aPC.

corresponding to the Set 1 of layer III (standard Set 1′ in layer I), the accuracy of estimation was reduced by 17%, and the probability of including the correct information for the two upper candidates was reduced by 25% (to 75%) (data not shown). By using single standard brain waves (standard Set s1r), the accuracy and probability were slightly reduced compared to those of the standard Set s1 (data not shown).

Finally, it was examined whether the combination of data for two recording sites (layers I and III) affected the accuracy for the first candidates. Using this method, the accuracy (75%) of estimated information was maintained but not improved in standard Set 1 + 1′ and Set 1r + 1r′ (data not shown).

3.5. A new method of real-time estimation of in-brain information of single-trial brain waves

A new method is proposed for estimating the information of single-trial brain waves in fine temporal structures with a cross-trial variability by using a set of standard brain waves in a given category for each individual. In the oscillatory brain waves recorded in layer III or I of the aPC of the isolated whole brain of a guinea pig, the wavelet correlation analysis provided a 75% accuracy for the first candidate and a > 92% probability of including the correct information for the two upper candidates (**Tables 5** and **7**). The results support the validity of the proposed criteria for selecting standard brain waves with a wide variation for estimating different information in a given category.

The accuracy of this method was not affected by the information redundancy of signal sources, such as those resulting from olfactory receptors with overlapping tuning specificities and an

Standard brain waves	Second Lav	Eighth Lav	Second Lina	Third Lina	First mc468	Fourth mc468	First mc4	Fifth mc4	Second RN	Highest corr.		
Second Lav	1.00	0.56	0.64	0.73	0.63	0.56	0.59	0.42	0.47	Lina		
Eighth Lav	0.56	1.00	0.47	0.50	0.57	0.63	0.66	0.76	0.09	mc4		
Second Lina	0.64	0.47	1.00	0.67	0.85	0.68	0.76	0.49	0.48	mc468		
Third Lina	0.73	0.50	0.67	1.00	0.69	0.75	0.74	0.48	0.25	mc468		
First mc468	0.63	0.57	0.85	0.69	1.00	0.79	0.81	0.54	0.35	Lav		
Fourth mc468	0.56	0.63	0.68	0.75	0.79	1.00	0.87	0.58	0.23	mc4		
First mc4	0.59	0.66	0.76	0.74	0.81	0.87	1.00	0.69	0.30	mc468		
Fifth mc4	0.42	0.76	0.49	0.48	0.54	0.58	0.69	1.00	0.20	Lav		
Second RN	0.47	0.09	0.48	0.25	0.35	0.23	0.30	0.20	1.00	—		
Single-trial brain waves										**Estimated information**	**Second candidate (>0.6)**	**Third candidate (>0.6)**
First Lav	0.77	0.70	0.66	0.52	0.59	0.55	0.63	0.70	0.43	Lav	Lav	mc4
Third Lav	0.79	0.43	0.74	0.60	0.72	0.49	0.51	0.38	0.48	Lav	Lina	mc468
Fourth Lav	0.83	0.68	0.58	0.70	0.66	0.58	0.61	0.53	0.20	Lav	Lina	Lav
Sixth Lav	0.81	0.52	0.79	0.85	0.78	0.73	0.75	0.52	0.37	Lina	Lav	mc468
Seventh Lav	0.56	0.91	0.45	0.47	0.53	0.59	0.64	0.76	0.06	Lav	mc4	mc4
Ninth Lav	0.56	0.83	0.62	0.48	0.68	0.67	0.72	0.76	0.28	Lav	mc4	mc468
First Lina	0.73	0.44	0.78	0.66	0.69	0.55	0.54	0.38	0.63	Lina	Lav	mc468
Fourth Lina	0.59	0.40	0.7758	0.84	0.85	0.75	0.7756	0.42	0.27	mc468	Lina	Lina
Third mc468	0.64	0.55	0.77	0.837	0.83	0.89	0.839	0.51	0.36	mc468	mc4	Lina
Second mc4	0.52	0.62	0.69	0.77	0.84	0.891	0.887	0.64	0.27	mc468	mc4	mc468
Third mc4	0.53	0.779	0.71	0.64	0.775	0.778	0.84	0.80	0.18	mc4	mc4	Lav
Fourth mc4	0.59	0.65	0.57	0.67	0.65	0.79	0.85	0.70	0.21	mc4	mc468	mc4
								Correct rate		75%	100%	100%

Table 7. Estimated information of single-trial brain waves in layer I of the aPC by ranking of wavelet correlations using two standard brain waves (set 1r).

Information	Recoding sites	Estimated information			
		Lav	Lina	mc468	mc4
Lav	Layer I (input)	62.8%	20.9% (e)	9.3% (e)	7.0% (e)
	Layer III (output)	57.9%	18.4% (e)	1.3% (e)	22.4% (e)
Lina	Layer I (input)	0.0% (e)	53.3%	46.7% (e)	0.0% (e)
	Layer III (output)	7.7% (e)	73.1%	19.2% (e)	0.0% (e)
mc468	Layer I (input)	0.0% (e)	25.0% (e)	75.0%	0.0% (e)
	Layer III (output)	0.0% (e)	26.7% (e)	73.3%	0.0% (e)
mc4	Layer I (input)	4.5% (e)	0.0% (e)	13.6% (e)	81.8%
	Layer III (output)	0.0% (e)	0.0% (e)	30.8% (e)	69.2%

Table 8. Correct and error rates (e) of estimated information in single-trial brain waves recorded in layers I and III of the aPC by the wavelet correlation analysis.

experience dependency in layer I or from pyramidal cells with a stimulus dependency after the integration of signals from multiple cognate olfactory receptors in layer III (**Table 8**). Layer I brain waves comprising redundant signals exhibited a similar accuracy of estimated information and a slightly increased probability of including the correct information for the two upper candidates compared to layer III brain waves.

The redundancies of brain waves are attributable to two origins: information and signaling. In the olfactory system, the information redundancy changes through the signal pathway from the receptors to the higher cortical areas via signal integration in the third- or higher-order neurons and/or mutual inhibition [1, 11–13] for category [14] or elemental odor representation [15]. Unlike the >80% overlap of about 70 receptors for carvone enantiomers having similar odors [16], the quite different odors of Lav and mc468 evoked different amplitude receptor potentials in the olfactory epithelium and dissimilar brain waves in the anterior piriform cortex [1]. Nevertheless, the wavelet correlation analysis sometimes produced the highest correlation coefficients of Lav for mc468. The error rate of Lav for mc468 was 9.3% in layer I brain waves but was reduced to 1.7% in layer III brain waves (**Table 8** and **Figure 10**), which is consistent with the change in the information redundancy from high to low stages between layers I and III. On the other hand, the error rate of mc468 for Lav was 0% in both layers I and III. For the single-compound odors, Lina and mc4 exhibited odor similarity-dependent changes in the error rates of the estimated information between layers I and III. The error rates of the single compounds for their original mixture odors (partially similar odor) increased between layers I and III (0 → 7.7% in Lina and 13.6 → 30.8% in mc4) and those of single compounds for their nonrelative mixture odors (dissimilar odor) decreased between layers I and III (46.7 → 19.2% in Lina and 4.5 → 0% in mc4). Notably, the error rates between these single compounds were 0% in both layers I and III. These results suggest a partial overlap of the elemental odors that are represented in the pyramidal cells in the aPC and are recorded in layer III as brain waves. The total error rates of Lina decreased in layer III compared to those of layer I (and vice versa for the correct rate), whereas those of mc4 increased.

http://dx.doi.org/10.5772/intechopen.74810

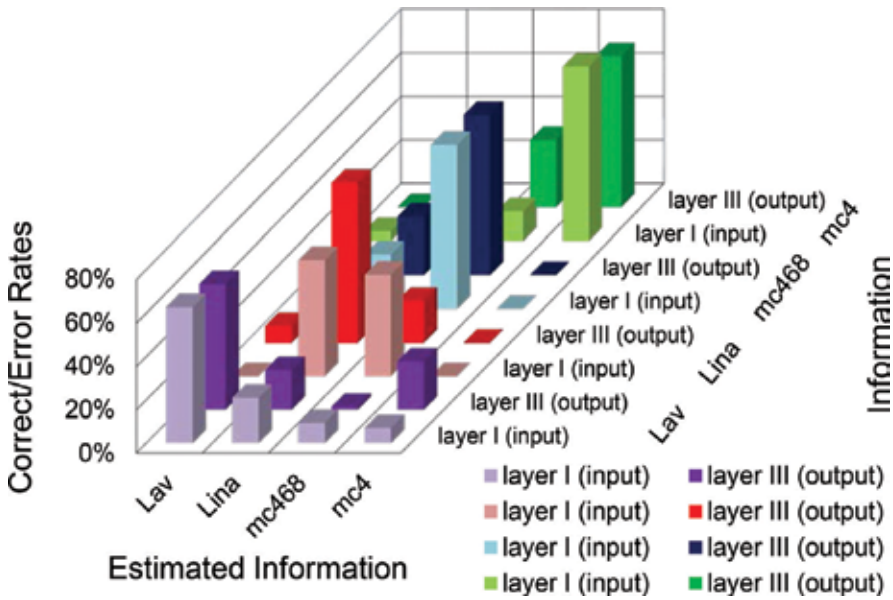


Figure 10. Correct and error rates of estimated information in single-trial brain waves recorded in layers I and III of the aPC by the wavelet correlation analysis. These values are listed in **Table 8**.

The signaling redundancy originates from an identical temporal profile of different subsets of neurons tuned to distinct or shared information or from identical temporal profiles that are composed of multiple different profiles of various different subsets of neurons tuned to multiple distinct or shared information. The constant error rates of mc468 for Lina between layers I and III (both ~25%, **Table 8** and **Figure 10**) are likely attributable to the signaling redundancy rather than the information similarity or information redundancy. Moreover, in the increased case, there was a threefold higher error rate of Lav for mc4 in layer III than layer I, whereas the error rates of Lav for Lina were almost constant between layers I and III.

3.6. Applicable examples of estimated in-brain information in humans using the wavelet correlation analysis

Each brain system (e.g., a sensory, memory, decision, or motor system) is organized in a hierarchical manner from simple to complicated matters. The sensory system generates oscillatory activities between the related cortical regions and the thalamus, and the latter acts (except in the olfactory system) to gate the sensory input to the cortex and provides feedback from the cortical pyramidal neurons. In olfaction, transient oscillatory brain waves are observed in the aPC [5, 17–21]. Strong feed-forward inhibition [5, 22, 23] via the sensitive pathway from the olfactory bulb [24] and the other sensory thalamocortical circuit [25, 26] or higher olfactory centers [27] could induce oscillatory brain waves that would contribute to parts of the EEGs recorded at the respective positions on the human scalp, in analogy to these experimental animals. Such information-dependent temporal profiles of the EEGs may enable us to estimate in-brain information by comparison with a set of standard time-frequency

power profiles of EEGs in each individual. To this aim, a wavelet correlation analysis of the brain waves in a guinea pig was conducted using standard brain waves with the proposed criteria and achieved an accuracy of 75% for the first candidates. This accuracy is attributable to the comparisons with standard single-trial responses in the wavelet time-frequency power profiles.

Conventional methods have focused only on some parts of the brain wave characteristics. For example, the FFT power spectra of sensorimotor EEGs [28, 29] or auditory EEGs [30] in specific frequency bands at a specific recording position were analyzed for the development of brain-computer interfaces. The Morlet wavelet convolutions for four-frequency band powers of the single-trial EEGs were analyzed to understand the cognitive control system via a priori estimation of information across three tasks [31]. By using the wavelet correlation analysis in the time-frequency power profiles at nine frequencies, these analyses could be improved in their subprocesses. Odor sensation [32, 33] and color-opponent responses [34] were also recorded in humans at Fz and an intermediate position between Oz and the inion, respectively, and they demonstrated informational differences in response amplitudes or profiles. Like EEGs in object recognition and those responsible for mental states, these EEGs are also subjects for the application of the wavelet correlation analysis for estimating in-brain fine information. Pain-related alpha-band desynchronization at contralateral-central electrodes (C2, C4, CP2, and CP4) and gamma-band synchronization at the ipsilateral-posterior electrodes (P3, P5, and so on) [35] are also good candidates for application. In animal models, the neural pathways of innate and learned fear responses have been revealed [36], and different pathways of stress relaxation using rose and hinokitiol odors were found [37, 38]. Therefore, determining their differing time-frequency power profiles would enable us to estimate the strengths of stress or relaxation in EEGs in humans. Future studies will focus on programming the wavelet correlation analysis for real-time estimates of in-brain information in humans.

4. Conclusions

We developed a new method for a similarity analysis and real-time estimates of in-brain information in single-trial brain waves by ranking the correlation coefficients in the wavelet correlation analysis. The wavelet correlation analysis with a set of standard brain waves provided the first candidate of estimated information with an accuracy of 75% with a > 92% probability of including the correct information for the two upper candidates, regardless of the information redundancy of signal sources. This method may be also useful for its applications to brain-machine interfaces or medical/research tools.

Acknowledgements

We would like to thank Dr. Mutsumi Matsukawa for his contributions to the development of the isolated whole-brain experimental system that enabled the recordings of odor-induced and

nonolfactory origin-free brain waves. We are also grateful to Kiyo Murano for writing the computer software for wavelet transformation. This work was supported by grants (T.S.) from METI, Japan, and Grant-in-Aids for Scientific Research (B) #15H02730 (T.S.) from the MEXT, Japan.

Abbreviations

aPC	anterior piriform cortex
aPCvr	ventro-rostral region of the aPC
EEG	electroencephalography
EOG	electro-olfactogram
FFT	fast Fourier transform
LFP	local field potential
LOT	lateral olfactory tract
OR	olfactory receptor
osci-LFP	oscillatory local field potential

Author details

Takaaki Sato[1]*, Riichi Kajiwara[2], Ichiro Takashima[3] and Toshio Iijima[4]

*Address all correspondence to: taka-sato@aist.go.jp

1 Biomedical Research Institute, National Institute of Advanced Industrial Science and Technology (AIST), Ikeda, Japan

2 School of Science and Technology, Meiji University, Kawasaki, Japan

3 Human Technology Research Institute, AIST, Tsukuba, Japan

4 Graduate School of Life Sciences, Tohoku University, Sendai, Japan

References

[1] Sato T, Kajiwara R, Takashima I, Iijima T. A novel method for quantifying similarities between oscillatory neural responses in wavelet time-frequency power profiles. Brain Research. 2016;**1636**:107-117. DOI: 10.1016/j.brainres.2016.01.054

[2] Myrden A, Chau T. A passive EEG-BCI for single-trial detection of changes in mental state. IEEE Transactions on Neural Systems and Rehabilitation Engineering. 2017;**25**:345-356

[3] Doesburg SM, Bedo N, Ward LM. Top-down alpha oscillatory network interactions during visuospatial attention orienting. NeuroImage. 2016;**132**:512-519. DOI: 10.1016/j.neuroimage.2016.02.076

[4] Wilsch A, Obleser J. What works in auditory working memory? A neural oscillations perspective. Brain Research. 2016;**1640**:193-207. DOI: 10.1016/j.brainres.2015.10.054

[5] Ishikawa T, Sato T, Shimizu A, de Curtis M, Kakei T, Iijima T. Odour-driven activity in the olfactory cortex of an in vitro isolated guinea-pig whole brain with olfactory epithelium. Journal of Neurophysiology. 2007;**97**:670-679

[6] Torrence C, Compo GP. A practical guide to wavelet analysis. Bulletin of the American Meteorological Society. 1998;**79**(61-78)

[7] Capurro A, Baroni F, Kuebler LS, Kárpáti Z, Dekker T, Lei H, Hansson BS, Pearce TC, Olsson SB. Temporal features of spike trains in the moth antennal lobe revealed by a comparative time-frequency analysis. PLoS One. 2014;**9**:e84037. DOI: 10.1371/journal.pone.0084037

[8] Pardo-Bellver C, Martínez-Bellver S, Martínez-García F, Lanuza E, Teruel-Martí V. Synchronized activity in the main and accessory olfactory bulbs and vomeronasal amygdala elicited by chemical signals in freely behaving mice. Scientific Reports. 2017;**7**:9924. DOI: 10.1038/s41598-017-10089-4

[9] Jiang H, Schuele S, Rosenow J, Zelano C, Parvizi J, Tao JX, Wu S, Gottfried JA. Theta oscillations rapidly convey odor-specific content in human piriform cortex. Neuron. 2017;**94**:207-219

[10] Vinograd A, Livneh Y, Mizrahi A. History-dependent odor processing in the mouse olfactory bulb. The Journal of Neuroscience. 2017;**37**:12018-12030

[11] Desmaisons D, Vincent JD, Lledo P-M. Control of action potential timing by intrinsic subthreshold oscillations in olfactory bulb output neurons. The Journal of Neuroscience. 1999;**19**:10727-10737

[12] Kashiwadani H, Sasaki YF, Uchida N, Mori K. Synchronized oscillatory discharges of mitral/tufted cells with different molecular receptive ranges in the rabbit olfactory bulb. Journal of Neurophysiology. 1999;**82**:1786-1792

[13] Stettler DD, Axel R. Representations of odor in the piriform cortex. Neuron. 2009;**63**:854-864. DOI: 10.1016/j.neuron.2009.09.005

[14] Yoshida I, Mori K. Odorant category profile selectivity of olfactory cortex. The Journal of Neuroscience. 2007;**27**:9105-9114

[15] Sato T, Kawasaki T, Mine S, Matsumura H. Functional role of the C-terminal amphipathic helix 8 of olfactory receptors and other G protein-coupled receptors. International Journal of Molecular Sciences. 2016;**17**(pii):E1930. DOI: 10.3390/ijms17111930

[16] Hamana H, Hirono J, Kizumi M, Sato T. Sensitivity-dependent hierarchical receptor codes for odours. Chemical Senses. 2003;**28**:87-104

[17] Bressler SL, Freeman WG. Frequency analysis of olfactory system EEG cat, rabbit, and rat. Electroencephalography and Clinical Neurophysiology. 1980;**50**:19-24

[18] Ketchum KL, Haberly LB. Synaptic events that generate fast oscillations in piriform cortex. The Journal of Neuroscience. 1993;**13**:3980-3985

[19] de Curtis, M, Biella, G, Forti, M, Panzica, F. Multifocal spontaneous epileptic activity induced by restricted bicuculline ejection in the piriform cortex of the isolated guinea pig brain. Journal of Neurophysiology 1994;**71**:2463–2476

[20] Chapman CA, Xu Y, Haykin S, Racine RJ. Beta-frequency (15-35 Hz) electroencephalogram activities elicited by toluene and electrical stimulation in the behaving rat. Neuroscience. 1998;**86**:1307-1319

[21] Chabaud P, Ravel N, Wilson DA, Mouly AM, Vigouroux M, Farget V, Gervais R. Exposure to behaviourally relevant odour reveals differential characteristics in rat central olfactory pathways as studied through oscillatory activities. Chemical Senses. 2000;**25**:561-573

[22] Sato T, Hirono J, Hamana H, Ishikawa T, Shimizu A, Takashima I, Kajiwara R, Iijima T. Architecture of odour information processing in the olfactory system. Anatomical Science International. 2008;**83**:195-206. DOI: 10.1111/j.1447-073X.2007.00215.x

[23] Sato T, Kobayakawa R, Kobayakawa K, Emura M, Itohara S, Kizumi M, Hamana H, Tsuboi A, Hirono J. Supersensitive detection and discrimination of enantiomers by dorsal olfactory receptors: Evidence for hierarchical odour coding. Scientific Reports. 2015;**5**: 14073. DOI: 10.1038/srep14073

[24] Igarashi KM, Ieki N, An M, Yamaguchi Y, Nagayama S, Kobayakawa K, Kobayakawa R, Tanifuji M, Sakano H, Chen WR, Mori K. Parallel mitral and tufted cell pathways route distinct odor information to different targets in the olfactory cortex. The Journal of Neuroscience. 2012;**32**:7970-7985. DOI: 10.1523/JNEUROSCI.0154-12.2012

[25] Bruno RM. Synchrony in sensation. Current Opinion in Neurobiology. 2011;**21**:701-708. DOI: 10.1016/j.conb.2011.06.003

[26] Kajiwara R, Tominaga T, Takashima I. Olfactory information converges in the amygdaloid cortex via the piriform and entorhinal cortices: Observations in the guinea pig isolated whole-brain preparation. The European Journal of Neuroscience. 2007;**25**:3648-3658. DOI: 10.1111/j.1460-9568.2007.05610.x

[27] Schoenbaum G, Eichenbaum H. Information coding in the rodent prefrontal cortex. I. Single-neuron activity in orbitofrontal cortex compared with that in pyriform cortex. Journal of Neurophysiology. 1995;**74**:733-750

[28] Cincotti F, Mattia D, Aloise F, Bufalari S, Astolfi L, De Vico Fallani F, Tocci A, Bianchi L, Marciani MG, Gao S, Millan J, Babiloni F. High-resolution EEG techniques for brain-computer interface applications. Journal of Neuroscience Methods. 2008;**167**:31-42

[29] Cincotti F, Mattia D, Aloise F, Bufalari S, Schalk G, Oriolo G, Cherubini A, Marciani MG, Babiloni F. Non-invasive brain-computer interface system: Towards its application as assistive technology. Brain Research Bulletin. 2008;**75**:796-803. DOI: 10.1016/j.brainresbull.2008.01.007

[30] Nijboer F, Furdea A, Gunst I, Mellinger J, McFarland DJ, Birbaumer N, Kübler A. An auditory brain-computer interface (BCI). Journal of Neuroscience Methods. 2008;**167**:43-50

[31] Cooper PS, Darriba Á, Karayanidis F, Barceló F. Contextually sensitive power changes across multiple frequency bands underpin cognitive control. NeuroImage. 2016;**132**:499-511. DOI: 10.1016/j.neuroimage.2016.03.010

[32] Tonoike M. Emotional analysis of odors of equivalent sensory intensity. Bulletin Electro-Technical Laboratory. 1984;**48**:796-808

[33] Tonoike M, Seta N, Maetani T, Koizuka I, Takebayashi M. Measurements of olfactory evoked potentials and event related potentials using odorant stimuli. In: Proceedings of the Twelfth Annual International Conference of the IEEE Engineering in Medicine and Biology Society; Nov. 1–4, 1990. New York, NY: IEEE; 1990;**2**:912-913. DOI: 10.1109/IEMBS.1990.691487

[34] Yamanaka T, Sobagaki H, Nayatani Y. Opponent-colors responses in the visually evoked potential in man. Vision Research. 1973;**13**:1319-1333

[35] Peng W, Babiloni C, Mao Y, Hu Y. Subjective pain perception mediated by alpha rhythms. Biological Psychology. 2015;**109**:141-150. DOI: 10.1016/j.biopsycho.2015.05.004

[36] Isosaka T, Matsuo T, Yamaguchi T, Funabiki K, Nakanishi S, Kobayakawa R, Kobayakawa K. Htr2a-expressing cells in the central amygdala control the hierarchy between innate and learned fear. Cell. 2015;**163**:1153-1164. DOI: 10.1016/j.cell.2015.10.047

[37] Matsukawa M, Imada M, Murakami T, Aizawa S, Sato T. Rose odour can innately counteract predator odour. Brain Research. 2011;**1381**:117-123. DOI: 10.1016/j.brainres.2011.01.053

[38] Murakami T, Matsukawa M, Katsuyama N, Imada M, Aizawa S, Sato T. Stress-related activities induced by predator odor may become indistinguishable by hinokitiol odor. Neuroreport. 2012;**23**:1071-1076. DOI: 10.1097/WNR.0b013e32835b373b

Chapter 3

Wavelets in ECG Security Application

Seedahmed S. Mahmoud and Jusak Jusak

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/intechopen.74477

Abstract

Wavelet packet transform has been used in many applications of biomedical signal processing, for example, feature extraction, noise reduction, data compression, electrocardiogram (ECG) anonymisation and QRS detection. The wavelet analysis methods, in these applications, represent the temporal characteristics of a biological signal by its spectral components in the frequency domain. Furthermore, it has been shown in many works that the ECG signal can be used as a biometric method for robust human identification and authentication. In this case, it is necessary to anonymise the ECG data during the distribution and storage of the signal in a public repository. A neglectful system leads to an eavesdropper recording the ECG data and uses it as recognition data to gain access via an ECG biometric system. This chapter discusses and reviews recent researches on ECG anonymisation wavelets-based techniques. These techniques use discrete wavelet transform and wavelet packet transform. A comparative study between the wavelets-based methods will be presented.

Keywords: anonymisation, biometric, electrocardiogram, encryption, steganography, wavelet analysis

1. Introduction

In June 2006, Cisco released a virtual network index (VNI) forecast that projects global IP traffic over the next 5 years [1]. According to Cisco's paper, there has been quantitative evidence that proliferation of global IP traffic will exchange data to reach the order of zettabyte (ZB) by 2021. This massive amount of data will be driven mainly by the number of connected devices to IP networks, such as smart phones, tablets, sensors and machine-to-machine (M2M) applications that are estimated to be more than three times the global population. Hence, in this era, just about every physical object we see (e.g. health-care monitoring apparatus, machinery, appliances, autonomous cars and intelligent transportation, etc.) will be connected,

forming the Internet of Things (IoT) [2]. In order to handle the countless number and various types of devices as well as linking the existing radio-access technologies, a new architecture that will increase data rate, lower end-to-end latency and improve the coverage is urgently required. Therefore, to meet with this demand, a new standard on the fifth-generation (5G) networks is currently under consideration [3].

Health and medical care are considered as one of the most fascinating applications that can fully benefit from IoT deployment. The IoT that employs various sensor and smart medical devices may serve in, for example, tele-auscultation, remote health monitoring, remote diagnostics and possibly treatment as well as elderly care [4–6]. Such Internet of Medical Things (IMedT) is expected to reduce consultation and transportation cost and to shrink the gap for those who live in the isolated/remote areas where the presence of doctors is void. Nevertheless, transmitting medical data to health-care providers through the public networks require high data security as public networks are somehow vulnerable to spoof attack. In this chapter, two anonymisation techniques based on wavelet decomposition and wavelet packet (WP) transform for securing ECG signals will be discussed.

2. Motivations

An electrocardiogram (ECG) signal contains important health information of a patient. It is used to detect abnormal heart rhythms by measuring the electrical activity generated by the heart as it contracts. Recent studies show that an ECG signal can be used as a biometric method for robust human identification and authentication [7–9]. The ECG signal was found to be unique for each individual over a long period of time [10, 11]. An ECG biometric system consists of feature extraction and classifiers to identify and recognise a person. The selection of appropriate features is crucial for successful individual identification. In [12], ECG-based biometric features were grouped as fiducial based, non-fiducial based or hybrid.

An unsecure ECG signal can be subjected to *man in the middle* attack where fraudsters can use the spoofed recorded ECG data to gain access to a secured service [13–15]. A scenario where a man in the middle attack can be a real threat for health information transmission is presented in **Figure 1**. The figure illustrates possible attack points that include (1) wireless links between sensor nodes that collect health information data from wireless body area networks (WBAN) and gateways, (2) wire/wireless links between the gateway and the edge router, (3) wire/wireless links between the other side of the edge router and health-care provider router and (4) repository in the data centre/public server or health-care provider. In order to minimise such security threat to a system, a health-care provider needs to comply with certain widely accepted standards to protect medical records safely. For example, US government passed the Health Insurance Portability and Accountability Act (HIPAA) in 1996 for protecting medical privacy users [16], the European Union adopted the Directive on Data Protection in 1995 [17], the Health Information Privacy Code was passed by New Zealand government in 1994, which sets specific rules for agencies in the health sector to ensure protection of individual privacy [18] and the personally controlled electronic health record (PCEHR) eHealth system was launched by Australian government in 2012 [19].

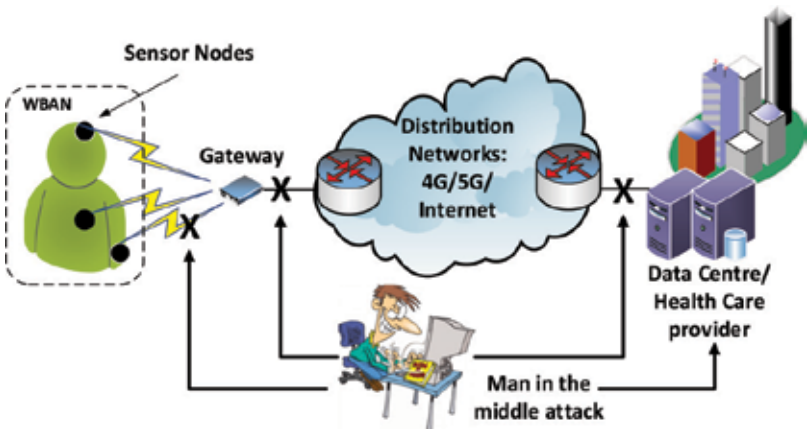


Figure 1. Possible attack points for unsecure ECG signals subjected to man in the middle attack.

There have been several proposed security techniques including image [20] and ECG steganography [21–24] to secure confidential patient information. In the steganography techniques, sensitive patient information is concealed inside public host data without incurring huge computational overhead or any increase in the size of the host data [21]. ECG data is used as the host signal to embed secret patient information and physiological readings. This may create watermarked ECG signals that is then transferred to a remote hospital server for further diagnosis. The effectiveness of ECG watermarking is dependent on the difference between the original host data and the watermarked data, that is, greater differences point to an ineffective steganography process. Unfortunately, all steganography methods bear some degree of information loss. This severe loss of information contributes to smeared/incorrect signal features and in some cases can lead to the failure of reconstructing the original ECG signal from the watermarked ECG signal [22]. However, even effective ECG watermarking can result in the delectability of ECG fiducial and non-fiducial features, which may allow for patient identification according to research in [7–9]. Therefore, a method combining the advantages of steganography with a technique that hides ECG fiducial and non-fiducial features is required. In this chapter, a review between two ECG anonymisation methods based on wavelet decomposition and wavelet packet transform (WPT) is presented.

3. Wavelet decomposition-based ECG anonymisation approach

Recent ECG anonymisation approaches based on wavelet decomposition were proposed in [13, 14]. During the wavelet decomposition process, filters of different cut-off frequencies were used to analyse the ECG signal at different scales (frequencies). It can be done by passing the ECG signal through a series of high-pass filters (i.e. the detail coefficients) for examining the high-frequency bands. The ECG signal was also passed through a series of low-pass filters (i.e. the approximation coefficients) to evaluate the low-frequency bands. Wavelet decomposition at level 3 was used during signal evaluation in the chapter [13]. Moreover, in the order to

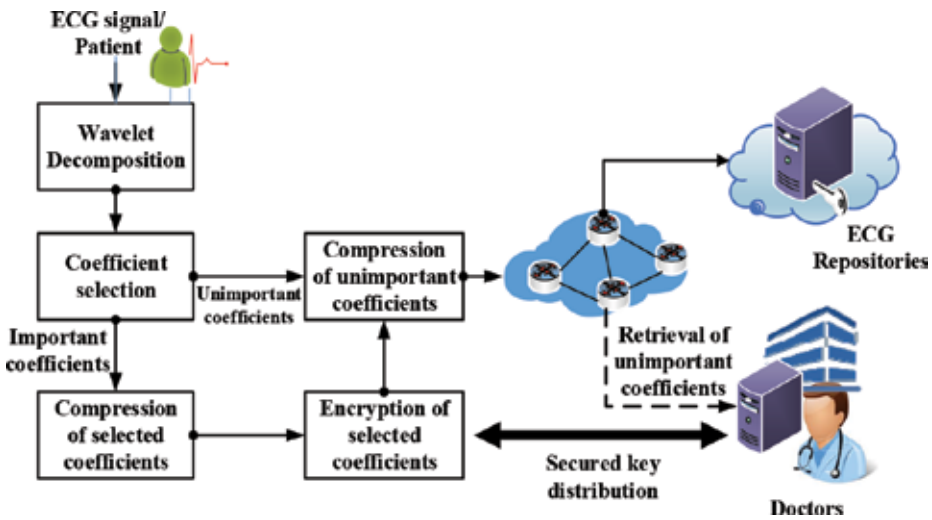


Figure 2. ECG anonymisation using wavelet decomposition.

construct a complete evaluation, two individual methods were studied during the experimentation [13]. Block diagram for the wavelet decomposition can be seen in **Figure 2**.

3.1. Method 1: discrete wavelet base anonymisation

In the first method, approximation (cA3) and detail (cD3) coefficients were removed after level 3 decomposition. Subsequently these nodes were encrypted using the well-known RSA symmetric cryptography. On the other hand, the remaining nodes, that is, cD1 and cD2, were compressed and transmitted to the ECG repository. **Figure 3** shows that without knowledge of nodes cA3 and cD3, the newly constructed signal in the repository completely hides P wave and T wave of the original ECG. It can be concluded that the first method hides most of the features required to reconcile the identity of a patient [7]. On the contrary, this method is not able to provide complete obfuscation of the cardiovascular conditions. This is mainly because

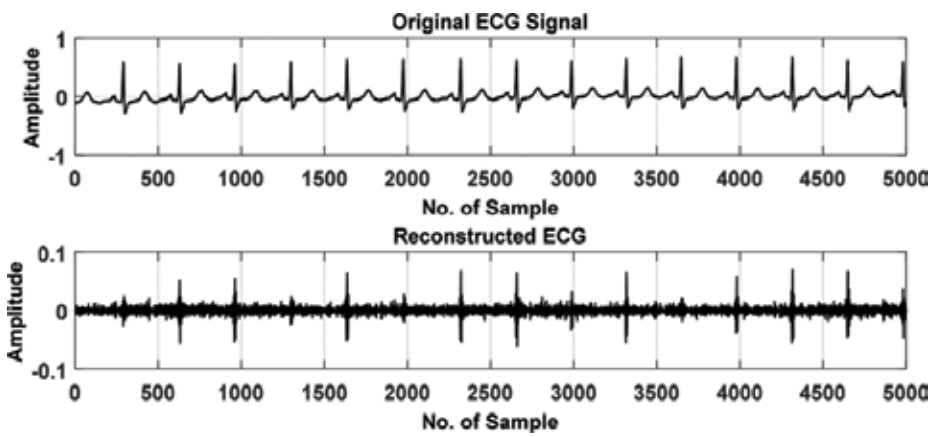


Figure 3. Normal ECG signal (top) and reconstructed annonymised ECG signal without nodes cA3 and cD3.

http://dx.doi.org/10.5772/intechopen.74477

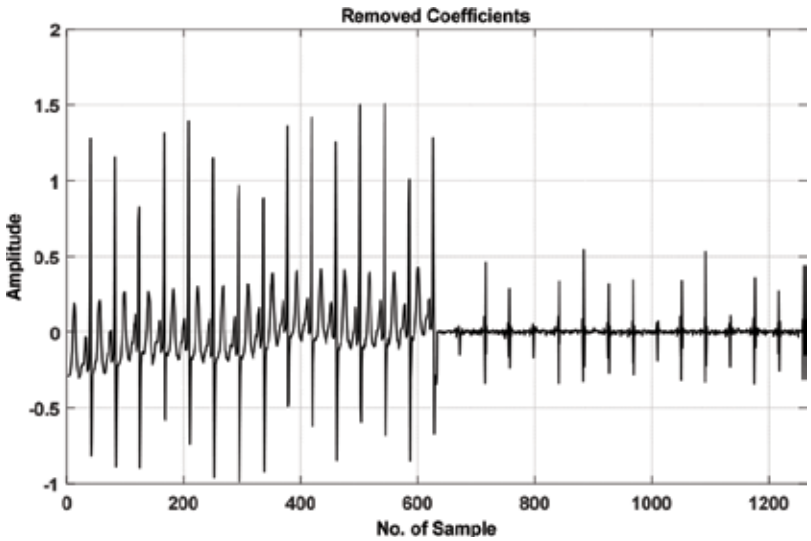


Figure 4. Removed (Selected for Encryption) Coefficients for Method 1.

the RR interval and certain types of arrhythmias are visible [25] as obvious in **Figure 3**. However, this method only required minimal selection of coefficient (approximately 25%) for encryption and key distribution. This is the main advantage of the first method. This method will perform well when faster distribution of key is priority and strong security is not deemed necessary. The removed coefficients are shown in **Figure 4**.

3.2. Method 2: discrete wavelet base anonymisation

In the second method, nodes cA3, cD3 and cD2 were selected for encryption, while the remaining coefficients cD1 were transmitted to the ECG repository. In contrast to the previous method, **Figure 5** shows that the reconstructed ECG from the coefficients that are extracted

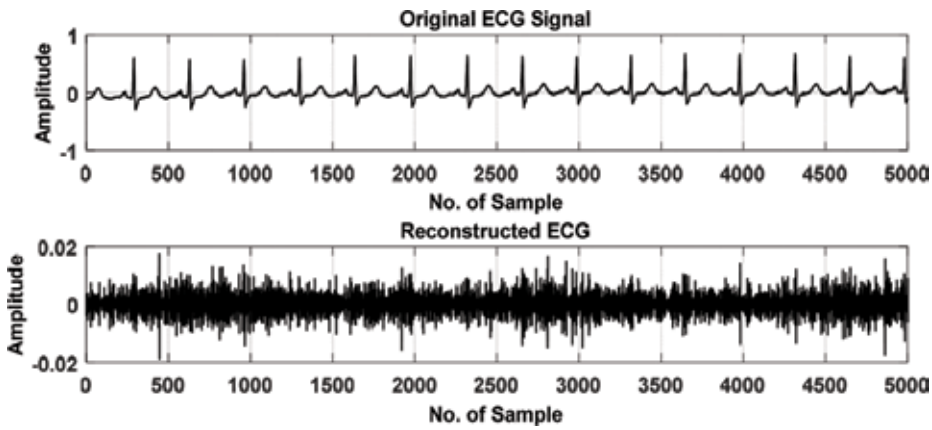


Figure 5. Normal ECG signal (top) and reconstructed annonymised ECG signal without nodes cA3, cD3 and cD2.

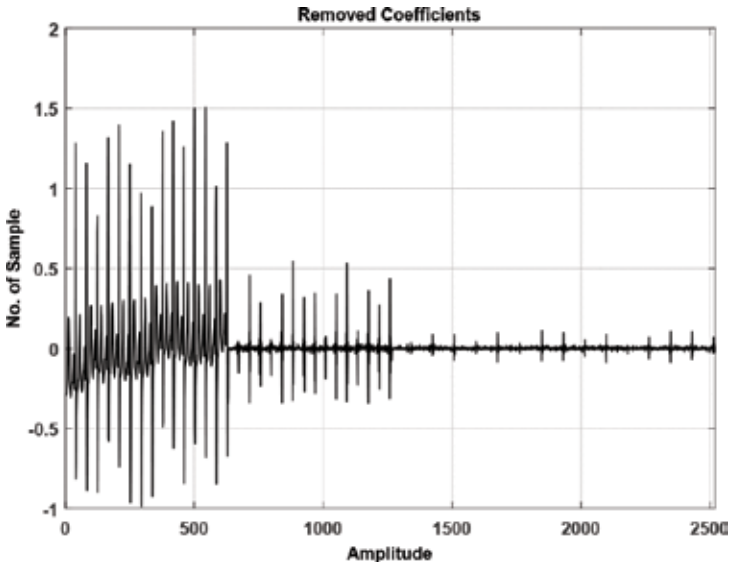


Figure 6. Removed (selected for encryption) coefficients for method 2.

from the repository is completely able to obfuscate features related to cardiovascular condition and person identification.

Therefore, this method provides higher ECG security by compromising larger key size (approximately 50%) as can be seen in **Figure 6**. **Figure 5** shows that the reconstructed ECG signal does not contain any ECG features.

Both methods described above suffer from long key size and lack of complete obfuscation to the ECG data. The long key size requires wider bandwidth during transmission process of the key to the ECG repository. On the other hand, lack of complete obfuscation results in trivial interpretation of the anonymised ECG signal. Therefore, due to these two main reasons, other methods based on the wavelet packet were proposed and developed.

4. Wavelet packet-based ECG anonymisation approach

4.1. Overview of wavelet packet transform

Wavelet packet transform has been used in many applications of biomedical signal processing, for example, feature extraction, noise reduction, data compression and QRS detection. Furthermore, wavelet packet transform has long been used for ECG signal analysis. A wavelet packet function [18] is defined as

$$\varphi_{j,k}^{n}(t) = 2^{\frac{1}{2}}\varphi^{n}\left(2^{j}t - m\right), \tag{1}$$

where j and m are the scale (frequency) and the translation (time) parameters, respectively, and $n = 0, 1, 3, \ldots$ is the oscillation parameter. The structure of wavelet packet (WP) decomposition

is described as a binary tree structure E; each node is described as (j, n), where j is a node's scale level and n is a node's number on the corresponded level. The root node $(0, 0)$ of the WP tree corresponds to the entire frequency range, $\left[0, \frac{f_s}{2}\right]$, where f_s is the ECG sampling frequency of the ECG signal. Each internal node of the WP tree $(j, n) \in E$ is called a parent node that is divided into two child nodes: the first and the second nodes are associated with low-pass $h(m)$ and high-pass $g(m)$ filters. These nodes forms a quadrature mirror filter (QMF) pair [19].

The scaling function $\omega(t)$ and the mother wavelet $\varphi(t)$ for the wavelet packet when $n = 0, 1$ and $j = m = 0$ are given by

$$\varphi^0(t) = \omega(t), \varphi^1(t) = \psi(t). \tag{2}$$

The other wavelet packet functions for $n = 2, 3, \ldots$ and $j = 1$ are shown as follows:

$$\varphi^{2n}(t) = \sum_m h(m)\varphi^n_{j,m}(t), \tag{3}$$

$$\varphi^{2n+1}(t) = \sum_m g(m)\varphi^n_{j,m}(t). \tag{4}$$

By substituting Eq. (1) into Eq. (3) and (4), we can get

$$\varphi^{2n}(t) = \sqrt{2}\sum_m h(m)\varphi^n(2t - m), \tag{5}$$

$$\varphi^{2n+1}(t) = \sqrt{2}\sum_m g(m)\varphi^n(2t - m), \tag{6}$$

where the low-pass filter gives $h(m) = \frac{1}{\sqrt{2}}\langle\omega(t), \omega(2t - m)\rangle$, and the high-pass filter gives $g(m) = \frac{1}{\sqrt{2}}\langle\psi(t), \psi(2t - m)\rangle$. The operator $\langle ., . \rangle$ stands for the inner product. The wavelet packet coefficients of the ECG signal, $x(t)$, are expressed as follows:

$$Q^n_j(m) = \left\langle x, \psi^n_{j,m} \right\rangle = \int_{-\infty}^{\infty} x(t)\psi^n_{j,m}(t)\, dt \tag{7}$$

Each coefficient measures a specific sub-band frequency content, controlled by the scaling parameter, j, and the oscillation parameter, n. The ECG signal, $x(t)$, can be decomposed into a different time-frequency space with Eq. (6) and Eq. (7). By computing the full wavelet packet decomposition on the ECG signal, for the jth level of decomposition, we have 2^j sets of sub-band coefficients of length $\frac{N}{2^j}$, where N is the ECG signal length [20]. Each sub-band coefficient, node, has a frequency range in the interval $\left[\frac{n}{2^{j+1}}, \frac{n+1}{2^{j+1}}\right]$, $n = 0, 1, \ldots, 2^j - 1$. This is how wavelet packet decomposes the original ECG signal into two or more coefficients.

4.2. The generalised framework for the ECG anonymisation method

In this section, a generalised framework for the ECG anonymisation using wavelet packet transform (WPT) will be introduced. The proposed framework for ECG anonymisation can be seen in **Figure 7**, while its pseudo-code is listed in Algorithm 1. This framework comprises the following steps:

Step 1: Perform wavelet packet decomposition of the ECG signal, $x(t)$, at level j. The signal coefficients at this level are given by

$$C = \{c(j,n): n = 0, 1, \ldots, 2^j - 1\} \tag{8}$$

where $c(j,n)$ represents the coefficients of the nth node at level j.

Step 2: Exclude the first node, $c(j,0)$, from C in Eq. (8) to get

$$\overline{C} = \{c(j,n): n = 1, 2, \ldots, 2^j - 1\}. \tag{9}$$

The excluded node is set to

$$k = c(j,0) \tag{10}$$

where k is an unencrypted and uncompressed key that includes the low-frequency components of the ECG signal, $x(t)$.

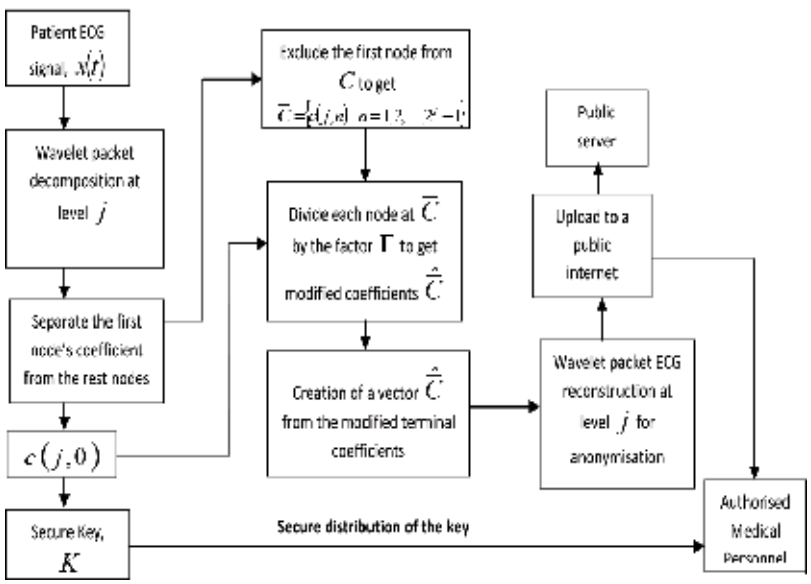


Figure 7. Wavelet packet-based ECG anonymisation process.

Step 3: Modify each node in $\overline{C}$, Eq. (9), using a reversible function/operation such as logarithm or division. In this chapter each node in $\overline{C}$ is divided by Γ. Γ is a reversible function driven from the key coefficients. Hence, the modified coefficients in $\overline{C}$ are given by

$$\hat{\overline{C}} = \frac{\overline{C}}{\Gamma} = \left\{ \frac{c(j,n)}{\Gamma} : n = 1, 2, \ldots, 2^j - 1 \right\}, \tag{11}$$

where $\Gamma = k + \text{offset}$, $\text{offset} = |\min(k)| + \eta$, η is a constant and $|.|$ is the absolute operator. The offset term in Γ is used to prevent division by zero.

Step 4: Securely distribute the key, K, and the offset value to medical personnel. The key security will be achieved by compressing and encrypting the first node, k, and the offset as follows:

$$K = E(\Delta(k, \text{offset})) = E(\Delta(c(j,0), \text{offset})), \tag{12}$$

where $\Delta(.)$ is the compression operator and $E(.)$ is the encryption operator [9]. Compression and encryption are beyond the scope of this chapter.

Step 5: Perform wavelet packet reconstruction to the modified terminal nodes' coefficient, $\hat{\overline{C}}$, to get the anonymised ECG, $y(t)$.

Step 6: Upload the anonymised ECG, $y(t)$, to the repository.

4.3. The ECG reconstruction method

The proposed reconstruction process for the anonymised ECG signal is shown in **Figure 8**, while the pseudocode is shown in Algorithm 2. The authorised personnel receives the secure key, K, and the anonymised ECG, $y(t)$, and performs the reconstruction process by the following steps:

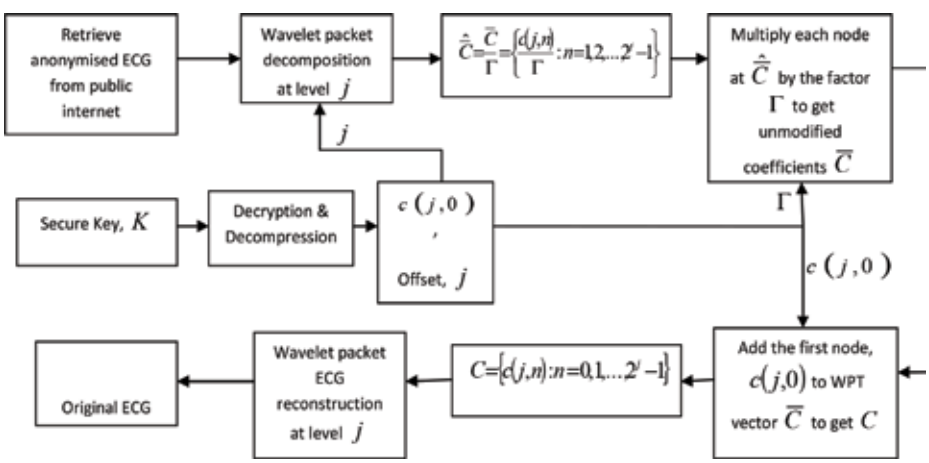


Figure 8. Wavelet packet-based reconstruction process for the anonymised ECG.

Step 1: Perform decryption and decompression to the key, K, to get Γ

$$\Gamma = \Lambda(D(K)), \tag{13}$$

where $\Lambda(.)$ and $D(.)$ are the decryption and decompression operators, respectively. Decryption and decompression are beyond the scope of this chapter.

Step 2: Perform wavelet packet decomposition of the ECG signal, $y(t)$, at level j to get $\widehat{\overline{C}} = \frac{\overline{C}}{\Gamma}$ as in Eq. (11).

Step 3: Multiply each node at $\widehat{\overline{C}}$ by the factor Γ to get

$$\overline{C} = f^{-1}\left(\widehat{\overline{C}}\right) = \widehat{\overline{C}} \times \Gamma = \left\{c(j,n) : n = 1, 2, \ldots, 2^j - 1\right\} \tag{14}$$

Algorithm 1: Wavelet packet-based ECG anonymisation process

1: Begin
2: $\boldsymbol{x(t)} \leftarrow$ ***ECG_signal***
3: $\boldsymbol{C} \leftarrow$ ***wpacket_decomposition***$\left(\boldsymbol{x(t)}, j\right)$
4: $\boldsymbol{k} \leftarrow \boldsymbol{c(j,0)}$ *// exclude the first node as a key*
5: $\overline{\boldsymbol{C}} \leftarrow \boldsymbol{c(j,n)}$
6: **offset** $\leftarrow |\mathbf{min}(\boldsymbol{k})| + \boldsymbol{\eta}$
7: $\boldsymbol{\Gamma} \leftarrow \boldsymbol{k} +$ **offset**
8: $\widehat{\overline{\boldsymbol{C}}} \leftarrow \frac{\overline{\boldsymbol{C}}}{\boldsymbol{\Gamma}}$
9: $\boldsymbol{K} \leftarrow \boldsymbol{E}\left(\boldsymbol{\Delta}\left(\boldsymbol{k}, \boldsymbol{offset}\right)\right)$ *//compression and encryption*
10: Send $\boldsymbol{K}$ to healthcare providers or doctors as a key
11: $\boldsymbol{y(t)} \leftarrow$ ***wpacket_reconstruction***$\left(\widehat{\overline{\boldsymbol{C}}}, j\right)$
12: Upload $\boldsymbol{y(t)}$ to public server
13: Save $\boldsymbol{y(t)}$ with unique ID for a particular individual
14: End

Algorithm 2: Wavelet packet-based reconstruction process

1: **Begin**
2: $\boldsymbol{k} \leftarrow \boldsymbol{\Lambda}(\boldsymbol{D}(\boldsymbol{K},$ **offset**$))$ *// decryption and decompression*
3: $\boldsymbol{y(t)} \leftarrow$ ***Anonymised_ECG_signal***
4: $\widehat{\overline{\boldsymbol{C}}} \leftarrow$ ***wpacket_decomposition***$\left(\boldsymbol{y(t)}, j\right)$
5: $\overline{\boldsymbol{C}} \leftarrow \widehat{\overline{\boldsymbol{C}}} \times \boldsymbol{\Gamma}$
6: $\boldsymbol{c(j,n)} \leftarrow \overline{\boldsymbol{C}}$
7: $\boldsymbol{k} \leftarrow \boldsymbol{\Gamma} -$ **offset**
8: $\boldsymbol{c(j,0)} \leftarrow \boldsymbol{k}$
9: $\boldsymbol{C} \leftarrow$ ***add_first_node***$\left(\boldsymbol{c(j,0)}, \boldsymbol{c(j,n)}\right)$
10: $\boldsymbol{x(t)} \leftarrow$ ***wpacket_reconstruction***$\left(\boldsymbol{C}, j\right)$
11: **End**

Step 4: Add the first node, $c(j,0) = \Gamma -$ offset, to the WPT vector $\overline{C}$ at Eq. (14) to get the WPT vector coefficients, C, of the original ECG signal, $x(t)$.

Step 5: Perform wavelet packet reconstruction of the coefficients vector, C, at level j to recover the original unanonymised ECG signal, $x(t)$.

5. Algorithm validation

Two types of electrocardiogram (ECG) signals were used to validate and investigate the performance and the effectiveness of the generalised ECG anonymisation framework. These signals are

1. normal ECG signal for a healthy subject, and
2. abnormal ECG signals for a patient with supraventricular arrhythmia and a patient with ventricular tachyarrhythmia.

The normal and abnormal ECG signals with different sampling frequencies were used in this chapter to study the robustness of the proposed anonymisation approach in concealing and smearing the ECG's fiducial and non-fiducial features. The normal and abnormal ECG data were obtained from the PTB ECG database [26] and the MIT-BIH arrhythmia database [27], respectively. These databases are publically available [26, 27].

In the evaluation process in the latter sub-section, bior5.5 wavelet was used. Besides this type of mother wavelet resembling the shape of an ECG signal, it is widely used for speech, video and biomedical signals providing that bior5.5 inherited linear phase. Nevertheless, it should be noted that for ECG anonymisation in this chapter, mother wavelet will not impact the anonymisation result since the ECG signal will be constructed back to its original at the receiver side.

The security of the proposed scheme depends on the following parameters that are required at the receiver side:

1. the encrypted security key which should be shared secretly,
2. the reversible function that should be used to reconstruct the original ECG information from the anonymised ECG, and
3. the type of transformation and the level of decomposition (wavelet packet transform at level 2 is used in this study).

An attacker with stolen key (i.e. able to decrypt the secure key) using brute force or any other method will require the knowledge of the reversible function and the level of decomposition. This information will be stored inside a patient/medical personnel PC and will not be transmitted under any circumstance. In this case, brute force attack is infeasible for the attack.

In the following sections, performance analysis using cross-correlation of normal and anonymised ECG signals, power spectral density of anonymised ECG signal and percentage residual difference (PRD) methods will be examined.

5.1. Performance evaluation over normal electrocardiogram

An electrocardiogram (ECG) signal has a well-defined P, QRS and T signature that is represented with each heartbeat. The P-wave arises from the depolarisation of the atrium. The QRS complex arises from depolarisation of the ventricles and T-wave arises from repolarisation of the ventricle muscles. The duration, shape and amplitude of these waves are considered as major features in time-domain analysis. Sometimes the time morphologies of these waves are similar.

The normal ECG was obtained from the PTB database (patient247, signal s0479). The sampling frequency, f_s, for this signal was 1 kHz. A total of 10 s of this signal was transformed by wavelet packet decomposition at level 2, $j = 2$. Decomposition level, j, depends on the ECG sampling frequency. Higher sampling frequency requires a low value of j to conceal all features in the anonymised signal. Node $c(2,0)$ of size $\frac{N}{4}$ (N = 10,000 samples) was removed from the wavelet packet coefficients of the normal ECG signal. This node was used to generate the key, K, which was distributed securely to medical personnel. The anonymised ECG is reconstructed

from the rest nodes, three nodes, using the anonymisation algorithm in Section II (B) and transmitted confidently over the public internet, since the anonymised ECG does not impose any threat to privacy.

Figure 9 (a) and **(b)** shows the time-domain representation of the 10-s normal ECG signal (patient247, signal s0479) and its anonymisation version, respectively. The frequency range for the anonymised ECG after node $c(2,0)$ removal is 125 and 500 Hz. From the time-domain representation of the ECG signal and its anonymisation in **Figure 9 (a)** and **(b)**, the proposed anonymisation algorithm conceals all fiducial features from the reconstructed ECG signal (**Figure 9(b)**). **Figure 10 (a)** and **(b)** shows the frequency representation of the 10-s normal ECG signal (**Figure 9 (a)**) and its anonymisation version, respectively. The non-fiducial features were also concealed as shown in the frequency-domain representation of the anonymised version of the normal ECG signal.

Figure 11 shows the time-domain representation of the coefficients $c(2,0)$ which was used to create the secure key, K. The frequency range for $c(2,0)$ in this data is between 0 and 125 Hz. This node preserves all fiducial features in the original ECG signal. **Figure 12 (a)** and **(b)** shows the reconstructed ECG signal at the medical personnel side and its cross-correlation with the original ECG signal at the patient side, respectively. From **Figure 12 (b)**, both signals are highly correlated, which guarantees a lossless reconstruction.

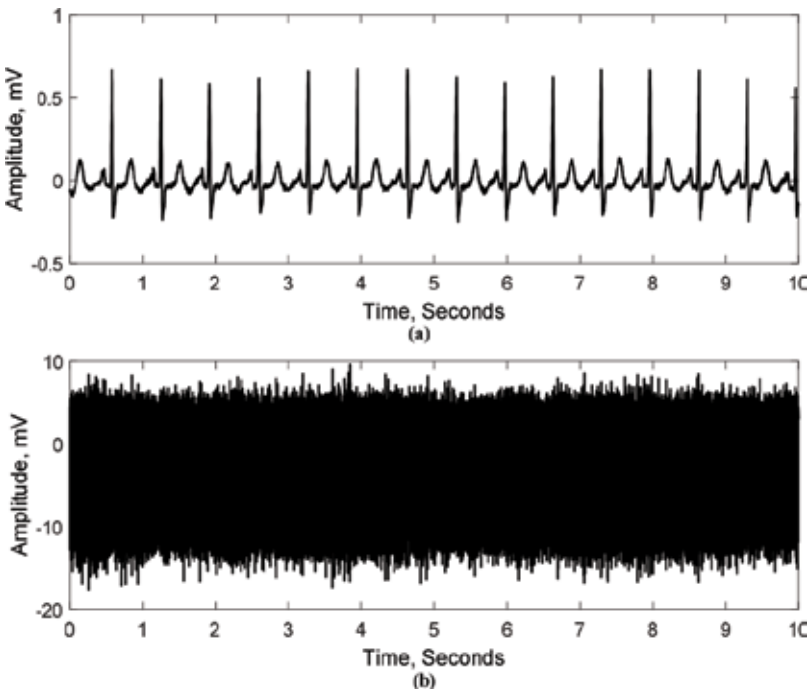


Figure 9. Time-domain representation of 10-s normal ECG signal, (a) unanonymised ECG signal and (b) anonymised ECG signal. The sampling frequency was f_s = 1 kHz.

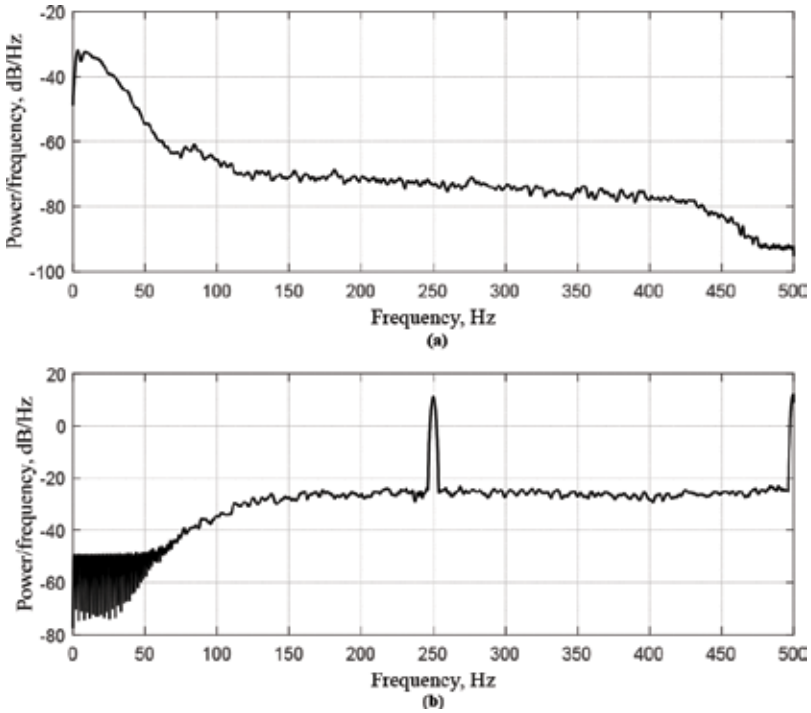


Figure 10. Power spectral density of 10-s normal ECG signal, (a) unanonymised ECG signal and (b) anonymised ECG signal. The sampling frequency was f_s = 1 kHz, the power spectral method was Welch periodogram.

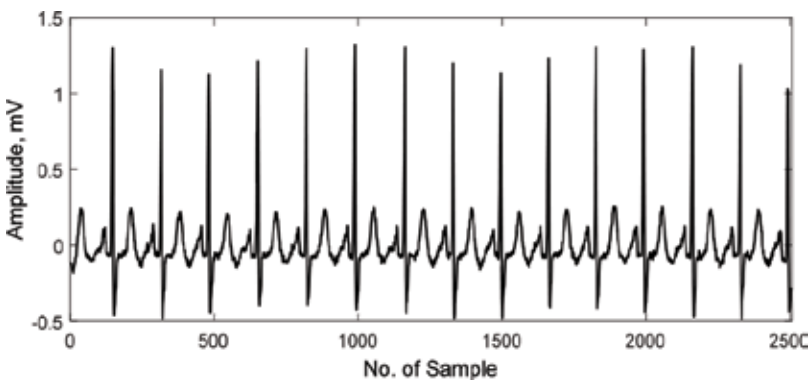


Figure 11. Time-domain representation of the first node $c(2,0)$ coefficients for the 10-s normal ECG signal in **Figure 9(a)**. This node was used to create the secure key.

5.2. Performance evaluation over abnormal electrocardiogram

An arrhythmia is an abnormality in the heart's rhythm or heartbeat pattern. The heartbeat can be too slow, too fast, have extra beats or otherwise beat irregularly [28]. The types of abnormal ECG signals investigated in this study were supraventricular arrhythmia and ventricular tachyarrhythmia. Supraventricular arrhythmia occurs in the upper areas of the heart and

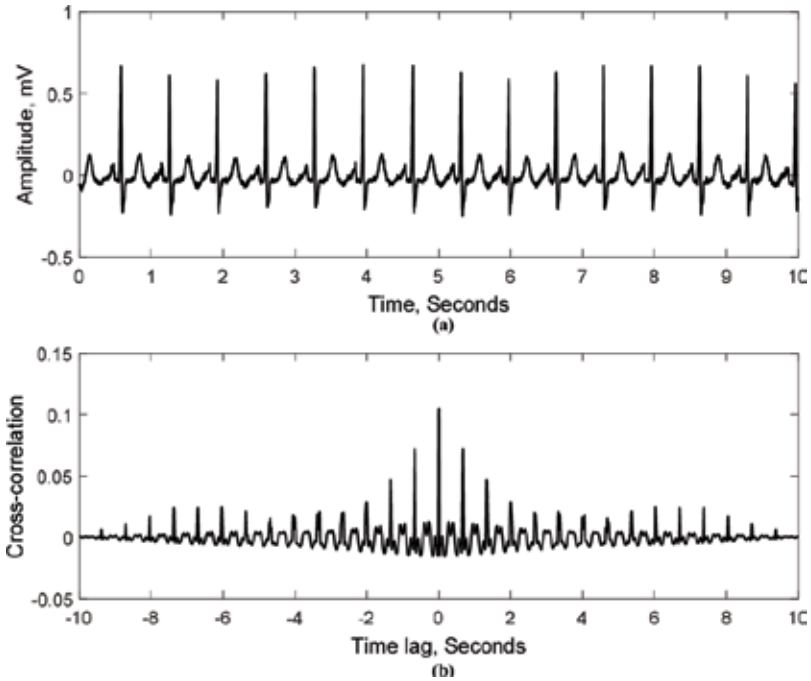


Figure 12. Ten seconds reconstructed ECG signal, (a) time domain representation of the reconstructed ECG signal, (b) cross correlation between the normal ECG signal in **Figure 9(a)** and its reconstructed version.

is less serious than ventricular arrhythmia. It has irregular shapes of QRS complexes [28]. These arrhythmia data—supraventricular arrhythmia and ventricular tachyarrhythmia—were obtained from the MIT-BIH arrhythmia database [26].

5.2.1. Supraventricular arrhythmia

The sampling frequency, f_s, for this signal was 128 Hz. A total of 10 s of this signal was transformed by wavelet packet decomposition at level 2, $j = 2$.

Node $c(2,0)$ of size $\frac{N}{4}$ (N = 1280 samples) was removed from the wavelet packet coefficients of the supraventricular arrhythmia signal. This node was used to generate the key, K, which was distributed securely to medical personnel. The frequency range for $c(2,0)$ in this data is between 0 and 16 Hz. The other nodes at level 2 with the frequency range between 16 and 64 were used to construct the anonymised signal.

Figure 13 (a) and **(b)** shows the time-domain representation of the 10-s ECG signal of a patient with supraventricular arrhythmia and its anonymisation version, respectively. The frequency-domain representation for both signals is shown in **Figure 14 (a)** and **(b)**. The fiducial and non-fiducial features were concealed in the time-domain and frequency-domain representation of the anonymised supraventricular arrhythmia signal.

Figure 15 shows the time-domain representation of the coefficients $c(2,0)$, which was used to create the secure key, K. This node preserves all fiducial features in the original supraventricular arrhythmia signal.

http://dx.doi.org/10.5772/intechopen.74477

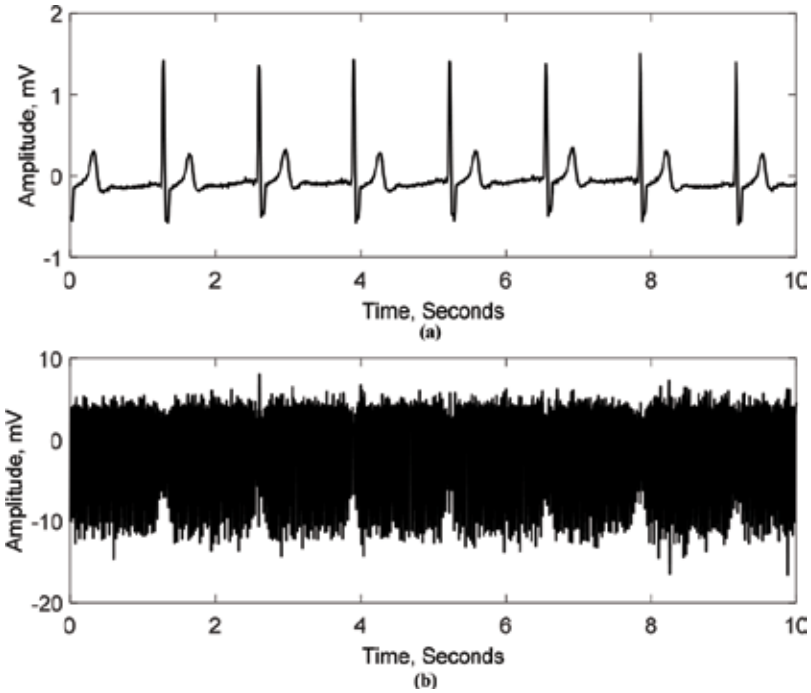


Figure 13. Time-domain representation of 10-s ECG signal of a patient with supraventricular arrhythmia, (a) unanonymised ECG and (b) anonymised CG. The sampling frequency was f_s = 128 Hz.

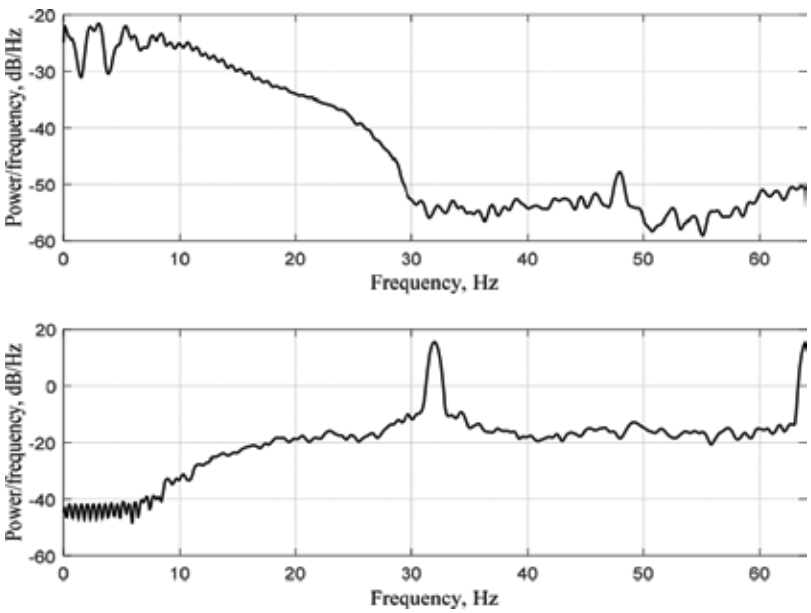


Figure 14. Power spectral density of ten seconds normal ECG signal of a patient with supraventricular arrhythmia, (a) unanonymised ECG and (b) anonymised ECG. The sampling frequency was f_s = 128 Hz, and the power spectral method was Welch periodogram.

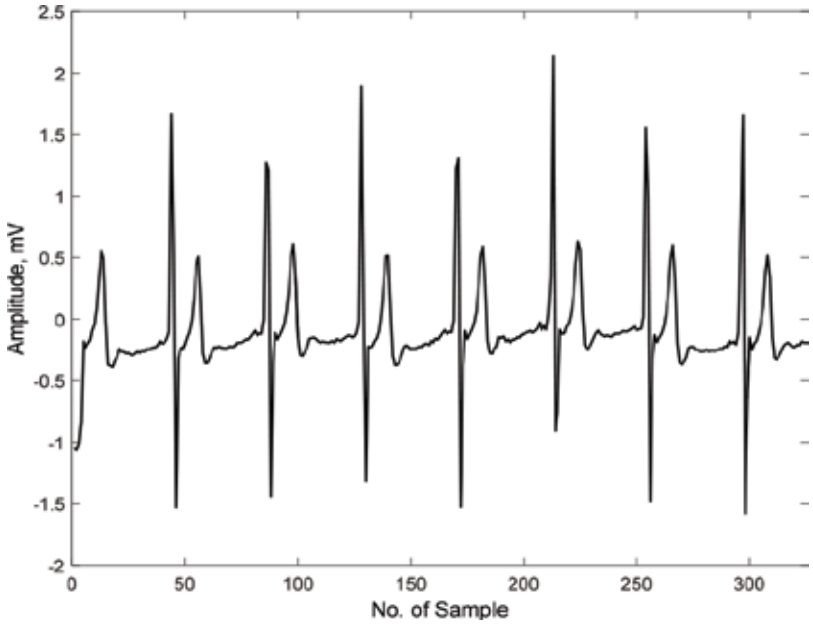


Figure 15. Time-domain representation of the first node $c(2,0)$ coefficients for the 10-s abnormal ECG signal in **Figure 7 (a)**. This node was used to create the secure key.

5.2.2. Ventricular tachyarrhythmia

The sampling frequency, f_s, for this signal was 250 Hz. A total of 10 s of this signal was transformed by wavelet packet decomposition at level 2. Node $c(2,0)$ of size $\frac{N}{4}$ (N = 2500 samples) was removed from the wavelet packet coefficients of the ventricular tachyarrhythmia signal. This node was used to generate the key *K*, which was distributed securely to medical personnel. The other nodes were used to reconstruct the anonymised ventricular tachyarrhythmia signal.

Figure 16 (a) and **(b)** shows the time-domain representation of the 10-s ECG signal of a patient with ventricular tachyarrhythmia and its anonymisation version, respectively. The frequency-domain representation for both signals is shown in **Figure 17 (a)** and **(b)**. The fiducial and non-fiducial features were concealed in the time-domain and frequency-domain representation of the anonymised supraventricular arrhythmia signal.

Figure 18 shows the time-domain representation of the coefficients $c(2,0)$ which was used to create the secure key, *K*. This node preserves all fiducial features in the original supraventricular arrhythmia signal.

5.3. Performance evaluation with the PRD metric

The percentage residual difference (PRD) is used to measure the difference between the original ECG signal and the anonymised ECG signal using the following equation.

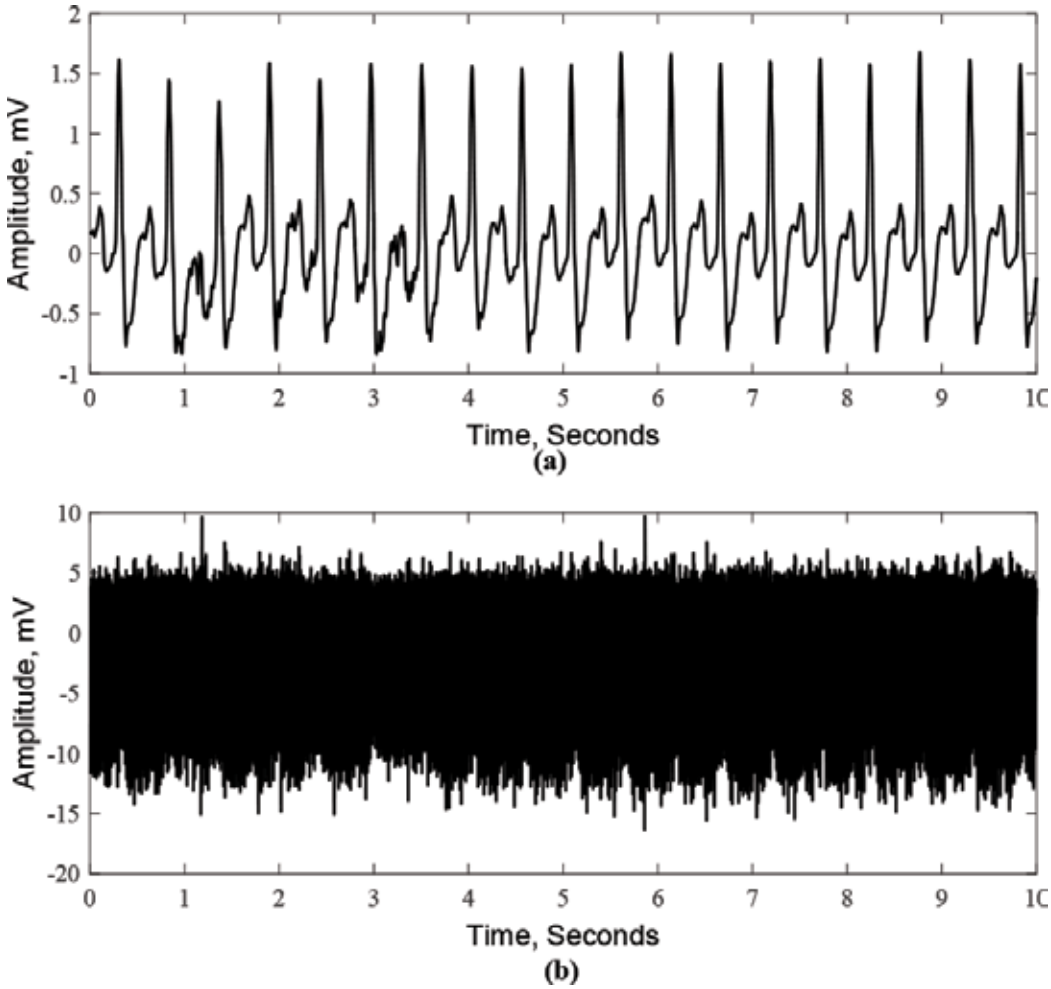


Figure 16. Time-domain representation of 10-s ECG signal of a patient with ventricular tachyarrhythmia, (a) unanonymised ECG and (b) anonymised ECG. The sampling frequency was f_s = 250 Hz.

$$\text{PRD} = \sqrt{\frac{\sum_{i=1}^{N}(x(i)-y(i))^2}{\sum_{i=1}^{N}x(i)^2}} \tag{15}$$

where $x(i)$ is the original ECG signal, $y(i)$ is the anonymised ECG signal and $i = 1\ldots N$, where N is the total number of the sample.

Performance of the proposed anonymisation algorithm using PRD metric is shown in **Table 1**. It can be seen from the table that the minimum and the maximum PRD measured were 14.8 and 70.6%, respectively. The PRD value depends on the ECG frequency bandwidth and its sampling frequency.

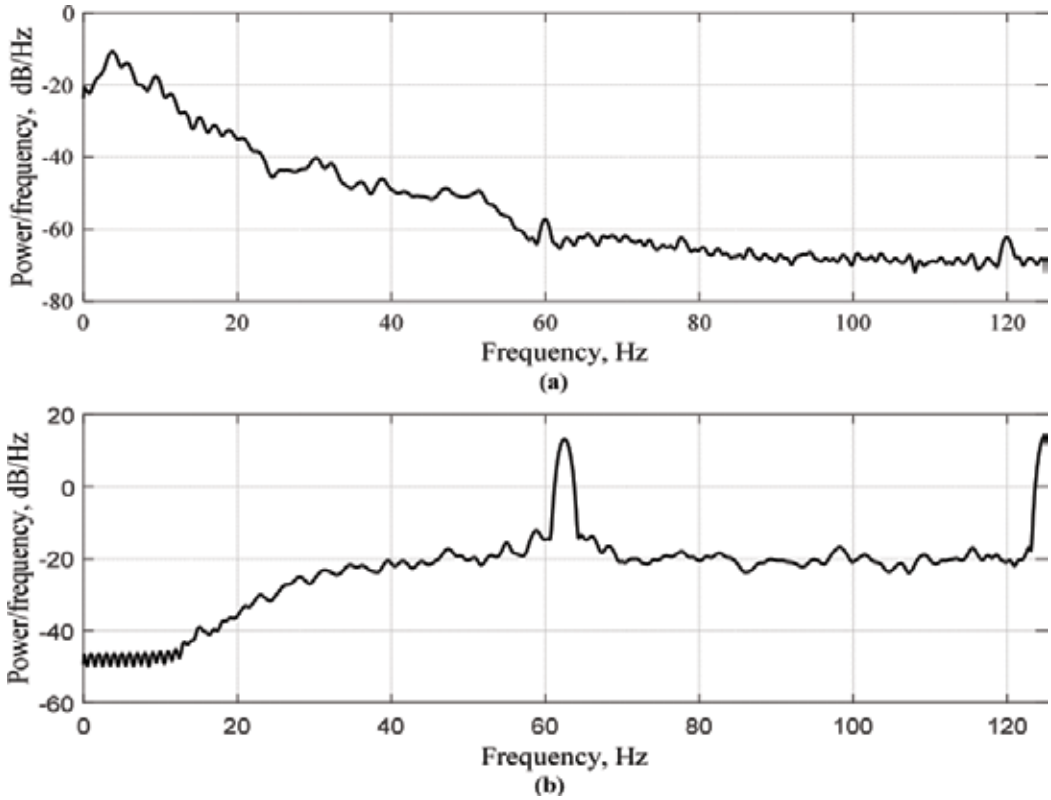


Figure 17. Power spectral density of 10-s ECG signal of a patient with supraventricular arrhythmia, (a) unanonymised ECG and (b) anonymised ECG. The sampling frequency was f_s = 250 Hz, the power spectral method was Welch periodogram.

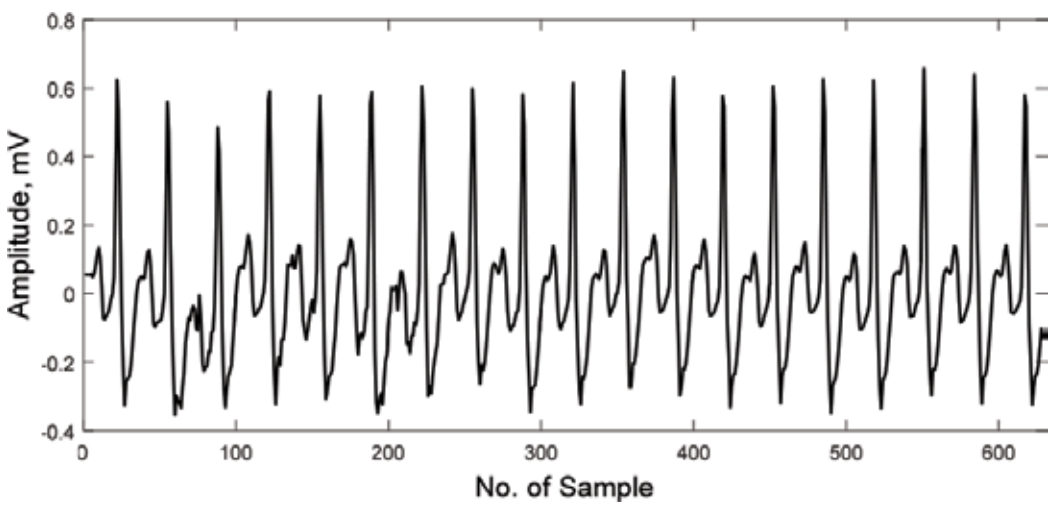


Figure 18. Time-domain representation of the first node $c(2,0)$ coefficients for the 10-s abnormal ECG signal in **Figure 10 (a)**. This node was used to create the secure key.

ECG type	Sampling frequency f_s, Hz	PRD %
Normal ECG	1000	70.6
Supraventricular arrhythmia	128	29.6
Ventricular tachyarrhythmia	250	14.8

Table 1. PRD performance results of normal and abnormal ECG signal for the proposed algorithm.

Comparing with ECG steganography methods, ECG steganography has a low PRD value between original and watermarked ECG signal. For example, in [14], the maximum PRD measured was 0.6%. Low PRD is essential in ECG steganography to guarantee correct diagnosis of the ECG watermarked signal. However, the lower value of PRD makes the ECG vulnerable to attack [1–3, 9].

6. Conclusions

A generalised wavelet packet-based ECG anonymisation framework has been presented in this chapter. This proposed anonymisation technique was used to conceal fiducial and non-fiducial features from normal and abnormal ECG signal for secure transmission over the public internet. Normal and abnormal ECG signals with different sampling frequencies have been investigated by the proposed method. Signal transformations other than wavelet packet transform can be used in this framework. Such transformations should have inverse property.

The performance analysis revealed that the proposed method is able to conceal both fiducial and non-fiducial features in normal and abnormal ECG signals under examination. Moreover, the analysis showed that the reconstructed ECG is highly correlated with the original ECG signal. It achieved a lossless reconstruction of the ECG data and proved the robustness of the proposed method. The security measures taken to secure the key and other information such as the level of decomposition and the knowledge of the reversible function make attacks using methods such as brute force is infeasible.

Author details

Seedahmed S. Mahmoud[1]* and Jusak Jusak[2]

*Address all correspondence to: seedahmed.sharif@gmail.com

1 Department of Electrical and Electronic Technology, Applied Engineering College, Lincoln College International, Buraidah, Riyadh, Kingdom of Saudi Arabia

2 Department of Computer Engineering, Institut Bisnis dan Informatika Stikom Surabaya, Surabaya, East Java, Indonesia

References

[1] Cisco. The Zettabyte Era: Trends and Analysis. White Paper at Cisco.Com, June 2016

[2] Atzori L, Iera A, Morabito G. Internet of things: A survey. Computer Networks. October 2010;**54**(15):2787-2805

[3] Andrews JG, Buzzi S, Choi W, Hanly SV, Lozano A, Soong ACK, Zhang JC. What will 5G be? IEEE Journal on Selected Areas in Communications. June 2014;**32**(6):1065-1082

[4] Islam SMR, Kwak D, Kabir MDH, Hossain M, Kwak KS. Internet of things for health care: A comprehensive survey. IEEE Access. 2015;**3**

[5] Jusak J, Puspasari I. Wireless tele-auscultation for phonocardiograph signal recording through the zigbee networks. IEEE Asia Pacific Conference on Wireless and Mobile (APWiMob), Bandung, Indonesia, August 27–29, 2015

[6] Jusak J, Pratikno H, Putra VH. Internet of medical things for cardiac monitoring: Paving the way to 5G mobile networks. IEEE International Conference on Communication, Networks and Satellite, Surabaya, Indonesia, Dec. 2016

[7] Biel L, Petersson O, Philipson L, Wide P. ECG analysis: A new approach in human identification. IEEE Transaction on Instrumentation and Measurement. 2001;**50**(3):808-812

[8] Irvine JM, Wiederhold BK, Gavshon LW, Israel SA, McGhee SB, Meyer R, Wiederhold MD. Heart rate variability: a new biometric for human identification. International Conference on Artificial Intelligence, Las Vegas, Nevada; 2001. pp. 1106-1111

[9] Israel SA, Scruggs WT, Worek WJ, Irvine JM. Fusing face and ECG for personal identification. In: Proc. 32nd IEEE Appl. Imagery Pattern Recog. Workshop; 2003. pp. 226-231

[10] Chan ADC, Hamdy MM, Badre A, Badee V. Wavelet distance measure for person identification using electrocardiograms. IEEE Transactions on Instrumentation and Measurement. 2008;**57**(2):248-253

[11] Wubbeler G, Stavridis M, Kreiseler D, Bousseljot R-D, Elster C. Verification of humans using the electrocardiogram. Pattern Recognition Letters. 2007;**28**:1172-1175

[12] Odinaka I, Lai P, Kaplan AD, O'Sullivan JA, Sirevaag EJ, Rohrbaugh JW. ECG biometric recognition: A comparative analysis. IEEE Transactions on Information Forensics and Security. 2012;**7**(6):1812-1824

[13] Sufi F, Mahmoud SS, Khalil I. A Wavelet Based Secured ECG Distribution Technique for Patient Centric Approach. In: The Proceedings of 5th International Workshop on Wearable and Implantable Body Sensor Networks, Hong Kong, China; 2008

[14] Fahim KS, Mahmoud SS, Khalil I. A novel wavelet packet-based anti-spoofing technique to secure ECG data. International Journal of Biometrics. 2008;**1**(2):191-208

[15] Reinsmidth E, Schwab D, Yang L. Securing a connected mobile system for healthcare. IEEE 17th International Symposium on High Assurance Systems Engineering (HASE), Orlando, FL, USA. Jan. 2016. pp. 19-22

[16] Department of Health & Human Services USA. Security 101 for covered entities. HIPAA Security Series. 2007;**2**:1-11

[17] European Parliament and of the Council. Directive 95/46/EC on the protection of individuals with regard to the processing of personal data and on the free movement of such data. Official Journal of the European Communities. 1995;**1**:281/31-281/50

[18] Privacy Commisioner. Health Information Privacy Code 1994. Ed. 2008. Auckland, New Zealand: KB Printed Ltd; 2008

[19] Pearce C, Bainbridge M. A personally controlled electronic health record for Australia. Journal of the American Medical Informatics Association. 2014;**21**(4):707-713

[20] Özkaynak F, Yavuz S. Analysis and improvement of a novel image fusion encryption algorithm based on DNA sequence operation and hyper-chaotic system. Nonlinear Dynamics. 2014;**78**(2):1311-1320

[21] Ibaida A, Khalil I. Wavelet-based ECG steganography for protecting patient confidential information in point-of-care systems. IEEE Transactions on Biomedical Engineering. 2013; **60**(12):3322-3331

[22] Tseng K, He X, Kung W, Chen S, Liao M, Huang H. Wavelet-based watermarking and compression for ECG signals with verification evaluation. Sensors. 2014;**14**(2): 3721-3736

[23] Liji CA, Indiradevi KP, Babu A. Integer-to-integer wavelet transform based ECG steganography for securing patient confidential information. Procedia Technology. 2016;**24**: 1039-1047

[24] Chen C-K et al. Personalized information encryption using ECG signals with chaotic functions. Information Sciences. 2012;**193**:125-140

[25] Bartolo A, Clymer BD, Bugess RC, Turnbull JP, Golish JA, Perry MC. An arrhythmia detector and heart rate estimator for overnight polysomnography studies. IEEE Transactions on Biomedical Engineering. 2001;**48**(5):513-521

[26] The PTB Diagnostic ECG Database, http://www.physionet.org/ physiobank/database/ ptbdb (viewed Dec. 2015)

[27] MIT-BIH Arrhythmia Database, http://physionet.org/physiobank/database/svdb/ (viewed Dec. 2015)

[28] Hebbar AK, Hueston WJ. Management of Common Arrhythmias: Part I. Supraventricular arrhythmias. Journal of American Family Physician. 2002;**65**(12):2479-2486

Section 3

Wavelets in Finance Application

Chapter 4

Empirical Support for Fundamental, Factor Models Explaining Major Capital Markets Using Wavelets

Michaela M. Kiermeier

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/intechopen.74725

Abstract

Factor models are used to explain asset returns on all major capital markets. We argue that standard econometric analyses implicitly assume that the relationships between prices, spreads, and interest rates and their respective risk factors are time-scale independent. Furthermore, by applying wavelet analysis, we do not have to assume capital market efficiency; in fact, we explicitly allow for inefficiencies such as noise trading, dispersed information, technical, feedback, fundamental, and rational trading to allow for typical characteristics of capital market data. We use wavelet analysis to decompose capital markets' developments, and the risk factors, using the maximal overlap discrete wavelet transform (MODWT). We proceed by estimating the relationships on a scale-by-scale basis. Our respective empirical analyses for stock and bond markets are summarized and new research is presented with regards to European corporate bonds markets. On stock market, this approach finds more stable relationships between risk factors and price movements. On the bond markets, we find empirical evidence for four significantly evaluated factors. For the European corporate bonds market, the results show that the amount of credit spreads explained by risk factors is in fact high for certain time scales only which is similar to the findings for the other capital markets.

Keywords: maximal overlap discrete wavelet transform, factor models, stock markets, term structure of interest rates, corporate bond spreads

1. Introduction

It has long been acknowledged that risk factors are important in explaining the development of asset prices on all major capital markets. Ross [1] states that the difference between expected and realized asset returns is due to the unexpected development of risk factors. In his arbitrage pricing theory, he derives a relationship between expected asset returns and the sum of assets'

sensitivities toward these risk factors. Similarities between equity and corporate bond markets' risks have long been recognized and risk factors similar to those applied in stock markets were included in the analysis of bond markets, and corporate bond spreads, for example, see [2]. Other empirical analyses present models for the simultaneous pricing of stock and bond returns [3]. Generally speaking, it has long been recognized that capital markets have similar characteristics [4]. Cutler et al. formulated four important characteristics of data concerning returns in the stock, bond, foreign exchange, and other capital markets. Using monthly return data, there appears to be a positive first-order auto-correlation from 1 month to the next. This does, however, change if the time horizon is medium or even long term. In those cases, the auto-correlation becomes negative. Finally, fundamental factors explain capital market movements significantly in the medium and/or long term which can be explained by allowing for capital markets' inefficiencies, for example, they postulate that the positive 1 month autocorrelation of data could be cause investors who only learn about relevant risk factors with a time lag. In addition to that also traders acting on the basis of technical analyses can cause a positive auto-correlation in the very short run. The negative medium, or long term, auto-correlation is then a direct result from misperceptions that are corrected on those time scales. In addition to those explanations, we consider that market participants have different objectives and therefore also different time horizons for their investments. Arbitrageurs seek to exploit mispricing in nanoseconds. Day-traders want to use knowledge derived from technical analysis on a daily or weekly basis. Although asset and wealth managers can represent investors with all sorts of investment horizons their performance is evaluated at least every month. To summarize, it is highly unlikely that the data generating process is the same for all investment horizons which is the reason why we apply wavelet analysis to allow for discrepancies at different time horizons.

We apply wavelet analysis to shed light on the applicability of factor models for stock, bond, and corporate bond markets. For this purpose, we shortly summarize our respective findings for stock and bond markets. We then present a detailed, exemplary, new analysis for European corporate bond markets and present general ideas why the use of wavelet analysis improves on the applicability of factor models in practice.

The wavelet decomposition we apply allows us to specifically distinguish short, medium, and long run periods and at the same time it is possible to investigate if information from past continues to be of importance for the following time period. There is little information about the frequency content of data if no frequency analysis is performed. The frequency analysis, however, is not able to maintain information about the time location of events. In our empirical analysis of these models, explanatory variables are selected according to general considerations which fundamental variable influence the capital markets and proceed by assuming that the identified k factors contain the important information, so that we assume an approximate factor structure to hold. We investigate if averaging over various time periods veils the fact that the risk factors are of importance in explaining capital markets' asset returns for certain time scales only, i.e., we investigate if risk factors are especially powerful in explaining asset returns at certain time horizons. For that purpose, we decompose asset returns and risk factors into their time-scale components using the maximal overlap discrete wavelet transform

(MODWT) thereby decomposing monthly data to their respective time scales (short term, medium term, and long term). We then proceed by estimating the impact of the risk factors on various capital markets on a scale-by-scale basis. We test for significance using the Fama/MacBeth approach.

Only recently researchers start to analyze relationships to hold for various time periods and not just for the short and long run. This is why wavelet analysis has been applied to macro-economic and financial theories, for example see [5–10].

This chapter is organized in the following way. First, we review shortly the underlying theoretical backgrounds for the various capital markets' factor models in Section 2. In Section 3, we introduce the basic ideas of wavelet analysis and motivate its use to test for significantly evaluated premiums for risk factors which we test for their significance on different time scales. Our previous results for the stock and bond market are shortly summarized. In Section 4, we describe the respective analysis performed for the European corporate bond market as an example in detail and Section 5 concludes.

2. Factor models in finance

Factor models have always been of great interest to explain price movements on all major capital markets. If risk factors can be identified that are significantly evaluated by the market, that information is valuable for the purpose of general management, determining fair values of firms, asset management, finance, and controlling.

2.1. Stock markets

One of the most important and general approaches to explain price movements on stock markets is the arbitrage pricing theory (APT) developed by Ross [1]. The advantage of the APT is its generality. Various factor models can be derived and require different estimation and testing techniques. A detailed overview of the various possibilities for factor models is given in [7]. The factor models can be distinguished according to the origin of the factors. Statistical factors can be derived from applying factor analysis. Factors can also be determined in advance—derived from theoretical considerations—and observable data of macro-economic variables can be investigated for being risk factors. Since the purpose is to identify risk factors and not to derive fair prices for financial derivatives, the relationship between asset prices and risk factors is restricted to be approximately linear [7].

Ross develops his theory in the context of neo-classical assumptions concerning capital markets without frictions. He assumes that investors differ in their opinion of the exact distribution of the risk factors, however they all agree on a linear k-factor structure. The main assumption is the following: the return at the end of the period is determined by the return that was expected at the beginning of the investment period (μ_i) but also by the returns of the common risk factors ($\tilde{\lambda}_k$). The importance of the risk factors for an asset i depends on how sensitive the asset

is with regards to the k risk factors (b_{ik}). Those sensitivities are called factor loadings. Last but not least there is a white noise error variable ($\tilde{\epsilon}_i$). The k factors are common factors, i.e., every asset reacts to the development of these factors.

$$\tilde{r}_i = \mu_i + b_{i1}\tilde{\lambda}_1 + \ldots + b_{ik}\tilde{\lambda}_k + \tilde{\epsilon}_i \qquad \forall i = 1, \ldots, n \tag{1}$$

with $\tilde{r}_i$ = realization of the random variable asset i's asset return at the end of the investment period; μ_i = expectation of asset i's return at the beginning of the investment period; b_{ik} = factor loading of asset i's return in relation to the risk factor k's realized end-of-period return; $\tilde{\lambda}_k$ = realization of the random variable risk factor k's end of period return; and $\tilde{\epsilon}_i$ = realization of the random variable asset i's idiosyncratic risk.

In matrix notation this becomes Eq. (2):

$$\underset{(nx1)}{\tilde{r}} = \underset{(nx1)}{\mu} + \underset{(nxk)}{B} * \underset{(kx1)}{\tilde{\lambda}} + \underset{(nx1)}{\tilde{\varepsilon}} \tag{2}$$

In this economy, systematic risk is represented through unexpected changes of common risk factor returns. Ross assumes that idiosyncratic risk is diversifiable and that there are no arbitrage opportunities. It is then possible to derive a relationship between asset i's expected return and the factor loadings multiplied by the risk premiums of the k risk factors (λ_1, …, λk). The exact APT equation is given by Eq. (3).

$$\mu_i = \lambda_0 + b_{i1}\lambda_1 + \ldots + b_{ik}\lambda_k \qquad \forall i = 1, \ldots, n \tag{3}$$

This is the APT equation which we use in the empirical analysis to identify statistically significant risk factors. Without idiosyncratic risk, Eq. (3) is an immediate result arising from the absence of arbitrage opportunities, because a riskless portfolio is then simply a combination of assets such that the portfolio is insensitive with regards to the risk of the risk factors and therefore orthogonal to the column space of the B-matrix.

The factor models based on the APT can be summarized by four different model types according to the different ways to choose risk factors. They can be macro-economic, fundamental, statistical or non-linear. Once the factors are determined the asset returns sensitivities toward them must be estimated. In the second step, the estimated sensitivities are incorporated in a cross-section regression and the risk premiums are estimated.

After some transformation, Fama/MacBeth derived an OLS-estimator for risk premiums at every point in time $\widehat{\lambda}_t = \left(\widehat{B}_k^{'}\widehat{B}_k\right)^{-1}\widehat{B}_k^{'}\tilde{r}_t$ for all t = 1, …, T in a cross-section regression [11]. This results in a time series of estimated risk premiums $\widehat{\lambda}_t$ to which they apply a test statistic that is t-distributed and that allows to test for significantly evaluated risk factors, see Eq. (4).

$$t\left(\overline{\lambda}_k\right) = \frac{\overline{\lambda}_k * T^{\frac{1}{2}}}{s\left(\widehat{\lambda}_k\right)} \tag{4}$$

with T = numbers of observations, $\overline{\lambda}_k$ = arithmetic mean of $\widehat{\lambda}_{kt}$, and s($\overline{\lambda}_k$) = standard deviation of the monthly estimates $\widehat{\lambda}_{kt}$.

Wavelet analysis is then applied to decompose the risk factors and asset returns. The test for significantly evaluated risk factors is not only performed on an aggregate level but also at different time scales that allow information to be of relevance for certain time periods only. Furthermore, we also apply wavelet to distinguish expected and unexpected components of the risk factors. This approach results in the identification of risk factors that remain significant over longer time periods, the problem of parameter constancy is therefore mitigated as well.

This approach reduces the variance of the estimated means of the risk premiums. Furthermore, it shows that only certain scale information of the risk factors remains important over time. We find that this approach improves on the findings which fundamental factors are significant in explaining stock market returns. For a detailed derivation of the estimation equations and the results in which fundamental factors are significantly evaluated in the stock market, see [7].

2.2. Term structure of interest rates

The models to explain the term structure of interest rates have been of interest to researchers for a long time. The models differ in the purpose they are built for. In our analysis, we assume that the data generating process for term structure of interest rates can be expressed as an approximate factor model as in the previous section. Those types of models are especially meaningful if the task at hand is to forecast future term structures of interest rates. The models that generate good forecasts and are equally satisfying from a theoretical, arbitrage-free viewpoint have been developed, for example, see [12–14]. The risk factors are found to represent information with regards to the level, slope, and curvature of the term structure of interest rates. We find that in this market too, for the same reasons as before, an analysis on an aggregate level can be misleading so that we perform our analysis on a scale-by-scale basis. We then apply the procedure of Fama/MacBeth to test for significance of risk factors. The Nelson-Siegel model approximates the actual yield curve observed in the market on any specific date t for zero rates y with maturity τ through the following Eq. (5):

$$y_t(\tau) = \beta_{0t} + \beta_{1t}\left(\frac{1 - e^{-\gamma\tau}}{\gamma\tau}\right) + \beta_{2t}\left(\frac{1 - e^{-\gamma\tau}}{\gamma\tau} - e^{-\gamma\tau}\right) \tag{5}$$

with β_{0t}, β_{1t}, β_{2t}, and γ as model parameters [15].

The respective β_i's can be viewed as dynamic factors that represent short-, medium-, and long-term behavior [12]. The factors level (β_0), slope (β_1), curvature (β_2), and γ the mean reversion rate are then identified as risk factors. The models parameters are then estimated by assuming an autoregressive, dynamic data generating process for the factors.

The dynamic generalized Nelson-Siegel [14] embeds the Nelson-Siegel approach in an arbitrage-free setting. In order to ensure the absence of arbitrage, the number of risk factors has to be increased to five.

The above models increase the number of explanatory factors according to theoretical considerations. In our analysis, we test whether there is statistical evidence for the proposed risk factors to be significantly evaluated by the market. As before, we acknowledge that there might be inefficiencies present in the market. Similar to stock markets, we then assume an approximate factor structure to hold in the bond markets. As before, we then test for significance using the Fama/MacBeth approach. The data used consist of European Zero Coupon Curves estimated by ICAP and provided by Thomson Reuters. We then determine whether risk factors are significant for every time scale and not only on an aggregate level. Similar to our analysis with regards to the stock markets, we find that the significance of the risk factors varies with different time scales. By reconstructing the time series using the significant time scales only, we concentrate on a relatively small number of wavelet functions. We then investigate the scaled and significantly evaluated risk factors for their ability to help forecast the term structure of interest rates. In our analysis, we can only detect four significantly evaluated risk factors for the term structure of interest rates [16].

2.3. Corporate bonds

Structural models based on the idea of Merton result in theoretical credit spreads that significantly deviate from observable corporate bond markets spread [17]. The models can only explain a limited proportion of corporate bond market spreads even if tax asymmetries, liquidity, and conversion options are considered. This empirical finding is referred to as the credit spread puzzle [18]. Similarities between equity and corporate bond market's risk have long been recognized and risk factors similar to those applied in stock markets are included in the analysis of corporate bond spreads, for example see [2]. The set of explanatory variables is enriched by other researchers to also account for market inefficiencies. For example, it can be assumed that there are limits to arbitrage which combined with noise leads to predictable deviations of market prices from the asset's fundamental value [19]. A solution could be a dynamic model with dispersed information in which noisy investors only learn about fundamental information with a time delay in order to solve the puzzle. Furthermore, it can be assumed that market participants develop habit formation [20]. Other researchers find that there are higher spreads for bonds for which analysts' forecasts are more diverse, i.e., that higher risk premiums are present for bonds where there is higher disagreement [21, 22]. Furthermore, the necessity to analyze varying frequency behavior in the data has been documented for credit markets, for example see [23]. In contrast to the stock and bond market, we do not impose Ross' approximate factor structure, but instead we use Merton's approach to postulate a straightforward relationship between credit spreads and risk factors that influence the corporate's ability to pay back its debt and credit spreads on corporate bond markets in general (fundamental factors). If the purpose is to analyze corporate bond markets jointly, the assumption of Ross's factor structure would become necessary.

To estimate the proportion of credit spreads (cs) explained by risk factors, Eq. (6) has to be analyzed econometrically.

$$cs_t = \mathrm{a} + \mathrm{b}\,(x_t) + u_t \tag{6}$$

with u_t being a white noise error term, and x_t being the risk factors.

http://dx.doi.org/10.5772/intechopen.74725

Risk factors represent risks arising from the possibility to default, term structure of interest rates, equity markets, liquidity from mutual funds, and business cycle. Huang and Kong find that for B (BB) rated corporate bonds approx. 68% (61%) of the variation in credit spreads can be explained by respective risk factors [2]. For investment grade bonds however they find that the proportion explained is much lower. Inefficiencies can lead to a higher proportion being explained by the models, for example see [19, 21]. Again we want to analyze the data at different time horizons and simultaneously allow for inefficiencies such as delayed learning about relevant information or other forms of feedback, or technical trading and account for different investment horizons of market participants.

We decompose the data with wavelet analysis. We then test for significantly evaluated risk factors on a scale-by-scale basis, we find that only four factors can be viewed as significantly evaluated by the market [16].

In the following section, we describe the respective analysis in detail for the European corporate bond market.

3. Estimation techniques

Wavelet analysis estimates the frequency structure of a time series and in addition to that it keeps the information when an event of the time series takes place. This way an event can be localized in the time domain with regards to its time of occurrence although frequencies are analyzed as well. The functions at the heart of our analyses are wavelets. In contrast to co-sine functions (waves), wavelets are not defined over the entire time axis but have limited support. In order to achieve the ability to analyze relationships for different time periods, the wavelets are moved over the time axis and at the various scales the support is accordingly. By doing so it is possible to allow for changing regime shifts and the problem of parameter constancy is less severe which removes the necessity to eliminate extreme market moves from a purely statistical point of view. The length (width) of a wavelet on a certain scale represents an investment period of interest. The maximal overlap discrete wavelet transform (MODWT) increases the support of the dilated wavelet with increasing scale, thereby increasing the investment period. The advantage of this form of discrete wavelet transform is that it can be applied to any number of observations of the time series of interest.

Wavelets ($\psi_{j,\,k}$ and $\phi_{J,k}$) when multiplied with their respective coefficients at a certain level "j" or "J" are called atoms $D_{j,k}$ and $S_{J,k}$ (i.e., $d_{j,k}{}^{*}\psi_{j,k} = D_{j,k}$ and $s_{J,k}{}^{*}\phi_{J,k} = S_{J,k}$) with $\psi_{j,k}$ and $\phi_{J,k}$ being the wavelet and scaling functions at level "j" or "J" and "k" indicating the location of the wavelet on the time axis. The sum of all atoms $S_{J,k}(t)$ and $D_{j,k}(t)$ over all locations on the time axis k = 1, …, $\frac{n}{2^j}$ at a certain level "j" or "J" are given by Eqs. (7) and (8).

$$S_J(t) = \phi_{J,k} \text{ at level J} \tag{7}$$

$$D_j(t) = \sum_{k=1}^{\frac{n}{2^j}} d_{j,k}\psi_{j,k} \forall j = 1, \ldots, J \tag{8}$$

Defining the importance of information to be valid for a specific time period only, the time series are decomposed into their respective resolutions in time (time scales). The time series are then approximated using only parts of the coefficients and their respective wavelets, i.e., the multiresolution decomposition is applied to the time series which are then in turn reconstructed using only the significant portions at the various scales.

The wavelets used in the analysis are "symmlets." Those wavelets are best suited for the analyses because their characteristics are closest to the functions used in the classical Fourier analysis in that they are symmetric and do not contain jumps. This makes most sense if our goal is to analyze the time series in the time and frequency domain. As co-sine functions, the chosen wavelets should not require an interpretation in itself. In that sense, those wavelets are the most "neutral" functions so that no other wavelet functions are considered that would require additional explanations. Our goal is to be able to allow for an analysis on different scales but we would like to keep as much structure of the original time series as possible. The decomposition of the data is done by identifying significant wavelets at certain scales, i.e., wavelets with a specific support on the time axis. The search for significant wavelets is then repeated on the next higher scale (lower frequency). With each increase of the wavelets´ widths a new scale is defined. The number of scales used in this analysis equals four (i.e., J = 4) which is a direct result of the number of observations available. For an explanation of how many levels are recommendable, see [24]. Level "j" wavelet coefficients are associated with periods $[2^j, 2^{j+1}]$. The sums of all atoms at all levels—one to four—result in the original time series.

We perform the regression analysis at each level. Asset returns are regressed on risk factors at different time scales, i.e., the factor pricing equations are estimated at every time scale 1, …, J using the reconstructed time series as outlined before (see Eqs. (9) and (10)):

$$(cs_t)\left[\mathrm{d_j}\right] = \mathrm{a} + \mathrm{b}\ (\mathrm{x}_t)\ \left[\mathrm{d_j}\right] + u_t \text{ for all d1 to d4} \tag{9}$$

$$(\mathrm{cs}_t)[\mathrm{s}_4] = \mathrm{a} + \mathrm{b}\ (\mathrm{x}_t)\ [\mathrm{s}_4] + u_t\ \mathrm{s4} \tag{10}$$

The proportion explained by the risk factors is therefore estimated at each time scale.

4. Empirical analysis

4.1. The data

The credit index data used in this analysis are taken from Bank of America/Merrill Lynch. We use monthly OAS spreads of corporate bond indexes for the time period January 2000 to January 2013. We analyze EMU corporates in the rating category BBB-A (all EMU Corporates). The analysis is performed by using the indexes for various times to maturities 1–3, 3–5 in case of investment grade corporates. The differentiation is necessary to address the phenomenon that short maturities of investment grade corporate bonds depict a higher extend of the credit spread puzzle. For the Euro high yield market we use the Euro high yield index which contains firms with credit ratings BB and lower. Due to concerns with regards to biases caused

by low liquidity we do not distinguish the high yield index with regards to time to maturity. As explanatory variables, the level, slope and volatility of the bond markets are calculated from the monthly time series for European government term structures of interest rates available from the same source. Data with regards to the stock index Dax are included in the analysis to capture risk characteristics present in the stock markets. Volatility for the stock market is calculated from that time series. Data with regards to European corporate default probabilities are taken from Moody's. Monthly 1-year and 5-year default rates for European investment grade and Caa-C rated companies are available from January 2000 to April 2012. Due to data availability and quality of the data, the 5-year default rates are combined with 1-year default rates.

4.2. Wavelet analysis

We decompose each time series using the maximal overlap discrete wavelet transform (MODWT), i.e. the time series European credit spreads, Government Yields, Slopes, Volatilities, monthly return of Dax, volatility of monthly return of Dax, and Moody's default rates for European Investment Grade Corporates and CCC-Lower-Rated Corporates are decomposed to their respective time and frequency domain components as explained in section 3. Calculating the volatility from the monthly return data, the number of monthly observations we are left to be able to use is 132 (January 2001 to April 2012). As a result of the number of observations the number is set to four. The MODWT estimates the wavelet coefficients "d1" to "d4" and "s4" scaling coefficients.

The decomposition of the time series and the amount of variation explained with Crystals (sum of wavelets and their estimated coefficients at levels j = 1, …, 4) are summarized in **Table 1**.

The risk factors are well explained by coarse scales (low frequencies, e.g., "s4"). The only variable that has different features is the return of the DAX index. In that case the high

	"d1"	"d2"	"d3"	"d4"	"s4"
EMU corporates all maturities spread	0.3	0.4	2.2	5.5	91.5
EMU corporates 1–3 year maturity spread	0.4	0.6	2.7	7.2	89
EMU corporates 3–5 year maturity spread	0.3	0.5	2.4	5.6	91.2
Euro high yield spread	0.4	1	2.1	4.6	92
DAX return	46	25	12	8	8
DAX volatility	0.1	0.4	2.4	3	94.1
Euro government 10-year yield	0.1	0.1	0.3	0.4	99
Euro government yield curve slope	0.4	0.5	1.4	1.7	96
Euro government yield volatility	0.1	0.3	1.9	4.4	93.2
European investment grade default rates	0.1	0.2	0.3	1.7	97.7
European high yield default rates	0.5	0.7	1.6	6.8	90.3

Table 1. Variation of the time series explained by crystals (in %).

frequencies contribute the most in explaining the variation of the time series. The other variables are best explained by time scales ranging from "d4" to "s4," whereas the return of the DAX is best explained by time scales "d1" to "d3."

At each scale "j" the coefficients are associated with time periods $[2^j, 2^{j+1}]$. The decomposition of the monthly data allows us to extract components of the data that prevail in the medium or long term. At the highest frequency of the monthly data—at scale "d1"—coefficients approximate reactions to information for the time period of 2–4 months. At scale two, three, and four, the respective time periods are 4–8 months, 8–16 months, 16–32 months. We associate the scales "d1," "d2," and "d3" with the medium term (short medium term equals 2–4 months, medium term 4–8 months, and longer medium term 8–16 months). The remaining two scales at the lower frequencies represent long-term behavior (1.3–2.6 years and longer). Extracting the components of the data that are influential in the medium or long term allows us to detect patterns that can be a result of different investment behavior or different information used in forming expectations, i.e., we are able to allow for inefficiencies in the credit market as outlined above.

In a next step we regress the credit spreads on the explanatory variables on a scale-by-scale basis, i.e., we restrict features of the data to be of importance in the medium ("d1" to "d3") or long term ("d4" to "s4"). After decomposing the regression variables, we reconstruct the time series using features of the time series at the respective resolutions 1, …, 4 only, thereby restricting their variation to the respective time scale. On the other hand, it allows for the possibility that information from more than just the previous period continues to be of influence in explaining credit spreads. By analyzing the amount of the variation explained in a regression (R^2) of the decomposed data at various time scales, we can infer which of the above outlined possible expectation formations is significant in the medium and long run. **Table 2** summarizes the regression results for regressing European investment grade and European high yield credit spreads on the explanatory variables when the data are decomposed, i.e., when the time series are reconstructed to represent behavior present at scales "d1" to "s4."

Determining significant components gives us insights into how long time periods are for processing information. For the short medium term (2–4 months), we find that the default rate is either not significant ("d1" for EMU Corporate all maturities, and "d1," and "d2" for EMU high yield) or even of negative influence. This is a strong indication that the fundamentals are influential for longer time periods only, but do not explain well the variation in investment grade credit spreads for shorter time periods. The credit spread puzzle therefore manifests itself if the data are analyzed on time scales and in that the default rate is not significant in explaining credit spreads at all at some time scales. At the time scales that carry most of the energy, the default rate is significantly positive in explaining the investment grade credit spreads, i.e., for time periods (1.3–2.6 years). For high yield spreads, the default rate is significant only for a longer time horizon (8 months and above). We find that the influence of other explanatory variables changes at the various time scales as well. At the coarsest scale "s4," we find all explanatory variables of significant influence for the credit spreads. However, at scale "d3" (i.e., for a time period of 8–16 months), the variables that capture the volatilities in the stock and bond markets cannot be viewed as being significant variables. The volatility of the DAX, although of importance in the aggregate data, loses its significance for investment grade credit spreads on several scales. It

	DAX	DAXV	Yield	Yield slope	Yield vola.	Default rate	R^2
EMU corporates all maturities credit spread							
d1	−275	2.3*	40.9	−131	890	257	0.1
d2	−16	0.1*	−21.3*	39.9*	−264.6*	−7079*	0.53
d3	−418*	0.03	−19.8	72.9*	19.3	−10662*	0.61
d4	−365.7*	0.01	−77*	21.7*	323.4*	14632*	0.85
s4	−2278*	0.4*	−114*	65.2*	−645.2*	5518*	0.98
EMU corporates maturity 1–3 years credit spread							
d1	1.1	−0.0	−52.2*	46.2*	−18.1	−4219*	0.48
d2	20.7	0.1*	−18.1*	57*	−317.2*	−8857*	0.55
d3	−336*	0.02	−34.7*	86*	12.7	−6722*	0.61
d4	−399*	0.0	−96.5*	36.8*	368.3*	16641*	0.85
s4	−2201*	0.45*	−112*	83.7*	−670.5*	5381*	0.98
EMU corporates maturity 3–5 years credit spread							
d1	−26	0.05	−42.9*	24.8*	−53.1	−3471*	0.44
d2	−27	0.1*	−29.9*	40.6*	−243.5*	−7071*	0.53
d3	−410*	0.03	−23.4*	76.1*	14.2	−10330*	0.64
d4	−348*	0.0	−79.5*	23.2*	306.5*	14859*	0.87
s4	−2610*	0.4*	−111*	59*	−520*	4130*	0.98
EMU corporates high yield credit spread							
d1	−275	2.2*	40.9	−132	890	257	0.09
d2	−792	2.3*	109.9	−333.5	−1025	1029	0.12
d3	−6267*	0.2	−71.1	700.7*	117.9	2622*	0.38
d4	−4765*	1.3*	−135.9	−404.3*	5096*	4224*	0.84
s4	−9471*	2.8*	−411*	340*	−4762*	6988*	0.96

*Indicates significance at a 5% confidence level.

Table 2. Regression results for the European investment grade and high yield credit spreads on explanatory variables using reconstructed time.

continues to be important in explaining the high yield spreads though (with the exception of "d3"), which is another indication for the fact that stock market characteristics are more influential in the high yield bond markets than in the investment grade bond markets. The R^2 supports the fact that has to be performed allowing for inefficiencies in the markets. We find that the amount of the variation in credit spreads explained by the identified risk factors is highest for time horizons from 1.3 to 2.6 years and above. The R^2 at these time horizons in case of the investment grade bonds is 85–98%. Similar results are achieved to the high yield spreads. For shorter time periods, the amount of variation explained is much lower.

We therefore conclude that if information from the fundamental risk factors is allowed to be of influencing longer time periods (1.3–2.6 years and above), then the variables from Eqs (9) and (10)

are significantly linked and the amount of variation explained is high. This means if we allow information from the previous 1.3–2.6 years at scales "d4"and "s4" to be relevant, the proportion of credit spreads explained by risk factors is higher. At the short horizon, technical trading is perceived to be the most important influence in forming expectations; therefore, the insignificance of the default rate to explain credit spreads for shorter time period is in line with previous results and market data.

We conclude that aggregating over time scales "d1" to "s4" results in misleading interpretations of the influence of the various risk factors in explaining credit spreads. Only at time scales that represent medium terms, the default rate is of significant, positive influence. The amount of variation explainable with the fundamental risk factors is highest at that time scales. This supports the fact that fundamental considerations are more important in longer time periods and that inefficiencies in the credit markets are present at shorter time periods.

5. Conclusion

In this chapter, we give an overview of factor models that are applied to major capital markets. Ross' arbitrage pricing theory is chosen as the theoretical background for the stock and bond markets, since it allows to test for significant risk factors even if there are non-stationary features present in the data. In case of the corporate bond markets, Merton's approach is used to motivate which fundamental factors are chosen to explain market observations. We argue that the assumptions made in standard econometric procedures to test for significantly evaluated risk factors are responsible for the failure of finding the risk factors explain a higher proportion of developments on those markets in practice. We use the maximal overlap discrete wavelet transform to decompose the data into their time-scale components to allow for inefficiencies on capital markets and to allow for different time periods for adjustments to new information. The decomposition of the time series with wavelets in the time domain enables us to interpret data having features at different investment periods. This way we analyze the influence of various variables at different time scales. We examine the significance of risk factors and evaluate the proportion of variation explained at various time scales and find that fundamental factors are especially significant at longer time periods. Wavelet application allows for a thorough discrimination of various time horizons. The analysis is performed by the author for all major capital markets and we present new empirical research with regards to the European corporate bond market in detail as an example. A high percentage of variation in credit spreads explained by fundamental factors can be found in the medium terms (1.3–2.6 years) for investment grade and high yield corporates. We conclude that the adjustment time period to new information is crucial for explaining the credit spreads by risk factors. Aggregating over the time scales veils the fact that a higher proportion in variation of credit spreads is explainable with the fundamental factors for the medium term and that the short term is driven by other factors. These findings confirm our previous findings for major capital markets where estimation and identification of significant fundamental risk factors improved when the analyses were done on a scale-by-scale basis.

Author details

Michaela M. Kiermeier

Address all correspondence to: michaela.kiermeier@h-da.de

Department of Economics and Business Administration, University of Applied Sciences, Darmstadt, Germany

References

[1] Ross S. The arbitrage theory of capital asset pricing. Journal of Economic Theory. 1976;**13**: 341-360. DOI: 10.1016/0022-0531(76)90046-6

[2] Huang JZ, Kong W. Explaining Credit Spread Changes: Some New Evidence from Option-Adjusted Spreads of Bond Indexes, NYU Stern School of Business, Department of Finance Working Paper Series FIN-03-013. 2003

[3] Bekaert G, Engstrom E, Grenadier SR. Stock and bond returns with moody investors. Journal of Empirical Finance. 2010;**17**:867-894

[4] Cutler DM, Poterba JM, Summers LH. Speculative dynamics and the role of feedback traders. American Economic Review. 1990;**80**:63-68

[5] Ramsey JB, Lampart C. The Decomposition of Economic Relationships by Time Scale Using Wavelets. New York: New York University; 1996

[6] Kim S, Haueck IF. The relationship between financial variables and real economic activity: Evidence from spectral and wavelet analysis. Studies in Nonlinear Dynamics and Econometrics. 2003;**7**:1-18

[7] Kiermeier MM. Essays on wavelet analysis and the arbitrage pricing theory [thesis]. Florence: European University Institute; 1998

[8] Raihan S, Wen Y, Zeng B. Wavelet: A New Tool for Business Cycle Analysis. The Federal Reserve Bank of St. Louis, Working Paper 2005-050A. 2005

[9] Gallegati M, Gallegati M, Ramsey JB, Semmler W. The US wage Phillips curve across frequencies and over time. Oxford Bulletins of Economics and Statistics. 2011;**73**:489-508

[10] Gencay R, Selçuk R, Whitcher B. An Introduction to Wavelets and Other Filtering Methods in Finance and Economics. Philadelphia: Academic Press; 2009

[11] Fama EF, McBeth J. Risk, return, and equilibrium: Empirical tests. Journal of Political Economy. 1973;**83**:607-636

[12] Diebold FX, Li C. Forecasting the term structure of government bond yields. Journal of Econometrics. 2006;**130**:337-364

[13] Christensen JHE, Diebold FX, Rudebusch GD. The Affine Arbitrage-Free Class of Nelson-Siegel Term Structure Models. NBER Working Paper 13611. 2007

[14] Christensen JHE, Diebold FX, Rudebusch GD. An arbitrage-free generalized Nelson-Siegel term structure model. Paper presented at Meeting of the Royal Economic Society 2008

[15] Nelson CR, Siegel AF. Parsimonious modeling of yield curves. Journal of Business. 1987; **60**:473-489

[16] Kiermeier MM. Essay on wavelet analysis and the European term structure of interest rates. Business and Economic Horizons. 2014;**9**:18-26. DOI: 10.15208/beh.2013.19

[17] Merton RC. On the pricing of corporate debt: The risk structure of interest rates. Journal of Finance. 1974;**29**:449-470

[18] Huang J, Huang M. How much of the corporate-treasury yield spread is due to credit risk? Review of Asset Pricing Studies. 2012;**2**:153-202

[19] Albagli E, Hellwig C, Tsyvinski A. Dynamic dispersed information and the credit spread puzzle. In: NBER Working Paper. 2014. p. 19788

[20] Chen L, Collin-Dufresne P, Goldstein RS. On the relation between the credit spread puzzle and the equity premium puzzle. The Review of Financial Studies. 2009;**22**:3367-3409

[21] Güntay L, Hackbarth D. Corporate bond credit spreads and forecast dispersion. Journal of Banking and Finance. 2010;**34**:2328-2345

[22] Buraschi A, Trojani F, Vedolin A. Economic uncertainty, disagreement, and credit markets. Management Science. 2014;**60**:1281-1296

[23] McQuade T. Stochastic volatility and asset pricing puzzles. In: Working Paper. Harvard University; 2013

[24] Crowley PM. An intuitive guide to wavelets for economists. In: Bank of Finland Research Discussion Papers. 2005

Chapter 5

An Application of Wavelets to Finance: The Three-Factor Fama/French Model

Bruce D. McNevin and Joan Nix

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/intechopen.74165

Abstract

We use multi-scale analysis and a rolling 250-day window to estimate a widely used standard for empirical asset pricing. The asset pricing model employed is the Fama-French three-factor model. The model is estimated using stock returns for 49 industry stocks of US industry portfolios for the period from July 1969 to September 2017. The rolling window estimation approach allows us to capture the behavior of an investor who periodically reallocates his portfolio. Employing periodic estimates of expected return, we implement a set of long/short investment strategies based on the standard Fama-French three-factor model, and scale versions of the model. We find that during recessions, the higher scale long/short strategies tend to outperform the standard approach. Our results suggest distinct risk dynamics at specific horizons during recessions. We conclude that the information content of the economic phenomena that generate the three-factor model does not follow strict periodicity during recessions, making the wavelet approach more suitable for portfolio managers who must be prepared to rebalance portfolios during official downturns.

Keywords: wavelets, portfolio returns, investment horizon

1. Introduction

The Holy Grail of finance is an empirical asset pricing model that explains stock returns. Most models fall under the risk/return umbrella where risk is positively related to return. There are two basic models in empirical asset pricing, the standard Capital Asset Pricing Model, CAPM [15, 17, 21] and the Fama/French three-factor model, FF3 [5]. The basic idea behind the CAPM is that market movements matter a lot for capturing the relationship between risk and return. The systematic risk measure, beta, is an estimate of the sensitivity of a security or portfolio's

returns to market movements. In the risk/return world, the CAPM is considered a one-factor model in that a single factor, the market return, does all the heavy lifting. The model specification is as follows:

$$r_{it} - rf_t = \alpha_i + \beta_i^* \left(rm_t - rf_t \right) + e_t \tag{1}$$

where r_{it} = return of firm i at time t, rm_t = market return at time t, and rf_t = risk free rate at time t. The slope term, β_{it}, estimates systematic risk. The intercept, α_i, measures abnormal returns, or returns not explained by market exposure of the security or portfolio. In the context of the CAPM, α_i is expected to be zero since only non-diversifiable, also referred to as systematic or market risk, represents the risk that matters for explaining returns.

While the CAPM remains a cornerstone of financial theory, numerous empirical studies have called into question the ability of the CAPM to explain the cross-section of expected stock returns (see for instance, [3]). Several studies have used wavelets to examine the CAPM across scale. Gencay [7] first proposed the use of wavelets to estimate systematic risk in the Capital Asset Pricing Model. They estimate the beta of each stock annually for 6 wavelet scales using daily returns for the period January 1973 to November 2000 for stocks that were in the S&P 500. They find a positive relationship between portfolio returns and beta. Gencay et al. [8] extend their 2003 study by including stocks from the Germany and UK. They find that scale matters in other markets in that the relationship between portfolio returns and beta becomes stronger at high scales. Fernandez [6] applies wavelet analysis to a model of the international CAPM using a data set that consists of daily aggregate equity returns for seven emerging markets for the period 1990–2004.[1] The ICAPM[2] was estimated at 6 scales (2–128 day dynamics). Fernandez finds that market sensitivities are generally greatest at the higher scales of 5 and 6. In addition, the R^2 peaked at scales 5 and 6. She concludes that the ICAPM does its best at capturing the relationship between risk and return at the medium scale or long-term scale that for their data set is 32–128 days. An important takeaway from research employing wavelet measures of beta is that when the environment is distinguished by slowly changing features, or low frequency events the CAPMs' applicability in terms of providing a measure of systematic risk improves when using wavelets. This is consistent with the findings of Rua and Nunes [20] that employs wavelet methodology and provides evidence that market risk varies across time and over frequencies.[3]

The adage the proof of the pudding is in the eating is of particular relevance for empirical asset pricing models. Practitioners want to know if they employ a specific empirical asset pricing model will their investors benefit? The fierce competition to develop a winning model continues among various market players, especially hedge funds [2]. The prescription to basically accept that markets are efficient and form a portfolio that passively tracks the market has contributed to the growth of index investing, but has not slowed the search for a better model.

---

[1] Brazil, Chile, Mexico, Indonesia, South Korea, Malaysia, and Thailand.

[2] ICAPM for two countries $E(r_i - r) = \beta_1 \mathrm{cov}(r_i, r_w) + \beta_2 \mathrm{cov}(r_i, s)$, where r_i = returns for domestic asset, r_w = returns for world portfolio, s is the percent change in the exchange rate for domestic and foreign currency.

[3] Their application is to Emerging Markets.

http://dx.doi.org/10.5772/intechopen.74165

The idea of basically finding other factors besides the market that explain equity returns has generated many different versions of factor models. One that has gained widespread acceptance is the Fama and French three-factor model (FF3). The general consensus is that the FF3 has greater explanatory power than the CAPM. The Fama-French model adds to the explanatory power of the standard CAPM by including two additional factors, firm size and the book-to-market ratio. Both factors were found in previous research to matter for explaining equity returns. That small firms outperform large cap firms is found in Banz [1], while Barr Rosenberg, Kenneth Reid, and Ronald Lanstein [19] find a positive relationship between average stock returns and book-to-market ratio. Low B/M firms are considered "value stocks" while high B/M are "growth stocks." There is strong consensus around the idea that smaller cap firms are riskier and therefore, generating greater returns beyond what would be expected from simple market beta exposure is a widely accepted explanation for the size factor. There is less agreement for an explanation of the value premium, but one is rooted in behavior where basically relatively cheap stocks outperform relatively expensive stocks because optimism and pessimism persist among investors. Investors bid up growth stocks leading to future under performance, and keep down value stocks leading to future over-performance. Both size and B/M factors are added to their model as factors that account for returns, along with the market factor as found in the CAPM. The FF3 model is specified as follows:

$$r_{it} - rf_t = \alpha_i + \beta_i^* \left(rm_t - rf_t\right) beta_{2i}{}^* SMB_t + \beta_{3i}{}^* HML_t + e_t \tag{2}$$

where SMB_t and HML_t are the size and book-to-market factors, respectively. The book-to-market ratio is intended to capture the difference between value and growth stocks in the sense that the book-to-market ratio is high for value stocks and low for growth stocks.[4]

Several studies have examined the Fama-French 3-factor model at the scale level. Kim and In [10–14] apply wavelets to the Fama-French 3-factor model using monthly data from 1964 to 2004 for 12 industry portfolios. They find that the market variable plays an important role in explaining stock returns across all scales. In addition, they find that the estimated coefficients for the SMB and the HML are significant in specific time scales, depending on the industry. Trimtech et al. [22] apply wavelet analysis to the Fama-French model to study monthly returns for the French stock market for the period 1985. They find that the r-square of the medium and high scale versions of the Fama-French model exceed that of the standard model. They also find that the risk sensitivity of the factors depends on the time scale with the magnitude and sign of the size and book-to-market factors varying across scale.

We use multi-scale analysis and a rolling 250-day window to estimate the Fama-French 3-factor model of stock returns for 49 industry stocks of US industry portfolios. The data set, which consists of daily observations, covers the period from July 1, 1969 to September 29, 2017. We find through risk-adjusting the portfolios using the FF3 model that there are distinct risk dynamics during recessions. The rolling window estimation approach allows us to capture the behavior of an investor who periodically reallocates his portfolio. Using periodic estimates of expected return we implement a set of out-of-sample long/short investment strategies based on

[4]Book-to-market is defined and total assets less total liabilities.

the standard Fama-French model, and also the scale versions of the model. We find that for the sample as a whole the strategy based on the standard model outperforms each of the scale based strategies. In other words, frequency-based information does not appear to matter for portfolio performance when spanning the entire time period. However, during the majority of recessions, the higher scale long/short strategies tend to outperform the standard approach. The frequency content of information does appear to matter during recessions. We conclude that most recessions reflect a time-varying market regime where scale dynamics matter for portfolio performance. In terms of practioners the results suggest that an avenue for potential improvement in portfolio performance is found by taking scale into consideration when faced with potential recessionary periods.

The remainder of this chapter is organized as follows: Section 2 presents the data and basic statistics. Section 3 describes the methodology. Section 4 presents the empirical findings, and Section 5 follows with our concluding comments.

2. Data discussion

Our analysis uses daily equity returns for 49 value-weighted industry portfolios for the period July 1, 1967 to September 29, 2017. The portfolios, which are made available by Kenneth French at his website,[5] are defined by assigning each NYSE, AMEX, and NASDAQ stock to an industry at the end of June in year t, using Compustat 4 digit SIC codes for the fiscal year ending in calendar year t−1. The industry definitions, along with basic statistics for daily returns, are provided in **Table 1**. The returns, which are shown in excess of the risk free rate, range from a low of 0.002% for Real Estate to a high of 0.0522% for Tobacco. The sign of the skewness varies across industries, but the returns for all industries are leptokurtotic.

The period of analysis cover five recessions, which are listed in **Table 2**. Our analysis of the performance of the long/short portfolios across scale focuses on these five recessions.

Excess market returns (Mkt), the risk free rate (RF), and the 2 Fama-French factors (SMB and HML) are also from Kenneth French's website. Excess market returns include all NYSE, AMEX, and NASDAQ firms. The risk free rate is the 1-month Treasury bill rate. The two Fama-French factors are constructed using 6 value-weighted portfolios formed on size and book-to-market. The size factor, SMB (small minus big) is the average return on the three small portfolios minus the average return on the three big portfolios. Similarly, HML (high minus low) is the average return on the three value portfolios minus the average return on the three growth portfolios. **Table 3** contains summary statistics for Mkt, RF, SMB, and HML. The average return for the HML portfolio exceeds that of the SMB portfolio. The HMB portfolio has a small negative skew while Mkt and SMB each have a positive skew. The kurtosis for the SMB portfolio is relatively large.

[5] http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.

Sector	Name	Industry	Mean	Std.Dev	Skewness	Kurtosis
Business Equipment	Chips	Electronic equipment	0.0302	1.6051	0.415	6.48
Business Equipment	Hardw	Computers	0.0247	1.6601	0.048	11.34
Business Equipment	Softw	Computer software	0.0271	2.2342	0.023	7.55
Chemicals	Chems	Chemicals	0.0310	1.2503	−0.142	8.77
Consumer Durables	Hshld	Consumer goods	0.0227	1.0970	−0.236	9.76
Consumer Non-Durables	Agric	Agriculture	0.0305	1.4128	0.390	14.57
Consumer Non-Durables	Beer	Beer & liquor	0.0359	1.1521	−0.401	14.72
Consumer Non-Durables	Books	Printing and publishing	0.0219	1.2104	−0.284	10.71
Consumer Non-Durables	Clths	Apparel	0.0266	1.2706	−0.057	6.84
Consumer Non-Durables	Food	Food products	0.0334	0.9178	−0.044	9.93
Consumer Non-Durables	Smoke	Tobacco products	0.0523	1.4031	−0.366	7.11
Consumer Non-Durables	Soda	Cand & soda	0.0350	1.4357	−0.284	10.67
Consumer Non-Durables	Toys	Recreation	0.0163	1.4800	−0.035	18.54
Consumer Non-Durables	Txtls	Textiles	0.0268	1.3612	−0.843	22.88
Energy	Coal	Coal	0.0318	2.4043	−0.187	8.55
Energy	Mines	Non-metallic and metal	0.0274	1.6273	−0.355	9.47
Energy	Oil	Petroleum and natural gas	0.0306	1.3598	−0.126	3.94
Health	Drugs	Pharmaceutical products	0.0341	1.1545	0.205	10.03
Health	Hlth	Healthcare	0.0243	1.5270	0.471	12.22
Health	MedEq	Medical equipment	0.0309	1.1857	0.117	7.20
Manufacturing	Aero	Aircraft	0.0368	1.3506	−0.197	9.90
Manufacturing	Autos	Automobiles and trucks	0.0213	1.4643	−0.269	6.62
Manufacturing	Boxes	Shipping containers	0.0294	1.2786	−0.319	10.36
Manufacturing	ElcEq	Electrical equipment	0.0355	1.3878	−0.380	10.49
Manufacturing	FabPr	Fabricated products	0.0150	1.5073	−0.441	9.15
Manufacturing	Guns	Defense	0.0424	1.3798	0.246	16.64
Manufacturing	LabEq	Measuring and control equip.	0.0290	1.4337	−0.307	10.14
Manufacturing	Mach	Machinery	0.0272	1.3123	−0.122	7.25
Manufacturing	Paper	Business supplies	0.0276	1.1143	−0.291	14.22
Manufacturing	rubbr	Rubber and plastic products	0.0279	1.1525	−0.131	6.20
Manufacturing	Ships	Shipbuilding, railroad equip.	0.0322	1.5089	−0.296	10.80
Manufacturing	Steel	Steel works, etc.	0.0165	1.6334	−0.236	9.17
Money	Banks	Banking	0.0295	1.4384	−0.184	6.51
Money	Fin	Trading	0.0351	1.4694	−0.564	14.32

Sector	Name	Industry	Mean	Std.Dev	Skewness	Kurtosis
Money	Insur	Insurance	0.0313	1.1682	−0.484	11.24
Money	RlEst	Real estate	0.0022	1.5172	−0.355	8.30
Other	BldMt	Construction materials	0.0282	1.2248	−0.306	7.41
Other	BusSv	Business services	0.0242	1.1153	−0.196	10.40
Other	Cnstr	Construction	0.0246	1.5836	−0.175	6.79
Other	Fun	Entertainment	0.0429	1.6688	0.342	20.74
Other	Gold	Precious metals	0.0244	2.3694	−0.018	16.42
Other	Meals	Restaurants, hotels, motels	0.0301	1.2684	0.299	16.98
Other	Other	Almost nothing	0.0030	1.4295	0.226	15.55
Other	Trans	Transportation	0.0273	1.2429	−0.161	11.68
Shops	PerSv	Personal services	0.0089	1.3192	−0.092	13.46
Shops	Rtail	Retail	0.0305	1.1679	−0.438	6.14
Shops	Whlsl	Wholesale	0.0245	1.0612	−0.528	9.46
Telecommunications	Telcm	Communication	0.0266	1.1191	−0.163	13.71
Utilities	Util	Utilities	0.0239	0.8743	0.012	21.07

Table 1. Daily return statistics (%), July 1, 1967 to September 29, 2017.

Period	Duration (mos)
Nov 1973–Mar 1975	16
Jan–July 1980	6
July 1981–Nov 1982	16
July 1990–Mar 1991	8
Mar 2001–Nov 2001	8
Dec 2007–June 2009	18

Table 2. Recessions and duration in data sample.

	MKt	SMB	HML
Mean	0.0253	0.0032	0.0171
Std. Dev.	1.0248	0.5435	0.5217
Skewness	−0.5049	−1.0605	0.3507
Kurtosis	14.8116	23.2372	9.9377

Table 3. Summary statistics for model factors, daily data, July 1, 1967–September 29, 2017.

http://dx.doi.org/10.5772/intechopen.74165

Figures 1 and **2** contain the continuous wavelet power plots and time series plots of returns for Mkt, SMB, and HMB, respectively. For all three series the power tends to be highest for periods less than 256 days. Of the three series, HMB has the highest volatility of returns, and it tends to cluster around the recessionary periods. This is particularly true for the last two recessions. The SMB series has the lowest volatility, however, its power also tends to be highest during recessions.

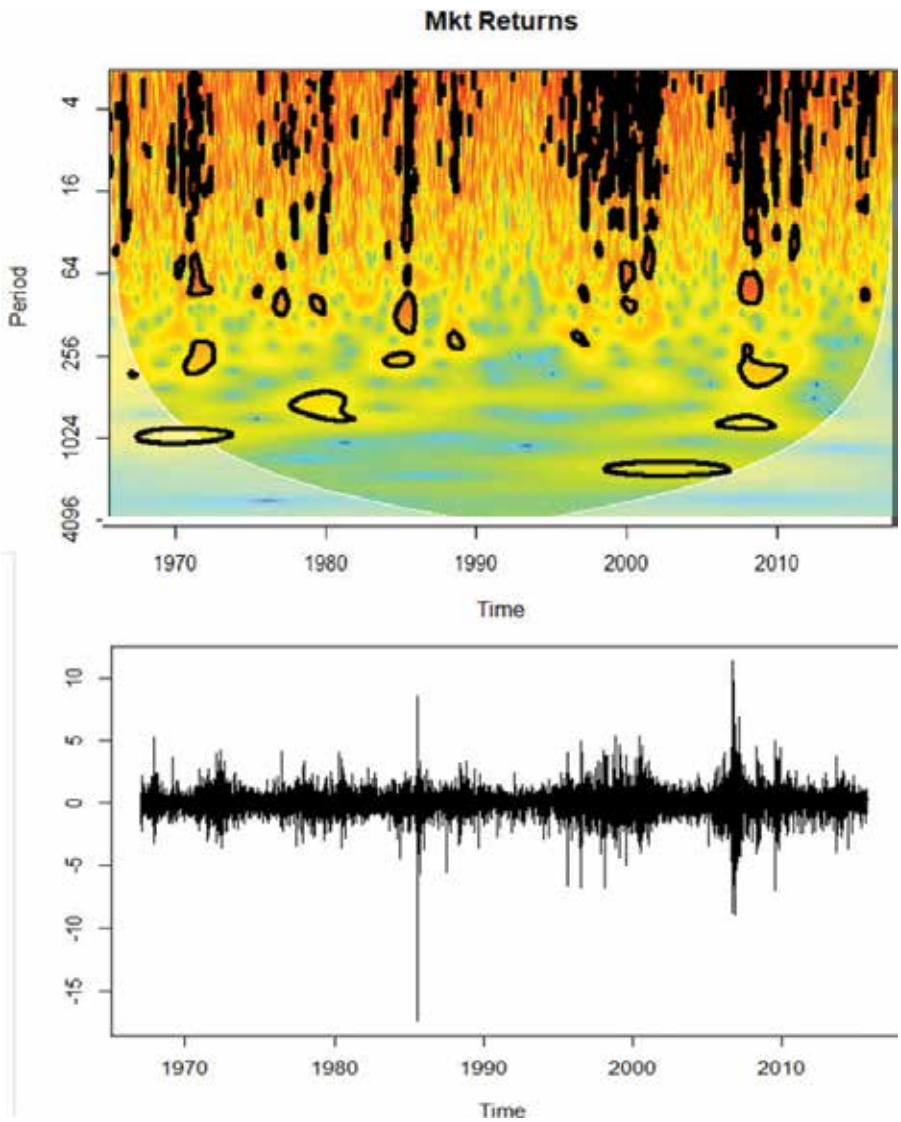


Figure 1. Mkt returns and wavelet power.

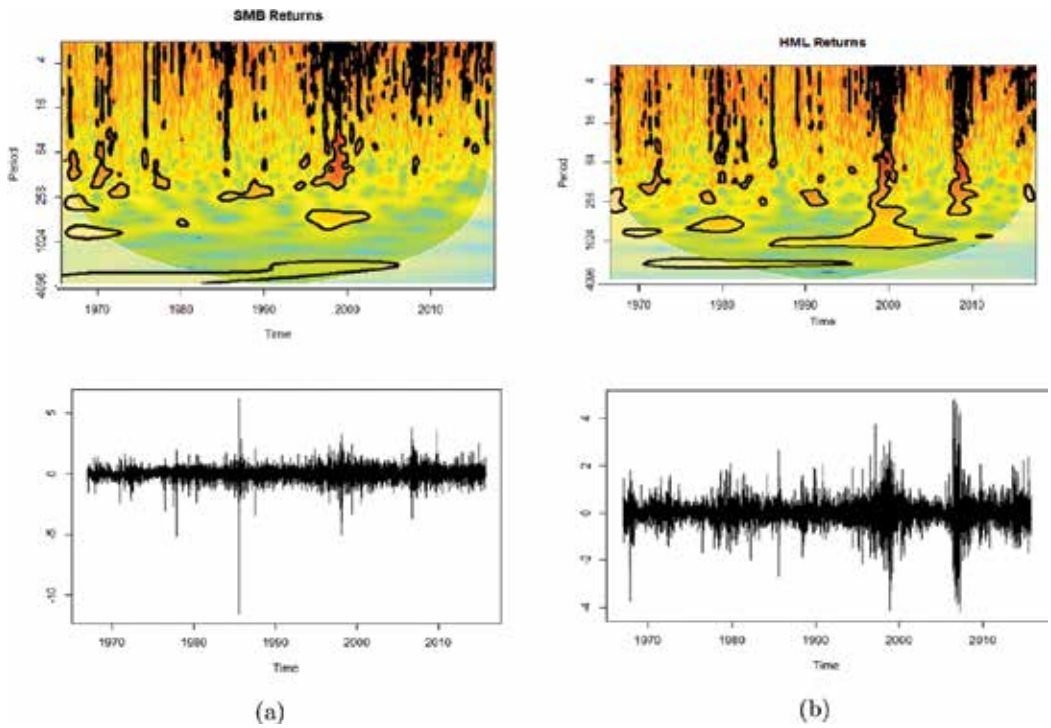


Figure 2. SMB and HML, returns and wavelet power.

3. Methodology

Our analysis of industry returns uses the Maximal Overlap Discrete Wavelet Transform (MODWT). The MODWT is calculated using a pyramid algorithm. Given a data series x_t, a high pass wavelet filter $\tilde{h_1}$, and a low pass scaling filter $\tilde{g_1}$ are applied to obtain wavelet coefficients $\tilde{w_1}$, and scaling coefficients $\tilde{v_1}$. In the second step of the pyramid, the original data series x_t is replaced by $\tilde{v_1}$ which is passed a high pass filter $\tilde{h_2}$ and a low pass filter $\tilde{g_2}$ to obtain wavelet and scaling coefficients, $\tilde{w_2}$, and $\tilde{v_2}$, respectively. This procedure is repeated up to J times where $J = \log_2(N)$. An important feature of the MODWT is that it can be applied to any sample size, while the Discrete Wavelet Transform (DWT) can only be applied to series of size 2^J.[6]

We apply MODWT to each portfolio of industry returns, as well as, the market returns (MKT), the size returns (SMB), and the book-to-market returns. For a filter we choose the Daubechies orthonormal compactly supported wavelet of length L = 8 [4], least asymmetric family. We selected J = 6, common practice in wavelet applications to empirical asset pricing models for providing a good balance in the time and frequency localization. The investment horizons we evaluate cover 2–4 days (J = 1) to 64–128 days (J = 6).

[6]See Chapter 4 of Gencay et al. [9] for additional detail.

3.1. Selecting a filter

In this section, we briefly discuss the process involved in selecting a filter. While our empirical analysis is primarily focused on results using a Daubechies Least Asymmetric filter of length L = 8, LA(8), we also provide results for two other filters to reflect the sensitivity of our results to the filter choice. These two alternative filters are the Daubechies extremal phase filter of length L = 4, DB(4), and the Coiflet filter of length L = 6, C(6).

Percival and Walden [18] point out that in selecting a filter there are two primary considerations, (1) if the filter length is too short it may introduce undesirable anomalies into the results; (2) if the filter is too long more coefficients will be affected by the boundary condition, and there will also be a decrease in the localization of the coefficients. They suggest using the smallest possible filter length that gives reasonable results. They also suggest that if one requires the filter coefficients to be aligned in time, as we do in or analysis, then the LA(8) is generally a good choice. It is not surprising that the LA(8) filter is a very common filter choice in research that applies wavelet methodology to finance.

Figure 3 compares the LA(8) wavelet filter with the two alternative filters used in our analysis. The filter lengths range from 4 to 8. The DB(4) filter has two vanishing moments; the Coiflet(6) has two vanishing moments and is nearly symmetric; the LA(8) has four vanishing moments. The greater the number of vanishing moments the smoother is the scale function.

Since our analysis employs the MODWT, we expect the results to be less sensitive to the filter choice than if we had used a DWT. As discussed in [18] MODWT details and smooths can be generated by averaging circularly shifted DWT details and smooths generated from circularly shifted time series. The averaging smooths out some of the choppiness that is found in DWT MRAs.[7]

3.2. Model specification

The specification of the Fama-French model that we estimated is as follows:

$$\begin{aligned} r_{it}(\lambda_j) - rf_t(\lambda_j) = a_i(\lambda_j) + \beta_i(\lambda_j) * \left(RM_t(\lambda_j) - RF_t(\lambda_j)\right) \\ + \beta_{2i}(\lambda_j) * SMB_t(\lambda_j) + \beta_3 i * (\lambda_j) * HML_t(\lambda_j) + e_{it}(\lambda_j) \end{aligned} \tag{3}$$

where $\lambda = 2^{j-1}$, for $j = 1, \ldots, 6$. $r_{it}(\lambda_j) - rf_t(\lambda_j)$ is the excess return for industry portfolio i and time t, and scale j. $RM_t\lambda_j$, $RM_t\lambda_j$, $SMB_t(\lambda_j)$, and $HML_t(\lambda_j)$ are the Fama-French factor for scale, j.

After we disaggregate the series to scale we use a rolling 250-day window to estimate the standard model, and each of the six scale level models. Each time we estimate the models we calculate the expected return for each industry as of the last day of the estimation period.

[7] Percival and Walden provide a comparison of DWT and MODWT smooths for various filters which shows that MODWT MRAs are less sensitive to the filter type than DWT MRAs. See pp. 195–200 in Percival and Walden for a discussion on the practical considerations of the MODWT.

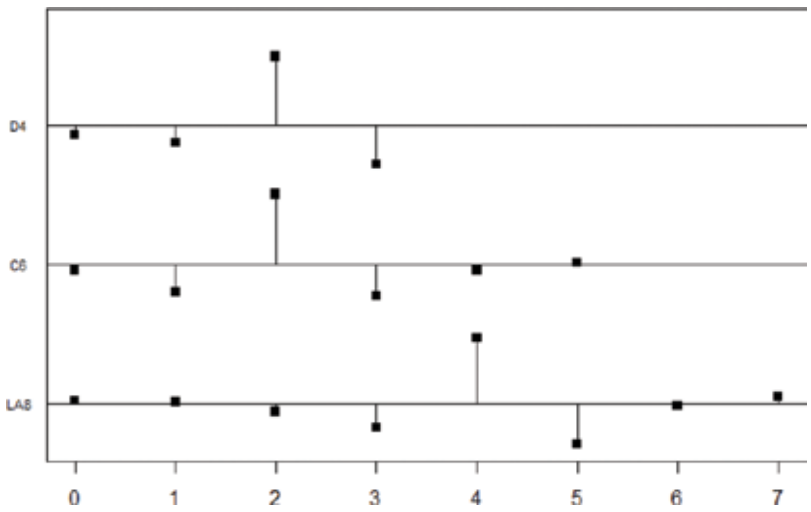


Figure 3. Three wavelet filters—DB(4), C(6), and LA(8).

We then rank the expected returns for that estimation period and assign a decile. The long-short strategy that we employ consists of going long (buying) the top decile, and going short (selling) the bottom decile. This position is held for 20 days. At the end of the 20 days period we re-estimate the models using the previous 250 days and repeat the investment selection process. Since there are 49 industry portfolios, this means that every 20 days we create a portfolio that is long 5 industries and short 5 industries. We calculate the out-of-sample cumulative returns for each 20-day period. We roll this process forward for the entire sample period.

4. Empirical findings

Our discussion of the empirical findings consists of four parts. We begin with a comparison of the parameters for the standard model parameters and the 6 scale models for the LA(8) filter. We discuss both sector averages, and industry results. Next, we examine parameter estimates for the alternative filters, DB(4) and C(6). We then discuss the returns for the long/short strategy at each scale over the entire sample period. Finally, we turn our focus to the performance of the strategies during periods of recession.

4.1. Parameter estimates

4.1.1. LA(8) filter

Table 4 contains sector level averages of the industry 'beta' parameter estimates. The difference between the standard model and the scale models for the industries tends to be modest. This is generally consistent with studies that have used monthly data to evaluate sector returns across scale. For instance using the CAPM, McNevin and Nix [16] found only small differences between the standard beta and wavelet betas for scales 1 and 2.

http://dx.doi.org/10.5772/intechopen.74165

Sector	Standard	Scale 1	Scale 2	Scale 3	Scale 4	Scale 5	Scale 6
Business Equipment	1.184	1.215	1.162	1.138	1.139	1.149	1.147
Chemicals	1.083	1.049	1.086	1.102	1.136	1.176	1.084
Consumer Durables	0.849	0.874	0.831	0.822	0.764	0.769	0.857
Consumer Non-Durables	0.889	0.887	0.891	0.890	0.897	0.879	0.907
Energy	1.108	1.092	1.140	1.141	1.102	1.189	1.018
Health	0.955	0.963	0.986	0.960	0.945	0.948	0.866
Manufacturing	1.061	1.051	1.046	1.083	1.092	1.110	1.070
Money	1.091	1.062	1.086	1.123	1.133	1.109	1.214
Other	1.015	0.996	1.022	1.051	1.043	1.033	0.982
Shops	1.014	1.020	1.015	1.012	1.031	1.038	0.987
Telecommunications	0.888	0.941	0.881	0.867	0.830	0.849	0.805
Utilities	0.709	0.707	0.713	0.729	0.739	0.708	0.728

Table 4. Average Beta parameter by sector—LA(8).

Table 15 (in Appendix) contains the industry level parameter estimates of the market variable, or the 'betas'. These parameters are averages of the rolling window estimates. There were a total of 597 rolling window regressions. On average, all of the parameter estimates in **Table 15** are significant at the 95% level of confidence. **Table 16** contains the corresponding t-statistics. There is no definitive pattern to the parameters across scale, though they tend to increase with scale.

Table 5 contains average sector parameters for the size variables. The range of parameters for the Business Equipment sector is the greatest, ranging from 0.092 for scale 1 to 0.463 for scale 6. Most of the other sectors do not exhibit a strong pattern across scale. The parameter estimates for utilities change sign across scale. In this case the sector and industry parameters are the same. An examination of **Table 18** indicates that the standard model size parameter is insignificant for the utilities, but the parameters for scales 4–6 are all negative and significant. As shown in **Table 17**, the size parameter at the industry level can vary quite a bit across scale and in comparison to the standard model indicating that in some industries investors require a premium for investing in small firm stocks over longer investment horizons. Some examples include Chips, Software, Mines, Steel, Gold, and Lab. equipment.

Table 6 contains the average sector parameter estimates for the book-to-market factor. Two sectors with notable differences across scale are Chemicals and Energy. The Chemical sector only contains a single industry. **Table 20** shows the t-statistics for the HML parameter at the industry level. On average, for the standard model the HML parameter is not statistically significant. However, it is positive and significant at scales 3–6. **Table 19** contains the industry level parameters for the HML risk factor. As is the case with SMB, the importance of the HML factor across scale varies widely by industry. Notable difference across scale can be seen in Coal, Lab. Equipment, and Construction.

Sector	Standard	Scale 1	Scale 2	Scale 3	Scale 4	Scale 5	Scale 6
Business Equipment	0.175	0.092	0.182	0.242	0.312	0.412	0.463
Chemicals	0.066	0.041	0.090	0.110	0.088	0.029	0.041
Consumer Durables	−0.270	−0.289	−0.290	−0.275	−0.175	−0.132	−0.207
Consumer Non-Durables	0.153	0.166	0.147	0.134	0.135	0.188	0.146
Energy	0.247	0.231	0.248	0.291	0.348	0.275	0.303
Health	0.158	0.186	0.164	0.139	0.114	0.120	0.179
Manufacturing	0.260	0.249	0.242	0.297	0.281	0.276	0.291
Money	0.286	0.304	0.291	0.249	0.216	0.190	0.244
Other	0.362	0.342	0.375	0.382	0.366	0.354	0.408
Shops	0.353	0.369	0.358	0.318	0.268	0.320	0.384
Telecommunications	−0.196	−0.168	−0.223	−0.207	−0.211	−0.249	−0.091
Utilities	−0.031	0.029	−0.018	−0.031	−0.147	−0.231	−0.337

Table 5. Average size parameter by sector—LA(8).

Sector	Standard	Scale 1	Scale 2	Scale 3	Scale 4	Scale 5	Scale 6
Business Equipment	−0.661	−0.640	−0.686	−0.712	−0.672	−0.645	−0.583
Chemicals	0.193	0.145	0.189	0.228	0.270	0.242	0.347
Consumer Durables	−0.231	−0.235	−0.231	−0.222	−0.260	−0.388	−0.227
Consumer Non-Durables	−0.021	−0.013	−0.016	−0.037	−0.042	−0.069	0.028
Energy	0.444	0.355	0.489	0.495	0.554	0.578	0.653
Health	−0.342	−0.268	−0.328	−0.395	−0.394	−0.411	−0.489
Manufacturing	0.188	0.202	0.176	0.197	0.189	0.103	0.152
Money	0.392	0.380	0.380	0.360	0.367	0.394	0.442
Other	0.077	0.070	0.077	0.095	0.074	0.071	0.155
Shops	−0.014	0.025	−0.003	−0.039	−0.035	−0.072	−0.114
Telecommunications	0.253	0.305	0.275	0.226	0.196	0.274	0.097
Utilities	0.418	0.372	0.422	0.435	0.467	0.513	0.395

Table 6. Average book-to-market parameter by sector—LA(8).

4.1.2. Alternative filter parameter estimates: DB(4), C(6) filters

In this section, we provide sector averages of parameter estimates for the Fama-French model based on two alternative filters.[8] **Tables 7** and **8** contain the average sector betas for the DB(4) and C(6) filters, respectively. The sector level averages for the two alternative filters are quite similar. What is important for our analysis is that they are similar to the results for the LA(8)

[8]Industry level parameter estimates and t-statistics for the alternative filters are available from the authors upon request.

http://dx.doi.org/10.5772/intechopen.74165

Sector	Scale 1	Scale 2	Scale 3	Scale 4	Scale 5	Scale 6
Business Equipment	1.210	1.162	1.141	1.141	1.143	1.137
Chemicals	1.051	1.086	1.103	1.131	1.164	1.109
Consumer Durables	0.869	0.835	0.824	0.774	0.781	0.849
Consumer Non-Durables	0.887	0.892	0.889	0.896	0.882	0.905
Energy	1.095	1.133	1.139	1.108	1.161	1.053
Health	0.964	0.983	0.961	0.946	0.949	0.879
Manufacturing	1.050	1.050	1.079	1.091	1.107	1.085
Money	1.063	1.086	1.119	1.130	1.116	1.201
Other	0.996	1.021	1.048	1.043	1.026	1.000
Shops	1.018	1.016	1.012	1.027	1.034	1.002
Telecommunications	0.937	0.883	0.863	0.835	0.848	0.814
Utilities	0.708	0.713	0.727	0.735	0.701	0.714

Table 7. Average Beta parameter by sector—DB(4).

Sector	Scale 1	Scale 2	Scale 3	Scale 4	Scale 5	Scale 6
Business Equipment	1.211	1.162	1.142	1.141	1.146	1.139
Chemicals	1.051	1.086	1.103	1.132	1.168	1.105
Consumer Durables	0.870	0.835	0.825	0.773	0.780	0.848
Consumer Non-Durables	0.887	0.891	0.889	0.896	0.882	0.907
Energy	1.095	1.133	1.141	1.109	1.162	1.043
Health	0.964	0.983	0.961	0.947	0.948	0.878
Manufacturing	1.050	1.049	1.079	1.091	1.108	1.082
Money	1.063	1.086	1.119	1.131	1.117	1.202
Other	0.996	1.021	1.048	1.043	1.028	1.000
Shops	1.018	1.016	1.012	1.028	1.035	1.002
Telecommunications	0.937	0.883	0.864	0.835	0.847	0.814
Utilities	0.708	0.713	0.727	0.735	0.701	0.716

Table 8. Average Beta parameter by sector—C(6).

filter (**Table 4**). **Tables 9** and **10** contain the sector parameter estimates for the firm size variable for the DB(4) and C(6) filters, respectively. These parameter estimates are also similar across filters. **Tables 11** and **12** show the parameters for the book-to-market variable for the alternative filters. In summary, there is very little difference in parameter estimates across the different filters.

Our comparison of paramater estimates across filters provides support that our parameter estimates based on the MODWT are not over sensitive to the choice of a filter. The remainder of

Sector	Scale 1	Scale 2	Scale 3	Scale 4	Scale 5	Scale 6
Business Equipment	0.094	0.176	0.241	0.302	0.406	0.447
Chemicals	0.043	0.083	0.108	0.089	0.026	0.042
Consumer Durables	−0.290	−0.288	−0.274	−0.187	−0.143	−0.211
Consumer Non-Durables	0.165	0.148	0.134	0.135	0.175	0.131
Energy	0.234	0.245	0.294	0.334	0.286	0.309
Health	0.185	0.164	0.140	0.121	0.122	0.147
Manufacturing	0.249	0.244	0.292	0.284	0.275	0.292
Money	0.303	0.291	0.253	0.221	0.190	0.232
Other	0.343	0.370	0.380	0.365	0.360	0.395
Shops	0.369	0.355	0.318	0.278	0.318	0.371
Telecommunications	−0.172	−0.215	−0.214	−0.212	−0.221	−0.101
Utilities	0.027	−0.015	−0.039	−0.138	−0.215	−0.315

Table 9. Average size parameter by sector—DB(4).

Sector	Scale 1	Scale 2	Scale 3	Scale 4	Scale 5	Scale 6
Business Equipment	0.094	0.176	0.242	0.301	0.404	0.446
Chemicals	0.043	0.083	0.108	0.089	0.028	0.043
Consumer Durables	−0.290	−0.287	−0.274	−0.186	−0.142	−0.209
Consumer Non-Durables	0.165	0.148	0.136	0.136	0.178	0.135
Energy	0.233	0.245	0.292	0.332	0.279	0.305
Health	0.185	0.164	0.140	0.120	0.125	0.156
Manufacturing	0.249	0.244	0.291	0.284	0.274	0.296
Money	0.303	0.291	0.255	0.222	0.188	0.231
Other	0.343	0.371	0.381	0.366	0.357	0.397
Shops	0.369	0.355	0.319	0.280	0.319	0.375
Telecommunications	−0.172	−0.215	−0.213	−0.213	−0.223	−0.105
Utilities	0.027	−0.015	−0.037	−0.139	−0.217	−0.319

Table 10. Average size parameter by sector—C(6).

the chapter focuses on the results for the LA(8) filter—a filter that is widely used in finance research employing wavelet methodology.

4.2. Long-short strategy

In this section, we review the results of the long/short strategies applied over time. We begin by examining the average statistics for the out-of-sample results for both the standard Fama-French model and each of the scales. **Table 13** presents a summary of the results.

http://dx.doi.org/10.5772/intechopen.74165

Sector	Scale 1	Scale 2	Scale 3	Scale 4	Scale 5	Scale 6
Business Equipment	−0.640	−0.685	−0.700	−0.682	−0.634	−0.611
Chemicals	0.147	0.190	0.227	0.260	0.252	0.348
Consumer Durables	−0.236	−0.226	−0.221	−0.261	−0.356	−0.225
Consumer Non-Durables	−0.014	−0.017	−0.036	−0.042	−0.061	0.022
Energy	0.366	0.475	0.500	0.552	0.570	0.624
Health	−0.271	−0.327	−0.395	−0.388	−0.419	−0.465
Manufacturing	0.201	0.181	0.195	0.184	0.117	0.160
Money	0.381	0.382	0.364	0.375	0.393	0.447
Other	0.070	0.078	0.093	0.075	0.076	0.146
Shops	0.020	−0.002	−0.036	−0.038	−0.070	−0.085
Telecommunications	0.302	0.273	0.221	0.206	0.252	0.105
Utilities	0.376	0.420	0.435	0.468	0.500	0.385

Table 11. Average book-to-market parameter by sector—DB(4).

Sector	Scale 1	Scale 2	Scale 3	Scale 4	Scale 5	Scale 6
Business Equipment	−0.640	−0.684	−0.701	−0.682	−0.637	−0.609
Chemicals	0.147	0.190	0.227	0.261	0.256	0.349
Consumer Durables	−0.236	−0.227	−0.220	−0.262	−0.358	−0.224
Consumer Non-Durables	−0.014	−0.017	−0.036	−0.042	−0.061	0.021
Energy	0.365	0.476	0.498	0.553	0.573	0.636
Health	−0.271	−0.328	−0.392	−0.389	−0.419	−0.464
Manufacturing	0.201	0.180	0.195	0.183	0.116	0.160
Money	0.381	0.381	0.363	0.373	0.391	0.442
Other	0.070	0.078	0.093	0.075	0.076	0.148
Shops	0.021	−0.002	−0.037	−0.040	−0.072	−0.090
Telecommunications	0.302	0.273	0.220	0.207	0.258	0.107
Utilities	0.376	0.420	0.435	0.469	0.502	0.391

Table 12. Average book-to-market parameter by sector—DB(4).

On average the cumulative 20-day return for the standard model (2.47%) exceeds all of the scale models. The scale 4 model has the second highest average cumulative returns (1.71%). The standard deviations are quite similar for all 7 models. The minimum and maximum cumulative returns are both quite high for all 7 models. This reflects the fact that there are only 10 positions in the out-of-sample portfolio at any point in time. It may also reflect the fact that the positions in the portfolio have equal weights (in absolute value). Finally, the Sharpe ratio for each of the models, even the standard model, is close to zero.

	Standard	Scale 1	Scale 2	Scale 3	Scale 4	Scale 5	Scale 6
Mean	2.47	−0.60	0.40	−0.54	1.71	−0.70	0.85
Std. dev	29.75	28.28	28.50	27.24	26.92	25.81	23.45
Skewness	−0.34	0.18	−0.09	−0.07	−0.09	−0.29	−0.20
Kurtosis	2.48	1.60	1.73	1.68	2.50	1.62	0.90
Minimum	−166.32	−112.08	−110.44	−117.63	−109.45	−127.75	−97.01
Maximum	106.20	115.68	105.22	100.11	125.54	87.21	74.49
Median	3.82	−0.35	1.21	−0.03	−0.07	0.35	1.08
Sharpe Ratio	0.08	−0.02	0.01	−0.02	0.06	−0.03	0.04

Table 13. Average 20-day cumulative returns for long-short strategy—LA(8).

4.3. Strategy performance during economic recessions

While the scale level model does not seem to improve the long/short strategy overall, an examination of the returns during recessions tells a different story. As shown in **Table 14** and **Figures 4–6** for four of six recessions the returns at scale level exceed those using the standard model.

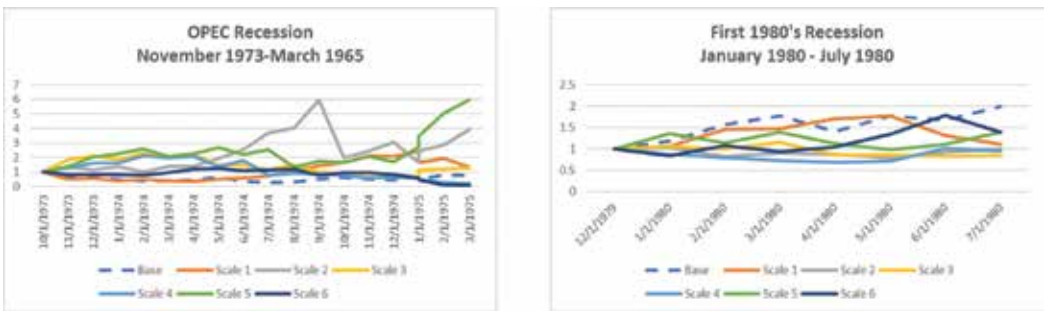


Figure 4. Out-of-sample returns—long-short strategy LA(8).

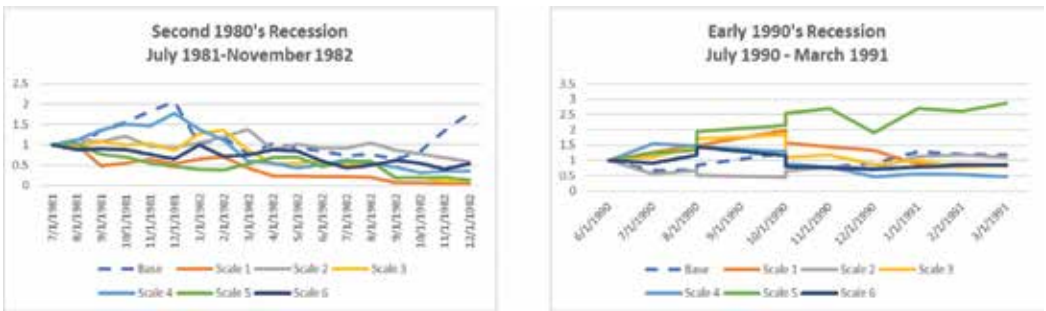


Figure 5. Out-of-sample returns—long-short strategy LA(8).

Recession	Base	Scale 1	Scale 2	Scale 3	Scale 4	Scale 5	Scale 6
Nov 1973–Mar 1975	0.78	1.35	3.94	1.28	0.22	5.97	0.10
Jan–July 1980	2.0	1.11	1.00	0.84	0.96	1.39	1.39
July 1981–Nov 1982	1.53	0.07	0.55	0.09	0.24	0.10	0.54
July 1990–Mar 1991	1.17	0.81	1.11	0.84	0.46	2.88	0.84
Mar 2001–Nov 2001	1.99	0.30	12.34	0.05	0.08	0.03	5.35
Dec 2007–June 2009	0.00	0.01	0.00	−0.41	0.11	0.00	2.35

Table 14. Cumulative out-of-sample returns during recessions—LA(8).

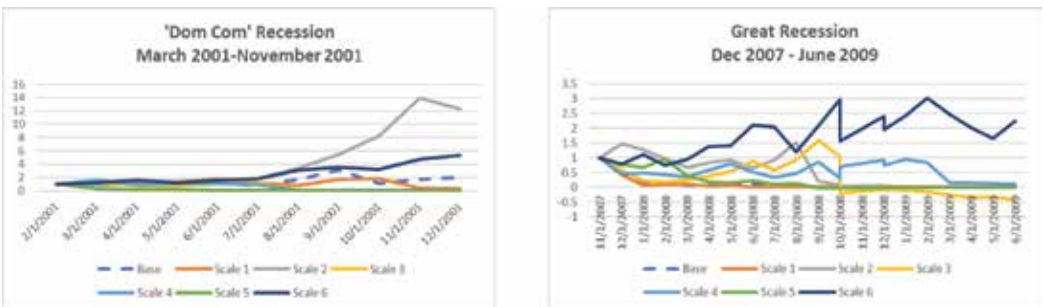


Figure 6. Out-of-sample returns—long-short strategy LA(8).

In particular, the deep recession of the 1970s, as well as, the more recent financial crisis, illustrates how scale effects matter for designing portfolios that maximize returns (**Figures 4–6** and **Table 14**).

5. Conclusion

The focus of this chapter is on whether adding wavelet methodology to the FF3 model is really "worth it." We attempt to show why it makes sense to add this methodology to the empirical asset pricing toolkit, and ultimately why practitioners should also consider including wavelet methodology in the mix of empirical asset pricing techniques used to provide advice and select portfolios for clients. The most fundamental reason for answering in the affirmative regarding whether wavelet methodology should have a seat at the table of empirical asset pricing models is that when an identified risk "signal" shows different behavior at different time periods, wavelet analysis, capable of decomposing data into several time scales, allows the researcher an opportunity to investigate the behavior of the risk factor/signal over various time scales. The exploration is richer because it allows windows to vary. Of course, allowing for risk measures that vary over time and across frequencies is not the same as finding that it will always matter for the results when compared to a standard approach devoid of such possibilities. Consistent with other research employing scale versions of the FF3 model, we find

industry-specific effects on size and HML factors that are absent using the standard model. The large-scale versus fine-scale information distinction that the scale version of the FF3 model is capable of capturing is found significant for portfolio performance during the majority of recessions included in our data. Finding that the wavelet-based version of the FF3 model produces better portfolio outcomes is of importance to practioners, as well as, researchers. Our main conclusion based on the inter-temporal behavior of financial characteristics estimated with the FF3 model is that risk measures that vary over time and across frequencies are needed to capture the risk dynamics associated with most downturns. The importance of scale effects during periods defined as recessions leads us to conclude that the distinct risk dynamics during recessions are better captured with a methodology that allows for scale effects, providing yet another reason why wavelet methodology is a worthwhile tool that belongs in the methodological toolbox of practitioners in finance.

A. Appendix

See **Tables 15–20**.

Sector	Industry	Standard	Scale 1	Scale 2	Scale 3	Scale 4	Scale 5	Scale 6
Business Equipment	Chips	1.190	1.192	1.160	1.145	1.183	1.193	1.309
Business Equipment	Hardw	1.045	1.045	0.988	0.974	1.031	1.054	1.017
Business Equipment	Softw	1.318	1.408	1.337	1.295	1.204	1.200	1.114
Chemicals	Chems	1.083	1.049	1.086	1.102	1.136	1.176	1.084
Consumer Durables	Hshld	0.849	0.874	0.831	0.822	0.764	0.769	0.857
Consumer Non-Durables	Agric	0.863	0.894	0.844	0.851	0.894	0.850	0.826
Consumer Non-Durables	Beer	0.749	0.749	0.762	0.748	0.689	0.734	0.827
Consumer Non-Durables	Books	0.913	0.872	0.916	0.929	0.932	1.020	1.071
Consumer Non-Durables	Clths	1.041	1.023	1.041	1.080	1.064	1.023	0.982
Consumer Non-Durables	Food	0.732	0.759	0.741	0.724	0.701	0.722	0.718
Consumer Non-Durables	Smoke	0.815	0.814	0.818	0.828	0.859	0.756	0.817
Consumer Non-Durables	Soda	0.810	0.846	0.798	0.741	0.775	0.742	0.892
Consumer Non-Durables	Toys	1.072	1.050	1.066	1.063	1.138	1.053	1.017
Consumer Non-Durables	Txtls	1.011	0.977	1.031	1.047	1.018	1.012	1.011
Energy	Coal	1.211	1.212	1.266	1.238	1.127	1.269	1.091
Energy	Mines	1.108	1.064	1.138	1.172	1.192	1.229	1.060
Energy	Oil	1.004	1.001	1.016	1.012	0.988	1.071	0.902

http://dx.doi.org/10.5772/intechopen.74165

Sector	Industry	Standard	Scale 1	Scale 2	Scale 3	Scale 4	Scale 5	Scale 6
Health	Drugs	0.841	0.850	0.874	0.837	0.815	0.867	0.793
Health	Hlth	1.086	1.090	1.112	1.093	1.108	1.073	1.030
Health	MedEq	0.938	0.949	0.972	0.949	0.911	0.903	0.776
Manufacturing	Aero	1.128	1.120	1.137	1.159	1.150	1.146	1.129
Manufacturing	Autos	1.216	1.208	1.199	1.220	1.269	1.193	1.151
Manufacturing	Boxes	0.962	0.963	0.960	1.005	0.960	1.000	0.986
Manufacturing	ElcEq	1.052	1.046	1.024	1.078	1.063	1.091	1.030
Manufacturing	FabPr	1.043	1.045	1.014	1.073	1.145	1.059	1.080
Manufacturing	Guns	0.868	0.883	0.816	0.848	0.863	0.900	0.858
Manufacturing	LabEq	1.113	1.096	1.107	1.105	1.126	1.125	1.108
Manufacturing	Mach	1.146	1.116	1.145	1.170	1.189	1.217	1.094
Manufacturing	Paper	0.983	0.971	0.982	0.992	1.021	1.059	1.100
Manufacturing	Rubbr	0.938	0.927	0.920	0.967	0.958	0.970	0.979
Manufacturing	Ships	0.973	0.949	0.932	1.014	1.018	1.167	1.045
Manufacturing	Steel	1.309	1.284	1.316	1.362	1.348	1.392	1.275
Money	Banks	1.149	1.102	1.157	1.177	1.194	1.163	1.269
Money	Fin	1.179	1.135	1.174	1.227	1.242	1.187	1.337
Money	Insur	1.015	0.990	1.024	1.036	1.061	1.057	1.128
Money	RlEst	1.021	1.021	0.988	1.050	1.035	1.028	1.122
Other	BldMt	1.056	1.024	1.066	1.100	1.088	1.057	1.085
Other	BusSv	1.028	1.021	1.033	1.051	1.064	1.075	1.030
Other	Cnstr	1.277	1.239	1.313	1.335	1.343	1.324	1.249
Other	Fun	1.155	1.162	1.182	1.141	1.131	1.161	1.206
Other	Gold	0.418	0.337	0.381	0.560	0.551	0.508	0.303
Other	Meals	1.006	1.011	1.002	0.990	1.010	0.967	0.926
Other	Other	1.030	1.022	1.045	1.080	0.979	1.019	1.038
Other	Trans	1.149	1.151	1.151	1.156	1.177	1.155	1.016
Shops	PerSv	1.045	1.057	1.039	1.029	1.073	1.109	1.080
Shops	Rtail	1.015	1.016	1.038	1.008	1.022	0.969	0.944
Shops	Whlsl	0.982	0.988	0.968	1.000	0.998	1.037	0.936
Telecommunications	Telcm	0.888	0.941	0.881	0.867	0.830	0.849	0.805
Utilities	Util	0.709	0.707	0.713	0.729	0.739	0.708	0.728

Table 15. Average betas by industry.

Sector	Industry	Standard	Scale 1	Scale 2	Scale 3	Scale 4	Scale 5	Scale 6
Business Equipment	Chips	20.058	18.676	18.987	20.548	22.601	24.972	31.644
Business Equipment	Hardw	14.283	13.223	13.377	14.701	16.357	17.872	20.137
Business Equipment	Softw	16.634	16.472	15.829	16.850	17.844	18.151	22.771
Chemicals	Chems	21.925	19.738	21.799	23.713	25.057	28.151	27.495
Consumer Durables	Hshld	17.380	16.028	16.498	18.202	19.852	22.056	27.869
Consumer Non-Durables	Agric	8.821	8.099	8.490	9.475	10.377	11.386	14.378
Consumer Non-Durables	Beer	10.932	9.917	10.975	11.545	11.508	13.507	17.300
Consumer Non-Durables	Books	16.103	13.937	15.865	18.200	19.563	22.655	27.925
Consumer Non-Durables	Clths	16.411	14.663	16.097	18.237	19.715	21.163	21.696
Consumer Non-Durables	Food	16.342	15.236	16.662	18.015	18.750	19.287	18.581
Consumer Non-Durables	Smoke	9.323	8.550	9.024	9.906	11.156	11.715	12.592
Consumer Non-Durables	Soda	8.714	8.010	8.347	9.161	10.212	10.908	14.548
Consumer Non-Durables	Toys	12.408	10.779	12.248	13.785	16.075	16.381	16.935
Consumer Non-Durables	Txtls	14.762	12.818	14.965	16.364	17.847	19.035	19.842
Energy	Coal	8.114	7.372	8.192	8.795	8.884	10.583	10.666
Energy	Mines	12.208	10.836	12.522	13.965	13.910	15.223	15.581
Energy	Oil	15.789	15.538	15.541	16.156	16.021	18.747	16.642
Health	Drugs	16.934	16.820	16.901	17.337	17.415	18.129	18.117
Health	Hlth	12.257	11.240	12.195	13.315	14.145	14.559	15.944
Health	MedEq	15.091	13.963	15.091	16.873	16.795	18.946	17.974
Manufacturing	Aero	16.333	14.903	15.914	18.339	19.158	20.647	22.643
Manufacturing	Autos	17.121	15.669	16.714	18.524	20.197	21.464	22.892
Manufacturing	Boxes	13.697	12.244	13.521	15.476	15.640	20.005	22.820
Manufacturing	ElcEq	17.729	16.077	17.329	19.733	20.469	23.810	25.480
Manufacturing	FabPr	11.949	10.657	11.535	13.547	15.453	16.114	18.031
Manufacturing	Guns	9.236	8.155	8.858	9.972	11.309	13.892	14.047
Manufacturing	LabEq	19.115	17.755	18.607	20.774	21.530	23.382	27.771
Manufacturing	Mach	25.353	23.134	25.554	27.713	28.412	31.287	30.476
Manufacturing	Paper	19.577	18.314	19.622	20.704	21.706	25.180	26.474
Manufacturing	Rubbr	15.763	14.015	15.287	18.207	19.327	22.154	26.683
Manufacturing	Ships	10.150	9.042	9.782	11.332	12.203	16.396	17.330
Manufacturing	Steel	18.723	17.538	18.439	20.513	20.903	22.971	23.042
Money	Banks	23.902	23.097	23.248	24.515	24.899	27.025	32.536
Money	Fin	28.065	24.606	27.841	32.581	34.215	35.430	42.598
Money	Insur	22.791	21.579	22.820	24.652	25.690	25.437	27.779
Money	RlEst	13.792	12.192	13.428	15.351	16.482	18.424	23.825

Sector	Industry	Standard	Scale 1	Scale 2	Scale 3	Scale 4	Scale 5	Scale 6
Other	BldMt	23.160	20.729	23.102	25.692	27.509	28.187	30.079
Other	BusSv	29.792	26.771	29.044	34.174	36.579	43.397	49.071
Other	Cnstr	16.446	14.352	16.451	18.835	20.246	22.905	22.818
Other	Fun	12.920	11.579	13.058	14.071	15.282	16.892	20.068
Other	Gold	2.172	1.752	1.859	2.948	3.147	2.995	2.716
Other	Meals	15.406	14.145	14.902	16.216	18.574	20.730	21.741
Other	Other	19.085	17.702	18.328	20.316	21.834	24.979	28.458
Other	Trans	20.572	19.167	19.705	22.067	23.519	26.414	23.600
Shops	PerSv	14.241	12.781	13.989	15.642	17.178	19.807	22.745
Shops	Rtail	21.002	20.071	20.629	21.615	24.246	24.585	25.547
Shops	Whlsl	24.826	22.538	24.086	27.632	30.981	36.594	37.185
Telecommunications	Telcm	17.784	17.634	16.979	18.254	18.723	20.464	21.987
Utilities	Util	17.385	17.873	18.222	18.486	17.156	17.952	20.056

Table 16. Average t-statistics for the Mkt risk parameter.

Sector	Industry	Standard	Scale 1	Scale 2	Scale 3	Scale 4	Scale 5	Scale 6
Business Equipment	Chips	0.1998	0.1560	0.1574	0.2144	0.2877	0.3544	0.3876
Business Equipment	Hardw	−0.2445	−0.4060	−0.2962	−0.1405	0.0260	0.1262	0.1714
Business Equipment	Softw	0.5691	0.5265	0.6843	0.6507	0.6234	0.7560	0.8310
Chemicals	Chems	0.0664	0.0408	0.0903	0.1104	0.0882	0.0286	0.0412
Consumer Durables	Hshld	−0.2695	−0.2885	−0.2902	−0.2755	−0.1750	−0.1320	−0.2072
Consumer Non-Durables	Agric	0.3758	0.4000	0.3346	0.4365	0.3317	0.4373	0.4856
Consumer Non-Durables	Beer	−0.1894	−0.1742	−0.1779	−0.2261	−0.2481	−0.1270	−0.1546
Consumer Non-Durables	Books	0.3068	0.3146	0.3086	0.2664	0.2520	0.2068	0.1749
Consumer Non-Durables	Clths	0.4596	0.4600	0.4668	0.4470	0.3915	0.4289	0.4416
Consumer Non-Durables	Food	−0.0900	−0.0347	−0.0801	−0.1009	−0.1251	−0.1603	−0.2084
Consumer Non-Durables	Smoke	−0.2571	−0.2403	−0.2177	−0.3084	−0.2380	−0.2543	−0.2380
Consumer Non-Durables	Soda	−0.1360	−0.0748	−0.1717	−0.2282	−0.2255	0.0180	−0.0552
Consumer Non-Durables	Toys	0.2947	0.2337	0.2728	0.3009	0.3984	0.6079	0.3526
Consumer Non-Durables	Txtls	0.6164	0.6125	0.5859	0.6208	0.6822	0.5305	0.5190
Energy	Coal	0.5192	0.5225	0.4946	0.5778	0.7556	0.5183	0.2613
Energy	Mines	0.4235	0.4023	0.4291	0.4710	0.3921	0.4007	0.7195
Energy	Oil	−0.2021	−0.2319	−0.1798	−0.1753	−0.1043	−0.0953	−0.0733
Health	Drugs	−0.2174	−0.1567	−0.2341	−0.2440	−0.2329	−0.2875	−0.3023

Sector	Industry	Standard	Scale 1	Scale 2	Scale 3	Scale 4	Scale 5	Scale 6
Health	Hlth	0.5992	0.6107	0.6146	0.6126	0.4997	0.5490	0.6607
Health	MedEq	0.0919	0.1046	0.1111	0.0494	0.0739	0.0974	0.1784
Manufacturing	Aero	0.1087	0.0759	0.1182	0.1603	0.1775	0.1178	0.1247
Manufacturing	Autos	−0.0033	−0.1007	−0.0018	0.0975	0.1487	0.1189	0.2935
Manufacturing	Boxes	0.1144	0.1614	0.1109	0.1248	0.1137	0.0077	−0.1318
Manufacturing	ElcEq	0.0835	0.0799	0.0182	0.1073	0.1636	0.1478	0.1187
Manufacturing	FabPr	0.7137	0.7314	0.7014	0.7544	0.7237	0.6921	0.6956
Manufacturing	Guns	0.0159	−0.0055	−0.0323	0.0950	0.0378	0.0723	0.0229
Manufacturing	LabEq	0.3092	0.2681	0.2857	0.2968	0.3211	0.4689	0.5188
Manufacturing	Mach	0.3527	0.3419	0.3153	0.3799	0.4194	0.4127	0.3551
Manufacturing	Paper	0.1091	0.1384	0.1146	0.1011	0.0161	0.0108	−0.0158
Manufacturing	Rubbr	0.5363	0.5332	0.4960	0.5663	0.5263	0.5307	0.3955
Manufacturing	Ships	0.3484	0.3504	0.3106	0.4258	0.2808	0.2984	0.3986
Manufacturing	Steel	0.4330	0.4086	0.4614	0.4548	0.4483	0.4327	0.7188
Money	Banks	0.0651	0.0740	0.0425	0.0219	0.0581	−0.0553	0.0099
Money	Fin	0.2349	0.2459	0.2314	0.1781	0.1735	0.1765	0.1781
Money	Insur	0.1104	0.1417	0.1584	0.0839	−0.0180	−0.0752	−0.0715
Money	RlEst	0.7355	0.7525	0.7310	0.7134	0.6523	0.7150	0.8606
Other	BldMt	0.3607	0.3596	0.3574	0.3410	0.3269	0.3968	0.2694
Other	BusSv	0.4879	0.5111	0.4951	0.4573	0.4539	0.4040	0.4887
Other	Cnstr	0.6069	0.5831	0.6786	0.5967	0.6344	0.5671	0.5042
Other	Fun	0.2688	0.2297	0.2801	0.2886	0.2084	0.4270	0.5197
Other	Gold	0.4373	0.3336	0.3802	0.6402	0.5836	0.2535	0.7318
Other	Meals	0.1580	0.1199	0.1692	0.1784	0.1175	0.2247	0.2498
Other	Other	0.3038	0.3289	0.3575	0.2765	0.3156	0.1769	0.2268
Other	Trans	0.2698	0.2699	0.2786	0.2747	0.2905	0.3822	0.2701
Shops	PerSv	0.5884	0.6341	0.6142	0.4737	0.4157	0.4531	0.7126
Shops	Rtail	0.0531	0.0586	0.0721	0.0506	0.0031	0.0346	0.0209
Shops	Whlsl	0.4161	0.4150	0.3876	0.4291	0.3849	0.4720	0.4194
Telecommunications	Telcm	−0.1961	−0.1676	−0.2229	−0.2066	−0.2111	−0.2495	−0.0907
Utilities	Util	−0.0315	0.0290	−0.0185	−0.0309	−0.1474	−0.2308	−0.3373

Table 17. Average parameters for SMB by industry.

http://dx.doi.org/10.5772/intechopen.74165

Sector	Industry	Standard	Scale 1	Scale 2	Scale 3	Scale 4	Scale 5	Scale 6
Business Equipment	Chips	2.0270	1.6195	1.5146	2.0304	2.8404	3.9464	4.8717
Business Equipment	Hardw	−2.1448	−3.1135	−2.2787	−1.4227	−0.1158	1.1290	1.8938
Business Equipment	Softw	2.4408	2.2554	2.7149	2.4298	2.6073	3.4493	5.0468
Chemicals	Chems	0.8049	0.4926	1.0285	1.1941	0.9490	0.3295	0.7415
Consumer Durables	Hshld	−3.5220	−3.3883	−3.5238	−3.5921	−2.8983	−2.1829	−4.8518
Consumer Non-Durables	Agric	2.5105	2.2923	2.1719	2.8455	2.6389	3.4121	5.8980
Consumer Non-Durables	Beer	−1.7689	−1.6097	−1.7492	−2.0344	−2.1755	−0.8634	−2.7206
Consumer Non-Durables	Books	3.2699	3.0658	3.0600	3.0247	2.8598	2.3327	2.7036
Consumer Non-Durables	Clths	4.6459	4.2592	4.4068	4.4906	4.3727	4.9556	6.6808
Consumer Non-Durables	Food	−0.7780	−0.2173	−0.7536	−0.6514	−1.4482	−1.7753	−2.1650
Consumer Non-Durables	Smoke	−1.9182	−1.6316	−1.6903	−2.2337	−1.8367	−1.8603	−3.0293
Consumer Non-Durables	Soda	−1.2139	−0.7835	−1.2947	−1.6556	−2.0505	−0.5054	−1.6006
Consumer Non-Durables	Toys	2.2761	1.8055	2.1091	2.3271	3.2539	5.5906	2.8274
Consumer Non-Durables	Txtls	5.8602	5.3274	5.3666	5.8874	6.9126	6.0350	7.7287
Energy	Coal	2.1650	2.0755	1.9526	2.3896	3.6066	2.7436	2.3730
Energy	Mines	3.1107	2.7770	3.0812	3.2862	2.9662	2.7582	5.6042
Energy	Oil	−2.0905	−2.3782	−1.7681	−1.6474	−1.0455	−0.8614	−0.0473
Health	Drugs	−2.4303	−1.6248	−2.5461	−2.7498	−2.7144	−3.3888	−4.1032
Health	Hlth	4.0472	3.8694	3.9082	3.8369	3.8314	4.3567	6.1027
Health	MedEq	1.0414	1.1860	1.1565	0.5778	0.6632	1.1281	2.7113
Manufacturing	Aero	0.7459	0.2913	0.6606	1.1866	1.6213	1.3320	2.0286
Manufacturing	Autos	0.0875	−0.5022	0.1685	0.8437	0.8694	0.8037	3.2068
Manufacturing	Boxes	0.9961	1.2750	0.9464	1.0394	0.8415	−0.1866	−2.0069
Manufacturing	ElcEq	1.3074	1.1908	0.7994	1.6185	2.1009	2.3364	1.8252
Manufacturing	FabPr	5.1538	4.7242	4.9314	5.4761	6.3364	6.1554	7.7016
Manufacturing	Guns	−0.1197	−0.4346	−0.4361	0.5318	0.2840	0.5103	0.2687
Manufacturing	LabEq	3.4862	3.0238	3.1742	3.5787	3.7878	5.3567	8.5789
Manufacturing	Mach	4.9237	4.6258	4.3522	4.9186	5.8787	5.6268	5.8431
Manufacturing	Paper	1.3213	1.5565	1.2804	1.0202	0.5431	0.0406	0.6256
Manufacturing	Rubbr	5.4436	4.8395	4.8241	5.9850	6.1595	7.2427	7.7446
Manufacturing	Ships	2.2170	2.0087	2.0331	2.6819	2.0796	2.0421	3.0639
Manufacturing	Steel	3.8506	3.5880	4.0202	3.9397	4.2276	3.9885	6.3386
Money	Banks	0.8555	1.1092	0.5180	0.2203	0.4598	−1.4619	−0.9630
Money	Fin	4.0807	3.8641	3.9287	3.5456	3.3965	2.9959	2.9406
Money	Insur	1.5832	1.9682	2.0359	1.1199	−0.1656	−0.8606	−1.7632

Sector	Industry	Standard	Scale 1	Scale 2	Scale 3	Scale 4	Scale 5	Scale 6
Money	RlEst	5.9944	5.4956	5.7325	5.9163	6.1071	7.1971	10.0976
Other	BldMt	5.3071	5.0776	5.0110	4.7677	4.9501	5.8093	4.5240
Other	BusSv	8.6517	8.1784	8.3038	8.4158	9.0863	9.5592	13.4606
Other	Cnstr	5.0753	4.4826	5.3446	4.8955	5.8743	6.6856	6.7519
Other	Fun	1.9042	1.6188	1.9848	2.1492	1.5732	3.7308	4.8763
Other	Gold	1.4068	0.9538	1.1572	2.0191	2.3464	1.1064	2.9292
Other	Meals	1.5560	1.3460	1.6073	1.5225	0.8809	2.6134	2.4165
Other	Other	3.6119	3.2794	3.6755	3.2138	4.4955	3.2650	4.9368
Other	Trans	3.0650	2.9089	3.0021	3.0266	3.4103	4.8166	3.8428
Shops	PerSv	5.0862	5.1274	4.9562	4.2433	3.8334	4.6278	9.3979
Shops	Rtail	0.8020	0.8529	0.9481	0.7217	0.0397	0.3771	−0.9400
Shops	Whlsl	6.3787	5.7385	5.6478	6.6311	7.2358	10.2466	10.6375
Telecommunications	Telcm	−2.4523	−1.9666	−2.4536	−2.4007	−3.0064	−3.6550	−1.8704
Utilities	Util	−0.2203	1.1265	0.0408	−0.2535	−1.9483	−3.0557	−5.3771

Table 18. Average t-statistics for the size parameter, SMB.

Sector	Industry	Standard	Scale 1	Scale 2	Scale 3	Scale 4	Scale 5	Scale 6
Business Equipment	Chips	−0.5377	−0.5285	−0.5752	−0.6469	−0.5214	−0.4729	−0.3322
Business Equipment	Hardw	−0.8098	−0.8607	−0.8808	−0.8076	−0.7578	−0.7139	−0.6310
Business Equipment	Softw	−0.6364	−0.5303	−0.6027	−0.6813	−0.7367	−0.7481	−0.7859
Chemicals	Chems	0.1927	0.1449	0.1889	0.2278	0.2695	0.2417	0.3472
Consumer Durables	Hshld	−0.2313	−0.2352	−0.2307	−0.2223	−0.2604	−0.3878	−0.2267
Consumer Non-Durables	Agric	0.0095	−0.0165	−0.0158	0.0992	0.0344	−0.0904	0.0845
Consumer Non-Durables	Beer	−0.2143	−0.2016	−0.1889	−0.2460	−0.2633	−0.2624	−0.0864
Consumer Non-Durables	Books	0.0862	0.0981	0.0933	0.0987	0.0568	0.0769	0.2010
Consumer Non-Durables	Clths	0.1122	0.1421	0.1134	0.0999	0.1115	−0.0405	−0.0016
Consumer Non-Durables	Food	−0.0502	−0.0040	−0.0396	−0.1035	−0.1418	−0.0782	−0.0692
Consumer Non-Durables	Smoke	−0.1436	−0.1884	−0.1401	−0.1125	−0.0478	−0.0634	−0.0241
Consumer Non-Durables	Soda	−0.1529	−0.1422	−0.1075	−0.2756	−0.2496	−0.2288	−0.0798
Consumer Non-Durables	Toys	−0.1494	−0.1311	−0.1945	−0.2309	−0.0376	−0.0763	−0.1821
Consumer Non-Durables	Txtls	0.3158	0.3260	0.3353	0.3410	0.1587	0.1391	0.4091
Energy	Coal	0.5011	0.3591	0.5568	0.5143	0.6019	0.8063	0.8292
Energy	Mines	0.3804	0.3116	0.4216	0.4245	0.5255	0.3970	0.5823
Energy	Oil	0.4514	0.3939	0.4899	0.5467	0.5338	0.5298	0.5488

http://dx.doi.org/10.5772/intechopen.74165

Sector	Industry	Standard	Scale 1	Scale 2	Scale 3	Scale 4	Scale 5	Scale 6
Health	Drugs	−0.5554	−0.5052	−0.5491	−0.5895	−0.5725	−0.6042	−0.7282
Health	Hlth	−0.1292	−0.0484	−0.1273	−0.1609	−0.2080	−0.1178	−0.2578
Health	MedEq	−0.3412	−0.2495	−0.3072	−0.4346	−0.4019	−0.5114	−0.4813
Manufacturing	Aero	0.1564	0.1499	0.1667	0.1944	0.2285	0.0375	0.1123
Manufacturing	Autos	0.5113	0.5198	0.4965	0.5262	0.5037	0.4832	0.5802
Manufacturing	Boxes	0.0993	0.1472	0.0730	0.0643	0.1148	0.0842	−0.1027
Manufacturing	ElcEq	−0.0827	−0.0828	−0.1275	−0.0159	−0.1266	−0.2090	−0.1646
Manufacturing	FabPr	0.2800	0.3523	0.2280	0.2061	0.2466	0.1813	0.1398
Manufacturing	Guns	0.1831	0.2354	0.1139	0.1712	0.1130	0.0979	0.2811
Manufacturing	LabEq	−0.3127	−0.2836	−0.2694	−0.3621	−0.3496	−0.4837	−0.5052
Manufacturing	Mach	0.1358	0.1126	0.1292	0.1661	0.1470	0.1298	0.2192
Manufacturing	Paper	0.2291	0.2275	0.2272	0.2515	0.2642	0.1278	0.3562
Manufacturing	Rubbr	0.2029	0.2476	0.2163	0.1876	0.1510	0.0427	0.0903
Manufacturing	Ships	0.2085	0.2072	0.2072	0.2523	0.1929	0.1642	0.0738
Manufacturing	Steel	0.6477	0.5966	0.6558	0.7183	0.7811	0.5844	0.7460
Money	Banks	0.5708	0.5330	0.5573	0.5681	0.5618	0.6051	0.6895
Money	Fin	0.3545	0.3418	0.3303	0.3266	0.3530	0.3187	0.3266
Money	Insur	0.3219	0.3246	0.3257	0.2904	0.3173	0.3464	0.3412
Money	RlEst	0.3227	0.3210	0.3064	0.2561	0.2363	0.3058	0.4090
Other	BldMt	0.2165	0.1679	0.2457	0.2584	0.2389	0.1483	0.2445
Other	BusSv	−0.0287	0.0044	−0.0173	−0.0309	−0.0591	−0.0452	−0.1486
Other	Cnstr	0.3317	0.3605	0.3720	0.2881	0.2687	0.4498	0.5322
Other	Fun	−0.1184	−0.0801	−0.1530	−0.1116	−0.1949	−0.2539	−0.0978
Other	Gold	0.1358	0.0125	0.0852	0.2177	0.2851	0.2729	0.7532
Other	Meals	−0.2193	−0.2227	−0.2296	−0.2285	−0.2175	−0.3029	−0.2790
Other	Other	0.0386	0.0365	0.0558	0.1012	0.0031	0.0910	0.0652
Other	Trans	0.2608	0.2783	0.2557	0.2630	0.2673	0.2069	0.1664
Shops	PerSv	0.0106	0.0754	0.0059	−0.0664	0.0108	−0.0061	−0.0195
Shops	Rtail	−0.1028	−0.0794	−0.0716	−0.1005	−0.1233	−0.2047	−0.3034
Shops	Whlsl	0.0510	0.0798	0.0577	0.0503	0.0075	−0.0043	−0.0177
Telecommunications	Telcm	0.2525	0.3053	0.2749	0.2256	0.1955	0.2742	0.0969
Utilities	Util	0.4179	0.3718	0.4217	0.4349	0.4668	0.5126	0.3949

Table 19. Average parameters for HML by industry.

Sector	Industry	Standard	Scale 1	Scale 2	Scale 3	Scale 4	Scale 5	Scale 6
Business Equipment	Chips	−3.8316	−3.6320	−3.7992	−4.6833	−4.4293	−4.9713	−3.3804
Business Equipment	Hardw	−4.8485	−4.7393	−5.0597	−5.2962	−6.1684	−6.4471	−5.7814
Business Equipment	Softw	−4.9513	−4.2904	−4.7301	−5.5315	−6.4523	−6.0185	−8.9519
Chemicals	Chems	1.5807	1.0305	1.5022	2.0940	2.7400	2.9636	4.0330
Consumer Durables	Hshld	−2.3809	−2.0684	−2.3437	−2.4775	−3.7354	−5.6384	−4.9896
Consumer Non-Durables	Agric	0.0740	−0.1157	0.0024	0.4892	0.3488	−0.5291	−0.0022
Consumer Non-Durables	Beer	−1.5311	−1.3189	−1.5000	−1.8565	−2.1101	−2.3938	−1.5355
Consumer Non-Durables	Books	0.9144	0.9247	0.9670	1.1610	0.5773	1.1321	2.4769
Consumer Non-Durables	Clths	0.9467	0.9735	0.9728	0.9999	1.2157	0.3739	0.5315
Consumer Non-Durables	Food	−0.4112	0.0323	−0.3587	−0.8661	−1.6998	−0.7677	−0.8502
Consumer Non-Durables	Smoke	−0.9420	−1.0478	−0.9154	−0.9083	−0.7504	−0.5228	−0.6924
Consumer Non-Durables	Soda	−0.8866	−0.7657	−0.7021	−1.2258	−1.5206	−1.9345	−2.9555
Consumer Non-Durables	Toys	−0.6113	−0.3523	−0.7208	−1.2232	−0.2576	−0.4538	−1.7159
Consumer Non-Durables	Txtls	2.3341	2.0792	2.3535	2.7470	1.8222	2.0718	4.8854
Energy	Coal	1.4393	0.9462	1.5578	1.7379	2.4123	2.6359	3.6478
Energy	Mines	2.0426	1.5405	2.2404	2.5477	3.3198	2.5847	3.8822
Energy	Oil	2.6581	2.3133	2.7992	3.2025	3.9142	4.3390	4.3149
Health	Drugs	−5.0907	−4.3717	−4.8950	−5.6002	−6.1844	−6.6855	−8.9212
Health	Hlth	−0.6160	−0.1308	−0.6923	−0.9001	−1.3981	−0.7229	−2.5513
Health	MedEq	−2.7898	−1.9996	−2.4451	−3.7228	−4.1971	−5.7059	−5.6987
Manufacturing	Aero	0.9648	0.6773	1.0958	1.5436	1.9928	0.6222	1.5235
Manufacturing	Autos	3.2835	2.9080	3.1879	3.9504	4.1288	4.5340	5.8731
Manufacturing	Boxes	0.6010	0.7789	0.5898	0.4788	0.7280	0.3423	−2.5781
Manufacturing	ElcEq	−0.3872	−0.3341	−0.4123	0.0537	−1.0260	−2.1100	−1.3577
Manufacturing	FabPr	1.4724	1.6384	1.1517	1.2185	1.6245	1.3601	1.6359
Manufacturing	Guns	0.6623	0.7106	0.3766	0.9148	0.6122	0.6942	1.9906
Manufacturing	LabEq	−2.2409	−1.9613	−1.9003	−2.3894	−3.1570	−4.9967	−6.9546
Manufacturing	Mach	1.4261	1.1052	1.5513	1.7953	1.6190	1.2927	2.5047
Manufacturing	Paper	1.8917	1.6192	1.9279	2.3544	2.4226	1.2835	4.0063
Manufacturing	Rubbr	1.5629	1.6763	1.5835	1.5672	1.4813	0.5814	1.0477
Manufacturing	Ships	1.1188	1.0080	1.1144	1.4306	1.5571	1.2472	0.2335
Manufacturing	Steel	4.4322	3.7856	4.3600	5.3271	6.1578	5.2726	6.4820
Money	Banks	6.5329	6.1530	6.1435	6.6598	6.7497	7.9405	10.1990
Money	Fin	4.4794	3.9067	4.2282	4.8100	5.5739	5.1944	6.3825
Money	Insur	3.8657	3.7464	3.8372	3.8679	4.0888	4.6434	5.1487
Money	RlEst	2.2449	2.0399	2.1946	1.8982	2.0844	2.7702	5.0913

Sector	Industry	Standard	Scale 1	Scale 2	Scale 3	Scale 4	Scale 5	Scale 6
Other	BldMt	2.3264	1.7688	2.4867	2.8982	2.8896	2.1000	3.7157
Other	BusSv	−0.5466	−0.1084	−0.4198	−0.5681	−1.3059	−1.3618	−4.3800
Other	Cnstr	2.1102	2.0812	2.2255	1.9296	1.9579	4.0280	5.2924
Other	Fun	−0.5212	−0.2189	−0.7152	−0.4888	−1.3764	−1.9931	−1.2353
Other	Gold	0.1164	−0.1277	−0.0720	0.3751	0.7530	0.3273	1.9220
Other	Meals	−1.5515	−1.3257	−1.5194	−1.7373	−2.2743	−3.2196	−3.3617
Other	Other	0.4663	0.5129	0.5866	0.7305	0.0857	1.0849	2.0143
Other	Trans	2.1981	2.0402	2.1532	2.5688	2.7935	2.5544	1.9368
Shops	PerSv	0.2821	0.5863	0.2316	−0.1287	0.2114	0.1905	−0.2324
Shops	Rtail	−1.2744	−1.0379	−0.9744	−1.3434	−1.8573	−2.4995	−3.1044
Shops	Whlsl	0.3727	0.5534	0.4743	0.4695	−0.2664	−0.2654	−1.1328
Telecommunications	Telcm	2.0616	2.0806	2.1254	1.9210	2.2603	3.3976	2.7911
Utilities	Util	4.8274	4.3822	4.9051	5.3088	5.5951	6.5609	5.8998

Table 20. Average t-statistics for the book-to-market parameter, HML.

Author details

Bruce D. McNevin[1] and Joan Nix[2]*

*Address all correspondence to: joan.nix@qc.cuny.edu

1 Queens College, Flushing, NY, USA

2 Bank of America, New York, NY, USA

References

[1] Banz R. The relationship between return and market value of common stocks. Journal of Financial Economics. 1981;**9**(1):3-18

[2] Berger T, Fieberg C. On portfolio optimization. The Journal of Risk Finance. 2016;**17**(3): 295-309

[3] Bollerslev T, Engle RF, Wooldridge JM. A capital asset pricing model with time-varying covariances. Journal of Political Economy. 1988;**96**(1):116-131

[4] Daubechies I. Ten Lectures on Wavelets. CBMS-NSF Lecture Notes nr. 61. SIAM; 1992

[5] Fama E, French KR. The cross-section of expected stock returns. Journal of Finance. 1992; **47**(2):427-465

[6] Fernandez V. The international CAPM and a wavelet-based decomposition of value at risk. Studies in Nonlinear Dynamics & Econometrics Article 4 ser. 9.1. 2005

[7] Gencay R, Faruk S, Brandon W. Systematic risk and timescales. Quantitative Finance. 2003;**3**:108-116

[8] Gencay R, Faruk S, Brandon W. Multiscale systematic risk. Journal of International Money and Finance. 2005;**24**:55-70

[9] Gencay R, Faruk S, Brandon W. An Introduction to Wavelets and Other Filtering Methods in Finance and Economics. New York: Academic Press; 2010

[10] In F, Kim S. The relationship between Fama-French three risk factors, industry portfolio returns, and industrial production, Francis. 2006. Available at SSRN: http://ssrn.com/abstract 891567

[11] In F, Kim S. A note on the relationship between Fama-French risk factors and innovations of ICAPM state variables. Finance Research Letters, Francis. 2007;**4**(3):165-171

[12] In F, Kim S, Faff R. Explaining mispricing with Fama French factors: New evidence from the multiscaling approach. Applied Financial Economics, Francis. 2010;**20**(4):323-330

[13] In F, Kim S. Portfolio allocation and the investment horizon: A multiscaling approach. Quantitative Finance, Francis. 2010;**10**(4):443-453

[14] In F, Kim S. Investment horizon effect on asset allocation between value and growth strategies. Economic Modelling, Francis. 2011;**28**:1489-1497

[15] Lintner J. Security prices, risk, and maximal gains from diversification. The Journal of Finance. 1965;**20**(4):587-615

[16] McNevin B, Nix J. The beta heuristic from a time/frequency perspective: A wavelet analysis of the market risk of sectors. Economic Modeling. 2017. Forthcoming

[17] Mossin J. Equilibrium in a capital asset market. Econometrica. 1966;**34**(4):768-783

[18] Percival DB, Walden AT. Wavelet Methods for Time Series Analysis. New York: Cambridge University Press; 2000

[19] Rosenberg B, Reid K, Lanstein R. Persuasive evidence of market inefficiency. The Journal of Portfolio Management. 1985;**11**(Spring, 3):9-16

[20] Rua A, Nunes LC. A wavelet based assessment of market risk: The emerging markets case. The Quarterly Review of Economics and Finance. 2012;**52**:84-92

[21] Sharpe W. Capital asset prices: A theory of market equilibrium under conditions of risk. The Journal of Finance. 1964;**19**(3):425-442

[22] Trimech A, Kortas H, Benammou A, Benammou S. Multiscale Fama? French model: Application to the French market. The Journal of Risk Finance. 2009;**10**(2):179-192

Section 4

Wavelets in Contemporary Application

Chapter 6

A Comparative Performance of Discrete Wavelet Transform Implementations Using Multiplierless

Husam Alzaq and Burak Berk Üstündağ

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/intechopen.76522

Abstract

Using discrete wavelet transform (DWT) in high-speed signal-processing applications imposes a high degree of care to hardware resource availability, latency, and power consumption. In this chapter, the design aspects and performance of multiplierless DWT is analyzed. We presented the two key multiplierless approaches, namely the distributed arithmetic algorithm (DAA) and the residue number system (RNS). We aim to estimate the performance requirements and hardware resources for each approach, allowing for the selection of proper algorithm and implementation of multi-level DAA- and RNS-based DWT. The design has been implemented and synthesized in Xilinx Virtex 6 ML605, taking advantage of Virtex 6's embedded block RAMs (BRAMs).

Keywords: discrete wavelet transform (DWT), distributed arithmetic algorithm (DAA), field programmable gate array (FPGA), residue number system (RNS), multiplierless implementation

1. Introduction

The architecture of the embedded platform plays a significant role in ensuring that real-time systems meet the performance requirements. Moreover, software development suffers from increased implementation complexity and a lack of standard methodology for partitioning the implementation of signal-processing functionalities to heterogeneous hardware platforms. For instance, digital signal processor (DSP) is cheaper, consumes less power, and is easy to develop software applications, but it has a considerable latency and less throughput compared with field programmable gate arrays (FPGAs) [1]. For high-speed signal-processing (HSP) communication systems, such as cognitive radio (CR) [2, 3] and software-defined radio (SDR) [4], DSP may fail to capture and process the received data due to data loss. In addition, implementing

applications such as finite impulse response (FIR) filtering, discrete wavelet transform (DWT), or fast Fourier transform (FFT) by software application limits the throughput, which is not sufficient to meet the requirements of high-bandwidth and high-performance applications. As a result, HSP systems are enhanced by off-loading complex signal-processing operations to hardware platforms.

Although FPGAs exhibit an increased development time and design complexity, they are preferred to meet high-performance requirements for two reasons. First, they efficiently address signal-processing tasks that can be pipelined. Second, they have the capacity to develop a programmable circuit architecture with the flexibility of computational, memory, speed, and power requirements. However, FPGA has its own resources such as memory, configurable logic blocks (CLBs), and multipliers that influence on the performance and selected algorithm. As a consequence, the choice of algorithm is determined by the hardware resource availability and performance requirements. These factors have an impact on each other and create many challenges that need to be optimized.

As an example, the discrete wavelet transform (DWT) [5–9], a linear signal-processing technique that transforms a signal from the time domain to the wavelet domain [10], employs various techniques for signal decomposing into an orthonormal time series with different frequency bands. The signal decomposition is performed using a pyramid algorithm (PA) [10, 11] or a recursive pyramid transform (RPT) [12]. While the PA algorithm is based on convolutions with quadrature mirror filters, which is infeasible for HW implementation, RPT decomposes the signal x[n] into two parts using high- and low-pass filters, which can be implemented using FIR filter [13]. **Figure 1** shows a four-tap FIR filter with four multipliers, named as multiplier accumulator (MAC). By using the MAC structure, multipliers are involved in multiplying an input with filter coefficients, b_i. It is clear that the direct implementation of the N-tap filter requires N multipliers.

This work focuses exclusively on implementing a one-level multiplierless DWT for a pattern-based cognitive communication system receiver (PBCCS) [8] by means of FPGA. The DWT is required to extract the received signal's features. Then, the extracted features are fed into a multilayer perceptron (MLP) neural network (NN) to identify the received symbol. The most challenging part is that the NN could consume most of the available multipliers inside the FPGA. As an example, Ntoune et al. [14] have implemented a real-valued time-delay neural network (RVTDNN) and real-valued recurrent neural network (RVRNN) architecture with 600

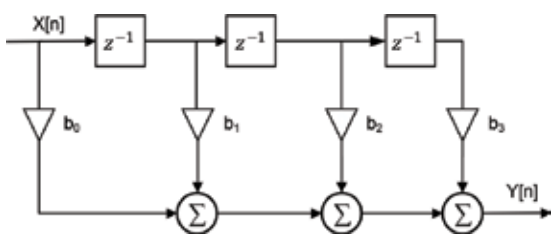


Figure 1. Four-tap finite impulse response filter.

and 720 multipliers, respectively, while ML605 [15], ZC706 [16], and VC707 [17] have 768, 900, and 2800 multipliers (DSP48Es), respectively.

Although the modern FPGAs come with a reasonable number of multipliers, designers prefer to implement multiplier-free DWT architecture for many reasons. First, a partial number of multipliers can be preserved for tasks, such as pulse shape filter, digital-up and digital-down converter that are used at SDR front-ends. Second, in contrast to DWT, the MLP weights depend on the learning step. Third, MLP weights could be frequently changed at runtime in an adaptive manner, whereas the DWT coefficients are fixed and known. Therefore, the multiplier-free DWT architecture could simplify the design process and allow the designers to focus on the MLP design.

In this work, we present the 1-D DWT implementation on FPGA by means of memory-based approaches. The aim is to compare different implementations in terms of system performance and resource consumption. We demonstrate the implementation of Daubechies wavelets (DB2, DB4, and DB5) using DAA and RNS approaches. These approaches do not employ explicit multipliers in the design. Because the main focus of this work is on extracting the key features of a signal via DWT, the inverse DWT (IDWT) and high-pass filter coefficients are not considered in this work.

1.1. Related work

Implementations of 1-D DWT for signal de-noising, feature extraction, and pattern recognition and compression can be found in [8, 9, 18, 19]. The conventional convolution-based DWT requires massive computations and consumes much area and power, which could be overcome by using the lifting-based scheme for the DWT, which is introduced by Sweldens [20]. Although, the lifting scheme is used to compute the output of low- and high-pass using fewer components, it may not be well suited to our application, owing to the PBCCS's nature, where the low-frequency components are much important than the higher ones. Therefore, in this study, 1-D DWT decomposition, which is implemented by means of filter banks, is considered. Another advantage of using convolution-based DWT over lifting approach is that they do not require temporary registers to store the intermediate results, and with an appropriate design strategy, they could have better area and power efficiency [21].

Rather than the simplest implementation of FIR filter via multipliers and an adder tree, a multiplier-free architecture is used because they result in low-complexity systems and for their high-throughput-processing capability [22]. Fundamentally, there are two techniques for facilitating parallel processing. They are the distributed arithmetic algorithm (DAA) and the residue number system (RNS). DAA is an algorithm that performs the inner product in a bit serial with the assist of a lookup table (LUT) scheme followed by shift accumulation operations [23, 24]. Several techniques have been proposed to improve the design, such as the partial sum technique [25], a multiple memory bank technique [26, 27], and an LUT-less adder-based [28]. The DAA approach has been adapted in many applications, such as least mean square (LMS) adaptive filter [29] and square-root-raised cosine filter [30].

On the other hand, RNS is an integer number system, in which the operations are performed based on the residue of division operation [31–33]. Eventually, the RNS-based results are converted back to the equivalent binary number format using a Chinese reminder theorem (CRT) [34]. The key advantage of RNS is gained by reducing an arithmetic operation to a set of concurrent, but simple, operations. Several applications, such as digital filters, benefit from the RNS implementation, for example, [35–37]. In addition, RNS was combined with DAA in one architecture, called RNS-DA [38, 39], which benefits from the advantages of both approaches.

In this chapter, three major 1-D DWT approaches are implemented on FPGA-based platforms and compared in terms of performance and energy requirements. The implementations are compared for different number of, multipliers, memory consumptions, number of taps (N), and levels (L) of the transform to show their advantages. To the best of our knowledge, no detailed comparisons of hardware implementations of the three major 1-D DWT designs exist in the study. This comparison will give significant insight on which implementation is the most suitable for given values of relevant algorithmic parameters. Although there are many efficient designs in the study, we did not optimize the number of memories in any approach, so that we have a fair comparison.

The remainder of this chapter is organized as follows. Section 2 presents the preliminaries information to understand DWT. It also reviews the theoretical background of DAA and RNS. Section 3 describes the implementation of discrete wavelet transform. We further show an analytical comparison between these approaches. Section 4 presents the performance results. Finally, this chapter concludes in Section 5.

2. Fundamentals and basic concepts

2.1. Discrete wavelet transform

The wavelet decomposition mainly depends on the orthonormal filter banks. **Figure 2** shows a two-channel wavelet structure for decomposition, where ***x[n]*** is the input signal, ***g[n]*** is the

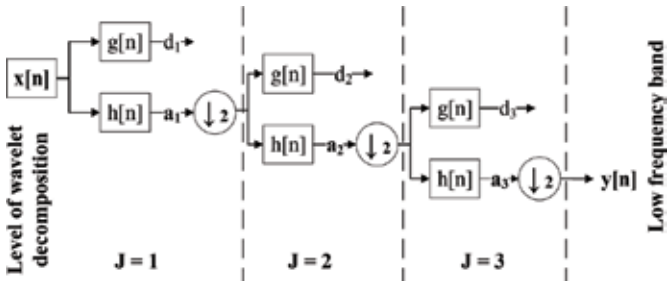


Figure 2. Multi-resolution wavelet decomposition. The block diagram of the two-channel four-level discrete wavelet transform decomposition (***J* = 3**) that decomposes a discrete signal into two parts. Note that **↓2** is maintaining one sample out of two, $\mathbf{a_i}$ and $\mathbf{d_i}$ are the approximation and details at level ***i***, respectively.

high-pass filter, ***h[n]*** is the low-pass filter, and ↓2 is the down-sampling by a factor of two. The output of each low-pass filter is fed to the next level, so that each filter creates a series of coefficients (a_i and d_i), which represent and compact the original signal information.

Mathematically, a signal ***y[n]*** consists of high- and low-frequency components, as shown in Eq. (1). It shows that the obtained signal can be represented by using half of the coefficients, because they are decimated by **2**

$$y[n] = y_{high}\,[n-1\,] + y_{low}\,[n-1] \tag{1}$$

The decimated low-pass-filtered output is recursively passed through identical filter banks to add the dimension of varying resolution at every stage. Eqs. (2) and (3) represent the filtering process through a digital low-pass filter ***h[k]*** and high-pass filter ***g[k]***, corresponding to a convolution with an impulse response of ***k***-tap filters

$$y_{low}[n] = \sum_{k} h\,[k].x\,[2n-k] \tag{2}$$

$$y_{low}[n] = \sum_{k} g\,[k].x\,[2n-k] \tag{3}$$

where ***2n*** is the down-sampling process. The outputs $\boldsymbol{y_{low}}$ **[n]** and $\boldsymbol{y_{high}}$ **[n]** provide an approximation signal and of the detailed signal, respectively [40].

2.2. Distributed arithmetic algorithm

The distributed arithmetic algorithm (DAA) gets rid of multipliers by performing the arithmetic operations in a bit-serial computation [13]. Because the down-sampling process follows each filter (as shown in **Figure 2**), Eq. (2) can be rewritten without the decimation factor as

$$y_{low}\,[n] = \sum_{k=0}^{N-1} x\,[k]\;.\;h\,[k] \tag{4}$$

Obviously, Eq. (4) requires an intensive operation due to multiplication of the real input values with the filter coefficients. Eq. (3) can be simplified by representing ***x[k]*** as a fixed point arithmetic of length ***L***:

$$x[k] = -\;x[k]_0 + \sum_{l=1}^{L-1} x\,[k]_l.2^{-l} \tag{5}$$

where $\boldsymbol{x[k]_l}$ is the l^{th} bit of ***x[k]*** and $\boldsymbol{x[k]_0}$ is the sign bit. Substituting Eq. (5) into Eq. (4), the output of the filter becomes

$$y\,[n] = \left[\sum_{l=1}^{L-1} 2^{-l}.\sum_{k=0}^{N-1} h[k].x[k]_l\right] + \sum_{k=0}^{N-1} h[k]\,\left(-x[k]_0\right) \tag{6}$$

Since $\boldsymbol{x[k]_l}$ takes the value of either **0** or **1,** $\sum_{k=0}^{N-1} h[k].x[k]_l$ may have only $\mathbf{2}^{\boldsymbol{N}}$ possible values. That is, rather than computing the summation at each iteration online, it is possible to pre-compute and store these values in a ROM, indexed by $\boldsymbol{x[k]_l}$. In other words, Eq. (6) simply realizes the sum of product computation by memory (LUT), adders, and shift register.

2.3. Residue number system

The RNS is a non-weighted number system that performs parallel carry-free addition and multiplication arithmetic. In DSP applications, which require intensive computations, the carry-free propagation allows for a concurrent computation in each residue channel. The RNS moduli set, $\boldsymbol{P = \{m_1, m_2, \ldots, m_q\}}$, consists of $\boldsymbol{q}$ channels. Each $\boldsymbol{m_i}$ represents a positive relatively prime integer; the greatest common divisor (GCD) $(\boldsymbol{m_i}, \boldsymbol{m_j}) = 1$ for $\boldsymbol{i \neq j}$.

Any number, $\boldsymbol{X \in Z_M = 0, 1, \ldots, M - 1}$, is uniquely represented in RNS by its residues $|X|_{m_i}$, which is the remainder of division $\boldsymbol{X}$ by $\boldsymbol{m_i}$ and $\boldsymbol{M}$ is defined in Eq. (7) as

$$M = \Pi_{i=1}^{q} m_i = m_1{}^{*}\ m_2{}^{*}\ \ldots\ {}^{*}\ m_q \tag{7}$$

where $\boldsymbol{M}$ determines the range of unsigned numbers in **[0, *M* - 1]**, and should be greater than the largest performed results. In addition, M uniquely represents any signed numbers. The implementation of RNS-based DWT obtained from Eq. (4) is given by Eq. (8) as follows:

$$y[n]_{m_i} = y_{m_i} = \left| \left(\sum_{k=0}^{N-1} |h\,[k]_{m_i}.x\,[n-k]_{m_i} \Big|_{m_i} \right) \right|_{m_i}. \tag{8}$$

for each $\boldsymbol{m_i \in P}$. This implies that a $\boldsymbol{q}$-channel DWT is implemented by $\boldsymbol{q}$ FIR filters that work in parallel.

Mapping from the RNS system to integers is performed by the Chinese reminder theorem (CRT) [34, 41, 42]. The CRT states that binary/decimal representation of a number can be obtained from its RNS if all elements of the moduli set are pairwise relatively prime.

Designing a robust RNS-based DWT requires selecting a moduli set and implementing the hardware design of residue to binary conversion. Most widely studied moduli sets are given as a power of two due to the attractive arithmetic properties of these modulo sets. For example, $\{2^n - 1\ ,\ 2^n,\ 2^{n+1} -\ 1\}$ [43], $\{2^n - 1\ ,\ 2^n,\ 2^n +\ 1\}$ [39], and $\{2^n,\ 2^{2n} -\ 1,\ \ 2^{2n} +\ 1\}$ [44] have been investigated.

For the purpose of illustrating, the moduli set $P_n = \{2^n - 1,\ \ 2^n,\ \ 2^{n+1} - 1\}$ is used for three reasons. First, the multiplicative adder (MA) is simple and identical for $m_1 = 2^n - 1$ and $m_3 = 2^{n+1} - 1$. Second, for small ($\boldsymbol{n = 7}$), the dynamic range of $\boldsymbol{P_7}$ is large, **M = 4,145,280**, which could efficiently express real numbers in the range $[-\mathbf{2.5, 2.5}]$ using a **16**-bit fixed-point representation, provided scaling and rounding are done properly. We assume that this interval is sufficient to map the input values, which does not exceed $\pm\mathbf{2}$. Third, the reverse converter unit is simple and regular [42] due to using simple circuits design.

http://dx.doi.org/10.5772/intechopen.76522

3. DWT implementation methodology

3.1. DWT implementation using DA

DAA hides the explicit multiplications with a ROM lookup table. The memory stores all possible values of the inner product of a fixed w-bit with any possible combination of the DWT filter coefficients. The input data, $x[n]$, are signed fixed-point of a **22**-bit width, with **16** binary-point bits ($Q_{5,16}$). We assumed that the memory contents have the same precision as the input, which is reasonable to give high enough accuracy for the fixed-point implementation. As a consequence, **22** ROMs, each consisting of **16** words, are required. Each ROM stores any possible combination of the four DWT filter coefficients, where the final result is a **22**-bit signed fixed-point ($Q_{5,16}$). In order to decrease the number of memory, the width should be reduced, which will have an impact on the output precision.

Figure 3 shows the block diagram of 1-bit DAA at position l. This block contains one ROM ($\mathbf{4} \times \mathbf{22}$) and one shift register. Because the word's length w of the input x is **22** bits, the actual design contains **22** memory blocks and **21** adders for summing up the partial results.

3.2. DWT implementation using RNS

The RNS-based DWT implementation has mainly three components. They are the forward converter, the modulo adders (MAs), and the reverse converter. The forward converter, which is also known as the binary-to-residue converter (BRC), is used to convert a binary input number to residue numbers. By contrast, the reverse converter or the residue-to-binary converter (RBC) is used to obtain the result in a binary format from the residue numbers. We refer to the RNS system, which does not include RBC, as a forward converter and modular-adders (FCMA), as illustrated in **Figure 4**.

3.2.1. The forward converter

The forward converter is used to convert the result of multiplying an input number by a wavelet coefficient to q residue numbers via LUT, shift, and modulo adders, where q is the number of channels.

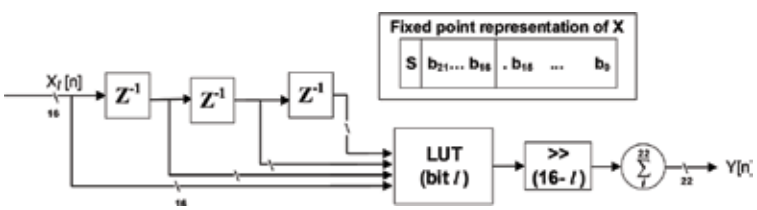


Figure 3. The block diagram of DAA-based architecture of the DB2. For simplicity, we showed one ROM and one shift register. In the actual design, there are **22** ROMs and shift registers. **>>** is a $\mathbf{16} - l$ shift operation, where **16** is the number of the binary point.

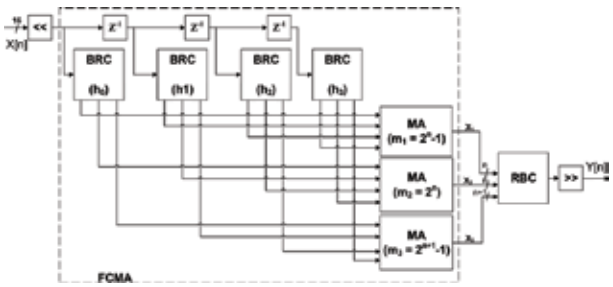


Figure 4. The block diagram of DB2 RNS-based architecture. BRC is an abbreviation for binary-to-residue converter, RBC for residue-to-binary converter, and MA for modulo adder.

Coefficient(h_k)	Real value	RNS-system value
h_0	−0.129409522550921	−266
h_1	0.224143868041857	459
h_2	0.836516303737469	1713
h_3	0.482962913144690	989

Table 1. The DB2 low-pass real and RNS-system number equivalent, multiplied by 2^{11}.

3.2.2. RNS-system number conversion

The received samples and wavelet coefficients span the real number and might take small values. One of the main drawbacks of RNS-number representation is that it only operates with positive integer numbers from [0, $\boldsymbol{M}$ **–** 1]. The DWT coefficients are generally between **1** and − **1**. As a possible solution, we have divided the range of RNS, [0, $\boldsymbol{M}$ **–** 1], to handle those numbers.

In addition, the received sample ***X[i]*** is scaled up by shifting $\boldsymbol{y}$ positions to the left (multiplying by $\mathbf{2}^{\boldsymbol{y}}$), which ensures that ***X[i]*** is a $\boldsymbol{y}$**-**bit fixed point integer. In a similar manner, the wavelet coefficients are scaled by shifting its z positions to the left. In our design, we set the filter scaling factor $\boldsymbol{z}$ to **11**. **Table 1** presents the low-pass filter of DB2 before and after scaling.

3.2.3. Modulo $\mathrm{m_i}$ *multiplier*

The multiplication of the received sample, ***X[i]***, by the filter coefficients, which are constants, is performed by indexing the ROM. As the word length, $\boldsymbol{w}$, of the received sample ***X[i]*** is increased, the memory size becomes $\mathbf{2}^{\boldsymbol{w}}$. In addition, $\boldsymbol{q}$ ROMs are required to perform the modulo multiplication.

We propose few improvements to this design. First, instead of preserving a dedicated memory for each modulo $\boldsymbol{m_i}$, a ROM that contains all module results is used. Thus, each word at location $\boldsymbol{j}$ contains the $\boldsymbol{q}$ modules of $h_k^*\ j^*\ 2^{11}$. **Figure 5** shows the internal BRC block design of the three-channel moduli set $\boldsymbol{P_7}$ **= {127, 128, 255}** with its memory map at the right top corner. It shows that, for a location $\boldsymbol{j}$, the least significant **8** bit contains $|h_k^* x|_{m_3}$, the next **7** bit

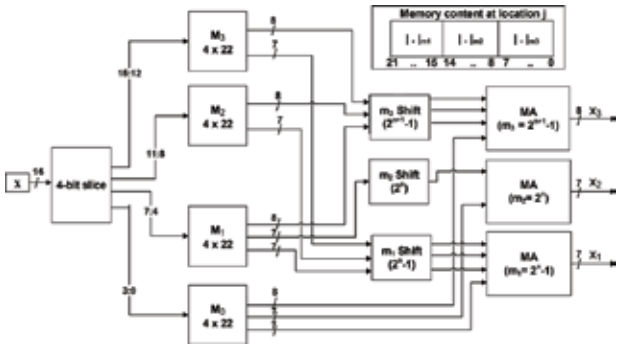


Figure 5. The block diagram of the binary-to-residue converter for the three-channel RNS-based DWT, $\mathbf{P_7 = \{127, 128, 255\}}$. Four identical memories are used for each tap. The upper corner shows the memory content at location $\mathbf{j \in [0, 15]}$.

contains $|h_k{}^* x|_{m_2}$ and the most significant **7** bit contains $|h_k{}^* x|_{m_1}$, which is generalized by Eq. (9). The advantage of this method is that no extra hardware is required to separate each module value.

$$ROM(j) = |h_k{}^* j{*}2^{11}|_{m_1}{}^* 2^{2n+1} + |h_k{}^* j{*}2^{11}|_{m_2}{}^* 2^{n+1} + \left|h_k{}^* j^* 2^{11}\right|_{m_3}, \quad j = [0, 2^w] \tag{9}$$

As with DAA-based approach, if the input word length is **16** bits, the ROM should contain $\mathbf{2^{16}}$ locations. One way to reduce the size of the memory is to divide it into four ROMs of $\mathbf{4 \times 22}$. **Figure 4** shows the block diagram of the binary-to-residue converter with four ROMs; each is indexed by four bits of $\boldsymbol{x}$. However, the output of each ROM should be combined, so that the final result can be corrected. It is worth noting that this division comes with a cost in terms of adders and registers.

According to the previous improvements, the RNS-based works are as follows. The input $X_{16-bit} = (x_1, x_2, x_3, x_4)$ is divided into four segments. Each of the **4**-bit segment is fed into one ROM, so that three outputs, corresponding to $\left|h_k{}^* x_l{*}2^{11}\right|_{m_i}$, are produced.

To obtain the final multiplications' result, each $\boldsymbol{m_i}$ output should be shifted by $\boldsymbol{l}$ positions, where $\boldsymbol{l}$ is the index of the lowest input bit (**4, 8,** or **12**). The modular multiplication and shift for ($\mathbf{2^n - 1}$) and ($\mathbf{2^{n+1} - 1}$) can be achieved by a left circular shift (left rotate) for $\boldsymbol{l}$ positions, whereas the modular multiplication and shift for $\mathbf{2^n}$ can be achieved by a left shift for $\boldsymbol{l}$ positions [17]. Finally, the modulo adder adds the corresponding output (**Figure 6**).

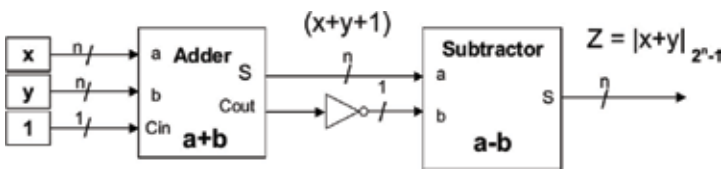


Figure 6. The block diagram of ($\mathbf{2^n - 1}$) modulo adder.

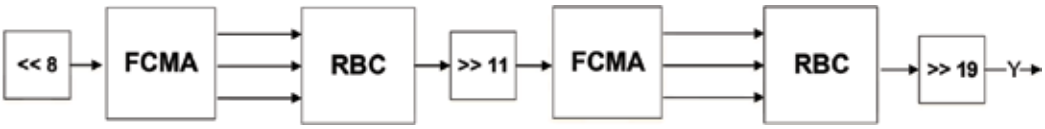


Figure 7. The block diagram of two-level RNS-based DWT design, and FCMA represents FIR-filtering process in RNS.

3.2.4. Modulo adder (MA)

The modulo adders are required for adding the results from a modular multiplier as well as for a reverse converter. In this work, we have two MAs—that is, one is based on $\mathbf{2^n}$ and the other is based on $\mathbf{2^n - 1}$. Modulo $\mathbf{2^n}$ adder is just the lowest $\boldsymbol{n}$ bits of adding two integer numbers, where the carry is ignored. **Figure 7** shows the block diagram of the $2^n - 1$ modulo adder.

3.2.5. The reverse converter

The Chinese remainder theorem (CRT) [34] provides the theoretical basis for converting a residue number into a natural integer. The moduli set $P_n = \{2^n - 1,\ 2^n,\ 2^{n+1} - 1\}$ can be efficiently implemented by four modulo adders and two multiplexers [42]. The output of the RBC is unsigned ($\mathbf{3 * n + 1}$)-bits integer number. The actual signed number can be found by shifting the result $\boldsymbol{y + z}$ positions to the left, which is equivalent to dividing by $\mathbf{2^{(y+z)}}$. $\boldsymbol{y}$ and $\boldsymbol{z}$ are the scaled values of the input and wavelet coefficients, respectively. Generally, the word length of one-level DWT is bounded by Eq. (10) and should not exceed ($\mathbf{3 * n - 2}$) bits

$$3^* \ n \ + \ 1 \ \geq \ y \ + \ z \ + \ 3 \tag{10}$$

As a consequence, the range of the moduli set should be greater than the maximum output, th_o, which can be computed as follows:

$$th_o = \left(\sum_{\mathrm{k}} \mathrm{h_k}\right)^2 * \max(\mathrm{x[n]}) * (2^{\mathrm{z}})^2 * 2^{\mathrm{y}} \leq M - 1 \tag{11}$$

where $\boldsymbol{h_k}$ is the $\boldsymbol{k^{th}}$ filter coefficient, $\boldsymbol{x[n]}$ is the input, y and z are the input and filter scaling factors, respectively, and $\boldsymbol{M}$ is the maximum range.

3.3. Two-level DWT implementation

The two-level discrete wavelet transform compromises two one-level DWTs, where the output of the first level is fed into the second level (as shown in **Figure 7**). The one-level DWT at each level is identical, but the output of each level is halved. For example, if a signal of 1800 samples is applied to the input, then 900 and 450 samples are produced by the first and second levels, respectively.

Figure 7 shows the design of two-level RNS-based DWT, which involves two FCMA blocks and two RBC blocks. Each FCMA requires converting the result of the first stage to binary, shifting the number by 11 and converting it to residue number again.

3.4. Hardware complexity

3.4.1. Memory usage

DAA and RNS techniques employ the memory as a key resource to avoid multiplying two input variables. In each approach, as the number of filter taps increases, both the size and the number of memories change. Assuming that the length of the received word is w-bit and there are N filter taps, the size of a memory element can be considered as $a \times b$, where a and b are the word length in bits of the input and output, respectively. The value of a determines the size of the memory, 2^a.

The total number of memory elements that are occupied by the DAA-based filter is $w * (N \times 22)$. The output is a fixed **16**-bit fixed point and the word length is **22** bits. The number of memory elements remains constant as the filter taps increase, whereas the size of the memory exponentially increases to 2^N.

On the other hand, the total number of memory elements occupied by an RNS-based filter is $N^*\lceil \log_2(w) \rceil * (4 \times 22)$. This equation shows that the number of memory elements increases linearly with the number of filter taps, while the memory size remains constant (4×22). **Table 2** shows a comparison of the memory usage with $w = 16$ for different DWT families.

3.4.2. Shift register and adder counts

DAA-based implementation employs shift registers and adders to sum the result at each bit level (**Figure 3**). For a word length w with m magnitude bits, we need $(w - 1)$ shift registers and $(w - 1)$ 2-input adders (data combined by a tree adder architecture). To handle the negative numbers, the two's complement operation requires additional $(m - 1)$ shift registers and $(m - 1)$ adders. Thus, for l-level DA-based implementation, a total of $l * (w - m - 2)$ shift registers and two-input adders is required.

On the other hand, for a word length w and **N**-tap filter, the q-channel FCMA implementation requires N BRC blocks and $(q*(N - 1))$ MA blocks to compute the final result. Each BRC block has $\left(\lceil \log_2 w \rceil\right) - 1, \left(\lceil \log_2 n \rceil - 1\right)$, and $\left(\lceil \log_2 w \rceil\right) - 1$ MA blocks for $2^n - 1$, 2^n, and $2^{n+1} - 1$ modulo, respectively. The modulo 2^n requires $log_2(n)$ because shifting operations is not circular and shifting **n**-bit numbers to the left by n positions or more is always zero. Likewise, the RBC has four MA blocks (for $2^{n+1} - 1$), two multiplexers, and two subtractors. Thus, the total number of MA blocks at one-level RNS-based is

	DB2	DB4	DB5
Number of filter taps	4	8	10
DA memory usage	22*(4 × 22)	22*(8 × 22)	22*(10 × 22)
RNS memory usage	16*(4 × 22)	32*(4 × 22)	40*(4 × 22)

Table 2. Occupied memories when DA- and RNS-based approaches are used. The word length, w, is 22 and 16 bits for DA- and RNS-based, respectively.

$$MA_t = 2N * \left(\left(\lceil \log_2 w \rceil - 1\right)_{2^n-1}\right) + \left(\lceil \log_2 n \rceil - 1\right)_{2^n} + \ q * (N-1) \ \ + \ 4 \tag{12}$$

For instance, three-channel DB2 implementation requires nine MA blocks to sum the result, and P_7 RNS-based implementation has a total of **45** MA blocks when $w = 16$.

Meanwhile, the number of RNS-based adders depends on the design of the MA block. For example, each MA block of ($2^n - 1$) and ($2^{n+1} - 1$) requires two adders, while each MA block of 2^n requires one adder. Thus, $a_t = \ 12\,N + \ N\left(\lceil \log_2 n \rceil - \ 1\right) + 5^*\,(N-1) + 10$ adders are required, which can be simplified as follows (summarized in **Table 3**):

$$a_t \ = \ 17\,N \ + \ N\left(\ \lceil \log_2 n \rceil - \ 1\ \right) \ + \ 10 \tag{13}$$

	DA-based	RNS-based
Memory usage	$w^*(N \times 22)$	$w - m - 2$
Number of adders	$N * \lceil \log_2 w \rceil * (4 \times 22)$	$17N + N(\lceil \log_2 n \rceil - 1) + 10$

Table 3. Memory usage and adders for **1-L N**-tap DA and RNS-based approaches DWT.

4. Performance analysis and validation

Hardware analysis was carried out by using a Xilinx System Generator for DSP (SysGen) [45], which is a high-level software tool that enables the use of MATLAB/Simulink environment to create and verify hardware designs for Xilinx FPGAs. It enables the use of the MathWorks model-based Simulink design environment for FPGA design. Furthermore, the hardware-software co-simulation design was synthesized and implemented on ML605 Xilinx Vertex 6 [15].

The implementation of RNS and DA is compared with the multiplier-accumulate-based DWT structure (MAC), as shown in **Figure 8**. We also consider the direct DWT implementation using an IP FIR Compiler 6.3 (FIR6.3) block [46], which provides a common interface to generate highly area-efficient and high-performance FIR filters (**Figure 9**).

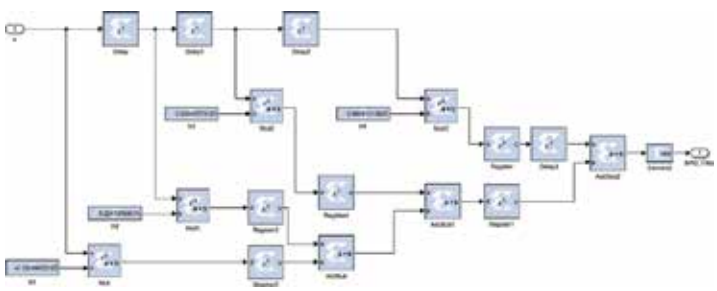


Figure 8. The Simulink model of MAC-based one-level DB2 discrete wavelet transform. Filter coefficients are stored as constants.

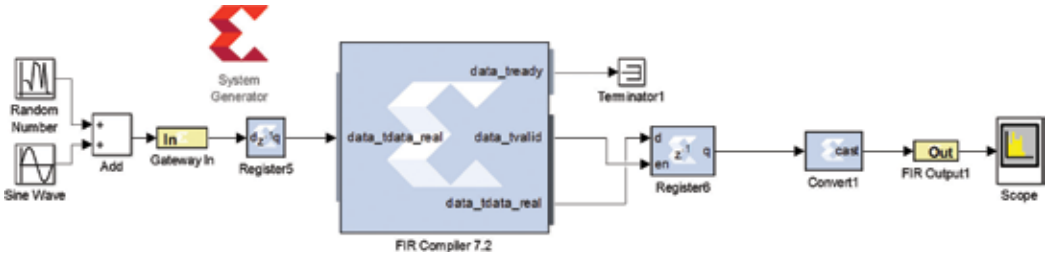


Figure 9. The Simulink model of FIR-based one-level DB2 discrete wavelet transform. The IP FIR compiler 6.3 of the system generator is used.

For RNS implementation, the moduli sets of $\boldsymbol{P_7}$ **= {127, 128, 255}** and $\boldsymbol{P_{10}}$ = **{1023, 1024, 2047}** were used. The dynamic ranges of these sets are M = 4,161,536 and 2,144,338,944, respectively. The moduli set of $\boldsymbol{P_{10}}$ is selected because its dynamic range is greater than th_o (Eq. (11) with $y = 6$, $z = 11$ and $\sum(hi) = 1.5436$). In all RNS-based implementations, the word length was set to 16 bits.

4.1. Resource utilization and system performance

Table 4 summarizes the resource use by RNS-based components—that is, FCMA and reverse converter (RBC). The RBC consumes fewer resources and less power. However, the operating frequency is equal in all models and greater than the entire RNS-based filter.

Table 5 summarizes the resource consumption of each filter implementation. It shows that the MAC and IP FIR-based implementations have four multiplier units (DSP48E1s) with maximum frequencies of **296** and **472** MHz, respectively. By contrast, the proposed approaches are more complex than MAC. However; DAA- and RNS-based implementations has **22** and **16** memory blocks (BRAMs) used to store the pre-calculated wavelet coefficients. It also shows that the number of slice registers, slice LUTs, and occupied slices of $\boldsymbol{P_{10}}$ RNS-based is greater than one of $\boldsymbol{P_7}$ because the former has **31** output signals, while the latter has **22** output signals. As a result, the number of flip-flops is increased and the number of resources is approximately

Resources	**RNS-based (n = 7)**		**RNS-based (n = 10)**	
	FCMA	**RBC**	**FCMA**	**RBC**
Number of slice registers	656	157	883	190
Number of slice LUTs	591	138	854	180
Number of RAMB18E1	16	0	16	0
Max. operating freq. (MHz)	291.2	311.62	283.85	298.67
Min. period (ns)	3.434	3.21	3.523	3.348
Estimated total power (mW)	40.5	6.59	43.08	7.33
Latency (clock cycle (CC))	6	6	6	6

Table 4. The resource use and system performance of the RNS components—that is, FCMA.

Resources	MAC	DA	FIR	RNS (n = 7)	RNS (n = 10)
Number of slice registers	282	661	167	767	1089
Number of slice LUTs	128	520	71	721	1055
Number of occupied slices	58	188	60	240	358
Number of DSP48E1s	4	0	4	0	0
Number of RAMB18E1	0	22	0	16	16
Max. operating freq. (MHz)	296.38	229.83	472.59	258.86	261.028
Min. period (ns)	3.374	4.351	2.030	3.863	3.831
Estimated total power (mW)	8.44	66.54	9.05	56.22	53.05

Table 5. The resource use and system performance of the DWT implementation for one-level DB2 implementation.

Resources	DA-based			RNS (n = 7)		
	DB2	DB4	DB5	DB2	DB4	DB5
Number of slice registers	650	737	780	767	1441	1898
Number of slice LUTs	521	539	568	721	132	1677
Number of RAMB18E1	22	22	22	16	32	40
Max. operating freq. (MHz)	232.72	205.55	223.31	258.87	265.32	258.80

Table 6. Resource use for the DWT implementation of DB2, DB4, and DB5.

increased by one-third, while the maximum frequency in both designs is greater than **235** MHz.

Table 6 shows a comparison between the DA- and RNS-based one-level DWT implementations when using larger filter banks—that is, DB4 and DB5. It shows that DAA-based implementation occupies a fixed number of RAMB18E1. The number of memory elements of DAA-based implementation is fixed and depends on the word length (**Table 2**).

However, as the number of filter taps increases, the memory size is exponentially increased to **2N**. By contrast, the number of memory elements that are used in RNS-based implementation is linearly increased as the number of filter taps is increased. Similarly, the number of memories that are used at multilevel DAA-based and RNS-based implementations with the l-level would be an aggregate of levels **1** through l.

4.2. Functionality verification

The discrete wavelet transform was simulated by means of ModelSim simulator. **Figure 10** shows that the MAC and DAA have lower latency than other approaches. It depicts that the FIR- and RNS-based of P_7 and P_{10} implementations lag behind MAC and DAA by four clock cycles.

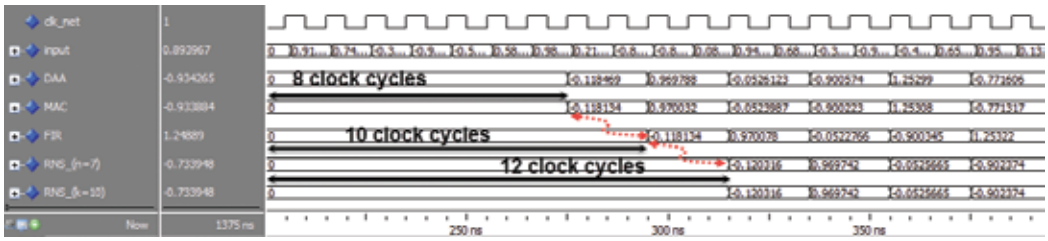


Figure 10. The output and latency of one-level DWT using a ModelSim simulator when a sin wave is applied. Each clock cycle is **10** ns.

4.3. Precision analysis

We carried out the precision analysis for the first and DWT levels, and the result is presented in **Table 7**. The output bit precision is set to $Q_{5,16}$ for all implementations. **Table 7** shows the maximum performance based on the signal-to-noise-ratio (SNR) and peak-signal-to-noise-ratio (PSNR). For P_7, we could not achieve a better accuracy with the specified scaling factors because $y + z = 19 < (3 * 7) + 1 = 22$. However, both DAA- and RNS-based approaches offer high-signal quality with a peak signal-to-noise ratio (PSNR) of **73.5** and **56.5** dB, respectively. **Figure 11** shows the effect of changing the scaling factors of P_{10} for DB2 RNS-based approach. The input scaling factor is increased from 8 to 13 bit and the filter scaling factor is increased from 11 to 18. As expected, lower scaler factors produce PSNR equal to 56 dB, whereas the maximum PSNR equal to 84 is obtained when $y = 12$ and $z = 16$.

4.4. Discussion

Hardware availability and system performance requirements are critical for selecting the appropriate architecture of the embedded platform. The number of DWT levels, filter taps, and word length has a substantial influence on the performance of the design and complexity.

Increasing the number of DWT levels has roughly the same effect on the operating frequency. Because the only change between the RNS-based with $\boldsymbol{P_7}$ or $\boldsymbol{P_{10}}$ implementations is the output signal width, and the maximum operating frequencies slightly change. Furthermore, the one-level DB2 filter bank was designed with maximum operating frequencies of **232** and **258** MHz for

Resources	**FIR**	**MAC**	**DAA-based**	**RNS-based**	
				P_7	P_7
Input precision	$Q_{5,16}$	$Q_{5,16}$	$Q_{5,16}$	$y = 8$	$y = 8$
Coefficients precision	$Q_{0,12}$	$Q_{1,15}$	$Q_{0,15}$	$z = 11$	$z = 11$
Internal word length	22 bit	22 bit	NA	22 bit	31 bit
SNR (dB)	83.2	78.7	70.4	53.41	54.78
PSNR (dB)	86.3	81.8	73.5	56.5	57.9

Table 7. The SNR and PSNR values of different DWT implementations.

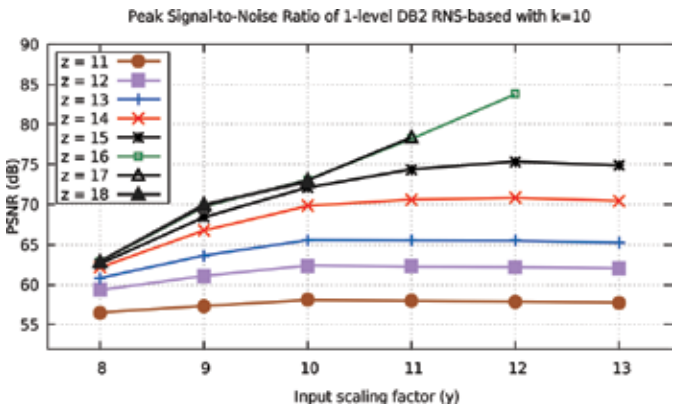


Figure 11. The impact of input and wavelet filter scaling factors of one-level RNS-based implementation with respect to P10 and P13 moduli sets on PSNR.

DAA- and RNS-based approaches, respectively. However, all high-frequency implementations introduce a latency of at least **10** clock cycles for one-level DA-based DWT.

Another critical parameter that affects the DWT performance is the filter order. DAA-based implementation outperforms the RNS-based with at most **10** taps. When the number of taps increases, the number of memory units and binary adders within the RNS-based implementation constantly increases, and the size is not affected as shown in **Table 2**. The memory requirement for DAA-based implementation is exponentially increased as the number of filter taps increases.

In addition, the two approaches have different memory content. Whereas the memory content of DAA-based implementation is consistent and identical, the memory content of RNS-based varies from tap to tap. This is obvious because each memory 590 stores the multiplication values of each filter coefficient by the moduli set.

The word length determines the number of occupied memory in both implementations. As the word length increases, the number of memory within the DAA- and RNS-based approaches increases linearly by w and $w*\log2(w)$, respectively. Furthermore, we could not neglect the effect of output word length on the accuracy and the internal structure. The DAA-based approach requires large memories to have high precision. By contrast, RNS-based approach could achieve high precision with adopting the scaling factors, which do not require any change to the design, except updating memory contents.

5. Conclusion

In this chapter, we addressed the effect of multiplierless DWT implementations, which have a substantial impact on the overall performance of the design and resource availability. We presented DAA- and RNS-based implementations of DWT and compared them with the

MAC-based approach. The former approaches are multiplierless architectures that intensively use memory to speed up the entire processing time.

Given implementation examples for experimental verifications and analysis, the approaches were simulated using Simulink and validated on a Xilinx Virtex 6 FPGA platform. The co-simulation results have also been verified and compared with the simulation environment. The complexity and optimization of multi-level DWT with respect to hardware structure provides a foundation for employing an appropriate algorithm for high-performance applications, such as in cognitive communication when combining the DWT analysis with machine-learning algorithms.

Author details

Husam Alzaq* and Burak Berk Üstündağ

*Address all correspondence to: alzaq@itu.edu.tr

Faculty of Computer Engineering, Department of Computer Engineering, Istanbul Technical University, Istanbul, Turkey

References

[1] Sklivanitis G, Gannon A, Batalama SN, Pados DA. Addressing next-generation wireless hallenges with commercial software-DefinedRadio platforms. IEEE Communications Magazine. 2016;**54**(1):59-67. DOI: 10.1109/MCOM.2016.7378427

[2] Mitola J, Maguire JGQ. Cognitive radio: Making software radios more personal. IEEE Personal Communications. 1999;**6**(4):13-18. DOI: 10.1109/98.788210

[3] Akyildiz IF, Lee WY, Vuran MC, Mohanty S. Next generation/dynamic Spectrum access/ cognitive radio wireless networks: A survey. Computer Networks. 2006;**50**(13):2127-2159. DOI: 10.1016/j.comnet.2006.05.001

[4] Mitola J. The software radio architecture. IEEE Communications Magazine. 1995;**33**(5):26-38. DOI: 10.1109/35.393001

[5] Yang P, Li Q. Wavelet transform-based feature extraction for ultrasonic flaw signal classification. Neural Computing and Applications. 2014;**24**(3–4):817-826. DOI: 10.1007/s00521-012-1305-7

[6] Madishetty SK, Madanayake A, Cintra RJ, Dimitrov VS. Precise VLSI architecture for AI based 1-D/ 2-D Daub-6 wavelet filter banks with low adder-count. IEEE Transactions on Circuits and Systems I: Regular Papers. 2014;**61**(7):1984-1993. DOI: 10.1109/TCSI.20142298283

[7] Martina M, Masera G, Roch MR, Piccinini G. Result-biased distributed-arithmetic-based filter architectures for approximately computing the DWT. IEEE Transactions on Circuits and Systems I: Regular Papers. 2015;**62**(8):2103-2113. DOI: 10.1109/TCSI.2015.2437513

[8] Alzaq H, Ustundag BB. Wavelet preprocessed neural network based receiver for low SNR communication system. In: European Wireless 2015; Proceedings of 21th European Wireless Conference; 2015. pp. 1-6

[9] Carta N, Pani D, Raffo L. Biomedical Engineering Systems and Technologies: 7th International Joint Conference, BIOSTEC 2014; 3–6 March 2014; Angers; Revised Selected Papers. Cham: Springer International Publishing; 2015. pp. 66-81. DOI: 10.1007/978-3-319-26129-4_5

[10] Mallat S. A Wavelet Tour of Signal Processing. The Sparse Way. 3rd ed. Philadelphia: Academic Press; 2008

[11] Mallat SG. A theory for multiresolution signal decomposition: The wavelet representation. IEEE Transactions on Pattern Analysis and Machine Intelligence. 1989;**11**(7):674-693. DOI: 10.1109/34.192463

[12] Vishwanath M. The recursive pyramid algorithm for the discrete wavelet transform. IEEE Transactions on Signal Processing. 1994;**42**(3):673-676. DOI: 10.1109/78.277863

[13] Vetterli M, Herley C. Wavelets and filter banks: Theory and design. IEEE Transactions on Signal Processing. 1992;**40**(9):2207-2232. DOI: 10.1109/78.157221

[14] Ntoune RSN, Bahoura M, Park CW. Power amplifier behavioral modeling by neural networks and their implementation on FPGA. In: 2012 IEEE Vehicular Technology Conference (VTC Fall); 2012. pp. 1-5

[15] Xilinx Inc. Virtex-6 FPGA ML605 Evaluation Kit. Available from: http://www.xilinx.com/products/boards-and-kits/ek-v6-ml605-g.html

[16] Xilinx Inc. Zynq-7000 All Programmable SoC ZC706 Evaluation Kit. Available from: https://www.xilinx.com/products/boards-and-kits/ek-z7-zc706-g.html

[17] Xilinx Inc. Virtex-7 FPGA VC707 Evaluation Kit. Available from: https://www.xilinx.com/products/boards-and-kits/ek-v7-vc707-g.html

[18] Chen SW, Chen YH. Hardware design and implementation of a wavelet De-noising procedure for medical signal preprocessing. Sensors. 2015;**15**(10):26396-26414. Available from: http://www.mdpi.com/1424-8220/15/10/26396

[19] Duan F, Dai L, Chang W, Chen Z, Zhu C, Li W. sEMG-based identification of hand motion commands using wavelet neural network combined with discrete wavelet transform. IEEE Transactions on Industrial Electronics. 2016;**63**(3):1923-1934

[20] Daubechies I, Sweldens W. Factoring wavelet transforms into lifting steps. Journal of Fourier Analysis and Applications. 1998;**4**(3):247-269. Available from: http://dx.doi.org/10.1007/BF02476026

[21] Meher PK, Mohanty BK, Swamy MMS. Low-area and low-power reconfigurable architecture for convolution-based 1-D DWT using 9/7 and 5/3 filters. In: 2015 28th International Conference on VLSI Design; 2015. pp. 327-332

[22] Madanayake A, Cintra RJ, Dimitrov V, Bayer F, Wahid KA, Kulasekera S, et al. Low-power VLSI architectures for DCT/DWT: Precision vs approximation for HD video, biomedical, and smart antenna applications. IEEE Circuits and Systems Magazine. 2015 Firstquarter; **15**(1):25-47

[23] Peled A, Liu B. A new hardware realization of digital filters. IEEE Transactions on Acoustics, Speech and Signal Processing. 1974;**22**(6):456-462

[24] Taylor FJ. Residue arithmetic a tutorial with examples. Computer. 1984;**17**(5):50-62

[25] White SA. Applications of distributed arithmetic to digital signal processing: A tutorial review. IEEE ASSP Magazine. 1989;**6**(3):4-19

[26] Jeng SS, Lin HC, Chang SM. FPGA implementation of fir filter using M-bit parallel distributed arithmetic. In: 2006 IEEE International Symposium on Circuits and Systems; 2006. pp. 4-878

[27] Meher PK, Chandrasekaran S, Amira A. FPGA realization of FIR filters by efficient and flexible Systolization using distributed arithmetic. IEEE Transactions on Signal Processing. 2008;**56**(7):3009-3017

[28] Yoo H, Anderson DV. Hardware-efficient Distributed Arithmetic Architecture for High-Order Digital Filters. In: Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'05). Vol. 5; 2005. pp. v/125-v/128

[29] Allred DJ, Huang W, Krishnan V, Yoo H, Anderson DV. An FPGA Implementation for a High Throughput Adaptive Filter using Distributed Arithmetic. In: 12th Annual IEEE Symposium on Field-programmable custom computing machines. FCCM 2004; 2004. pp. 324-325

[30] Srividya P, Nataraj KR, Rekha KR. FPGA implementation of multiplier less matched filters to transmit video signals over satellites. In: 2014 International Conference on Communications and Signal Processing (ICCSP); 2014. pp. 602-606

[31] Pontarelli S, Cardarilli G, Re M, Salsano A. Optimized implementation of RNS FIR filters based on FPGAs. Journal of Signal Processing Systems. 2012;**67**(3):201-212. Available from: http://dx.doi.org/10.1007/s11265-010-0537-y

[32] Jenkins W, Leon B. The use of residue number systems in the design of finite impulse response digital filters. IEEE Transactions on Circuits and Systems. 1977;**24**(4):191-201

[33] Chang CH, Molahosseini AS, Zarandi AAE, Tay TF. Residue number systems: A new paradigm to Datapath optimization for low-power and high-performance digital signal processing applications. IEEE Circuits and Systems Magazine. 2015 Fourthquarter;**15**(4): 26-44

[34] Rosen KH. Elementary Number Theory and Its Applications. 5th ed. Reading: Addison-Wesley; 2004

[35] Ramírez J, Meyer-Base U, Taylor F, García A, Lloris A. Design and Implementation of High-Performance RNS Wavelet Processors Using Custom IC Technologies. Journal of VLSI Signal Processing Systems for Signal, Image and Video Technology. 2003;**34**(3):227-237. Available from: http://dx.doi.org/10.1023/A:1023296218588

[36] Cardarilli GC, Nannarelli A, Petricca M, Re M. Characterization of RNS multiply-add units for power efficient DSP. In: 2015 IEEE 58th International Midwest Symposium on Circuits and Systems (MWSCAS); 2015. pp. 1-4

[37] Conway R, Nelson J. Improved RNS FIR filter architectures. IEEE Transactions on Circuits and Systems II: Express Briefs. 2004;**51**(1):26-28

[38] Ramírez J, García A, Meyer-Base U, Taylor F, Lloris A. Implementation of rns-based distributed arithmetic discrete wavelet transform architectures using field-programmable logic. Journal of VLSI Signal Processing Systems. 2002;**33**(1):171-120. Available from: http://dx.doi.org/10.1023/A:1021158221825

[39] Vun CH, Premkumar AB, Zhang W. A new RNS based DA approach for inner product computation. IEEE Transactions on Circuits and Systems I: Regular Papers. 2013;**60**(8): 2139-2152

[40] Daubechies I. Ten Lectures on Wavelets. Philadelphia: Society for Industrial and Applied Mathematics; 1992

[41] Mohan PVA. RNS-to-binary converter for a new three-moduli set $2n+1 - 1$, $2n$, $2n - 1$. IEEE Transactions on Circuits and Systems II: Express Briefs. 2007;**54**(9):775-779

[42] Lin SH, hwa Sheu M, Wang CH, Kuo YC. Area-time-power efficient VLSI design for residue-to-binary converter based on moduli set ($2n$, $2n+1 - 1$, $2n + 1$). In: IEEE Asia Pacific Conference on Circuits and systems. APCCAS 2008; 2008. pp. 168-171

[43] Reddy KS, Bajaj S, Kumar SS. Shift add approach based implementation of RNS-FIR filter using modified product encoder. In: TENCON 2014–2014 IEEE Region 10 Conference; 2014. pp. 1-6

[44] Hariri A, Navi K, Rastegar R. A New High Dynamic Range Moduli Set with Efficient Reverse Converter. Computers & Mathematics with Applications. 2008;**55**(4):660-668. Available from: http://www.sciencedirect.com/science/article/pii/S0898122107004993

[45] Xilinx Inc. System Generator for DSP. Available from: http://www.xilinx.com/products/design-tools/vivado/integration/sysgen.html

[46] Xilinx. LogiCORE IP FIR Compiler v6.3; 2011. DS795. Available from: http://www.xilinx.com/support/documentation/ip_documentation/fir_compiler/v6_3/ds795_fir_compiler.pdf

Chapter 7

Application of Wavelet Decomposition and Phase Space Reconstruction in Urban Water Consumption Forecasting: Chaotic Approach (Case Study)

Peyman Yousefi, Gholamreza Naser and Hadi Mohammadi

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/intechopen.76537

Abstract

The forecasting of future value of water consumption in an urban area is highly complex and nonlinear. It often exhibits a high degree of spatial and temporal variability. It is a crucial factor for long-term sustainable management and improvement of the operation of urban water allocation system. This chapter will study the application of two pre-processing phase space reconstruction (PSR) and wavelet decomposition transform (WDT) methods to investigate the behavior of time series to forecast short-term water demand value of Kelowna City (BC, Canada). The research proposes two pre-process technique to improve the accuracy of the models. Artificial neural networks (ANNs), gene expression programming (GEP) and multilinear regression (MLR) methods are the tools that considered for forecasting the demand values. Evaluation of the tools is based on two steps with and without applying the pre-processing methods. Moreover, autocorrelation function (ACF) is used to calculate the lag time. Correlation dimension is used to study the chaotic behavior of the dataset. The models' relative performance is compared using three different fitness indexes; coefficient of determination (CD), root mean square error (RMSE) and mean absolute error (MAE). The results showed how pre-processing combination of WDT and PSR improved the performance of the models in forecasting short-term demand values.

Keywords: artificial neural network, correlation dimension, chaos, gene expression programming, Kelowna, water demand, wavelet

1. Introduction

Climate change significantly affects the water availability all around the world. This effect plays a crucial role in arid and semiarid regions. On the other hand, urban development, population growth, industrial development and economic expansion also increases water scarcity concerns critically worldwide. Therefore, the governments have to be prepared beforehand for any consequences related to water problems, especially drinking water. The efficient operation and a management plan of urban water supply requires information about the value of consumption in the future. For using different standards to simulate hydraulic constitutions in pipeline systems (to improve the reliability of the system), it is necessary to have an accurate simulation of consumption value in a specific period. In other words, "The purpose of water demand forecast is to demonstrate futuristic information available for public water suppliers as they conduct their business" [1, 2]. Short-term (e.g., less than a week), mid-term (e.g., weekly to monthly) and long-term (e.g., greater than monthly) period forecast demand values are critical for daily operations and future management of the system. Long-term urban demand forecasting (up to 25 years), mid-term (up to 2 years) and short-term values (up to 2 days) depends upon vital factors such as water supply planning, pipeline maintenance, and water distribution system optimization (e.g. optimized pumping, pipeline maintenance, minimize energy cost and water supply cost, improving system reliability and water quality), respectively [3–5]. While studies have advanced the understanding of nonlinear characteristics and high complexity of water consumption factors, further research is still required. The present accepted knowledge for these factors is still limited and depends upon (1) accurate estimation and forecast water consumption and (2) determination of type and degree of nonlinearity among the effective variables [6]. Over the past decades, two groups of deterministic and probabilistic methods have been proposed to forecast urban water demand. The deterministic approach is solely based on the input variables and their initial conditions, whereas a probabilistic model relies on modeling uncertainties and randomness of the input variables.

Given the significant challenges and complexity of probabilistic methods and the fact that pre-processing methods can provide a useful approximation to their probabilistic counterparts, this research focused on the application of pre-processing to forecast short-term consumption.

2. Literature review

Midterm water demand forecast helps the water management authorities to develop an integrated plan which balances supply and demand in a given period. Water stress of an area can be reduced by accurate estimation of drinking water supply demand [3, 7–9]. Moreover, management can provide water sustainability based on their experience as well as the accurate and reliable value of future demand [10].

Compared to other hydrological forecast studies (e.g., river discharge, sedimentation, rainfall, etc.) water consumption is not as influenced by the input factors as other studies do. The

most significant input variables are temperature, precipitation, and past demand values that were popular in most of the studies [11–13]. Two different types of variables affecting water demand: climatic (e.g., temperature, relative humidity, rainfall, etc.) and socioeconomic (e.g., population and income) [14]. Climatic variables can affect short-term and mid-term values while socioeconomic variables are useful for long-term forecasting [11, 15, 16]. However, a few studies investigated the impact of climatic variables on demand forecasting [17–19]. Literature enlists various deterministic and probabilistic techniques for forecasting urban drinking water demand. In general, conventional methods were prevalent for a better understanding of determinants of water demand [20–22], which consider linear relationships between effective variables and water demand, which is nonlinear. The mentioned studies are broadly categorized into two-fold: physical based and black box models. Without analyzing the physical processes, the second one applies artificial intelligence techniques (artificial neural networks, genetic programming, etc.), fuzzy-based (fuzzy logic, neuro-fuzzy, etc.), soft computing (support vector machine, etc.), and nonlinear deterministic (nonlinear local approximation, etc.) to identify the relationship between the input and output variables. Conventional regression models [3], autoregressive integrated moving average (ARIMA) [23], autoregressive integrated moving average with explanatory variable (ARIMAX) [24, 25], artificial neural networks (ANN) [9, 26–29], a combination of conventional and ANN [11, 12, 30], feedforward neural networks [12, 31], general regression neural networks [32, 33], support vector machines [14, 9, 34–37], gene expression programming [14, 38], fuzzy regression [39], neuro-fuzzy systems [40, 41], Fourier analysis [4], hybrid models (e.g. combined wavelet-ANN and wavelet-GEP) [13, 38], fuzzy cognitive map learning method [42, 43]. This research applies probabilistic ANN, GEP approach and a conventional method (MLR) to determine the performance of the methods with/without phase space reconstruction and wavelet decomposition in the case.

The chaotic nature has been addressed for various systems [44–49]. Any chaotic system is deterministic in which minor changes in the initial conditions could lead to entire different behaviors in the next periods [44]. Chaos theory was successfully used to understand the nonlinear dynamic of the system. The models that are based on chaos theory and nonlinear dynamics are a better representative of the behavior of dynamic of observed data [50]. In general, chaos theory improves the understanding of nonlinear dynamics [51]. Ng et al. applied chaos theory on noisy time series of discharge in Saugeen River (Canada) [52]. They argued that noisy time series not only increase the complications of the data but also gave high embedding dimension. Sivakumar et al. utilized the concept of nonlinear dynamic behavior to classify rivers from phase-space data reconstruction perspective [53].

Genetic programming (GP) and gene expression programming (GEP) are among the heuristic algorithms based on Darwin's evolution theory [53]. GP was employed to complete missing data in wave records and forecasting [55–57]. Aytek and Kishi used GP model to suspended sediment in the Tongue River (United States) and found GP more accurate than sediment rating curves and multiple linear regressions (MLR) [58]. Ghorbani et al. investigated the chaos theory, artificial neural network (ANN) and GEP in estimating suspended sediment in the Mississippi River (United States) [59]. GEP is superior to GP as it is more convenient to interpret the results by a GEP tree that comes along with output results. GEP also performs

better at extracting a mathematical equation which shows the relation between input and output variables [59–61]. Nasseri et al. developed a hybrid model combining the extended Kalman filter with genetic programming for monthly water demand forecasting in Tehran [62]. Shabani et al. proposed a new rationale and a novel technique in forecasting water demand using lag time to feed the determinants of water demand by the development of GEP and SVM models [14]. Yousefi et al. implemented sophisticated mathematical models to forecast water demand of City of Kelowna in monthly temporal scale. Their study assessed the performance of GEP using wavelet decomposition [38].

Among the variety of examined methods Artificial Neural Networks (ANNs), have been applied to the various period in the wide variety of hydrological issues. The main reason of ANNs frequent usage is its ability to overcome the relationship in determining the complexity of time series, even with the shortage of amount of data available to train the models. Therefore, most of the studies applicable in area of water resources demand applies ANNs to forecast short, mid and long-term demand values [13, 30, 31].

Regarding the literature review reported by Nourani et al. concluded about the dominant application of wavelet-based models [63]. Moreover, Labat notified about the improving ability of wavelet in models' performance [64]. Therefore, the application of wavelet brought researchers attention into the area such as denoising [65]; stream flow and water resources [66]; evaporation and climatic models [67]; groundwater level modeling [68]; water demand forecasting [13, 38], where in most of the mentioned studies combination of Wavelet-ANNs performed accurately over conventional models without hybrid wavelet models (e.g. ARIMA, MLR, ANN and etc.).

The objectives of this study are four-fold: (1) to investigate chaotic behavior of case data and finding the proper lag time; (2) to find the accuracy of the forecasting for one-day ahead lead time with various input combination, and (3) to study if phase space reconstruction (PSR) based on optimum embedding dimension would improve the accuracy of the models, and 4) application of wavelet decomposition by five different transform functions combined with all the mentioned models with and without PSR.

3. Methodology

3.1. Case study and data information

3.1.1. Understandings

Unlike natural water resources like rainfall, the lower percentage of drinking water which is change to waste water after use, back to the cycle. Water pressure in a pipeline, water quality, supply peak consumption time, pipeline maintenance, maintenance cost, specialist and educated human resources, pipeline failure management, etc. are the variables that all of them should be under control at the same time. Also, to develop an integrated long-term plan, availability of resources is crucial. Therefore, knowing about the value of consumption in a

specific period is the first step for any management plan beyond urban drinking water supply and allocation. This chapter investigates the first step of every long-term plan development in urban drinking water as discussed below. Water utility management needs drinking water long-term forecasted values in several terms. (1) water distribution network design; (2) supply and consumption management; (3) efficient application of distribution network; (4) pipeline pressure management; (5) network development; (6) optimizing the cost of water supply and network maintenance.

3.1.2. Study area

The present research selected Water consumption of the City of Kelowna (BC, Canada) as the test case. The city of Kelowna water utility provides services for approximately 65,000 residents. Poplar Point, Eldorado, Cedar Creek and Swick Road pump stations cover services for 99% of the population of the area [69]. However, few areas in the boundary are named as "Future City" where does not contain any population yet, land development plan shows water servicing is considered in the area. Monitoring of water quality, the operation of the pumps, water level in reservoirs, and pipeline pressure are conducted by the use of Supervisory Control and Data Acquisition Software (SCADA).

3.1.3. Review of data records

Hourly water demand for the above-mentioned stations has been made available by the city utility of Kelowna. The data used 6 years (approximately 52,464 hourly consumption) starting from January 1st, 2011 to 30th December 2016. **Figure 1** shows the variation of daily and monthly water demand and the consumption pattern. Concerning the 6 years water demand samples of daily scale (2186 points), the first 5 years (1882 points) are used for calibrating the models and the last year (365 points – 2016) is considered as the test period. **Table 1** shows the characteristics of the dataset in the test case.

3.2. Phase space reconstruction (PSR)

Given a set of physical variables and their interactions, the dynamics of a system (e.g., water consumption) can be defined by a single point moving on a trajectory, where each of its points

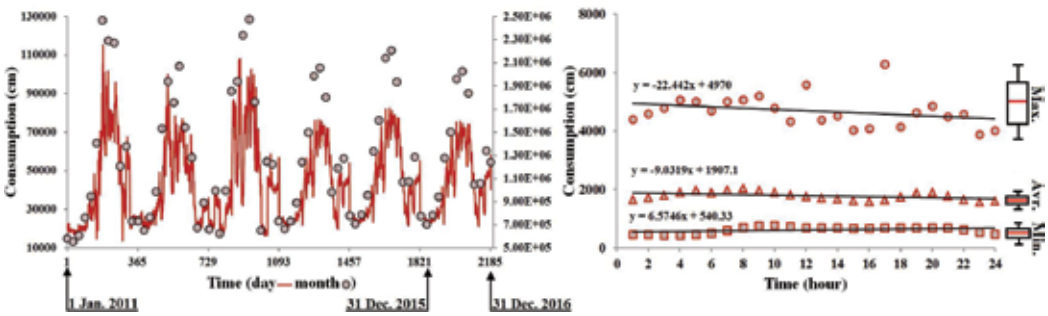


Figure 1. Time series plot of (a) daily water demand; (b) average of the consumption pattern in 24 h within 6 years.

Property	Number of Data	Max. Value*	Min. Value*	Average*	Standard deviation*	Coefficient of variation	Skew	Kurtosis
Data	2186	114597.2	14,124	43046.4	20074.5	0.46	0.73	−0.38

Table 1. Statistics of water consumption of Kelowna City in different temporal resolutions (*m^3).

represents a state of the system. The lag-embedding method reconstructs phase-space from a univariate or multivariate time series generated by a deterministic dynamic system [70]. The underlying dynamics can be studied by building an m-dimensional space X_t defined by [49]:

$$X_t = \left\{x_t, x_{t-\tau}, x_{t-2\tau}, \ldots, x_{t-(m-1)\tau}\right\}, t = 1, 2, \ldots, N \tag{1}$$

where X_t is a vector of the observed data of $\{x_t\}$ t = 1,…,N, N is the total number of observed data, τ is the lag time, and m is embedding dimension. The embedding dimension (m) is typically in the range of 1–10 [53, 54]. The lag-embedding method is sensitive to both embedding parameters of τ and m. Average mutual information (AMI) and autocorrelation function (ACF) are the two well-known methods for estimating the lag time [71, 72]. More details about ACF and different functions are available at [73].

3.3. Correlation dimension (Chaos investigation)

Correlation dimension is a nonlinear measure of the correlation between pairs lying on the attractor. The dimension of a system reveals the number of effective variables in the system. Kermani (2016) classified different dimensions in a system as topological, Hausdorf, box counting, point-wise, and correlation dimension. These dimensions are nearly equal in chaotic systems [52, 74]. This research employed correlation dimension, as it is a lower bound measure of the fractal dimension [59, 74]. For time series whose underlying dynamics is chaotic, the correlation dimension gets a finite fractional value, whereas it is infinite for stochastic systems. The later does not saturate to a specific amount of correlation exponent [75]. For an m-dimensional phase-space, the correlation function, $Cm(r)$, is defined as the fraction of states closer than r [76].

$$C_m(r) = \lim_{N_p \to \infty} \frac{2}{(N_p - w)(N_p - w - 1)} \sum_{i=1}^{N_p} \sum_{j=i+1+w}^{N_p} H\left(r - \left|X_i - X_j\right|\right) \tag{2}$$

where H is the Heaviside step function, X_i is the i_{th} state vector, N_p is the number of points on the reconstructed attractor, r is the radius of a sphere with the content of X_i or X_j. The Theiler window (w) is the correction needed to avoid spurious results due to temporal correlations instead of dynamical ones. $C_m(r)$ is proportional to r for stochastic time series, whereas for chaotic time series it scales with r as:

$$C_m(r) \propto r^{c_e}. \tag{3}$$

where c_e is correlation exponent defined by:

$$c_e = \lim_{r \to 0} \frac{\ln C_m(r)}{\ln r}. \tag{4}$$

The parameters m and C_e can be determined as the slopes of the lines when plotted $C_m(r)$ against r in logarithmic scale. In a deterministic system, C_e increases by increasing m until eventually remaining unchanged. The correlation dimension of time series is defined as the specific value of m after which C_e remains unchanged [54, 59].

3.4. Artificial neural networks

When ANN is based roughly on the neural layout of the human brain and is capable of non-linear modeling processes that can classify the patterns and recognize the capabilities [77]. Regarding the ability of multilayer perceptron (MLP)-ANN outperformance as a conventional ANN approaches [77, 78, 79], this research employed three-layer MLP-ANN (input, hidden and output layers) and the different number of neurons. In hidden layer, neurons are calculated by the summation of demand values (d_i) with the given weight for each value (w_{ij}) to determine the output signal as (u_j).

$$u_j = \sum_{i=1}^{t} w_{ij} d_i \tag{5}$$

$$O_j = \phi(u_j - \theta_j) \tag{6}$$

where ϕ is the transfer function and θ is a threshold limit [80, 81, 82]. Among various transfer functions (e.g., sigmoid shape, piecewise, step, linear and non-linear functions), the logistic sigmoid and Purelin (linear) transfer functions. Regarding the large number of input variables in the present study, no transfer function is applied to reduce the computationally demanding. While, the logistic sigmoid and Purelin transfer functions that are commonly used in literature [79, 81, 82] are provided at the output and hidden layers, respectively (further details about the bias and transfer functions are available at [79, 81]). Feed-forward multi linear perceptron is employed in this study containing input, hidden and output layers. The number of neurons in the input layer varies from 1 to 10 (without decomposition) and from 4 to 24 (with decomposition). Moreover, the neurons of the layers are connected with the neurons in the next layer by weights. Also, to consider all optimal solutions with the highest probable accuracy, this study investigated the number of HLN from 1 to 20 in 1 to 200 epochs.

3.5. Gene expression programming

Evolutionary computation has received significant attention among researchers for studying complex engineering systems. Genetic algorithm (GA), genetic programming (GP), and gene expression programming (GEP) were inspired by Darwin's theory of evolution [60, 61]. GEP defines an algorithm and equation which shows the relation between input and output variables. GA and GP rely on a string of numbers with defined length called "chromosomes", while GEP employs a set of nonlinear entities with different shapes and sizes, "expression/

parse trees". The expression tree accommodates the ease of a GA solution as well as the capability of accepting the nonlinear/complex behavior in a typical GP solution. The chromosome can have one or more genes of equal length. A gene represents a set of symbols containing two parts; a head which has functions and terminals and a tail which only has terminals. Initiating with the random generation of chromosomes, GEP is followed by different applications of genetic operators like replication, recombination, mutation, etc. The terminating condition for developing GEP depends upon the selection of maximum fitness. This research applied 30 chromosomes, eight head sizes, three genes, and arithmetic operators of {+,-, ×, x, x^2, $\sqrt{x}$}.

3.6. Multilinear regression

When MLR corresponds to a linear combination of the components of multiple signals x (e.g. recorded discharge, lag time discharge, or combination of both) to a single output signal y (Demand) by:

$$y = b + \sum_{i=0}^{N} a_i x_i \tag{7}$$

where x_i is the defined input (demand) and a_i is regression coefficient determined by the least square method with the residual r defined by:

$$r = y - a_1 x_1 - a_2 x_2 - \ldots - b. \tag{8}$$

3.7. Wavelet decomposition

Commonly wavelet transforms are used for decomposition, de-noising, and compression of the time series [83]. Time series have a combination of low and high frequency which represent improved features (e.g., cyclical trends) and chaotic element, respectively [84]. Considering these frequencies, separation of low and high frequency is helpful in studying the original pattern and behavior of the time series. One of the mentioned methods is discrete wavelet transform (DWT) to separate per level of frequencies in time series. One of the common discretion ways proposed by Mallat that this study used the mentioned DTW method to separate the frequencies of the applied data [85]. The level of the decomposition shows the subseries. For example, for level 1 decomposition, the number of subseries is two. Therefore, the number of levels indicates the number of subseries plus one. Level 3 is considered as suitable decomposition level in the present study regarding the number of data (2186 day) and following Nourani et al. (2009) that offered [83]:

$$L_n = \text{int}[\log(N)]. \tag{9}$$

where L_n is the number, the level of decomposition and N is the number of used data. Thus, the proper level in this study is considered as 3. However, increasing the level number does not necessarily improve the accuracy of the models. Therefore, the original data are discretized in a high-frequency subset (a_3) and three high frequencies as (d_1), (d_2) and (d_3), where the

summation of all is equal with the value of original data. This research employed Haar, the second and fourth order Daubechies (db2, db4), and the second and fourth order Symlets (Sym2, sym4) wavelets to decompose daily water demand time series into sub-series. The software MATLAB 2015 (https://www.mathworks.com) was employed for the analysis.

3.8. Evaluation of models' performance

This research measured the models' accuracy by coefficient of determination (CD), root mean squared error (RMSE) and mean absolute error (MAE) defined as:

$$\mathrm{CD} = \left[\frac{\sum_{i=1}^{N_t} \left(O_i - \overline{O}\right)\left(F_i - \overline{F}\right)}{\left[\sum_i^{N_t} \left(O_i - \overline{O}\right)^2\right]^{\frac{1}{2}} \left[\sum_i^{N_t} \left(F_i - \overline{F}\right)^2\right]^{\frac{1}{2}}} \right]^2. \tag{10}$$

$$MSE = \sqrt{\frac{\sum_{i=1}^{N_t} \left(O_i - F_i\right)^2}{N_t}}. \tag{11}$$

$$MAE = \frac{1}{n}\sum_{j=1}^{n} |O_i - F_i|. \tag{12}$$

where N_t is the number of values, O and F are the observed and forecasted values of demand, respectively. $\overline{O}$ and $\overline{F}$ are the mean of the observed and forecasted demand values, respectively. Note that the range of CD is between 0 and 1 with higher positive values indicate better agreement. A lower value of RMSE and MAE indicates better agreement between the observed and forecasted values.

4. Preliminary results

4.1. Phase space reconstruction and investigation of chaotic behavior

Existence of chaotic behavior in the time series is shown in **Figure 2**. However, the results are not entirely based on the proof of having chaotic behavior, as the figure only shows possible low-dimensional chaotic behavior. Theoretically, several methods are well known for investigating the chaotic behavior such as lag time calculation method (e.g., average mutual information (AMI), Autocorrelation function (ACF)), correlation dimension, largest Lyapunov exponent, etc.). This study investigates the chaotic behavior by applying ACF and correlation dimension. Having chaotic behavior allows using ACF to calculate the lag time of the time series. The value of lag time is considered as the first approach of ACF to 0 (**Figure 2**).

The results show 83-days as the lag time of the time series. Therefore, 83-day is used to design combination of inputs as phase space for the time series. In this study, the difference between 1st day and 83rd day is used as delay period for phase space reconstruction varying embedding dimensions from 1 to 10 (m_1: D_t; m_2: D_t,$D_{t\text{-}\tau}$; m_{10}: D_t,…,$D_{t\text{-}10\tau}$). It should be noticed that

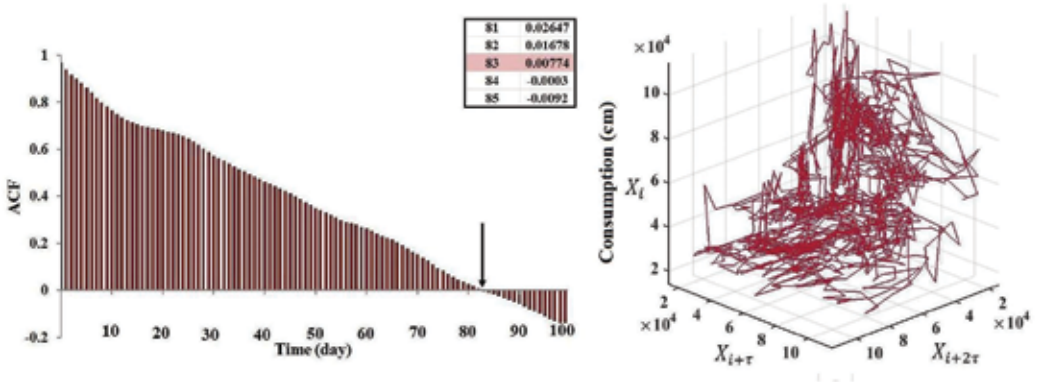


Figure 2. (a) Autocorrelation function (τ); (b) reconstructed phase space by (τ and 2τ -day lag time).

several methods were introduced in literature to calculate the value of optimum embedding dimension which may be more than 10 for the used time series in this study. This study aims at showing the performance of embedding dimension and reconstructed phase space, where m is only considered 1 to 10. **Figure 2** shows the value of ACF for the demand series and reconstructed phase space (τ = 83). **Figure 3a** shows the relation between $C(r)$ and r and (3b) correlation exponent by varying m. **Figure 3b** shows that the value of correlation exponent increases by m and as m = 17, the correlation exponent reaches a specific value (Ce = 3.41). This constant value of Ce at m = 17 indicates the existence of the deterministic behavior of the time series.

4.2. Multilinear regression

Excel 2010 was used to implement MLR model. The train period was used to derive regression coefficient from getting the value of variables in the linear equation. The availability of trained equation, helped in testifying the last year data as the test period. In the first fold, the 1-day delay was considered for m 1 to 10, and second fold applied 83-day delay. **Table 2** shows the results of both MLR and PSR-MLR in the test period.

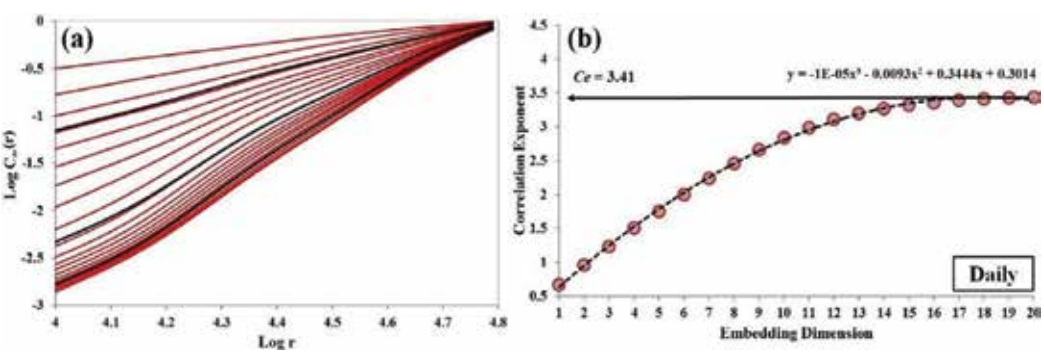


Figure 3. (a) The relation between correlation function $C(r)$ and r by various m; (b) Saturation of correlation dimension Ce (m) with embedding dimensions.

MLR, $\tau = 1$				PSR-MLR, $\tau = 83$			
m	CD	RMSE(m^3/day)	MAE	m	CD	RMSE(m^3/day)	MAE
1	**0.9565**	**3642.89**	**50.42**	**1**	0.9565	3642.89	50.42
2	0.9565	3804.14	52.14	**2**	0.9565	3804.14	52.14
3	0.9468	14106.70	112.82	**3**	0.9570	5319.51	66.90
4	0.9473	13174.97	108.82	**4**	**0.9572**	**3636.34**	**51.04**
5	0.9505	3724.99	49.81	**5**	0.9568	4167.55	56.45
6	0.9503	3746.33	50.09	**6**	0.9569	5907.90	71.65
7	0.9503	3747.49	50.10	**7**	0.9565	4370.03	58.86
8	0.9493	6058.34	70.88	**8**	0.9566	4581.10	60.89
9	0.9505	3736.33	50.02	**9**	0.9566	5023.16	64.71
10	0.9506	3738.35	50.07	**10**	0.9566	4327.34	58.48

Table 2. Fitness values for MLR and PSR-MLR methods in different embedding dimensions (bolded lines are the most accurate values).

Statistical indices for the fitness values showed $m = 1$ for 1-day delay and $m = 4$ for the reconstructed phase space with the value of (CD = 0.9565, RMSE = 3642.89 and MAE = 50.42) and (CD = 0.9572, RMSE = 3636.34 and MAE = 51.04), respectively. However, the difference between the two models is not considerable, in the large value of demand in long-term this difference can come into account. **Figure 4** shows the comparison of observed and demand values. Moreover, the suggested equation for the best result by MLR is given by:

$$D_{t+1} = -0.00854D_t - 0.0366D_{t+\tau} - 0.0128D_{t+2\tau} + 0.9427D_{t+3\tau}. \tag{13}$$

4.3. Performance of artificial neural network

ANN is another approach to model the demand values which represented in Section 3.4. ANN's structures have different hidden layer neurons (HLN) from 1 to 20 with 200 epochs

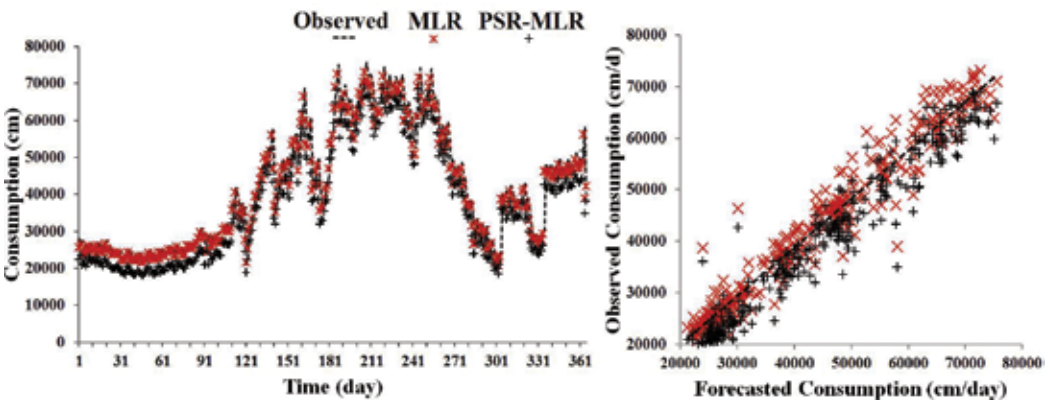


Figure 4. The performance of MLR and PSR-MLR in comparison with observed values.

for each model. **Table 3** represents the result of ANN for both 1-day delay and PSR values. The results in the table for each m, are extracted from the result of various HLN and epochs. **Figure 5** shows the example for selecting $m = 3$ among ($20 \times 200 = 4000$). This calculation has been done for all m from 1 to 10 for both 1-day delay and PSR. ($4000 \times 10 \times 2 = 80{,}000$) number of calculations where the best 10 values have been selected (**Table 3**).

Selection of ANN structures are represented in **Table 3** for the test period. Statistical indices for the fitness values showed $m = 6$ for 1-day delay and $m = 3$ for PSR, with the values of (CD = 0.9520, RMSE = 3535.66 and MAE = 47.58) and (CD = 0.9578, RMSE = 3330.53 and

	ANN, $\tau = 1$						PSR-ANN, $\tau = 83$				
m	Structure	Epoch	CD	RMSE*	MAE	m	Structure	Epoch	CD	RMSE*	MAE
1	1-5-1	110	0.9505	3611.56	48.34	**1**	1-6-1	150	0.9573	3369.55	47.63
2	1-3-1	140	0.9509	3602.25	48.25	**2**	1-4-1	20	0.9568	3369.50	47.83
3	1-16-1	20	0.9514	3554.13	48.06	**3**	**1-2-1**	**120**	**0.9578**	**3330.53**	**47.13**
4	1-16-1	170	0.9516	3550.99	47.90	**4**	1-2-1	70	0.9575	3333.67	47.25
5	1-3-1	160	0.9513	3561.33	47.98	**5**	1-3-1	110	0.9578	3340.36	47.15
6	**1-9-1**	**50**	**0.9520**	**3535.66**	**47.58**	**6**	1-3-1	40	0.9572	3340.16	47.46
7	1-3-1	100	0.9511	3563.14	48.08	**7**	1-3-1	100	0.9570	3348.68	47.80
8	1-8-1	20	0.9510	3570.70	47.84	**8**	1-2-1	150	0.9573	3333.88	47.24
9	1-4-1	200	0.9511	3566.00	47.69	**9**	1-2-1	140	0.9571	3338.53	47.89
10	1-3-1	100	0.9515	3546.72	47.94	**10**	1-4-1	10	0.9539	3518.86	49.28

Table 3. Fitness values for ANN and PSR-ANN in different embedding dimensions *m^3/day). (bolded lines are the most accurate values).

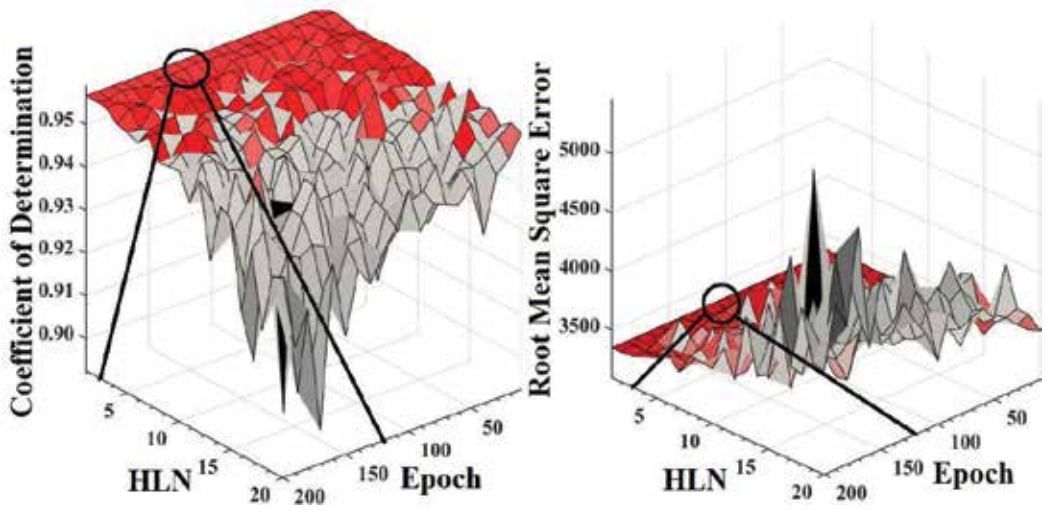


Figure 5. The results of ANN for $\tau = 83$ PSR by various HLN and epochs.

MAE = 47.13), respectively. Regarding the results, PSR-ANN mostly dominates in all embedding dimensions for the fitness accuracy indices. **Figure 6** shows the comparison of observed and demand values in the test period for both ANN and PSR-ANN in $m = 6$ and 3, respectively. The results showed (D_t, $D_{t+\tau}$, $D_{t+2\tau}$) as the best input combination for the models.

4.4. Performance of gene expression programming

GEP preliminarily investigates the relationship between input and output as discussed in Section 3.5. Unlike the other models in this study, 1-day ahead is output, and various combinations of input in terms of m are considered as input variables. The arithmetic operations used in this study are $\{+, -, \times, x, x_2, \sqrt{x}\}$, and GEP applies them to fit the best accuracy between input and output variables. Further details of GEP initial term values are in following of [14, 38, 59] to extract the GEP model for both 1-day delay and PSR. The results are shown in the **Table 4** for the test period.

According to the **Table 4**, there is not much difference among the different m. But the difference in PSR-GEP results can be considered as a proof of sensitivity to the initial values of specific time lags where the variations of the results for different m are more than 1-day delay. There is not a significant difference in the results in this study comparing to other alternative models, especially PSR-ANN is not an advantage of GEP. However, extracting the mathematical equation through GEP is one of advantage of GEP comparing to other artificial models. As a result of given model, equation for $m = 3$ (PSR-GEP) can calculate the demand value for 1-day ahead by:

$$D_{t+1} = 0.0529\sqrt{D_{t+\tau} + D_{t+2\tau}} + D_t - 7.0838 \tag{14}$$

Although, variety of other arithmetic operations may have been applied here but focusing on the aim of study, only simple known operations were applied to extract the GEP equation. The results of PSR-GEP and alternative ones prove the advantage of PSR to improve the accuracy of the models. Statistical indices for the fitness values showed $m = 2$ for 1-day delay and $m = 3$ for the reconstructed phase space with the value of (CD = 0.9497, RMSE = 3609.82, and

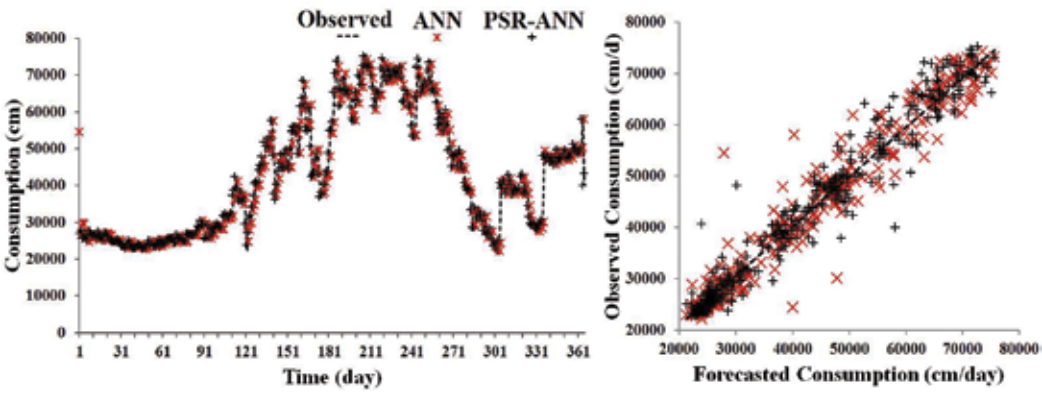


Figure 6. The performance of ANN and PSR-ANN in comparison with observed values.

GEP, τ = 1				PSR-GEP, τ = 83			
m	CD	RMSE(m^3/day)	MAE	*m*	CD	RMSE(m^3/day)	MAE
1	0.9494	3621.87	48.59	**1**	0.9565	3363.46	48.03
2	**0.9497**	**3609.82**	**48.37**	**2**	0.9565	3357.00	47.82
3	0.9494	3633.87	48.42	**3**	**0.9569**	**3343.36**	**47.50**
4	0.9494	3637.74	48.42	**4**	0.9566	3359.53	47.95
5	0.9494	3639.05	48.43	**5**	0.9562	3372.70	48.04
6	0.9494	3619.77	48.60	**6**	0.9566	3359.64	48.08
7	0.9495	3630.44	48.38	**7**	0.9564	3365.04	47.95
8	0.9494	3634.41	48.42	**8**	0.9567	3353.24	47.62
9	0.9494	3628.46	48.42	**9**	0.9562	3370.08	48.05
10	0.9494	3631.12	48.40	**10**	0.9565	3356.68	47.84

Table 4. Fitness values for GEP and PSR-GEP in different embedding dimensions (bolded lines are the most accurate values).

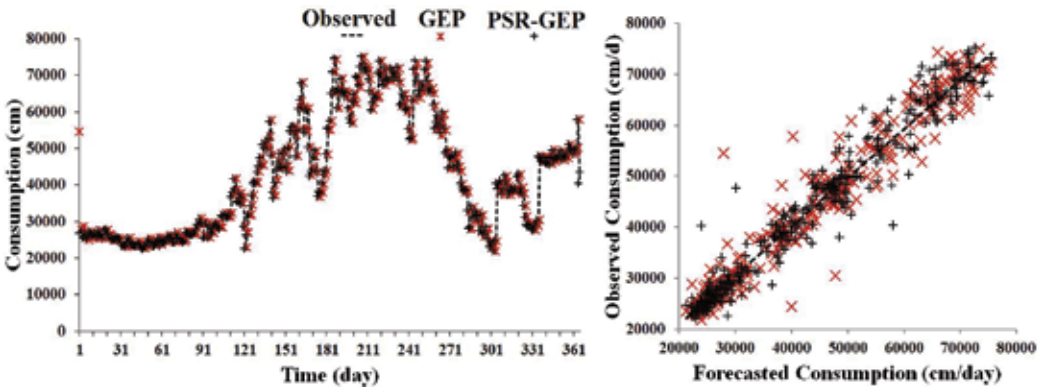


Figure 7. The performance of GEP and PSR-GEP in comparison with observed values.

MAE = 48.37) and (CD = 0.9569, RMSE = 3343.36, and MAE = 47.50), respectively. **Figure 7** shows the comparison of observed and demand values in the test period for both GEP and PSR-GEP in m = 2 and 3, respectively.

5. Wavelet decomposition and models' performance

The combination of models with wavelet decomposition is derived by adding the output of each wavelet to the input of the models. **Figure 8** shows the example of the decomposed values for water demand time series by db2 transform function. To discrete the demand values, five

wavelet transforms were applied (Section 3.7.). As suggested by Nourani et al. [83], 3rd level decomposition is recommended for 2186 point data.

Table 5 indicates the results of wavelet decomposition for the selected models in the previous section. As the table highlights, db4 and db2 are the transforms which resulted in the highest accuracy in W-MLR and W-PSR-MLR, with the value of (CD = 0.9697, RMSE = 2804.44 and MAE = 42.11) and (CD = 0.9745, RMSE = 2699.83 and MAE = 43.61), respectively. After implying the decomposed inputs for MLR and PSR-MLR for result comparison improved the results in both models. Also, sym4 and db2 are the transforms which resulted in the highest accuracy in W-ANN and W-PSR-ANN, with the value of (CD = 0.9915, RMSE = 1486.21 and MAE = 30.06) and (CD = 0.9756, RMSE = 2517.24, and MAE = 41.68), respectively. Also, calculations for W-ANN and W-PSR-ANN are done with HLN 1 to 20 and epochs 1 to 200, and the mentioned results in the table are selective of the highest among them. Unlike the results of MLR, W-ANN forecasted accurately than W-PSR-ANN which is the inversion of the results of ANN and PSR-ANN. However, wavelet decomposition improved the results of W-ANN and W-PSR-ANN comparing to the alternative without decomposition (**Table 3**). Moreover, db4 and db2 are the transforms which resulted in the highest accuracy in W-GEP and W-PSR- GEP, with the value of (CD = 0.9845, RMSE = 2027.28 and MAE = 36.62) and (CD = 0.9753, RMSE = 2532.21, and MAE = 41.69), respectively. Following the results of ANN method, W-GEP forecasted accurately than W-PSR-GEP. However, wavelet decomposition improved the results of W-GEP and W-PSR-GEP comparing to the alternative without decomposition (**Table 4**).

All PSR models resulted in the highest values which used the decomposed inputs by db2 transform. It is noticeable that PSR affects the inherent of the time series which the results of performance of all models are in common about improving the accuracy. Considering this fact,

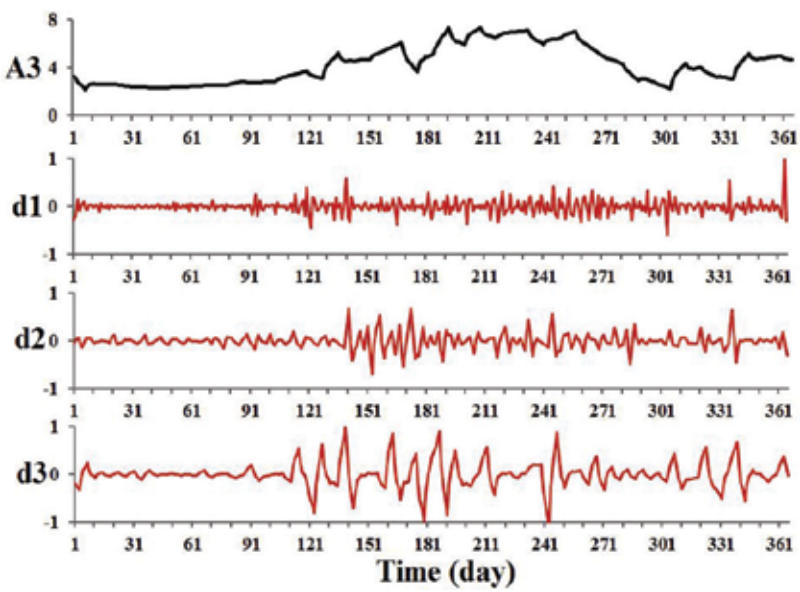


Figure 8. Three level DWT of daily water demand time series of Kelowna City in 2016.

Models	Fitness	Transform functions				
		haar	db2	db4	sym2	sym4
W-MLR	**CD**	0.9612	0.9477	**0.9697**	0.9677	0.9694
	RMSE(m^3/day)	3168.62	3681.17	**2804.44**	2893.60	2816.06
	MAE	44.48	49.04	**42.11**	43.54	42.24
W-PSR-MLR	**CD**	0.9670	**0.9745**	0.9719	0.9745	0.9712
	RMSE(m^3/day)	3008.34	**2699.83**	2811.58	2699.83	2845.69
	MAE	45.39	**43.61**	43.95	43.61	44.24
W-ANN	**CD**	0.9868	0.9816	0.9861	0.9856	**0.9915**
	RMSE(m^3/day)	1853.11	2189.15	2136.25	1948.28	**1486.21**
	MAE	33.78	36.91	39.50	33.86	**30.06**
W-PSR-ANN	**CD**	0.9685	**0.9756**	0.9723	0.9752	0.9715
	RMSE(m^3/day)	2867.87	**2517.24**	2677.44	2547.89	2724.56
	MAE	43.16	**41.68**	42.09	42.19	42.61
W-GEP	**CD**	0.9721	0.9766	**0.9845**	0.9297	0.9255
	RMSE(m^3/day)	2698.16	2492.46	**2027.28**	4311.60	4429.89
	MAE	41.21	39.05	**36.62**	54.23	55.25
W-PSR-GEP	**CD**	0.9667	**0.9753**	0.9721	0.9748	0.9704
	RMSE(m^3/day)	2937.76	**2532.21**	2689.20	2555.82	2770.80
	MAE	43.66	**41.69**	42.13	41.90	42.51

Table 5. Fitness values for decomposition of selection of models for the test period (bolded lines are the most accurate values).

PSR can be introduced as a pre-processing method like wavelet decomposition; however, complexity and accuracy of PSR cannot be compared with the higher result of wavelet decomposition. **Figure 9** shows the comparison of all selected models with highest accuracy (W-PSR-MLR, W-ANN, and W-GEP) in forecast of short-term water demand values.

The figure shows that the performance of W-ANN and W-GEP is better than W-PSR-MLR, while W-ANN's calculated values are more accurate than W-GEP in simulating peak points. This study eventually would suggest that these peak points are indication of critical issues related to water distribution system (pressure management, peak time demand, etc.) taking in account the performance of the models and simulations of highest and lowest values of demands. Therefore, it is recommended to evaluate models' performance in two separate parts as maximum values and minimum values along with evaluating criteria such as CD, RMSE, and MAE for the test period. The difference is not visible in **Figure 9**. Therefore focusing on **Figure 10**, it shows the performance of models by residual values in the test period.

In **Figure 10** the residual values show the remarkable difference of performance of models. W-ANN values distributed in the area of (−15%, +15%), unlike other two models. W-GEP dominates over W-PSR-MLR; however, the fitness criteria values for both are very close to each other (**Tables 2** and **4**).

http://dx.doi.org/10.5772/intechopen.76537

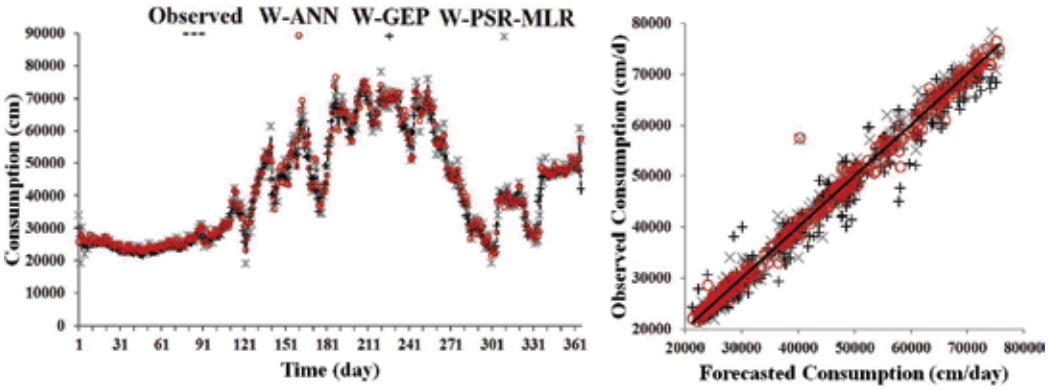


Figure 9. The performance of the W-models in comparison with observed values.

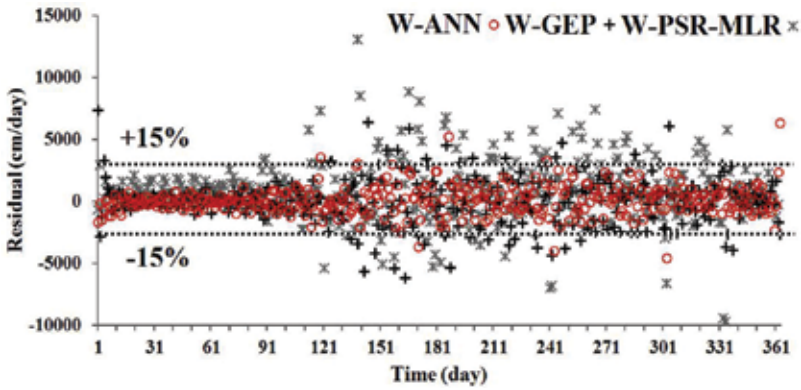


Figure 10. Residual values of the selected W-models.

This chapter presents the performance of two pre-processes methods in improving the accuracy of three models to forecast short-term urban water demand value in Kelowna City, BC, Canada. The first pre-process approach of PSR which is calculated by ACF method has improved the results of all models in this study. However, PSR does not improve the accuracy of models for entire dataset. Based on the behavior of time series, ACF or AMI (two lag time calculation methods) may have improved a non-deterministic dataset, but it seems in a chaotic dataset, PSR improves the performance of models in increasing accuracy with a proper number of embedding dimensions. Wavelet decomposition, the second pre-process method in the present study has also improved the accuracy of the models but, decomposition did not work on PSR based methods except MLR. It can be concluded that PSR and wavelet are in common with their outfits as two applicable pre-process methods. Also, PSR pre-processing is simpler than wavelet. Therefore, it is recommended to use PSR for the models. As per the results of this study it seems PSR works on a chaotic dataset which seems to be considered as disadvantage of PSR. Comparing the mentioned two pre-process methods, wavelet decomposition is significant to use, though, it is time-consuming and complex than PSR. Also, each transform functions have specific application where each of them can be used independently (e.g., seasonal, de-noising, peak points, etc.).

6. Conclusion

Over the past decades, hydrologists have paid attention to data-driven modeling techniques. City governments and WDS operators are always looking for an accurate estimation of water demand values not only for future but also focusing on probable failures like peak consumption and pressure values to manage the WDS pipelines. Therefore, the wide variety of modeling techniques such as artificial and evolutionary simulation methods are proposed by researchers. This chapter investigated the performance of three techniques (ANN, GEP, and MLR) in forecasting short-term water demand of Kelowna City (BC, Canada). About 6 years daily dataset was employed for training and testing the models. First 5 years were considered to train the model and the last year as the test period. All three techniques performed considerably accurate, while the focus of this chapter was on improving the accuracy of the models for the same dataset. Firstly, the model was calibrated by different input combination with 1-day lag time. Then, models were calibrated by the lag time of the data set (83-day) which was calculated by ACF method. WDT was combined with the models to capture multi-scale features of the signals by decomposing observed demand values into sub-series. Five WDT functions (haar, Db2, db4, Sym2, and sym4) were employed to decompose the dataset. The results were then compared with the MLR, ANN, and GEP when no pre-processing (PSR, WDT) was applied. The research results were accurate than PSR. WDT have also improved the accuracy of models with PSR and without PSR. However, the impact of wavelet on the models with PSR was not as considerable as without PSR. The lowest error was reported by W-ANN among all alternative models in this chapter. Regarding the improvement of all models combining WDT and PSR, it is recommended to use the method in modeling and forecasting issues, especially about the dataset that the peak points are very critical in the case. The inherent behavior of dataset (deterministic or stochastic) can affect the performance of the pre-processing methods. Therefore, behavior of datasets should be investigated before deciding to combine any pre-process methods.

Acknowledgements

The authors would like to thank the financial support from the Natural Sciences and Engineering Research Council (NSERC) of Canada. The Okanagan Basin Water Board and the City of Kelowna are thanked for providing water consumption data.

Author details

Peyman Yousefi, Gholamreza Naser* and Hadi Mohammadi

*Address all correspondence to: bahman.naser@ubc.ca

Okanagan School of Engineering, The University of British Columbia

References

[1] Billings RB, Jones CV. Forecasting Urban Water Demand. USA: American Water Works Association; 2011. ISBN: 1-58321-537-9

[2] Ghalehkhondabi I, Ardjmand E, Young WA, Weckman GR. Water demand forecasting: Review of soft computing methods. Environmental Monitoring and Assessment. 2017; **189**(7):313

[3] Ghiassi M, Zimbra DK, Saidane H. Urban water demand forecasting with a dynamic artificial neural network model. Journal of Water Resources Planning and Management. 2008;**134**(2):138-146

[4] Odan FK, Reis LF. Hybrid water demand forecasting model associating artificial neural network with Fourier series. Journal of Water Resources Planning and Management. 2012; **138**(3):245-256

[5] IwAnek M, Kowalska B, Hawryluk E, Kondraciuk K. Distance and time of water effluence on soil surface after failure of buried water pipe. Laboratory investigations and statistical analysis. Eksploatacja i niezawodnosc– Maintenance and Reliability. 2016;**18**(2):278-284

[6] Wang W, Vrijling JK, Van Gelder PH, Ma J. Testing for nonlinearity of streamflow processes at different timescales. Journal of Hydrology. 2006;**322**(1-4):247-268

[7] Jain A, Ormsbee LE. A decision support system for drought characterization and management. Civil Engineering Systems. 2001;**18**(2):105-140

[8] Kame'enui AE. Water Demand Forecasting in the Puget Sound Region: Short and Long-Term Models (Doctoral dissertation, University of Washington)

[9] Herrera M, Torgo L, Izquierdo J, Pérez-García R. Predictive models for forecasting hourly urban water demand. Journal of Hydrology. 2010;**387**(1-2):141-150

[10] Zhou SL, McMahon TA, Walton A, Lewis J. Forecasting operational demand for an urban water supply zone. Journal of Hydrology. 2002;**259**(1-4):189-202

[11] Jain A, Varshney AK, Joshi UC. Short-term water demand forecast modelling at IIT Kanpur using artificial neural networks. Water Resources Management. 2001;**15**(5):299-321

[12] Bougadis J, Adamowski K, Diduch R. Short-term municipal water demand forecasting. Hydrological Processes. 2005;**19**(1):137-148

[13] Adamowski J, Fung Chan H, Prasher SO, Ozga-Zielinski B, Sliusarieva A. Comparison of multiple linear and nonlinear regression, autoregressive integrated moving average, artificial neural network, and wavelet artificial neural network methods for urban water demand forecasting in Montreal, Canada. Water Resources Research. 2012;**48**(1)

[14] Shabani S, Yousefi P, Adamowski J, Naser G. Intelligent soft computing models in water demand forecasting. In: Water Stress in Plants. 2016. Rijeka, Croatia: InTech; ISBN: 978-953-51-2621-8

[15] Miaou SP. A class of time series urban water demand models with nonlinear climatic effects. Water Resources Research. 1990;**26**(2):169-178

[16] Gato-Trinidad S, Jayasuriya N, Roberts P. Understanding urban residential end uses of water. Water Science and Technology. 2011;**64**(1):36-42

[17] Zhou SL, McMahon TA, Walton A, Lewis J. Forecasting daily urban water demand: A case study of Melbourne. Journal of Hydrology. 2000;**236**(3-4):153-164

[18] Mukhopadhyay A, Akber A, Al-Awadi E. Analysis of freshwater consumption patterns in the private residences of Kuwait. Urban Water. 2001;**3**(1-2):53-62

[19] Dos Santos CC, Pereira Filho AJ. Water demand forecasting model for the metropolitan area of São Paulo, Brazil. Water Resources Management. 2014;**28**(13):4401-4414

[20] Brekke L, Larsen MD, Ausburn M, Takaichi L. Suburban water demand modeling using stepwise regression. American Water Works Association. Journal. 2002 Oct 1;**94**(10):65

[21] Polebitski AS, Palmer RN, Waddell P. Evaluating water demands under climate change and transitions in the urban environment. Journal of Water Resources Planning and Management. 2010;**137**(3):249-257

[22] Lee SJ, Wentz EA, Gober P. Space–time forecasting using soft geostatistics: A case study in forecasting municipal water demand for Phoenix, Arizona. Stochastic Environmental Research and Risk Assessment. 2010;**24**(2):283-295

[23] Adebiyi AA, Adewumi AO, Ayo CK. Comparison of ARIMA and artificial neural networks models for stock price prediction. Journal of Applied Mathematics. 2014;**2014**(1):1-7

[24] Young CC, Liu WC. Prediction and modelling of rainfall–runoff during typhoon events using a physically-based and artificial neural network hybrid model. Hydrological Sciences Journal. 2015;**60**(12):2102-2116

[25] Young CC, Liu WC, Hsieh WL. Predicting the water level fluctuation in an alpine lake using physically based, artificial neural network, and time series forecasting models. Mathematical Problems in Engineering. 2015:1-11

[26] Adamowski J, Karapataki C. Comparison of multivariate regression and artificial neural networks for peak urban water-demand forecasting: Evaluation of different ANN learning algorithms. Journal of Hydrologic Engineering. 2010;**15**(10):729-743

[27] Cutore P, Campisano A, Kapelan Z, Modica C, Savic D. Probabilistic prediction of urban water consumption using the SCEM-UA algorithm. Urban Water Journal. 2008;**5**(2):125-132

[28] Firat M, Yurdusev MA, Turan ME. Evaluation of artificial neural network techniques for municipal water consumption modeling. Water Resources Management. 2009;**23**(4):617-632

[29] Jentgen L, Kidder H, Hill R, Conrad S. Energy management strategies use short-term water consumption forecasting to minimize cost of pumping operations. American Water Works Association. Journal. 2007;**99**(6):86-94

[30] Jain A, Ormsbee LE. Short-term water demand forecast modeling techniques—Conventional methods versus AI. Journal (American Water Works Association). 2002;**94**(7):64-72

[31] Adamowski JF. Peak daily water demand forecast modeling using artificial neural networks. Journal of Water Resources Planning and Management. 2008;**134**(2):119-128

[32] Zhou J, Yang K. General regression neural network forecasting model based on PSO algorithm in water demand. In Knowledge Acquisition and Modeling (KAM), 2010 3rd International Symposium on 2010 Oct (pp. 51-54). IEEE

[33] Firat M, Turan ME, Yurdusev MA. Comparative analysis of neural network techniques for predicting water consumption time series. Journal of Hydrology. 2010;**384**(1-2):46-51

[34] Msiza IS, Nelwamondo FV, Marwala T. Water demand prediction using artificial neural networks and support vector regression. Journal of Computers. 2008;**3**(11):1-8

[35] Brentan BM, Luvizotto E Jr, Herrera M, Izquierdo J, Pérez-García R. Hybrid regression model for near real-time urban water demand forecasting. Journal of Computational and Applied Mathematics. 2017;**309**:532-541

[36] Msiza IS, Nelwamondo FV, Marwala T. Artificial neural networks and support vector machines for water demand time series forecasting. In: Systems, Man and Cybernetics, 2007. ISIC. IEEE International Conference on 2007 Oct (pp. 638-643). IEEE

[37] Trzęsiok M. Symulacyjna ocena jakości zagregowanych modeli zbudowanych metodą wektorów nośnych. Studia Ekonomiczne. 2013;**132**:115-126

[38] Yousefi P, Shabani S, Mohammadi H, Naser G. Gene expression programing in long term water demand forecasts using wavelet decomposition. Procedia Engineering. 2017;**186**: 544-550

[39] Azadeh A, Neshat N, Hamidipour H. Hybrid fuzzy regression–artificial neural network for improvement of short-term water consumption estimation and forecasting in uncertain and complex environments: Case of a large metropolitan city. Journal of Water Resources Planning and Management. 2011;**138**(1):71-75

[40] Atsalakis G, Minoudaki C, Markatos N, Stamou A, Beltrao J, Panagopoulos T. Daily irrigation water demand prediction using adaptive neuro-fuzzy inferences systems (anfis). In: Proceedings of international conference on energy, environment, Ecosystems and Sustainable Development; 2007 Jul

[41] Tabesh M, Dini M. Fuzzy and neuro-fuzzy models for short-term water demand forecasting in Tehran. Iranian Journal of Science and Technology. 2009;**33**(B1):61-77

[42] Papageorgiou EI, Poczęta K, Laspidou C. Application of fuzzy cognitive maps to water demand prediction. In: Fuzzy systems (FUZZ-IEEE), 2015 IEEE international conference on 2015 Aug (pp. 1-8). IEEE

[43] Ahmadi S, Alizadeh S, Forouzideh N, Yeh CH, Martin R, Papageorgiou E. ICLA imperialist competitive learning algorithm for fuzzy cognitive map: Application to water demand

forecasting. In: Fuzzy Systems (FUZZ-IEEE), 2014 IEEE International Conference on 2014 Jul, China: IEEE; (p. 1041-1048)

[44] Lorenz EN. Deterministic nonperiodic flow. Journal of the Atmospheric Sciences. 1963; **20**(2):130-141

[45] Lorenz EN. Atmospheric predictability as revealed by naturally occurring analogues. Journal of the Atmospheric Sciences. 1969;**26**(4):636-646

[46] Jayawardena AW, Lai F. Analysis and prediction of chaos in rainfall and stream flow time series. Journal of Hydrology. 1994;**153**(1-4):23-52

[47] Porporato A, Ridolfi L. Nonlinear analysis of river flow time sequences. Water Resources Research. 1997;**33**(6):1353-1367

[48] Krasovskaia I, Gottschalk L, Kundzewicz ZW. Dimensionality of Scandinavian river flow regimes. Hydrological Sciences Journal. 1999;**44**(5):705-723

[49] Sivakumar B, Berndtsson R, Persson M. Monthly runoff prediction using phase space reconstruction. Hydrological Sciences Journal. 2001;**46**(3):377-387

[50] Gan CB, Lei H. A new procedure for exploring chaotic attractors in nonlinear dynamical systems under random excitations. Acta Mechanica Sinica. 2011;**27**(4):593-601

[51] Casdagli M. Nonlinear forecasting, chaos and statistics. In: Modeling Complex Phenomena. New York, NY: Springer; 1992. pp. 131-152

[52] Ng WW, Panu US, Lennox WC. Chaos based analytical techniques for daily extreme hydrological observations. Journal of Hydrology. 2007;**342**(1-2):17-41

[53] Sivakumar B, Jayawardena AW, Li WK. Hydrologic complexity and classification: A simple data reconstruction approach. Hydrological Processes. 2007;**21**(20):2713-2728

[54] Khatibi R, Ghorbani MA, Aalami MT, Kocak K, Makarynskyy O, Makarynska D, Aalinezhad M. Dynamics of hourly sea level at Hillarys boat harbour, Western Australia: A chaos theory perspective. Ocean Dynamics. 2011;**61**(11):1797-1807

[55] Kalra R, Deo MC. Genetic programming for retrieving missing information in wave records along the west coast of India. Applied Ocean Research. 2007;**29**(3):99-111

[56] Ustoorikar K, Deo MC. Filling up gaps in wave data with genetic programming. Marine Structures. 2008;**21**(2-3):177-195

[57] Gaur S, Deo MC. Real-time wave forecasting using genetic programming. Ocean Engineering. 2008;**35**(11-12):1166-1172

[58] Aytek A, Kişi Ö. A genetic programming approach to suspended sediment modelling. Journal of Hydrology. 2008;**351**(3-4):288-298

[59] Ghorbani MA, Khatibi R, Asadi H, Yousefi P. Inter-comparison of an evolutionary programming model of suspended sediment time-series with other local models. In: Genetic

Programming-New Approaches and Successful Applications. Rijeka, Croatia: InTech; 2012. 978-953-51-0809-2

[60] Ferreira C. Gene expression programming in problem solving. In: Soft Computing and Industry. London: Springer; 2002. pp. 635-653

[61] Ferreira C. Function finding and the creation of numerical constants in gene expression programming. In: Advances in Soft Computing. London: Springer; 2003. pp. 257-265

[62] Nasseri M, Moeini A, Tabesh M. Forecasting monthly urban water demand using extended Kalman filter and genetic programming. Expert Systems with Applications. 2011;**38**(6):7387-7395

[63] Nourani V, Baghanam AH, Adamowski J, Kisi O. Applications of hybrid wavelet–artificial intelligence models in hydrology: A review. Journal of Hydrology. 2014;**514**:358-377

[64] Labat D. Recent advances in wavelet analyses: Part 1. A review of concepts. Journal of Hydrology. 2005 Nov 25;**314**(1-4):275-288

[65] Chou CM. Application of set pair analysis-based similarity forecast model and wavelet denoising for runoff forecasting. Water. 2014;**6**(4):912-928

[66] Labat D. Wavelet analysis of the annual discharge records of the world's largest rivers. Advances in Water Resources. 2008;**31**(1):109-117

[67] Partal T, Cigizoglu HK. Estimation and forecasting of daily suspended sediment data using wavelet–neural networks. Journal of Hydrology. 2008;**358**(3-4):317-331

[68] Adamowski J, Chan HF. A wavelet neural network conjunction model for groundwater level forecasting. Journal of Hydrology. 2011;**407**(1-4):28-40

[69] Water Quality and Customer Care Supervisor City of Kelowna. City of Kelowna 2016 Annual Water and Filtration Exclusion Report, Report Submitted: July 31, 2017

[70] Takens F. Detecting strange attractors in turbulence. In: Dynamical systems and turbulence. Warwick, Berlin, Heidelberg: Springer; 1980, 1981. pp. 366-381

[71] Fraser AM, Swinney HL. Independent coordinates for strange attractors from mutual information. Physical Review A. 1986;**33**(2):1134

[72] Holzfuss J, Mayer-Kress G. An approach to error-estimation in the application of dimension algorithms. In: Dimensions and Entropies in Chaotic Systems. Berlin, Heidelberg: Springer; 1986. 114-122

[73] Dunn PF, Davis MP. Measurement and Data Analysis for Engineering and Science. CRC Press; 2017

[74] Zounemat-Kermani M. Investigating chaos and nonlinear forecasting in short term and mid-term river discharge. Water Resources Management. 2016;**30**(5):1851-1865

[75] Dhanya CT, Kumar DN. Nonlinear ensemble prediction of chaotic daily rainfall. Advances in Water Resources. 2010;**33**(3):327-347

[76] Grassberger P, Procaccia I. Measuring the strangeness of strange attractors. In: The Theory of Chaotic Attractors. New York, NY: Springer; 2004. pp. 170-189

[77] Lek S, Guégan JF. Artificial neural networks as a tool in ecological modelling, an introduction. Ecological Modelling. 1999 Aug 17;**120**(2-3):65-73

[78] Nourani V, Entezari E, Yousefi P. ANN-RBF hybrid model for spatiotemporal estimation of monthly precipitation case study: Ardabil plain. International Journal of Applied Metaheuristic Computing (IJAMC). 2013;**4**(2):1-6

[79] Najah A, El-Shafie A, Karim OA, El-Shafie AH. Application of artificial neural networks for water quality prediction. Neural Computing and Applications. 2013;**22**(1):187-201

[80] Yousefi P, Naser G, Mohammadi H. Surface Water Quality Model: Impacts of Influential Variables. Journal of Water Resources Planning and Management. 2018 Feb 22;**144**(5): 04018015

[81] Haykin S. Neural Networks: A Comprehensive Foundation. 2nd ed. Upper Saddle River: Prentice Hall; 1999

[82] Melesse AM, Hanley RS. Artificial neural network application for multi-ecosystem carbon flux simulation. Ecological Modelling. 2005;**189**(3-4):305-314

[83] Nourani V, Alami MT, Aminfar MH. A combined neural-wavelet model for prediction of Ligvanchai watershed precipitation. Engineering Applications of Artificial Intelligence. 2009;**22**(3):466-472

[84] Zhou T, Wang F, Yang Z. Comparative analysis of ANN and SVM models combined with wavelet preprocess for groundwater depth prediction. Water. 2017 Oct 12;**9**(10):781

[85] Mallat SG. A theory for multiresolution signal decomposition: The wavelet representation. IEEE Transactions on Pattern Analysis and Machine Intelligence. 1989;**11**(7):674-693

Chapter 8

Wavelet Transform for Educational Network Data Traffic Analysis

Shwan Dyllon and Perry Xiao

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/intechopen.76455

Abstract

Network monitoring and analysis are very important, in order to understand the performance of the networks, the reliability of the networks, the security of the networks, and to identify potential problems. In this chapter, we present our latest work on university network data traffic analysis by using continuous wavelet transform (CWT). With CWT, you can analyse the data and show how the frequency content of the data changes over time. This time dependent frequency varying information, which is lacking in other techniques, such FFT, is very useful for network traffic analysis. A twelve month's network traffic data, include World Wide Web (WWW) data and Email data were presented in 3D format, by using wavelet transform, we can visualise the hourly, daily, weekly and monthly activities. We will first present the theoretical background, then show the experimental results.

Keywords: wavelet transform, educational network, data traffic analysis

1. Introduction

Network monitoring and analysis are very important, in order to understand the performance of the networks, the reliability of the networks, the security of the networks, and to identify potential problems. With network traffic analysis, network security staff would be able to identify any malicious or suspicious packets within the traffic, whilst network administrators could monitor the download/upload speeds, throughput, etc., and therefore to have a better understanding of network operations. To date, many techniques have been used in network data traffic analysis. The neural network (NN), also known as the artificial neural network (ANN), has been used for prediction, as well as to identify the presence of anomalies

[1–3]. Pattern recognition has been used for traffic data classification [4, 5], and chaos theory has been used for the correlation and the prediction of time series data, and to identify the nonlinear dynamical behaviour of real-time traffic data [6, 7]. The Fourier transform (FFT) and wavelet transform have been used to analyse the frequency components of the traffic data [8–10]. The main difference between FFT and wavelet transform is that wavelet transform is localised in both time and frequency whereas the standard Fourier transform is only localised in frequency. In other words, with wavelet transform, when a certain frequency event happened, we can know both what frequency component was, and when it happened. In this chapter, we will use wavelet transform to analyse the London South Bank University network traffic data, in order to understand and evaluate the network utilisation.

London South Bank University (LSBU) network is based on three-tier network architecture, i.e. CORE layer, distribution layer, and edge layer. The core layer is responsible for the routing protocols, the distribution layer responsible for all the VLAN management as well as spanning tree protocol and loop prevention as well as some level of security i.e. DOS protection, and finally the edge layer responsible for the end user connectivity. The LSBU network traffic raw data were first captured using the Paessler Router Traffic Grapher (PRTG) network-monitoring tool (Paessler AG, Germany), and then many numerical analysis algorithms, including wavelet transform, were developed to analyse the captured raw data.

2. Wavelet data analysis

A wavelet is a small wave. Wavelet data analysis is based on the wavelet transform, which has been used for numerous studies in geophysics, including tropical convection [11]. The wavelet transform can be used to analyse time series that contain nonstationary power at many different frequencies [12]. Unlike traditional sin(t) or cos(t) waves that go from negative infinity to positive infinity, wavelets always begin at zero, increases, and then decreases back to zero. Many types of wavelets exist, most of which are used for orthogonal wavelet analysis [13, 14], which purposefully crafted to have specific properties that make them useful for signal processing. **Figure 1** shows some examples of commonly used wavelets.

Wavelet transform is the convolution of time sequence data and wavelets, and can be generally expressed as:

$$F(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} f(t)\, \psi^*\left(\frac{t-b}{a}\right) dt \tag{1}$$

Here, a is the scale ($a > 0$), b is the translational value, t is the time, f(t) is the data, and $\psi(t)$ is the wavelet function, and the * is the complex conjugate symbol. Wavelet transform can be generally divided into discrete wavelet transform (DWT) and continuous wavelet transform (CWT).

The discrete wavelet transform (DWT) is an implementation of the wavelet transform using a discrete set of the wavelet scales. DWT decomposes the signal into mutually orthogonal set of wavelets, which is the main difference from the continuous wavelet transform (CWT). DWT can be used for wavelet decomposition and easy and fast denoising of a noisy signal.

http://dx.doi.org/10.5772/intechopen.76455

Continuous wavelet transform (CWT) is an implementation of the wavelet transform using arbitrary scales and almost arbitrary wavelets. The wavelets used are not orthogonal and the data obtained by this transform are highly correlated. To approximate the continuous wavelet transform, the convolution should be done N times for each scale, where N is the number of points in the time series [15]. The wavelet transform is similar to the Fourier transform, but unlike Fourier transform (which is localised only in the frequency space), the wavelet transform localised in both the time space and the frequency space, see **Figure 2**. CWT allows users to have variable resolutions, i.e. either high precision in time and low precision in frequency, or high precision in frequency and low precision in time. Although windowed transform,

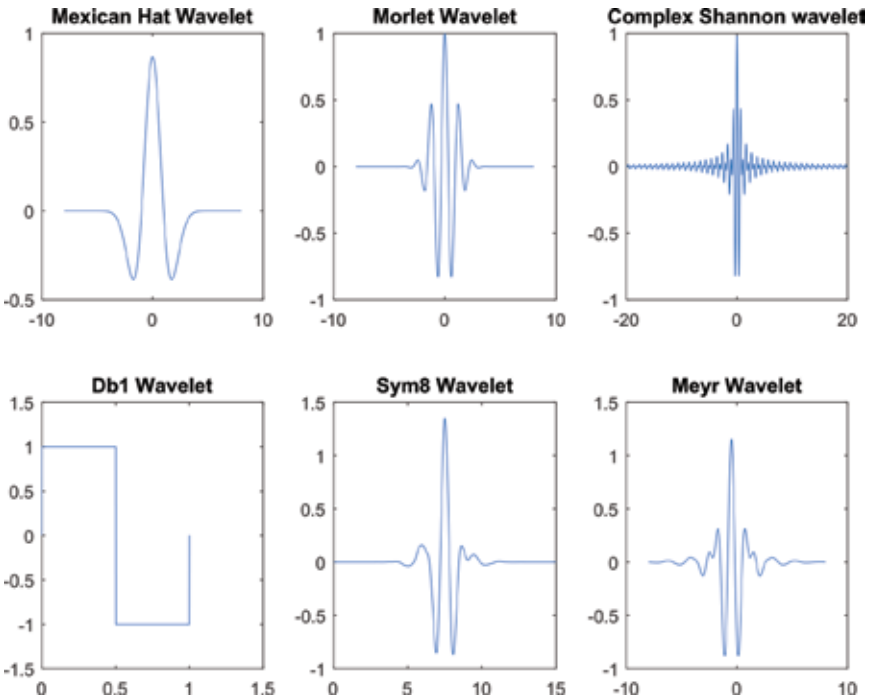


Figure 1. Examples of different types of wavelets.

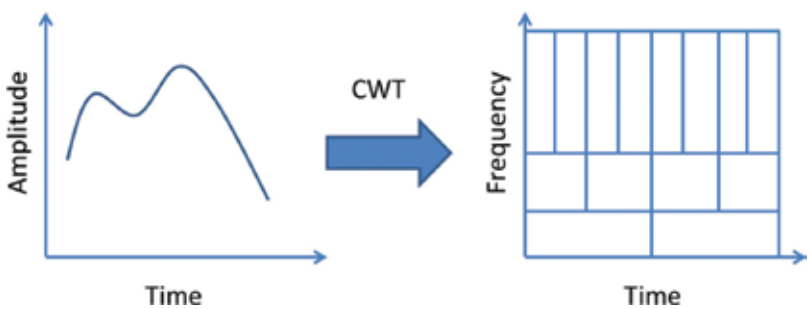


Figure 2. Continuous wavelet transform.

such as Short-Time Fourier Transform (STFT), which is also cable to create a local frequency analysis, the drawback of STFT is that the window size is fixed.

In this study, the wavelet transform used for data decomposition, data denoising and time dependent frequency components analysis by using continuous wavelet transform (CWT).

3. Results and discussions

3.1. Network traffic data 2D and 3D presentation

Figure 3 shows the LSBU 1-year total network traffic data, recorded at 1 h interval, in 3D format (top) and 2D format (bottom), which X axis represents the time of the day, from 01:00 to 24:00, and Y axis represents the day of the year, from 1 to 365, and Z axis represents the total traffic in Gbits per second. The data was recorded at 1 h interval for a period of 1 year, November 2016 to November 2017. By presenting the network traffic data in 2D and 3D formation we can better understand the network usage and characterisations.

The results show that the total network traffic varies from season to season throughout the year, and also varies from time to time throughout the day. By understanding the total network traffic pattern, we can plan better for the network operations, optimise the network usage, and identify potentially suspicious traffics.

Figure 4 shows the LSBU 1-year World Wide Web (WWW) traffic data in 3D format (top) and the corresponding 2D presentation (bottom). The results show that WWW traffic is highly seasonal. It has a strong week day and weekend effect, this agrees well with our previous studies [16, 17]. It also has a strong effect of Christmas, Easter and summer holiday periods. The WWW traffic data varies significantly within a day, with the highest between 10:00 am and 19:00 pm, and lowest between 06:00 am and 09:00 am, not at the midnight! Also, there seems more traffic during the autumn semester (September–January) than spring semester (February–June).

Figure 5 shows the 1-year Email traffic data in 3D format (top) and the corresponding 2D format (bottom). Similar to the WWW data, the Email data also shows week day and weekend effect, as well as seasonal effect. However, different from the WWW data, the major of the Email traffic was between 09:00 am and 18:00 pm, there is very little traffic in the evening and early in the morning. So in these periods, people browsed the web but did not send many emails. The massive peak at the middle of the graph is due to the Email upgrade, where a lot of emails have been sent and received.

3.2. Network traffic data and Fourier transform

Figure 6 shows the original 1-year LSBU total network traffic data (top) and the corresponding Fourier transforms (bottom). **Figure 7** shows the original 1-year LSBU WWW data (top) and the corresponding Fourier transforms (bottom). **Figure 8** shows the original 1-year LSBU Email data (top) and the corresponding Fourier transforms (bottom).

http://dx.doi.org/10.5772/intechopen.76455

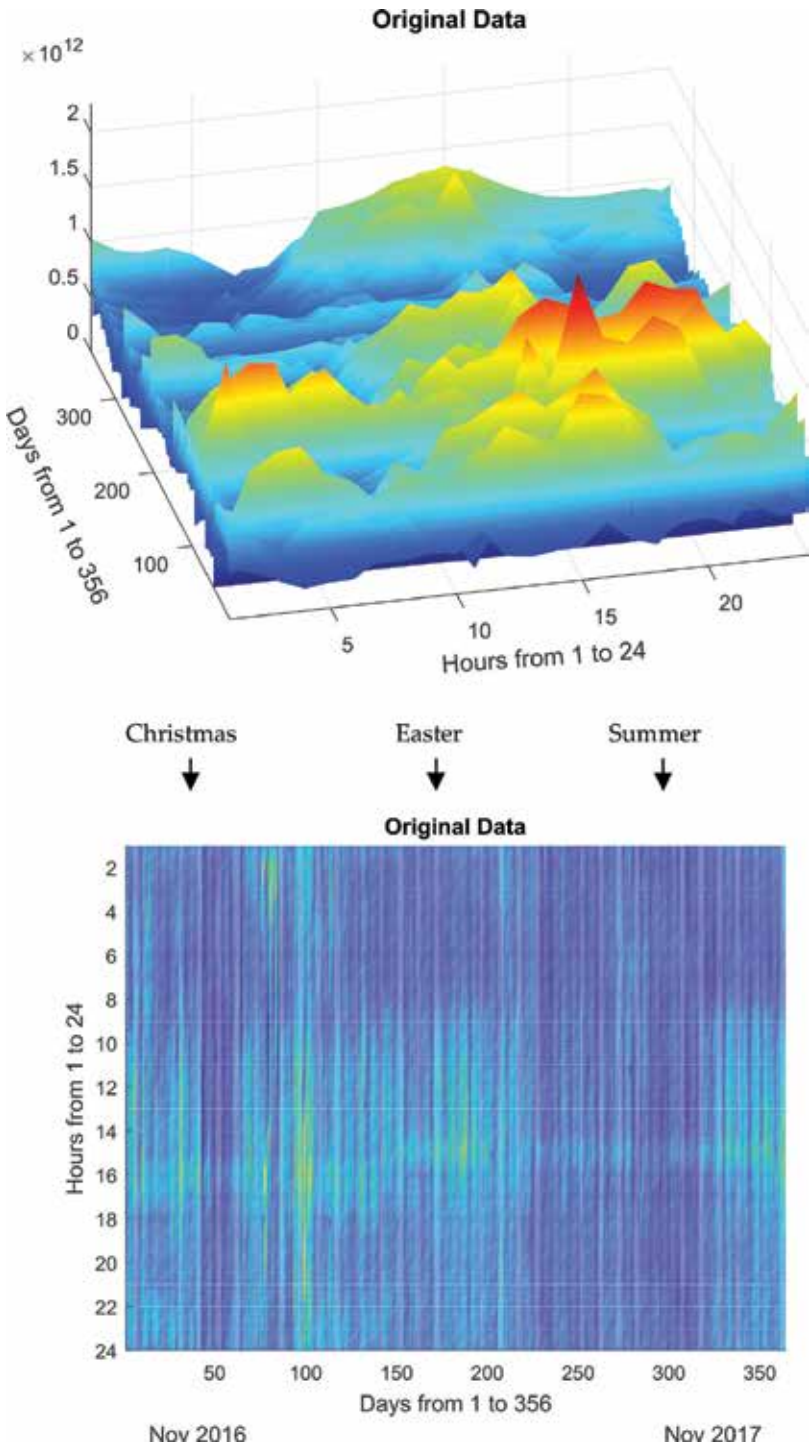


Figure 3. The 3D presentation (top) and the corresponding 2D presentation (bottom) of 1-year total network data in a daily usage pattern (Nov 2016–Nov 2017).

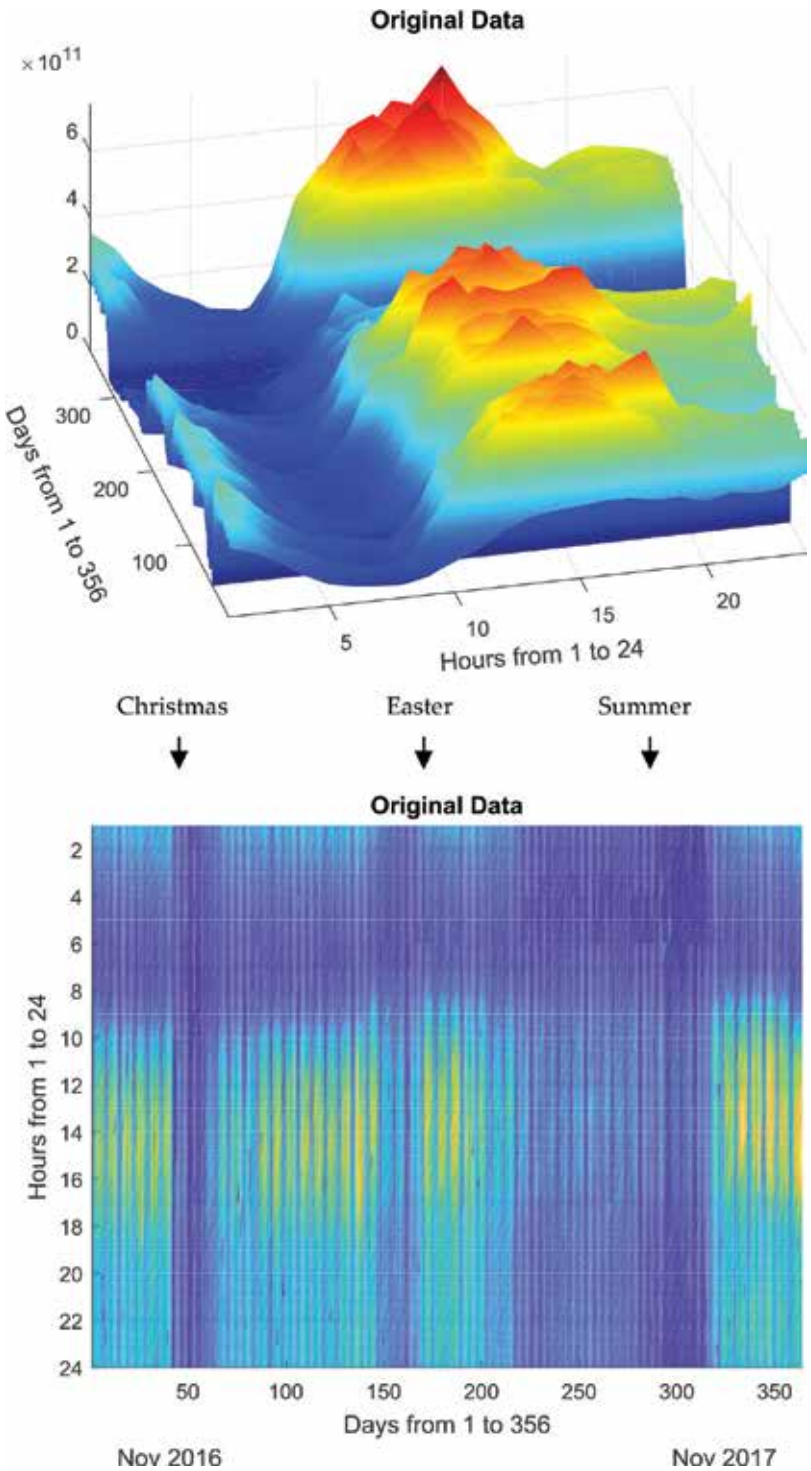


Figure 4. The 3D presentation (top) and the corresponding 2D presentation (bottom) of 1-year WWW traffic data in a daily usage pattern (Nov 2016–Nov 2017).

http://dx.doi.org/10.5772/intechopen.76455

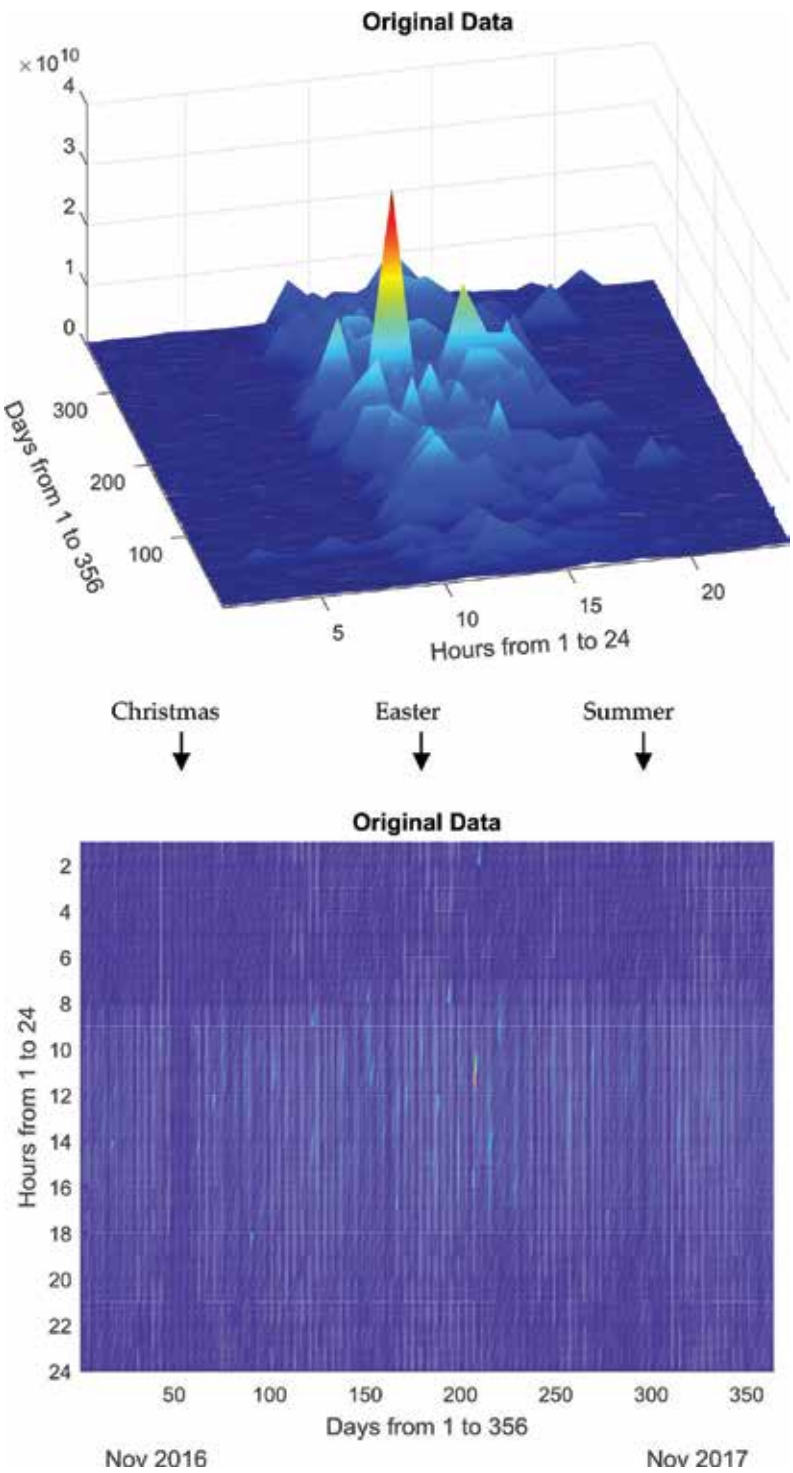


Figure 5. The 3D presentation (top) and the corresponding 2D presentation (bottom) of 1-year Email data in a daily usage pattern (Nov 2016–Nov 2017).

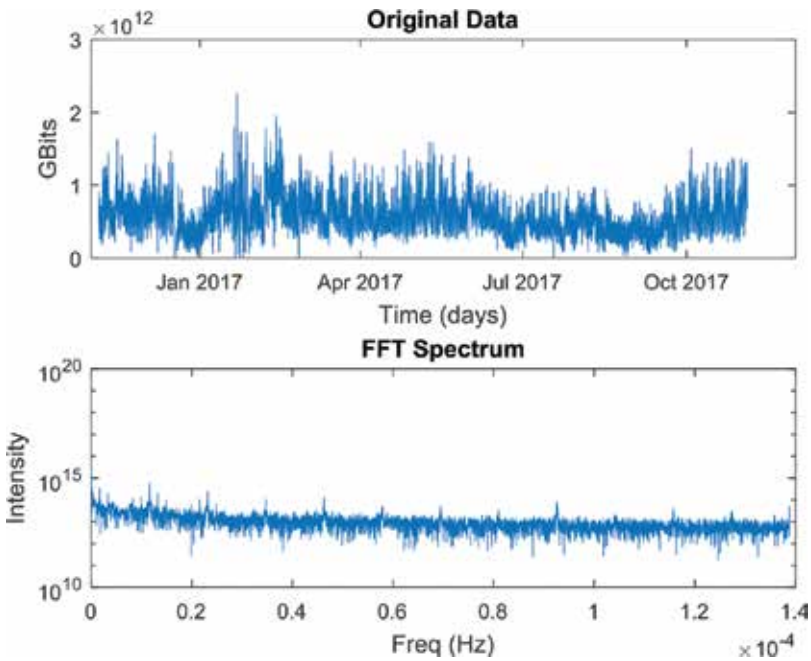


Figure 6. The original 1-year total network traffic data (top) and its corresponding FFT spectrum (bottom) (Nov 2016–Nov 2017).

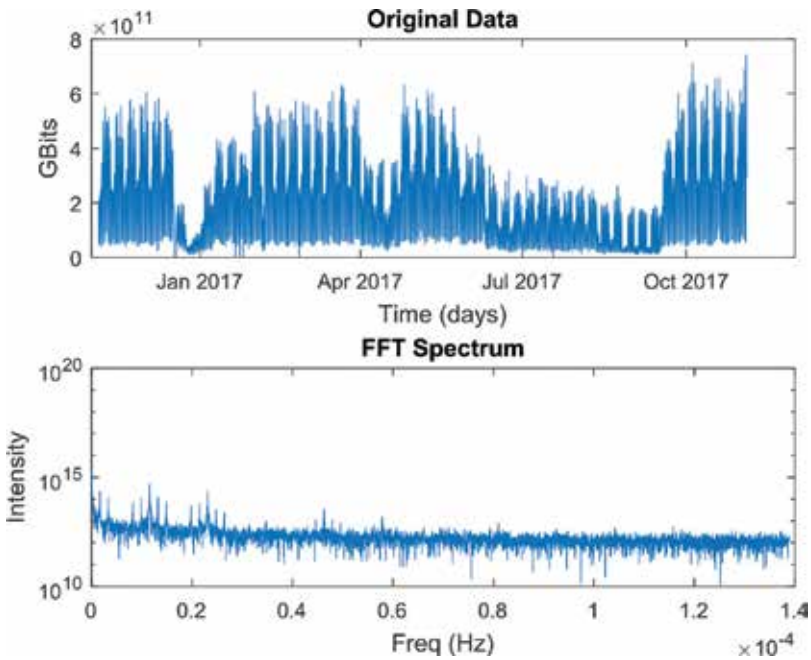


Figure 7. The original 1-year WWW data (top) and its corresponding FFT spectrum (bottom) (Nov 2016–Nov 2017).

The total network traffic data, the WWW data, and the Email data, have completely different patterns, and therefore different FFT spectrum. With the total network traffic data, there are a lot of small peaks throughout the FFT spectrum, indicating there are repeatedly happened events. But with the WWW data and Email, the FFT peaks mainly occurs at lower frequency range. But with FFT, it is not possible to identify when these events happened.

3.3. Wavelet decomposition

Wavelet decomposition [18] is a powerful tool that can decompose the original network traffic into low frequency component (A) and high frequency component (D). The low frequency component (A) is also called approximation coefficient, and the high frequency component (D) is called detail coefficient. By performing decomposition several times, we also have multilevel wave decomposition, see **Figure 9**. The multilevel wavelet decomposition allows us to gradually separate and to eliminate high frequency components, which is mostly noise. Through wavelet decomposition we can reduce the data noise, and therefore observe the data trend better.

Figure 10 shows the wavelet decomposition of the WWW network traffic data, wavelet used is "sym4" wavelet (see **Figure 1**) and the level of decomposition is level 4. The key in wavelet decomposition is to choose the right wavelet and to select the right level of decomposition. The results show that the low frequency component (A4) reflects better the trend of the network traffic data, where the high frequency component (D4) reflects more about the traffic noise.

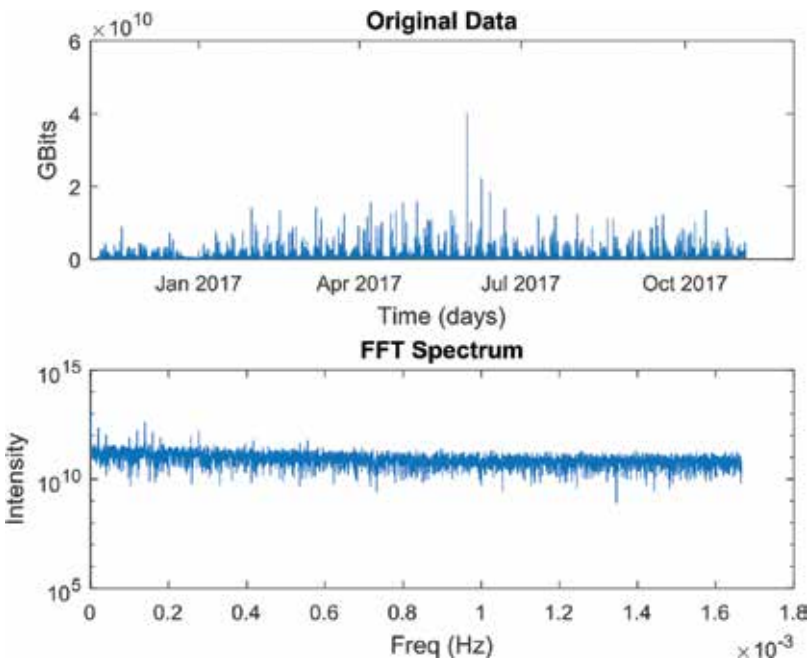


Figure 8. The original LSBU 1 year Email data (top, Nov 2016 – Nov 2017) and its corresponding FFT spectrum (bottom).

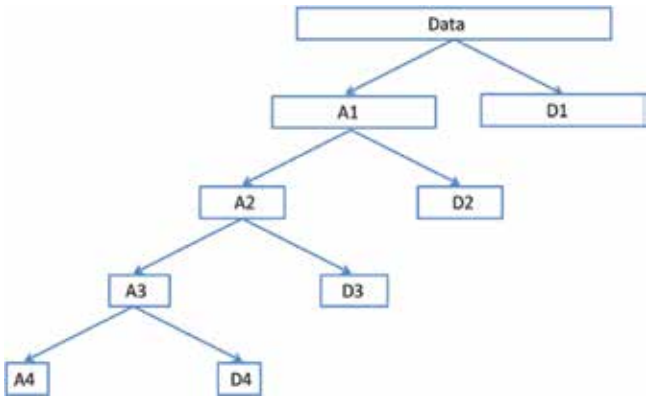


Figure 9. Multilevel wavelet decomposition, where approximation coefficient A is the low frequency component and detail coefficient D is the high frequency component.

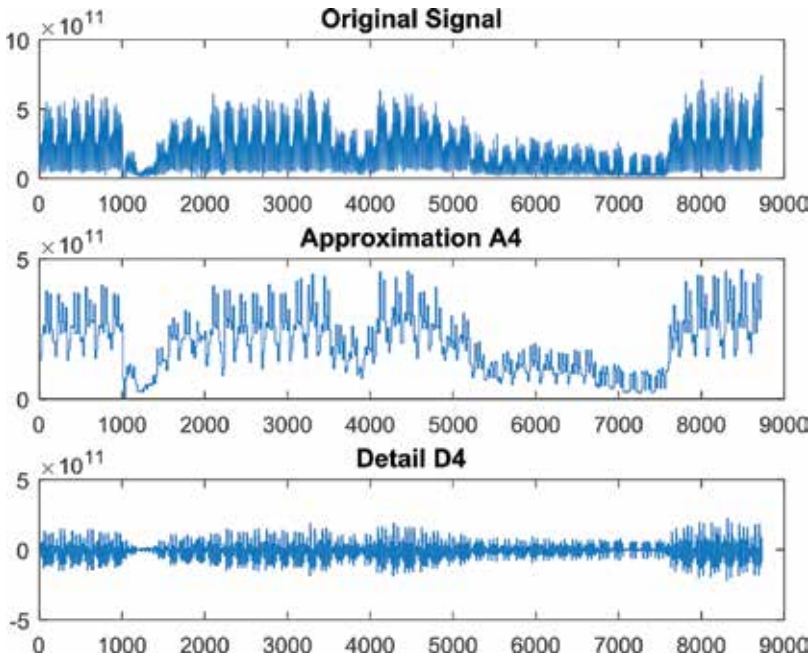


Figure 10. One-year WWW data in an hourly usage pattern (top, Nov 2016–Nov 2017), the corresponding level 4 low frequency component (A4) (middle) and level 4 high frequency component (D4) (bottom).

3.4. Wavelet denoising

Based on wavelet decomposition, a very useful feature of wavelet analysis is denoising, which is very useful for noisy data. The steps are as follows. First, choose a wavelet and a level of decomposition N, and then compute the wavelet decompositions of the data at levels 1 to N. For each level, a threshold is selected and the threshold applied to the detail coefficients (D). Finally, compute wavelet reconstructions using the original approximation coefficients (A) of level N and the modified detail coefficients (D) of levels 1 to N.

Figure 11 shows the 1-year total network traffic data in an hourly usage pattern (Nov 2016–Nov 2017) and the corresponding denoised results. In this case, the wavelet used is "sym8" wavelet (see **Figure 1**), the level of decomposition was chosen as N = 3. Similar to wavelet decomposition, the key in wavelet denoising is to choose the right wavelet and to select the right level of decomposition, in order balance the noise removal and signal integrity. The denoising of the WWW data and Email also yields similar results.

The quality of the denoised results is good. The trends of the original network traffic data are well preserved. To select the right wavelet and right level of decomposition is very important so that we can achieve maximum denoising and preserve the useful information.

3.5. Continuous wavelet analysis (CWT)

With continuous wavelet transform (CWT), we can analyse the data and show how the frequency content of the data changes over time. This time dependent frequency varying information, which is lacking in other techniques, such FFT, is very useful for network traffic analysis. In this CWT calculation, there are several parameters to choose from, i.e. the type of wavelet, the smallest scale (S_0), the space between scales (ds) and number of scales (N_s). The scales (S) can be converted to pseudo frequencies (f_p) by using the following formula,

$$f_p = \frac{f_c}{S\,dt} \tag{2}$$

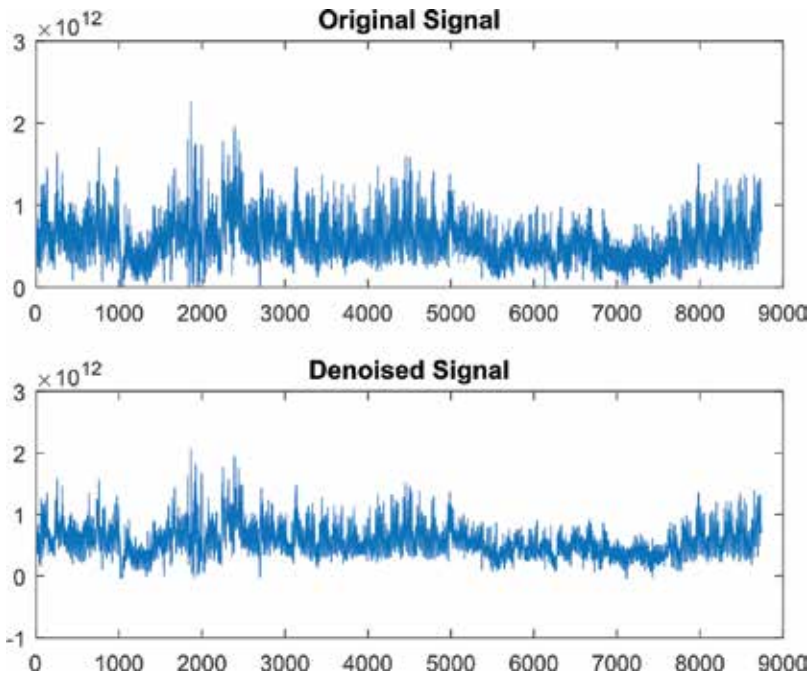


Figure 11. One-year total network traffic data in an hourly usage pattern (Nov 2016–Nov 2017) and the corresponding denoised results.

Where f_c is the centre frequency of the wavelet, and dt is the sampling time. Scales are inversely proportional to frequencies, i.e. small scales represents high frequencies, and vice versa.

Figure 12 shows the continuous wavelet transform (CWT) of 1 year total traffic using different wavelets, e.g. Morlet wavelet (analytic), Mexican hat wavelet (nonanalytic), bump wavelet (analytic), and Paul wavelet (analytic). The X axis is time of 1 year, and the Y axis is pseudo frequency. Different wavelet gives different results. Based on the results, we have decided to use Morlet wavelet to analyse the network traffic data, as it can provide more details on daily, weekly and monthly events, more details will be discussed later.

Figure 13 shows the continuous wavelet transform (CWT) using different parameters. The smallest scale (S_0) decides the highest frequency. The ds decides the resolution of the results, N_s decides the range of the frequency. By balancing the result resolution, frequency range and calculation time, we have decided to perform the CWT using the following values, $S_0 = 6 \times 3600 = 21{,}600$, i.e. six-hourly event, ds = 0.025, and $N_s = 300$.

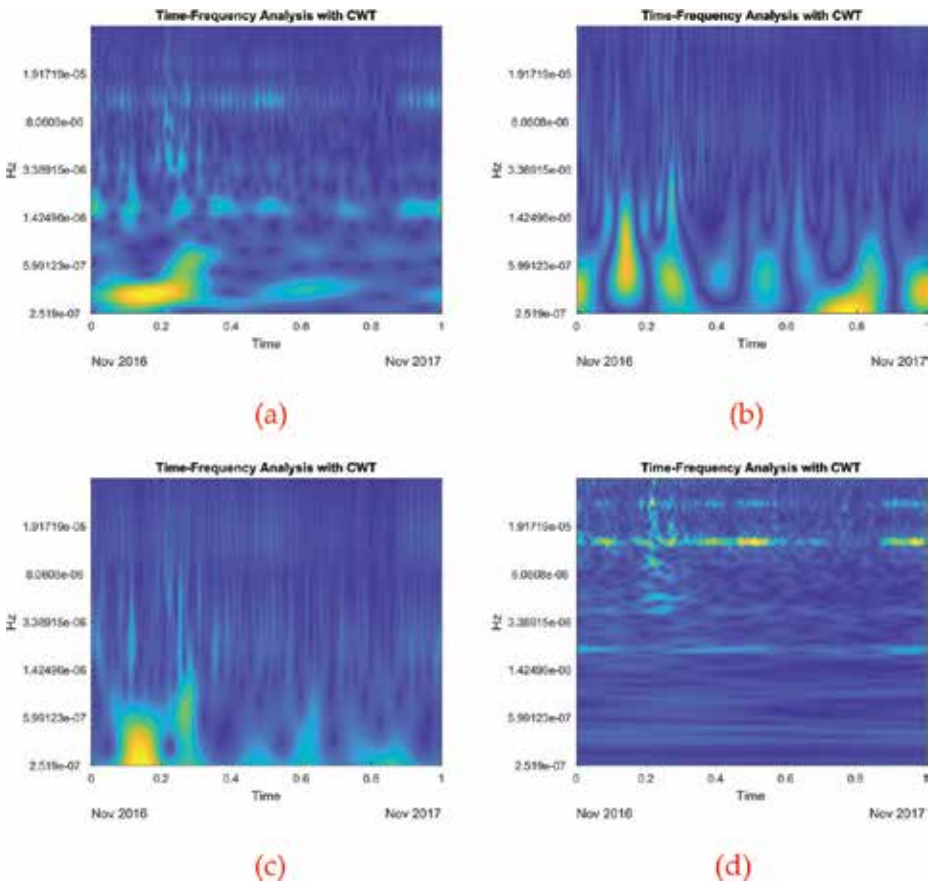


Figure 12. The continuous wavelet transform (CWT) using different wavelets. (a) CWT with Morlet wavelet, (b) CWT with Mexican hat wavelet, (c) CWT with Paul wavelet and (d) CWT with bump wavelet.

http://dx.doi.org/10.5772/intechopen.76455

Figure 14 shows the CWT results of the original 1 year total traffic data using Morlet wavelet, with S_0 = 21,600, ds = 0.025, and N_s = 300 as CWT parameters. The X axis is time of 1 year, and the Y axis is pseudo frequency. We can convert this pseudo frequency into event. **Table 1** shows the pseudo frequencies in Hz of hourly, daily, weekly, two weekly, monthly and quarterly events. Using these pseudo frequencies we can then identify the corresponding hourly, daily, weekly, two weekly, monthly and quarterly events in **Figure 11**. The hot spot at the lower left corner is the when the system is upgraded. By using CWT, we can easily identify the event which is otherwise difficult to identify in the original time domain.

Figure 15 shows the CWT results of the WWW traffic data using the same wavelet and same parameters. The results show that half daily and daily events happen throughout the year. They are highly seasonal, as you can clearly identify the summer, Christmas, and Easter gaps. The half daily and daily events also show clear day and night effects, as well as weekday and weekend effect, while weekly, two weekly and monthly events are patchy, with no seasonal effects.

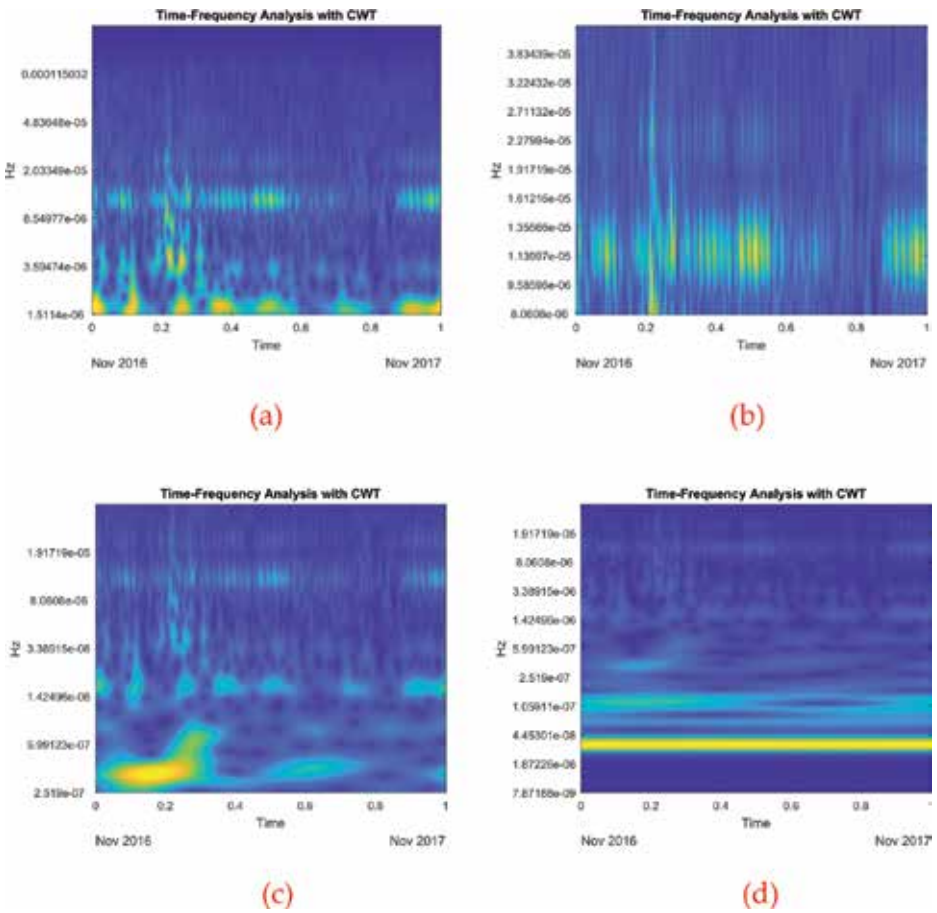


Figure 13. The continuous wavelet transform (CWT) using different parameters. (a) S_0 = 3600, ds = 0.025, N_s = 300, (b) S_0 = 21,600, ds = 0.025, N_s = 100, (c) S_0 = 21,600, ds = 0.025, N_s = 300 and (d) S_0 = 21,600, ds = 0.025, N_s = 500.

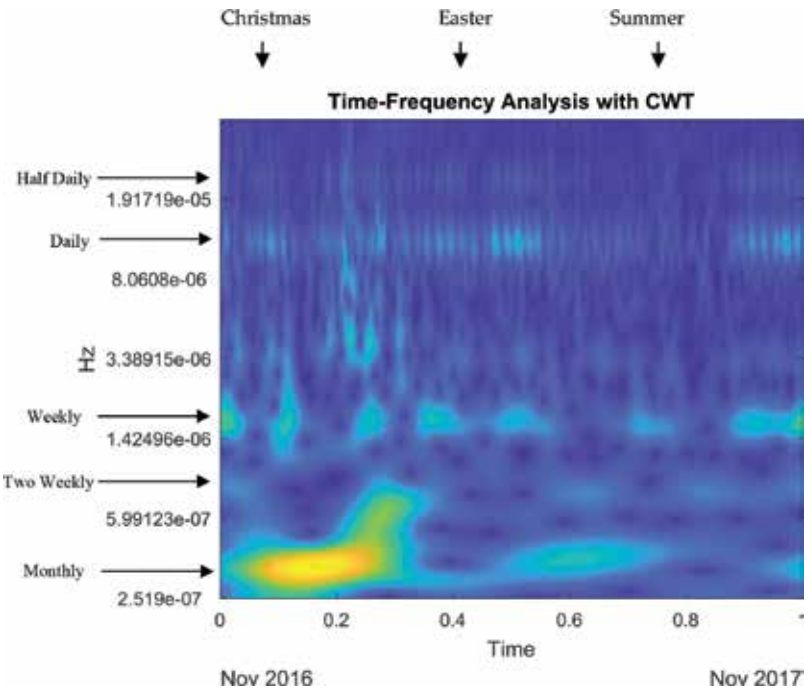


Figure 14. The CWT time-frequency 2D results of the 1-year total network traffic data (Nov 2016–Nov 2017). The hot spot at the lower left corner is the when the system is upgraded.

Time	Pseudo frequency (Hz)
Hourly	2.78E−04
Quarter daily	1.39E−04
Half daily	2.31E−05
Daily	1.16E−05
Weekly	1.65E−06
Two weekly	8.27E−07
Monthly	3.86E−07
Quarterly	1.29E−07

Table 1. The pseudo frequencies (Hz) of different events.

Figure 16 shows the CWT results of the Email traffic data using the same wavelet and same parameters. The results show that hourly event and quarter daily events happen throughout the year. The Christmas gap is obvious whilst the summer and the Easter gaps are not. They also show clear weekday and weekend effects. The half daily, daily and weekly events are very patchy, with no seasonal effects. This kind of time-frequency results can help us to understand the traffic characteristics better.

http://dx.doi.org/10.5772/intechopen.76455

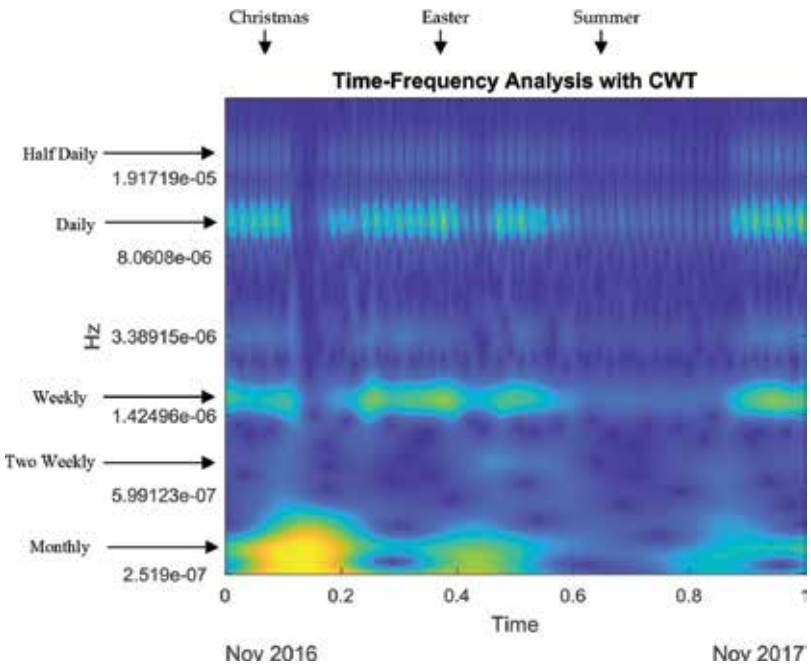


Figure 15. The CWT time-frequency 2D results of the 1-year WWW data (Nov 2016–Nov 2017).

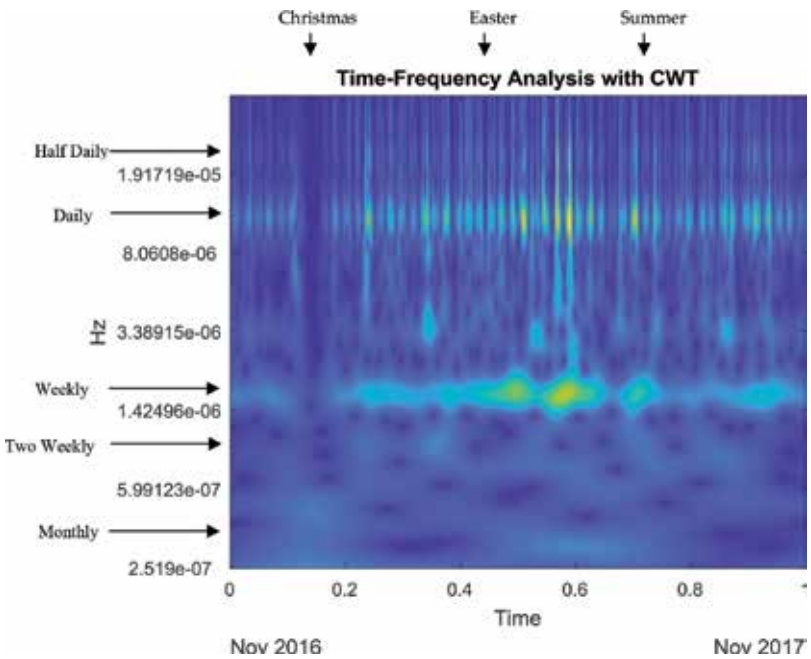


Figure 16. The CWT time-frequency 2D results of the 1-year Email data (Nov 2016–Nov 2017).

4. Conclusions

We present out latest study on using wavelet transform technique for analysing the educational network traffic data. The 2D and 3D presentation network traffic data, i.e. traffic in 24 h and 365 days, helps us to understand better the network traffic pattern. With wavelet transform, we are able to perform network traffic data decomposition and data denoising. With continuous wavelet transform (CWT), we can analyse the data and show how the frequency content of the data changes over time. The CWT analysis shows different characteristics of total traffic data, WWW data and Email data. This time dependent frequency varying information, which is lacking in other techniques, such FFT, is very useful for network traffic analysis. By using CWT, we can easily identify the event which is otherwise difficult to identify in the original time domain.

Acknowledgements

We thank London South Bank University for the financial support of this study.

Author details

Shwan Dyllon[1] and Perry Xiao[2]*

*Address all correspondence to: xiaop@lsbu.ac.uk

1 Department of ICT, London South Bank University, London, UK

2 School of Engineering, London South Bank University, London, UK

References

[1] Zhang J, Song C, Hu Y, Yu B. Improving robustness of robotic grasping by fusing multi-sensor. In: IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems. 2012. pp. 126-131

[2] Bermeo JP, Castillo H, Armijos X, Jara JD, Sanchez F, Bermeo H. Artificial Neural Network and Monte Carlo Simulation in a Hybrid Method for Time Series Forecasting with Generation of L-Scenarios. 2016 Intl IEEE Conferences on Ubiquitous Intelligence & Computing, Advanced and Trusted Computing, Scalable Computing and Communications, Cloud and Big Data Computing, Internet of People, and Smart World Congress (UIC/ATC/ScalCom/CBDCom/IoP/SmartWorld), Toulouse; 2016. pp. 665-670. DOI: 10.1109/UIC-ATC-ScalCom-CBDCom-IoP-SmartWorld.2016.0110

[3] Guo T, Xu Z, Yao X, Chen X, Aberer K, Funaya K. Robust online time series prediction with recurrent neural networks. In: 2016 IEEE International Conference on Data Science and Advanced Analytics (DSAA). 2016. pp. 816-825. Available from: http://ieeexplore.ieee.org/document/7796970/

[4] Rocha E, Salvador P, Nogueira A. A real-time traffic classification approach. In: 2011 International Conference for Internet Technology and Secured Transactions (ICITST); December. 2011. pp. 620-626

[5] Han L, Huang L, Hu Q, Han X, Shi J. Fast Fourier transform based IP traffic classification system for SIPTO at H(e)NB. In: 7th International Conference on Communications and Networking in China. 2012. pp. 430-435. Available from: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6417521

[6] Shang P, Li X, Kamae S. Chaotic analysis of traffic time series. 2005;**25**:121-128. http://doi.org/10.1016/j.chaos.2004.09.104

[7] Feng H, Shu Y, Yang OWW. Nonlinear analysis of wireless LAN traffic. Nonlinear Analysis: Real World Applications. 2009;**10**(2):1021-1028. Available from: http://www.sciencedirect.com/science/article/pii/S1468121807002386

[8] Drew M. A frequency analysis approach for categorizing air traffic behavior. In: 14th AIAA Aviation Technology, Integration, and Operations Conference; Atlanta, GA; 16-20 June. 2014. Available from: https://www.aviationsystemsdivision.arc.nasa.gov/publications/2014/AIAA-2014-2420.pdf

[9] Song R, Mason PC, Li M. Enhancement of frequency-based wormhole attack detection. In: Proceedings of the IEEE Military Communications Conference (MILCOM). 2011. pp. 1139-1145

[10] Kim SS, Reddy ALN, Vannucci M. Detecting traffic anomalies using discrete wavelet transform. 2004. pp. 951-961. Available from: http://link.springer.com/chapter/10.1007%2F978-3-540-25978-7_96

[11] Weng H, Lau K-M. Wavelets, period doubling, and time-frequency localization with application to organization of convection over the tropical western Pacific. Journal of the Atmospheric Sciences. 1994;**51**:2523-2541

[12] Daubechies I. The wavelet transform time-frequency localization and signal analysis. IEEE Transactions on Information Theory. 1990;**36**:961-1004

[13] Lindsay RW, Percival DB, Rothrock DA. The discrete wavelet transform and the scale analysis of the surface properties of sea ice. IEEE Transactions on Geoscience and Remote Sensing. 1996;**34**:771-787

[14] Mak M. Orthogonal wavelet analysis: Interannual variability in the sea surface temperature. Bulletin of the American Meteorological Society. 1995;**76**:2179-2186

[15] Kaiser G. A Friendly Guide to Wavelets. Cambridge, MA: Birkhäuser Boston; 1994. 300 pp

[16] Dyllon S, Dahimene1 H, Wright P, Xiao P. Analysis of HTTP and HTTPS usage on the university internet backbone links. Journal of Industrial and Intelligent Information. 2014;**2**(1):67-70. Available at: http://www.jiii.org/index.php?m=content&c=index&a=show&catid=37&id=88

[17] Dyllon S, Saravanan D, Xiao P. The Usage Analysis of Web and Email Traffic on the University Internet Backbone Links. In: Proceedings of the International MultiConference of Engineers and Computer Scientists. Hong Kong; March 18-20, 2015;**I**:18-21

[18] Wavedec. Multilevel 1-D wavelet decomposition. https://uk.mathworks.com/help/wavelet/ref/wavedec.html?searchHighlight=wavedec&s_tid=doc_srchtitle

Chapter 9

The 2D Continuous Wavelet Transform: Applications in Fringe Pattern Processing for Optical Measurement Techniques

José de Jesús Villa Hernández, Ismael de la Rosa, Gustavo Rodríguez, Jorge Luis Flores, Rumen Ivanov, Guillermo García, Daniel Alaniz and Efrén González

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/intechopen.74813

Abstract

Optical metrology and interferometry are widely known disciplines that study and develop techniques to measure physical quantities such as dimensions, force, temperature, stress, etc. A key part of these disciplines is the processing of interferograms, also called fringe patterns. Owing that this kind of images contains the information of interest in a codified form, processing them is of main relevance and has been a widely studied topic for many years. Several mathematical tools have been used to analyze fringe patterns, from the classic Fourier analysis to regularization methods. Some methods based on wavelet theory have been proposed for this purpose in the last years and have evidenced virtues to consider them as a good alternative for fringe pattern analysis. In this chapter, we resume the theoretical basis of fringe pattern image formation and processing, and some of the most relevant applications of the 2D continuous wavelet transform (CWT) in fringe pattern analysis.

Keywords: 2-D wavelets, fringe patterns, optical measurement techniques

1. Introduction

Fringe pattern processing has been an interesting topic in optical metrology and interferometry; owing to its relevance nowadays, it is a widely studied discipline. Digital fringe pattern processing is used in optical measurement techniques such as optical testing [1, 2], electronic

speckle pattern interferometry (ESPI), holographic interferometry, and moiré interferometry or profilometry [3–5]. They are quite popular for non-contact measurements in engineering and have been applied for measuring various physical quantities like displacement, strain, surface profile, refractive index, etc. In optical methods of measurement, the phase, which is related to the measured physical quantity, is encoded in an intensity distribution represented in an image which is, in general, the result of the interference phenomena. This phenomenon is used in classical interferometry, in holographic interferometry, and in electronic speckle pattern interferometry to convert the phase of a wave of interest into an intensity distribution. As the physical quantity to be measured is codified as the phase of a fringe pattern image, the main task of fringe pattern processing is to recover such phase.

The methods for phase recovery from fringe patterns can be classified mainly in three categories [2, 6]: (a) Phase-stepping or phase-shifting methods which require a series of fringe images to recover the phase information. (b) Spatial domain methods which can compute the phase from a single fringe pattern in the spatial domain. (c) Frequency domain methods which uses some kind of transformation to the frequency domain to compute the phase. In this category, the Fourier and Wavelet transforms are the most common mathematical tools to carry out the task.

Apart from the phase recovery, there are other important steps in fringe pattern processing. For example, many times the fringe patterns are corrupted by noise, such as the case of the electronic speckle pattern interferometry. Then, fringe image enhancement by means of low-pass filtering is usually required. Owing that most algorithms to retrieve the phase from a fringe pattern give the phase wrapped in the interval $[-\pi, \pi)$, other important step is the well-known phase-unwrapping process [6, 7]. In the field of fringe image enhancement, such as fringe image denoising or phase denoising, there has been a wide research activity in the last years. Researchers have realized that improving the quality of fringe images and wrapped phase fields is of main relevance for a successful phase recovery or phase unwrapping. However, enhancing fringe images or wrapped phase fields has resulted to be a task that must be realized in a special manner, so that ordinary techniques for image enhancement are not always adequate. Owing that frequencies of fringes and noise usually overlap and normally cannot be properly separated, common filters for image processing have blurring effects on fringe features, especially for patterns with high density fringes. For these cases, the use of anisotropic filters is a better way for removing noise without the harmful blurring effects.

In the fields of fringe pattern denoising and wrapped phase map denoising, there have been many proposals to realize these tasks. Some of the first contributions in this field were mainly based on convolution filters using different kinds of anisotropic filtering masks [8–12]. Other set of the main contributions in the last years is based on the variational calculus approach by solving partial differential equations [13–18], and by means of the regularization theory [19, 20]. The use of the Fourier transform for fringe or phase map denoising has also been proposed in [21, 22] (Localized Fourier transform filter and windowed Fourier transform, respectively). There have been other proposals that used different methodologies such as coherence enhancing diffusion [23], image decomposition [24], and multivariate empirical mode decomposition [25]. The great disadvantage of already reported methods for fringe and phase map denoising

http://dx.doi.org/10.5772/intechopen.74813

is that they require the previous estimation of the so-called fringe orientation which, as it uses the computation of the image gradient, could be an inaccurate procedure in the presence of noise and low modulation of fringes. This is not the case for the Fourier-based methods [21, 22]; however, as in the case of the Windowed Fourier transform technique, several parameters have to be adjusted depending on the particular image and it may require a long processing time.

In the field of phase recovery from fringe images, there have been a lot of researches along the last decades. For the case of phase-shifting algorithms, outstanding summaries of them can be found in [2, 26]. For the case of spatial and frequency domain methods from a single pattern image, two of the most popular techniques are the well-known Fourier Transform method reported by Takeda et al. [27] and the Synchronous detection method [28]. Other methods that use the regularization theory were also proposed [29, 30]. However, although these methods are efficient and easy to implement, they are limited to be used in fringe images with frequency carrier, which just in few experimental situations these kinds of images can be obtained. In most cases, experimental conditions in optical measurement techniques yields fringe images without a dominant frequency (i.e., closed fringes) which becomes the phase recovery problem difficult, therefore more complicated algorithms must be used. One of the first proposals for phase demodulation from single closed fringe images was reported by Kreis using a Fourier based approach [31]. In the last decade of the twentieth century, it was a boom in the research of closed fringe images, specially using the regularization theory. The Regularized phase-tracking technique was reported by Servín et al. [32]. Marroquín et al. reported the regularized adaptive quadrature filters [33] and the regularization method that uses the local orientation of fringes [34]. At the beginning of this century, Larkin et al. proposed the spiral-phase quadrature transform [35] and Servín et al. reported the General n-dimensional quadrature transform [36]. Also, we proposed the orientational vector-field-regularized estimator to demodulate closed fringe images [37].

As will be shown, closed fringe and wrapped-phase images have certain characteristics that make them to be treated in a special manner. First, it is common that this kind of images present structures with high anisotropy at the same time that many frequencies are dispersed over the entire image. For these reasons, in most situations, the use of linear-translation-spatial (LTI) filters, which are spatially invariant and independent of image content, do not give proper results. Furthermore, owing that the Fourier transform is a global operation, this technique is not always suitable for accurately model the local characteristics of closed fringe images.

It is widely known that the wavelet transform is a powerful tool that provides local, sparse, and decorrelated multiresolution analysis of signals. In the last years, 2D wavelets have been used for image analysis as a proper alternative to the weakness of LTI filters and linear transforms as the Fourier one. In particular, it has been shown that 1-D and 2D continuous wavelet transform (CWT) using Gabor atoms is a natural choice for proper analyses of fringe images. This kind of analysis has been used for fringe pattern denoising and fringe pattern demodulation showing several advantages, for example in laser plasma interferometry [38], in shadow moiré [39–41], in profilometry [42–44], in speckle interferometry [45], in digital holography [46], and other optical measurement techniques [47–55].

In this chapter, the theoretical basis of fringe pattern image formation and processing is described. Also, in general, the theory and advantages of the 2D continuous wavelet transform (CWT) for fringe pattern processing is described. We also explain some of the main applications in fringe pattern processing, such as phase recovery and wrapped phase map denoising, showing some examples of applications in different optical measurement techniques.

2. Digital fringe patterns

2.1. Elements of digital fringe image processing systems

Often, a digital fringe image processing system is represented by a sequence of devices, which typically starts with an imaging system that observes the target, a digitizer system which samples and quantizes the analog information acquired by the imaging system, a digital storage device, a digital computer that process the information, and finally, a displaying system to visualize the acquired and processed information (**Figure 1**).

A typical imaging system is composed by an objective lens to form images in a photosensitive plane which is commonly a CCD (charge couple devices) array.

2.2. Fringe image formation

Fringe pattern images are present in several kinds of optical tests for the measurement of different physical quantities. Such tests are examples for the quality measurement of optical devices using optical interferometry, photoelasticity for stress analysis, or electronic speckle pattern interferometry (ESPI) for the measurement of mechanical properties of materials. The interference phenomena are usually used in many optical methods of measurement. We now describe a classical way to form a fringe pattern image using the two-wave interference.

Two-wave interference can be generated by means of several types of interferometers, and the interferograms or fringe patterns are produced by superimposing two wavefronts. An interferometer can accurately measure deformations of the wavefront of the order of the wavelength. Considering two mutually coherent monochromatic waves, as depicted in **Figure 2**, $W(x, y)$ represents the wavefront shape under study (i.e., the wave that contains the information of the physical quantity to be measured). The sum of their complex amplitudes can be represented as

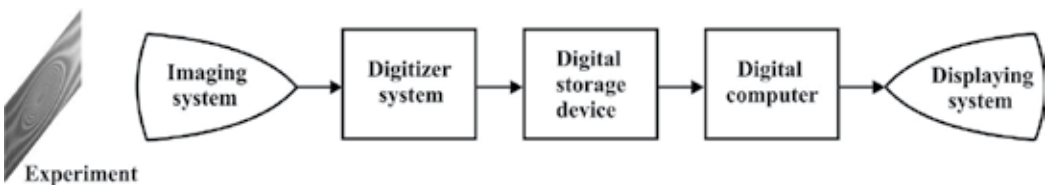


Figure 1. Typical sequence in a digital fringe pattern image processing system.

http://dx.doi.org/10.5772/intechopen.74813

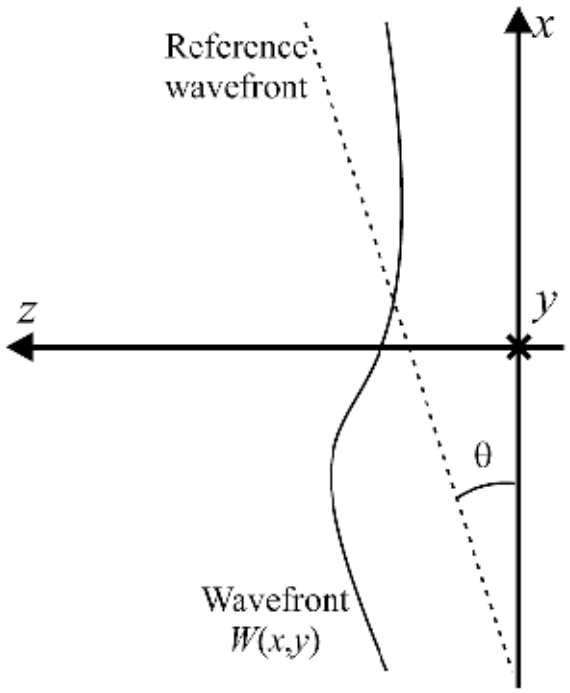


Figure 2. Interference of two wavefronts. Solid line represents the wavefront under test and dashed line represents the reference wavefront.

$$E(x,y) = A_1(x,y)e^{ikW(x,y)} + A_2(x,y)e^{ikx\sin\theta}, \tag{1}$$

where A_1 and A_2 are the amplitudes of the wavefront under test and the reference wavefront (a flat wavefront), respectively, and $k = \frac{2\pi}{\lambda}$, being λ the wavelength.

The irradiance at a given plane perpendicular to z-axis is then represented as

$$\begin{aligned} I(x,y) &= E(x,y)E^*(x,y) \\ &= A_1^2(x,y) + A_2^2(x,y) + 2A_1(x,y)A_2(x,y)\cos\left[kx\sin\theta + kW(x,y)\right]. \end{aligned} \tag{2}$$

For simplicity, Eq. (2) is usually written in a general form as:

$$I(x,y) = a(x,y) + b(x,y)\cos\left[u_0 x + \phi(x,y)\right], \tag{3}$$

where $a(x,y)$ and $b(x,y)$ are commonly called the background illumination and the amplitude modulation, respectively. The term $u_0 = k\sin\theta$ is the fringe carrier frequency and $\phi(x,y) = kW(x,y)$ is the phase to be recovered from the fringe pattern image. It must be noted that if the reference wavefront is perpendicular to z-axis (i.e., $\theta = 0$), the fringe carrier frequency is removed and Eq. (3) is simplified:

$$I(x,y) = a(x,y) + b(x,y)\cos\left[\phi(x,y)\right]. \tag{4}$$

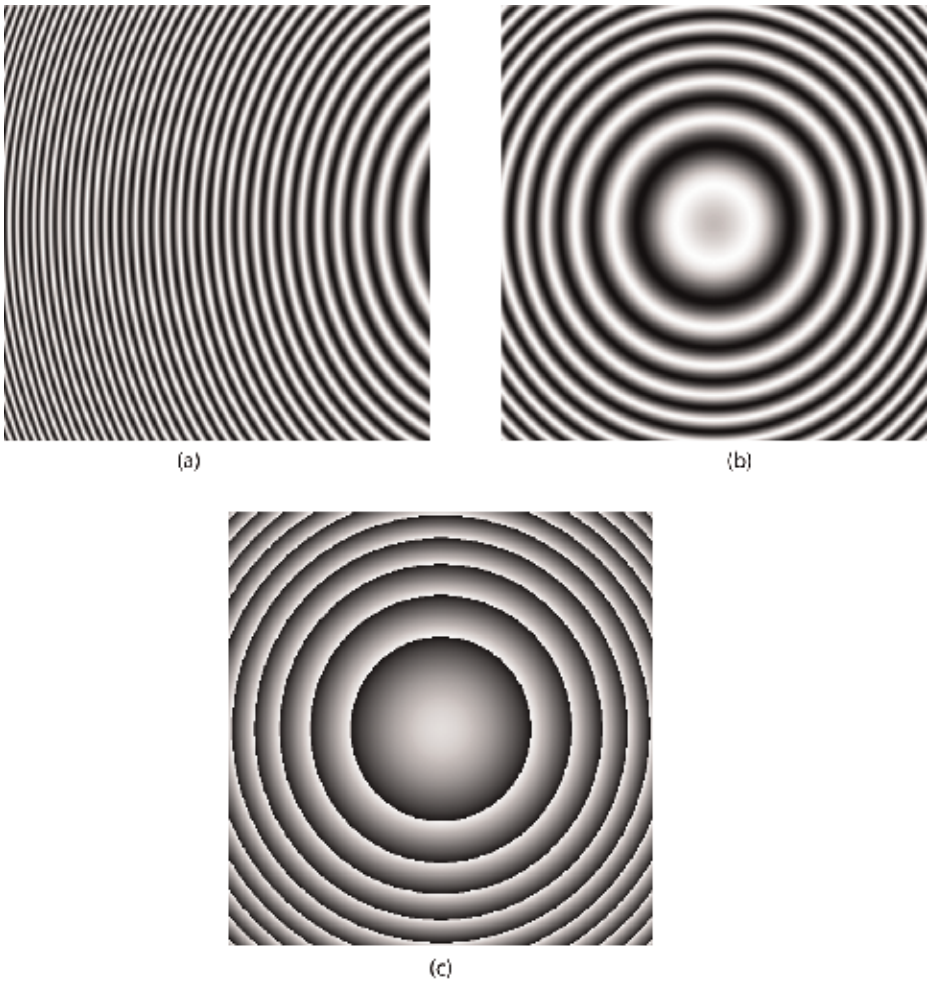


Figure 3. Examples of simulated fringe pattern images with (a) and without (b) fringe carrier frequency. The phase of modulation $\phi(x, y)$ (c) is the same for both fringe images (phase shown wrapped and codified in gray levels).

Equations (3) and (4) represent the mathematical expressions of fringe pattern images with and without fringe carrier frequency, respectively. Examples of these kinds of fringe images are shown in **Figure 3**.

3. Fringe pattern processing

3.1. Phase-shifting methods for phase recovery

One of the most popular methods for phase recovery is the well-known phase-shifting. This method requires a set of phase-shifted fringe patterns which are experimentally obtained in different ways depending on the optical measurement technique. For example, in interferometry

the phase shifting is realized by moving some mirrors in the optical interferometer. The set of N phase-shifted fringe patterns is defined as

$$I_n(x,y) = a(x,y) + b(x,y)\cos\left[\phi(x,y) + \alpha_n\right] \quad n = 1, 2, \ldots, N. \tag{5}$$

The pointwise solution for $\phi(x,y)$ from the non-linear system of equations is obtained by using the last-squares approach (see [2] for details):

$$W\{\phi(x,y)\} = \tan^{-1}\left(-\frac{\sum_{n=1}^{N} I_n \sin(\alpha_n)}{\sum_{n=1}^{N} I_n \cos(\alpha_n)}\right) \in [-\pi, \pi), \tag{6}$$

where W is the wrapping operator such that $W\{\phi(x,y)\} \in [-\pi, \pi)$. Several algorithms can be used that require three, four, up to eight images.

3.2. Phase recovery from single fringe patterns with carrier

As previously mentioned, processing fringe patterns with fringe carrier frequency may be simple to carry out. The key point in the demodulation of fringe patterns with carrier is that the total phase function $u_0 x + \phi(x,y)$ represents the addition of an inclined phase plane $u_0 x$ plus the target phase $\phi(x,y)$. In this case, a monotonically increasing (or decreasing) phase function has to be recovered. If we analyze the Fourier spectrum of Eq. (3), for a proper separation between spectral lobes in the Fourier space, the following inequality must be complied:

$$\max\{\|\nabla\phi\|\} < \|u_0\|. \tag{7}$$

The analytic signal $g(x,y)$ to recover the phase $\phi(x,y)$ can be computed with the Fourier transform method [27], which can expressed as

$$g(x,y) = \mathcal{F}^{-1}\{H(u,v)\mathcal{F}\{I(x,y)\}\} = e^{i2\pi\left[u_0 x + \phi(x,y)\right]}, \tag{8}$$

where $H(u,v)$ is a filter in the Fourier domain centered at the frequency u_0, u the frequency variable along x direction, and v the frequency variable along y direction. Finally, the wrapped phase is computed with

$$W\{\phi(x,y)\} = \tan^{-1}\left(\frac{\text{Real}\{g(x,y)e^{-i2\pi u_0}\}}{\text{Imag}\{g(x,y)e^{-i2\pi u_0}\}}\right) \in [-\pi, \pi). \tag{9}$$

Other technique to compute the phase from a carrier frequency fringe pattern is the synchronous detection technique [28], which is realized in the spatial domain. Using the complex notation, in this case, the analytic function $g(x,y)$ can be computed with

$$g(x,y) = h(x,y) * \left[I(x,y)e^{i2\pi u_0}\right] = e^{i2\pi\phi(x,y)}, \tag{10}$$

where $*$ represents the convolution operator and $h(x,y)$ a low-pass convolution filter in the spatial domain. The wrapped phase can be computed with

$$W\{\phi(x,y)\} = \tan^{-1}\left(\frac{\text{Real}\{g(x,y)\}}{\text{Imag}\{g(x,y)\}}\right) \in [-\pi,\pi). \tag{11}$$

3.3. Phase recovery from single fringe patterns without carrier

As described in [34–37], for the case in which $u_0 = 0$, the previous computation of the fringe direction is necessary to compute the analytic function $g(x,y)$, for example, using the quadrature transform [36]:

$$\text{Imag}\{g(x,y)\} = \sin\left[\phi(x,y)\right] = \mathbf{n}_\phi(x,y)\cdot\frac{\nabla I_n(x,y)}{\|\nabla\phi(x,y)\|}, \tag{12}$$

where $I_n(x,y) = \cos\left[\phi(x,y)\right] = \text{Real}\{g(x,y)\}$ is a normalized version of $I(x,y)$, and $\mathbf{n}_\phi$ is the unit vector normal to the corresponding isophase contour, which points to the direction of $\nabla\phi(x,y)$. It is well known that the computation of $\mathbf{n}_\phi$ is by far the most difficult problem to compute the phase using this method.

Also, the modulo-2π fringe orientation angle $\alpha(x,y)$ can be used to compute the quadrature fringe pattern by means of the spiral-phase signum function $S(u,v)$ in the Fourier domain [35]:

$$\text{Imag}\{g(x,y)\} = \sin\left[\phi(x,y)\right] = -ie^{-i\alpha(x,y)}\mathcal{F}^{-1}\{S(u,v)\mathcal{F}\{I_n(x,y)\}\}, \tag{13}$$

where

$$S(u,v) = \frac{u+iv}{\sqrt{u^2+v^2}}, \tag{14}$$

and $i = \sqrt{-1}$. However, the most difficult problem in this method is the computation of $\alpha(x,y)$. It can be deduced that Eqs. (12) and (13) are closely related because

$$\alpha(x,y) = \text{angle}\{\mathbf{n}_\phi(x,y)\} \in (0,2\pi]. \tag{15}$$

3.4. Wrapped phase maps denoising

The unwrapping process can be, in many cases, a difficult task due to phase inconsistencies or noise. In order to understand the phase unwrapping problem of noisy phase maps, we define the wrapped and the unwrapped phase as $\psi(x,y)$ and $\phi(x,y)$ respectively. As it is known that $\psi(x,y) \in [-\pi,\pi)$, the following relation can be established:

$$\psi(x,y) = \phi(x,y) + 2\pi k(x,y), \tag{16}$$

where $k(x,y)$ is a field of integers such that $\psi(x,y) \in [-\pi,\pi)$. The wrapped phase-difference vector field $\Delta\psi(x,y)$ which can be computed from the wrapped phase map, is defined as

$$\Delta\psi(x,y) = [\psi(x,y) - \psi(x-1,y), \psi(x,y) - \psi(x,y-1)], \tag{17}$$

where $(x-1,y)$ and $(x,y-1)$ are contiguous horizontal and vertical sites, respectively. In a similar manner, we can also define the unwrapped phase-difference field:

$$\Delta\phi(x,y) = \left[\phi(x,y) - \phi(x-1,y), \phi(x,y) - \phi(x,y-1)\right]. \tag{18}$$

It can be deduced that the problem of the recovery of ϕ from ψ can be properly solved if the sampling theorem is reached, that is, if the distance between two fringes is more than two pixels (the phase difference between two fringes is 2π). In phase terms, the sampling theorem is reached if the phase difference between two pixels is less than π or, in general

$$\|\Delta\phi\| < \pi, \quad \forall \quad (x,y). \tag{19}$$

If this condition is satisfied, the following relation can be established:

$$\Delta\phi = W\{\Delta\psi\} = \left[\psi_x, \psi_y\right], \tag{20}$$

where

$$\psi_x = W\{\psi(x,y) - \psi(x-1,y)\} \quad \text{and} \quad \psi_x = W\{\psi(x,y) - \psi(x,y-1)\}. \tag{21}$$

Note that $W\{\Delta\psi\}$ (the wrapped phase differences) can be obtained from the observed wrapped phase field ψ. Then, the unwrapped phase ϕ can be achieved by two-dimensional integration of the vector field $W\{\Delta\psi\}$.

A simple way to compute the unwrapped phase ϕ from the wrapped one ψ is by means of minimizing the cost function

$$U(\phi) = \sum_{(x,y)\in L} \left\{ \left[\psi_x(x,y) - \left(\phi(x,y) - \phi(x-1,y)\right)\right]^2 + \left[\psi_y(x,y) - \left(\phi(x,y) - \phi(x,y-1)\right)\right]^2 \right\}, \tag{22}$$

where L is the set of valid pixels in the image. Unfortunately, in most cases noise is present, therefore, inequality (19) is not always satisfied and the integration does not provide proper results. Therefore, denoising wrapped phase maps is a fundamental step before the phase unwrapping process.

4. The 2D continuous wavelet transform for processing fringe patterns

It is clear that the phase demodulation of fringe images with carrier may be easily realized. Owing that, in this case, the fringe image may represent a quasi-stationary signal along the direction of the frequency carrier, the use of classical linear operators such as the Fourier

transform may be adequate. It works well mainly for few components in the frequency domain (i.e., for narrow spectrums); however, this is not the case for many signals in the real world. This dependence is a serious weakness mainly in two aspects: the degree of automation and the accuracy of the method specially when fringes produce spread spectrums due to localized variations or phase transients. Additionally, in the case of closed fringes there may be a wide range of frequencies in all directions. Then, evidently standard Fourier analysis is inadequate for treating with this kind of images because it represents signals with a linear superposition of sine waves with "infinite" extension. For this reason, an image with closed fringes should be represented with localized components characterizing the frequency, shifting, and orientation. A powerful mathematical tool for signal description that has been developed in the last decades is the wavelet analysis. Fortunately, for our purposes, a key characteristic of this type of analysis is the finely detailed description of frequency or phase of signals. In consequence, it can have a good performance especially with fringes that produce spread spectrums. Additionally, one of the main advantages using wavelets compared with standard techniques is its high capability to deal with noise. In particular, the 2D continuous wavelet transform have recently been proposed for the processing of interferometric images. Advantages of denoising and demodulation of interferograms using the 2D CWT has been discussed in [44–55].

Considering an interferometric image (an interferogram or a wrapped-phase field) $G(r)$, where $r = (x, y) \in \mathbb{R}^2$, its 2D CWT decomposition can be defined as

$$G_{\mathcal{W}}(s, \theta, \eta) = \mathcal{W}\{G(r)\} = \int_{\mathbb{R}^2} G(r)\varphi^*_{s,\theta,\eta}(r)dr. \tag{23}$$

In Eq. (23), φ represents the 2D mother wavelet and * indicates the complex conjugated. The variable $s \in \mathbb{R}^2$ represents the shift, $\theta \in [0, 2\pi)$ the rotation angle, and η the scaling factor. It has been shown that a proper mother wavelet for processing interferometric images is the 2D Gabor wavelet (see **Figure 4**). The mathematical representation of this kind of wavelet can be defined as

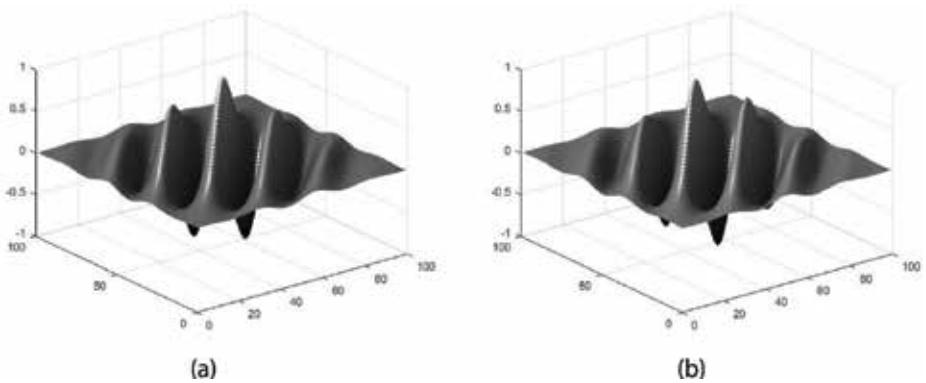


Figure 4. Example of a 2D Gabor wavelet. (a) Real part and (b) imaginary part.

http://dx.doi.org/10.5772/intechopen.74813

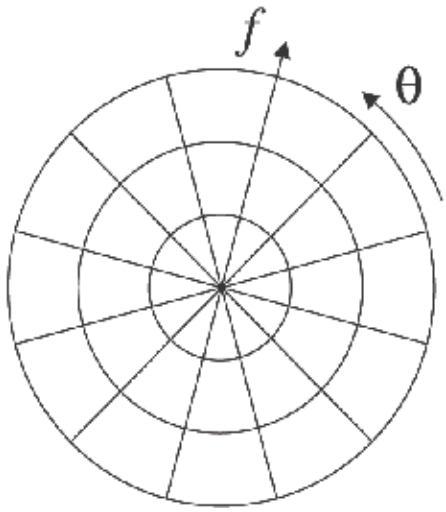


Figure 5. Frequency localization of the 2D wavelets in the Fourier domain ($f = \frac{\nu}{\eta}$).

$$\varphi_{s,\theta,\eta}(r) = \exp\left[-\pi \frac{\|r-s\|^2}{\eta}\right] \times \exp\left[i2\pi \frac{\nu}{\eta}((r-s)\cdot\Theta)\right], \tag{24}$$

where $\Theta = (\cos\theta, \sin\theta)$, $(\cdot)$ represents the dot product, and $\nu \in \mathbb{R}$ is the frequency variable.

Figure 5 shows that the 2D CWT is performed along different directions and frequencies.

4.1. Phase recovery with the 2D CWT

Owing that fringe pattern images with closed fringes generally contain elements with high anisotropy and sparse frequency components, the phase recovery is a complex procedure. Compounding the problem, the presence of noise makes the process even more complicated because noise and fringes are mixed in the Fourier domain.

Also, it has been shown that a single fringe pattern without carrier frequency, is not easy to deal with. Owing to ambiguities in the image formation process, a main drawback analyzing them is that several solutions of the phase function can satisfy the original observed image. Therefore, it is necessary to restrict the solution space of ϕ in Eq. (4). Fortunately, as in most practical cases the phase to be recovered is continuous, the algorithm to process the fringe pattern usually seeks for a continuous phase function. However, the recovery of the continuous phase function is not a simple task to carry out as occur with fringe patterns with carrier frequency. It can be observed that the phase gradient represents the local frequencies of the fringe pattern in the x and y directions; however, the sign of $\nabla\phi$ is ambiguous because negative and positive frequencies are mixed in the Fourier domain.

The following is a general description of the phase recovery method using the 2D CWT. First, it is necessary to consider a normalized version of the fringe pattern. The normalization

procedure can be carried out using the method proposed in [56]. Consider we represent the normalized fringe pattern in complex form:

$$G(r) = \cos\left[\phi(r)\right] = \frac{\exp\left[i\phi(r)\right]}{2} + \frac{\exp\left[-i\phi(r)\right]}{2}. \tag{25}$$

In this particular case, the 2D CWT of $G(r)$ is

$$\begin{aligned}\mathcal{W}\{G(r)\} = &\int_{\mathbb{R}^2} \frac{\exp\left[i\phi(r)\right]}{2} \exp\left[-\pi \frac{\|r-s\|^2}{\eta}\right] \times \exp\left[-i2\pi \frac{\nu}{\eta}((r-s)\cdot\Theta)\right] dx \\ &+ \int_{\mathbb{R}^2} \frac{\exp\left[-i\phi(r)\right]}{2} \exp\left[-\pi \frac{\|r-s\|^2}{\eta}\right] \times \exp\left[-i2\pi \frac{\nu}{\eta}((r-s)\cdot\Theta)\right] dx.\end{aligned} \tag{26}$$

Note that $\mathcal{W}\{G(r)\}$ represents a four-dimensional function depending on x, y, η, and θ. The process to recover the phase $\phi(r)$ using the 2D CWT consists on realizing the well-known ridge detection. To understand the phase recovery from the ridge detection, first it is necessary to know the meaning of Eq. (26). To do so, let $\tilde{r} = r - s$ and $\nu_\theta = \frac{\nu}{\eta}(\cos\theta, \sin\theta)$, where $\nu_\theta \in \mathbb{R}^2$. Using Taylor's expansion we know that

$$\phi(\tilde{r} + s) \approx \phi(s) + \nabla\phi(s)\cdot\tilde{r}. \tag{27}$$

Then, we can now rewrite Eq. (26) as

$$\begin{aligned}\mathcal{W}\{G(r)\} \approx &\frac{\exp\left[i\phi(s)\right]}{2} \int_{\mathbb{R}^2} \exp\left[i\left(\nabla\phi(s)\cdot\tilde{r}\right)\right] \times \exp\left[-\pi \frac{\|\tilde{r}\|^2}{\eta}\right] \exp\left[-i2\pi(\tilde{r}\cdot\nu_\theta)\right] d\tilde{r} \\ &+ \frac{\exp\left[-i\phi(s)\right]}{2} \int_{\mathbb{R}^2} \exp\left[-i\left(\nabla\phi(s)\cdot\tilde{r}\right)\right] \times \exp\left[-\pi \frac{\|\tilde{r}\|^2}{\eta}\right] \exp\left[-i2\pi(\tilde{r}\cdot\nu_\theta)\right] d\tilde{r},\end{aligned} \tag{28}$$

or, which is the same

$$\begin{aligned}\mathcal{W}\{G(r)\} \approx &\frac{\exp\left[i\phi(s)\right]}{2} \mathcal{F}\left\{\exp\left[i\left(\nabla\phi(s)\cdot\tilde{r}\right)\right] \times \exp\left[-\pi \frac{\|\tilde{r}\|^2}{\eta}\right]\right\} \\ &+ \frac{\exp\left[-i\phi(s)\right]}{2} \mathcal{F}\left\{\exp\left[-i\left(\nabla\phi(s)\cdot\tilde{r}\right)\right] \times \exp\left[-\pi \frac{\|\tilde{r}\|^2}{\eta}\right]\right\}.\end{aligned} \tag{29}$$

The two terms in (29) contains Fourier transforms of complex periodic functions of frequencies $\nabla\phi(s)/2\pi$ and $-\nabla\phi(s)/2\pi$. Then, applying the Fourier's similarity and modulation theorems this last equation can be finally written as

$$\begin{aligned}\mathcal{W}\{G(r)\} \approx &\eta \frac{\exp\left[i\phi(s)\right]}{2} \exp\left[-\eta\pi \left\| \nu_\theta - \frac{\nabla\phi(s)}{2\pi} \right\|^2\right] \\ &+ \eta \frac{\exp\left[-i\phi(s)\right]}{2} \exp\left[-\eta\pi \left\| \nu_\theta + \frac{\nabla\phi(s)}{2\pi} \right\|^2\right].\end{aligned} \tag{30}$$

In this case, ν_θ is the two-dimensional frequency variable. Note that for a fixed s, $\mathcal{W}\{G(r)\}$ represents two Gaussian filters in the Fourier domain localized at polar coordinates $\left(\frac{\nu}{\eta}, \theta\right)$. It can also be visualized as an orientation and frequency decomposition of the fringe pattern.

To detect the analytic function and consequently compute the phase $\phi(s)$ at a given pixel s (i.e., the ridge detection), we can choice one of two possibilities: at $\nu_\theta = \frac{\nabla\phi(s)}{2\pi}$ or $\nu_\theta = -\frac{\nabla\phi(s)}{2\pi}$. Owing that the sign of the phase gradient cannot be determined from the image intensity, there exists a sign ambiguity of the phase in the $\theta - \eta$ map. In **Figure 6**, it can be observed that in this situation, there are two maximum in each $\theta - \eta$ map. Also, it can be deduced that the magnitude of the coefficients map is periodic with respect to θ with period π. To solve the problem of sign ambiguity, Ma et al. [48] proposed a phase determination rule according to the phase distribution continuity. Also, Villa et al. [55] proposed a sliding 2D CWT method that assumes that the phase is continuous and smoothly varying, in this way, the ridge detection is realized assuming that the coefficient maps are similar in adjacent pixels, reducing the processing time too.

Once detected the ridge $\mathcal{W}\{G(r)\}_{ridge}$ that represents a 2D function, the wrapped phase can be computed with

$$W\{\phi(r)\} = \tan^{-1}\left(\frac{\text{Real}\left\{\mathcal{W}\{G(r)\}_{ridge}\right\}}{\text{Imag}\left\{\mathcal{W}\{G(r)\}_{ridge}\right\}}\right). \tag{31}$$

Figures 7 and **8** show examples of fringe pattern phase recovery using the 2D CWT method reported in [55]. It is important to remark that this method is highly robust against noise.

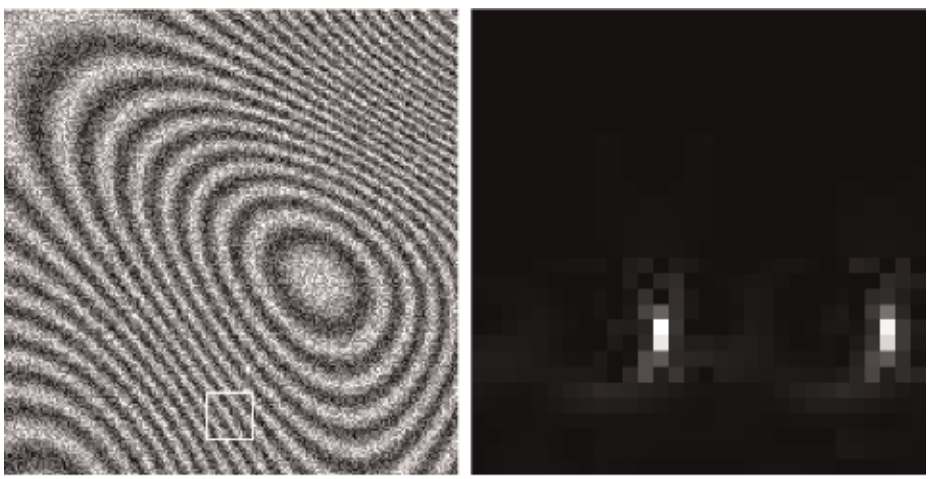


Figure 6. (a) Example of noisy simulated fringe pattern. The square indicates a region around a pixel s where the phase is estimated. (b) Magnitude of the $\theta - \eta$ map at the pixel s, codified in gray levels. Horizontal direction represents the rotation angle while the vertical direction represents the scale. The two white regions represent the two terms in Eq. (30).

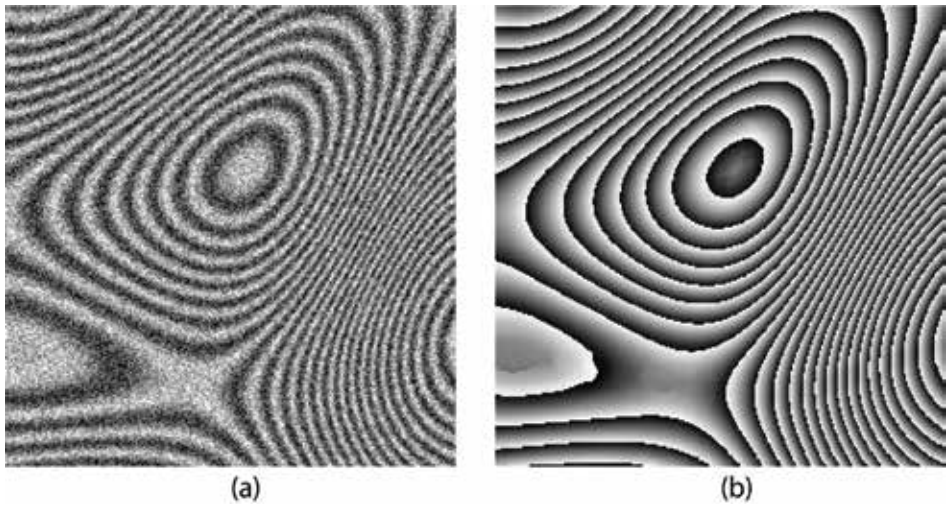


Figure 7. Example of the 2D CWT method applied to phase recovery. (a) Synthetic noisy fringe pattern. (b) Recovered phase.

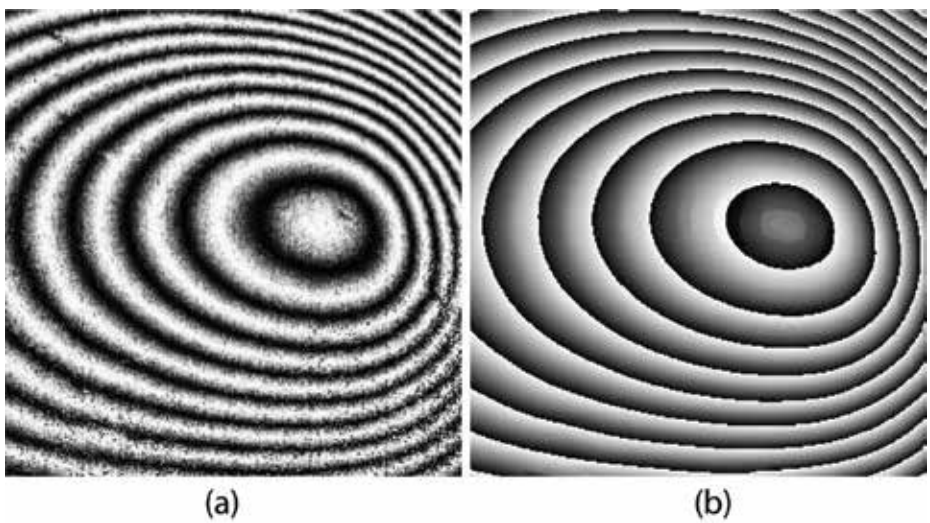


Figure 8. Example of the 2D CWT method applied to phase recovery. (a) Experimentally obtained moiré fringe pattern. (b) Recovered phase.

A big advantage of using the 2D CWT method to compute the phase from fringe patterns without carrier is that the sign ambiguity of $\nabla\phi$ can be easily solved, for example, with the method reported in [55]. The key idea of the method is the assumption that the phase ϕ is smooth; in other words, the fringe frequency and fringe orientation are very similar in neighbor pixels, hence the ridge detection at each $\theta - \eta$ map is simplified registering the previous computation of neighbor pixels.

http://dx.doi.org/10.5772/intechopen.74813

4.2. The 2D CWT for wrapped phase maps denoising

Other of the most relevant tasks in fringe pattern processing is the wrapped phase maps denoising. Owing that the phase unwrapping is a key step in fringe pattern processing for optical measurement techniques, the previous denoising of the wrapped phase is crucial for a proper measurement. Several optical measurement techniques, such as the electronic speckle pattern interferometry, use different phase recovery methods, inherently produces highly noisy wrapped phase maps. In these situations, the phase map denoising is a crucial pre-process for a successful phase unwrapping. Considering the problem of denoising wrapped phase maps, the drawback is that owing to 2π phase jumps of the wrapped phase ψ, direct application of any kind of filter is not always a proper procedure to solve it. For example, the application of a simple mean filter may smear out the phase jumps. In order to avoid this drawback, the wrapped phase filtering must be realized computing the following complex function:

$$G(r) = \exp\left[i\psi(r)\right], \tag{32}$$

where $i = \sqrt{-1}$. As both imaginary and real parts are continuous functions, we can properly apply a filter over $G(r)$, and the argument of the filtered complex signal will contain the denoised phase map. Again, substituting (32) in (23), we now obtain

$$\mathcal{W}\{G(r)\} = \int_{\mathbb{R}^2} \exp\left[i\psi(r)\right] \exp\left[-\pi \frac{\|r-s\|^2}{\eta}\right] \times \exp\left[-i2\pi \frac{\nu}{\eta}((r-s)\cdot\Theta)\right] dx. \tag{33}$$

Following the same reasoning to obtain Eq. (30), for this case, we obtain:

$$\mathcal{W}\{G(r)\} \approx \eta \exp\left[i\psi(s)\right] \exp\left[-\eta\pi \left\| \nu_\theta - \frac{\nabla\psi(s)}{2\pi} \right\|^2\right]. \tag{34}$$

The difference of this equation with the result shown in Eq. (30) is that at each $\theta - \eta$ map, there is only one maximum: at $\nu_\theta = \frac{\nabla\psi(s)}{2\pi}$ (see **Figure 9**). Thus, in this case, the ridge detection is simpler and the filtered wrapped phase map $\psi_f(r)$ can be computed with

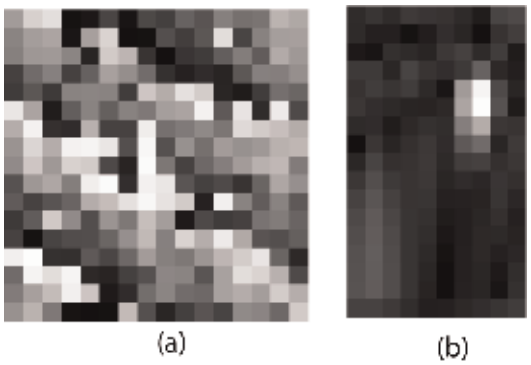


Figure 9. (a) Zoom of a small square region in a noisy wrapped phase map (around some pixel s). (b) Magnitude of the $\theta - \eta$ map at the pixel s, codified in gray levels. Horizontal direction represents the rotation angle while the vertical direction represents the scale.

$$\psi_f(r) = \tan^{-1}\left(\frac{\text{Real}\left\{\mathcal{W}\{G(r)\}_{ridge}\right\}}{\text{Imag}\left\{\mathcal{W}\{G(r)\}_{ridge}\right\}}\right) \in [-\pi, \pi). \tag{35}$$

Figures 10 and **11** are examples of the results applying the 2D CWT in wrapped phase map denoising. Note the outstanding performance removing the structures due to the gratings in the experimentally obtained wrapped phase map with moire deflectometry (**Figure 11**).

The key step in the 2D CWT method for phase map denoising is the ridge detection. In this way, all the coefficients in the $\theta - \eta$ map contributed by the noise and spurious information are

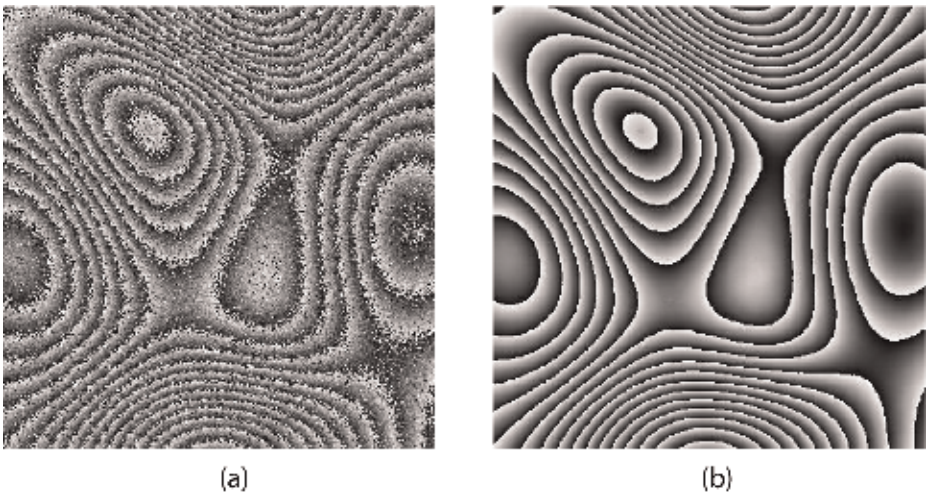


Figure 10. (a) Simulated noisy wrapped phase map. (b) Filtered wrapped phase map.

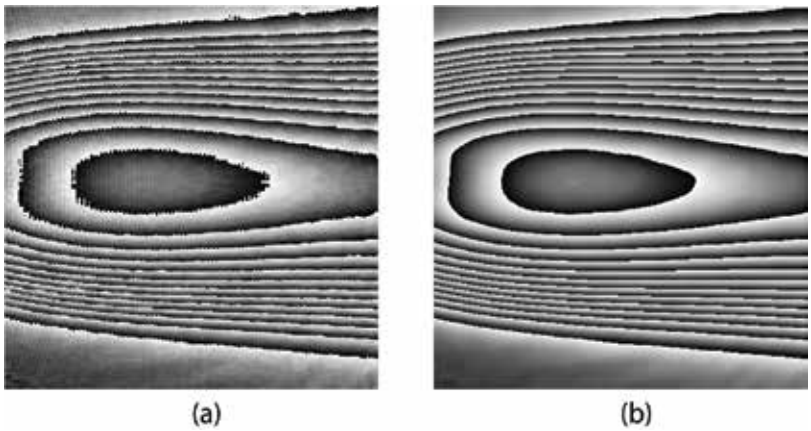


Figure 11. (a) Experimentally obtained moiré noisy wrapped phase map. (b) Filtered wrapped phase map.

2D-CWT	WFT	LFT
0.0692	0.0521	0.0747

Table 1. Performance comparison of the 2D CWT, WFT, and LFT methods, using the NMSE.

removed. A comparison of the performance of this method compared with the windowed Fourier transform method [22] and the localized Fourier transform method [21] is shown in **Table 1**. In this case, the normalized-mean-square-error (NMSE) was used as the metric applied over a synthetic noisy phase map ψ (**Figure 10**). Although the performance against noise of the WFT is better that the 2D CWT method, this last is much simpler to implement, as discussed in [53].

$$\text{NMSE} = \frac{\|\psi - \psi_f\|^2}{\|\psi\|^2}. \tag{36}$$

5. Conclusions

It can be obviously deduced that often fringe patterns contain elements with high anisotropy, sparse frequency components, and noise, which makes the processing of this kind of images by means of classical LTI methods inadequate. Several authors have shown that the use of multiresolution analysis by means of the 2D CWT for processing fringe patterns has resulted a proper and interesting alternative for this task. The 2D CWT methods present some attractive advantages compared with other commonly used techniques. (1) The use of the Gabor mother wavelet for processing this kind of images is a natural choice to model them, as can be obviously deduced analyzing the physical theory of fringe image formation. (2) In most classical methods for processing fringe images, the previous estimation of the fringe direction or orientation is a must, especially for fringe patterns without a fringe carrier frequency. Owing that the multiresolution analysis using the 2D CWT methods models the image by means of the angle θ, fringe direction or orientation is inherently computed through the ridge detection. (3) As the 2D CWT methods models the interferograms by means of scale and orientation, all spurious information and noise contributing in the $\theta - \eta$ map is efficiently removed through the ridge detection, resulting a powerful tool to remove the noise.

Author details

José de Jesús Villa Hernández[1]*, Ismael de la Rosa[1], Gustavo Rodríguez[1], Jorge Luis Flores[2], Rumen Ivanov[3], Guillermo García[2], Daniel Alaniz[1] and Efrén González[1]

*Address all correspondence to: jvillah@uaz.edu.mx

1 Unidad Académica de Ingeniería Eléctrica, Universidad Autónoma de Zacatecas, Zacatecas, México

2 Departamento de Electrónica, Universidad de Guadalajara, Guadalajara, Jalisco, México

3 Unidad Académica de Física, Universidad Autónoma de Zacatecas, Zacatecas, México

References

[1] Malacara D. Optical Shop Testing. Third ed. Hoboken: Wiley-Interscience; 2007

[2] Malacara D, Servín M, Malacara Z. Interferogram Analysis for Optical Testing. Second ed. Boca Raton: Taylor and Francis; 2005

[3] Cloud G. Optical Methods of Engineering Analysis. New York: Cambridge University Press; 1995

[4] Sirohi S. Optical Methods of Measurement, Wholefield Techniques. Second ed. Boca Raton: CRC Press; 2009

[5] Leach R. Optical Measurement of Surface Topography. Berlin, Heidelberg: Springer-Verlag, Springer; 2011

[6] Servín M, Quiroga JA, Padilla JM. Fringe Pattern Analysis for Optical Metrology. Weinheim: Wiley-VCH; 2014

[7] Ghiglia DC, Pritt MD. Two Dimensional Phase Unwrapping. New York: John Wiley & Sons, Wiley-Interscience; 1998

[8] Yu Q, Liu X, Andresen K. New spin filters for interferometric fringe patterns and grating patterns. Applied Optics. 1994;**33**:3705-3711

[9] Yu Q, Liu X, Sun X. Generalized spin filtering and improved derivative-sign binary image method for extraction of fringe skeletons. Applied Optics. 1998;**37**:4504-4509

[10] Yu Q, Sun X, Liu X, Qiu Z. Spin filtering with curve windows for interferometric fringe patterns. Applied Optics. 2002;**41**:2650-2654

[11] Villa J, Quiroga JA, de la Rosa I. Directional filters for fringe pattern denoising. SPIE Proceedings. 2009;**7499**:74990B

[12] Tang C, Gao T, Yan S, Wang L, Wu J. The oriented spatial filter masks for electronic speckle pattern interferometry phase patterns. Optics Express. 2010;**18**:8942-8947

[13] Tang C, Zhang F, Yan H, Chen Z. Denoising in electronic speckle pattern interferometry fringes by the filtering method based on partial differential equations. Optics Communication. 2006;**260**:91-96

[14] Tang C, Han L, Ren H, Zhou D, Chang Y, Wang X, Cui X. Second-order oriented partial-differential equations for denoising in electronic-speckle-pattern interferometry fringes. Optics Letters. 2008;**33**:2179-2181

[15] Tang C, Han L, Ren H, Gao T, Wang Z, Tang K. The oriented-couple partial differential equations for filtering in wrapped phase patterns. Optics Express. 2009;**17**:5606-5617

[16] Cheng L, Tang C, Yan S, Chen X, Wang L, Wang B. New fourth-order partial differential equations for filtering in electronic speckle pattern interferometry fringes. Optics Communication. 2011;**284**:5549-5555

[17] Xu W, Tang C, Gu F, Cheng J. Combination of oriented partial differential equation and shearlet transform for denoising in electronic speckle pattern interferometry fringe patterns. Applied Optics. 2017;**56**:2843-2850

[18] Zhang F, Xiao Z, Wu J, Geng L, Li H, Xi J, Wang J. Anisotropic coupled diffusion filter and binarization for the electronic speckle pattern interferometry fringes. Optics Express. 2012;**20**:21905-21916

[19] Villa J, Quiroga JA, De la Rosa I. Regularized quadratic cost function for oriented fringe-pattern filtering. Optics Letters. 2009;**34**:1741-1743

[20] Villa J, Rodríguez-Vera R, Antonio Quiroga JA, De la Rosa I, González E. Anisotropic phase-map denoising using a regularized cost-function with complex-valued Markov-random-fields. Optics and Lasers in Engineering. 2010;**48**:650-656

[21] Li C, Tang C, Yan H, Wang L, Zhang H. Localized Fourier transform filter for noise removal in electronic speckle pattern interferometry wrapped phase patterns. Applied Optics. 2011;**50**:4903-4911

[22] Kemao Q. Two-dimensional windowed Fourier transform for fringe pattern analysis: Principles, applications and implementations. Optics and Lasers in Engineering. 2007;**45**:304-317

[23] Wang H, Kemao Q, Gao W, Lin F, Seah HS. Fringe pattern denoising using coherence enhancing diffusion. Optics Letters. 2009;**34**:1141-1143

[24] Fu S, Zhang C. Fringe pattern denoising via image decomposition. Optics Letters. 2012; **37**:422-425

[25] Zhou X, Yang T, Zou H, Zhao H. Multivariate empirical mode decomposition approach for adaptive denoising of fringe patterns. Optics Letters. 2012;**37**:1904-1906

[26] Servín M, Estrada JC, Quiroga JA. The general theory of phase shifting algorithms. Optics Express. 2009;**17**:21867-21881

[27] Takeda M, Ina H, Kobayashi S. Fourier-transform method of fringe-pattern analysis for computer-based topography and interferometry. Applied Optics. 1982;**72**:156-160

[28] Womack KH. Interferometric phase measurement using spatial synchronous detection. Optical Engineering. 1984;**23**:391-395

[29] Marroquin JL, Figueroa JE, Servín M. Robust quadrature filters. JOSA-A. 1997;**14**:779-701

[30] Villa J, Servín M, Castillo L. Profilometry for the measurement of 3-D object shapes based on regularized filters. Optics Communications. 1999;**161**:13-18

[31] Kreis T. Digital holographic interference-phase measurement using the Fourier-transform method. JOSA-A. 1986;**3**:847-855

[32] Servín M, Marroquín JL, Cuevas FJ. Demodulation of a single interferogram by use of a two-dimensional regularized phase-tracking technique. Applied Optics. 1997;**36**:4540-4548

[33] Marroquín JL, Servín M, Rodríguez-Vera R. Adaptive quadrature filters and the recovery of phase from fringe pattern images. JOSA-A. 1997;**14**:1742-1753

[34] Marroquín JL, Rodríquez-Vera R, Servín M. Local phase from local orientation by solution of a sequence of linear systems. JOSA-A. 1998;**15**:1536-1544

[35] Larkin KG, Bone DJ, Oldfield MA. â€˜Natural demodulation of two-dimensional fringe patterns. I. General background of the spiral phase quadrature transform. JOSA-A. 2001; **18**:1862-1870

[36] Servín M, Quiroga JA, Marroquín JL. General n-dimensional quadrature transform and its application to interferogram demodulation. JOSA-A. 2003;**20**:925-934

[37] Villa J, de la Rosa I, Miramontes G, Quiroga JA. Phase recovery from a single fringe pattern using an orientational vector-field-regularized estimator. JOSA-A. 2005;**22**:2766-2773

[38] Tomassini P, Giulietti A, Gizzi LA, Galimberti M, Giulietti D, Borghesi M, Willi O. Analyzing laser plasma interferograms with a continuous wavelet transform ridge extraction technique: The method. Applied Optics. 2001;**40**:6561-6568

[39] Liu H, Cartwright AN, Basaran C. Sensitivity improvement in phase-shifted moiré interferometry using 1-D continuous wavelet transform image processing. Optical Engineering. 2003;**42**:2646-2652

[40] Liu H, Cartwright AN, Basaran C. Moiré interferogram phase extraction: A ridge detection algorithm for continuous wavelet transforms. Applied Optics. 2004;**43**:850-857

[41] Tay CJ, Quan C, Fu Y, Huang Y. Instantaneous velocity displacement and contour measurement by use of shadow moiré and temporal wavelet analysis. Applied Optics. 2004; **43**:4164-4171

[42] Quan C, Fu Y, Tay CJ, Tan JM. Profiling of objects with height steps by wavelet analysis of shadow moiré fringes. Applied Optics. 2005;**44**:3284-3290

[43] Zhong J, Weng J. Phase retrieval of optical fringe patterns from the ridge of a wavelet transform. Optics Letters. 2005;**30**:2560-2562

[44] Gdeisat MA, Burton DR, Lalor MJ. Spatial carrier fringe pattern demodulation by use of a two-dimensional continuous wavelet transform. Applied Optics. 2006;**45**:8722-8732

[45] Niu H, Quan C, Tay C. Phase retrieval of speckle fringe pattern with carrier using 2D wavelet transform. Optics and Lasers in Engineering. 2009;**47**:1334-1339

[46] Weng J, Zhong J, Hu C. Phase reconstruction of digital holography with the peak of the two-dimensional Gabor wavelet transform. Applied Optics. 2009;**48**:3308-3316

[47] Li S, Su X, Chen W. Wavelet ridge techniques in optical fringe pattern analysis. JOSA-A. 2010;**27**:1245-1254

[48] Ma J, Wang Z, Pan B, Hoang T, Vo M, Lu L. Two-dimensional continuous wavelet transform for phase determination of complex interferograms. Applied Optics. 2011;**50**: 2425-2430

[49] Ma J, Wang Z, Vo M. Hybrid two-dimensional continuous wavelet transform for analysis of phase-shifted interferograms. Optics Communication. 2012;**285**:3917-3920

[50] Wang Z, Ma J, Vo M. Recent progress in two-dimensional continuous wavelet transform technique for fringe pattern analysis. Optics and Lasers in Engineering. 2012;**50**:1052-1058

[51] Watkins LR. Review of fringe pattern phase recovery using the 1-D and 2-D continuous wavelet transforms. Optics and Lasers in Engineering. 2012;**50**:1015-1022

[52] Zhang Z, Jing Z, Wang Z, Kuang D. Comparison of Fourier transform, windowed Fourier transform, and wavelet transform methods for phase calculation at discontinuities in fringe projection profilometry. Optics and Lasers in Engineering. 2012;**50**:1152-1160

[53] Escalante N, Villa J, de la Rosa I, de la Rosa E, González-Ramirez E, Gutierrez O, Olvera C, Araiza M. 2-D continuous wavelet transform for ESPI phase-maps denoising. Optics and Lasers in Engineering. 2013;**51**:1060-1065

[54] Ma J, Wang Z, Pan B. Two-dimensional continuous wavelet transform algorithm for phase extraction of two-step arbitrarily phase-shifted interferograms. Optics and Lasers in Engineering. 2014;**55**:205-211

[55] Villa J, de la Rosa I, Ivanov R, Alaniz D, González E. Demodulation of single interferograms using a sliding 2-D continuous wavelet transform method. Journal of Modern Optics. 2015;**62**:633-637

[56] Quiroga JA, Gómez-Pedrero JA, García-Botella A. Algorithm for fringe pattern normalization. Optics Communication. 2001;**197**:43-51

Chapter 10

Applications of Wavelet Transforms to the Analysis of Superoscillations

Yossef Ben Ezra, Boris I. Lembrikov,
Moshe Schwartz and Segev Zarkovsky

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/intechopen.76333

Abstract

The phenomenon of superoscillation is the local oscillation of a band limited function at a frequency ω higher than the band limit. Superoscillations exist during the limited time intervals, and their amplitude is small compared to the signal components with the frequencies inside the bandwidth. For this reason, the wavelet transform is a useful mathematical tool for the quantitative description of the superoscillations. Continuous-time wavelet transform (CWT) of a transient signal $f(t)$ is a function of two variables: one of them represents a time shift, and the other one is the scale or dilation variable. As a result, CWT permits the simultaneous analysis of the transient signals both in the time and frequency domain. We show that the superoscillations strongly localized in time and frequency domains can be identified by using CWT analysis. We use CWT with the Mexican hat and Morlet mother wavelets for the theoretical investigation of superoscillation spectral features and time dependence for the first time, to our best knowledge. The results clearly show that the high superoscillation frequencies, time duration, and energy contours can be found by using CWT of the superoscillating signals.

Keywords: wavelet transform, superoscillations, transient signals, low-pass filter

1. Introduction

Superoscillating signals are band-limited signals that oscillate in some region faster than their largest Fourier component [1]. Superoscillatory functions may have interesting applications in quantum mechanics, signal processing, and optics (see, for instance, [1] and references therein). However, the superoscillation amplitude is usually so small compared to the typical values of the amplitude in non-superoscillating regions that the practical applications of the

superoscillating functions depend on tailoring the functions in order to reduce such an effect [1]. It has been shown that the superoscillations amplitude decreases exponentially with the length of the superoscillating stretch [2]. Nevertheless, the existence of superoscillations and the possibility of encoding arbitrary amounts of information into an arbitrary short segment of a low-bandwidth signal do not contradict the information theory [2]. Taking into account the Shannon's theorem concerning the information channel capacity, it appeared to be that the superoscillatory information can be compressed to an arbitrary extent under the condition that the signal power increases exponentially with the length of the superoscillatory part of the message [2]. Superoscillations can be designed by prescribing their amplitude and/or their derivative on a grid which is denser than the Nyquist density [3]. Four different ways to constrain the signal in order to render it superoscillatory have been described in Reference [3]: (1) amplitude constraints, without any restriction on the derivative; (2) derivative constraints, without restrictions on the amplitude; (3) the amplitude and the derivative constraints on staggered grids; and (4) the amplitude and the derivative constraints on aligned grids at one half density [3]. When a set of constraints is chosen to ensure a required superoscillation, the signal is optimized by minimizing its total energy within the subspace of all the superoscillatory functions obeying the same set of constraints [2]. Superoscillations can be constructed also by using the so-called direct approach. This approach is based on a signal that is a superposition of time shifted *SINC* functions which ensures its band limitation [4]. Then the coefficients of the superposition are chosen by specification of the signal values on a relatively dense set of points, which forces the required superoscillations yet leaves some degrees of freedom for optimization [4]. The propagation of the temporal optical signals with a superoscillation at an absorbing resonance of a dielectric medium has been studied theoretically [5]. The absorption acts only on the Fourier components of the band limited signals, while the superoscillation is not absorbed [5]. When the signal propagates through the medium, the superoscillation revives periodically or quasi-periodically, and a superoscillatory signal may be used in order to deliver fast oscillations to a target in a dielectric medium in the frequency bands characterized by a high absorption [5].

It should be noted that it is impossible to infer the bandwidth of a finite energy signal $f(t)$ from a sampled segment of length T even for a sufficiently large T because there exist signals of an arbitrary small bandwidth oscillating throughout an interval of a length T with an arbitrary small period [2]. The meaning is that we can make the superoscillatory part of a signal, T, and the corresponding frequency inside, ω, arbitrarily large. This comes at a price, that the amplitude of the signal outside the superoscillatory portion is exponentially large in the number of superoscillations present in the time interval, T, compared to the amplitude of superoscillation [1, 2]. In such a case, the standard Fourier analysis is not sufficient because we cannot locate the sharp pulses in the signal spectrum caused by the sharp changes of the signal in the time domain [6]. A small perturbation of the sinusoidal function $\sin(\omega t)$ or $\cos(\omega t)$ at any point of the time axis influences every point of the frequency axis and vice-versa [6]. The Fourier transform integral can be evaluated at only one frequency at a time which is not convenient for the signal processing [6]. In particular, the so-called time-frequency analysis combining both the frequency domain and the time domain analyses is necessary for superoscillation studies [2, 6]. The short-time Fourier transform (STFT) can be used for the time-frequency analysis because it permits to obtain the

approximate frequency contents of the time-dependent signal in the vicinity of a desired location in the time domain [6]. However, the fixed time-frequency resolution is a disadvantage of STFT because STFT can resolve properly either the low-frequency of the signal, or the high-frequency part of it [6]. The linear transform providing an efficient time-frequency resolution in any location of the time-frequency plane is the continuous-time wavelet transform (CWT) [6, 7]. CWT is defined by two real positive parameters: the scale or dilation variable a and the time shift or translation b [6, 7]. By changing the parameters a, b, CWT can be calculated on the entire time-frequency plane [6].

In this chapter, we constructed a family of complex valued superoscillating functions and investigated their behavior for different values of the maximum frequency ω_0 and amplitude by using CWT with the Mexican hat and Morlet mother wavelets. We have shown that the high superoscillation frequencies, short-time durations, and energy contours can be evaluated by using CWT.

The chapter is organized as follows. The superoscillating function properties are discussed in Section 2. Some possible applications of superoscillations are reviewed in Section 3. CWT and discrete wavelet transform (DWT) definition and fundamental features are presented in Section 4. The applications of wavelet transform for the optical signal processing are briefly discussed in Section 5. The simulation results are presented and discussed in Section 6. Conclusions are presented in Section 7.

2. The properties of the superoscillating functions

The frequency limited functions and superoscillations occur in a number of scientific and technological applications such as foundations of quantum mechanics, information theory, optics, and signal processing which led also to work on the optimization and stability of superoscillations [8–14]. Some examples of superoscillatory functions have been proposed and investigated in the past [1–4, 8, 10, 11]. In this section, we consider the generic example of the Aharonov, Popescu, Rohlich functions $f(x)$ given by [11]:

$$f(x) = (\cos x + ig \sin x)^n; g > 1; \mathrm{n} \gg 1 \tag{1}$$

Here, for general, $g f(x)$ is a periodic function with a period 2π. It is easy to see from Eq. (1) that for $g = 1$ we obtain from Eq. (1):

$$f(x) = \exp(inx) \tag{2}$$

For small $x \to 0$, Eq. (1) yields:

$$f(x) \approx \left(\exp(\ln(1 + igx))\right)^n \approx \exp(ignx) \tag{3}$$

Obviously, in the limiting case, the function determined by Eq. (3) is varying faster than the function determined by Eq. (2). Consider now the Fourier series for $f(x)$ given by [11]:

$$f(x) = \sum_{m=0}^{n} c_m \exp(ink_m x) \tag{4}$$

where

$$k_m = 1 - \frac{2m}{n}; c_m = \frac{n!}{2^n}(-1)^m \frac{\left(g^2-1\right)^{n/2}[(g-1)/(g+1)]^{nk_m/2}}{[n(1+k_m)/2]![n(1-k_m)/2]!} \tag{5}$$

Equations (4) and (5) contain only wavenumbers $|k_m| \leq 1$ [11]. Comparison of Eqs. (3–5) shows that the function $f(x)$ demonstrates superoscillatory behavior with the degree of superoscillation defined by g [11]. The function $f(x)$ Eq. (1) can be represented in an integral form [11]:

$$f(x) = \left(\frac{g}{k(x)}\right)^{n/2} \exp\left\{in\int_0^x dx' k(x')\right\} \tag{6}$$

Here, the local wavenumber $k(x)$ is introduced given by [11]:

$$k(x) \equiv \frac{1}{n}\mathrm{Im}\left(\frac{\partial \ln f(x)}{\partial x}\right) = \frac{g}{\cos^2 x + g^2 \sin^2 x} \tag{7}$$

The relationship Eq. (6) can be proved immediately taking into account that:

$$\int_0^x dx' k(x') = \arctan(g\tan x) \tag{8}$$

and using the identities $\cos x = \left(\sqrt{1+\tan^2 x}\right)^{-1}$ and $\sin x = \tan x / \left(\sqrt{1+\tan^2 x}\right)$. The wavenumber $k(x)$ determined by Eq. (7) is varying from the superoscillatory value $k(0) = g$ to the minimum value $k(\pi/2) = 1/g$. The superoscillatory region where $|k| > 1$ is determined by the following condition [11]:

$$|x| < x_s = arc\cot\left(\sqrt{g}\right) \tag{9}$$

The number of oscillations n_{osc} in the superoscillatory region is given by:

$$n_{osc} = \frac{n}{2\pi}\int_{-arc\cot(\sqrt{a})}^{arc\cot(\sqrt{a})} dx k(x) = \frac{n}{\pi}\arctan\left(\sqrt{g}\right) \tag{10}$$

Equation (6) shows in particular that in the superoscillatory region $|k| > 1$, the magnitude $|f(x)|$ is exponentially smaller than in the normal region $|k| < 1$ [11]. Consequently, n is the asymptotic parameter describing the number of oscillations in the superoscillatory region and the

corresponding exponential smallness of $|f(x)|$ [11]. More accurate approximation gives the following expressions for the region of fast superoscillations x_{fs} and the number of fast superoscillations n_{fs} [11]:

$$|x| < x_s = \frac{1}{\sqrt{n(g^2-1)}}; n_{fs} = \frac{g\sqrt{n}}{\pi\sqrt{g^2-1}} \tag{11}$$

Superoscillations is a week phenomenon such that there is no slightest indication of superoscillations in the power spectrum $P(k)$ of $f(x)$ [11]. Indeed, using the Fourier components Eq. (5) we obtain [11]:

$$P(k) = \frac{nc_m^2(m = n(1-k)/2)}{2\sum_0^n c_m^2} \approx \frac{1}{\sigma\sqrt{2\pi}}\exp\left\{-\frac{(k-\langle k\rangle)^2}{2\sigma^2}\right\} \tag{12}$$

where

$$\langle k\rangle = \frac{1}{g}; \sigma \equiv \sqrt{(k-\langle k\rangle)^2} = \sqrt{\frac{g^2-1}{2ng^2}} \tag{13}$$

The asymptotic spectrum Eq. (12) is a narrow Gaussian with the center at $k = g^{-1}$ representing the slow oscillations near $|x| = \pi$ [11]. The superoscillations gradually disappear getting slower and reducing to the region $|k| \leq 1$ according to the Fourier series Eqs. (4) and (5) [11].

The function $f(x)$ defined by Eq. (1) is periodic. Consequently, it may represent a diffraction grating with spatial period πd that transforms an incident light plane wave into a propagating series of diffracted beams [11]. Such a grating transforms the wave Eq. (1) into the superoscillatory function $\Psi(x,0) = f(x/d)$ under the following condition for the wavenumber K:$n/d < K < gn/d$ [11]. The grating produces a novel kind of super-resolution, that is, the subwavelength structure in the field with only propagating waves and without evanescent waves [11].

In the framework of the precise classical wave model, it has been shown how superoscillations can emerge and propagate into the far-field region [14]. The band-limited superoscillatory wave (the "red light") is propagating along the x axis of a unidimensional (1D) waveguide with the a segment of the x axis (the "window") which is opened and closed as the superoscillation passes by releasing the light pulse into the two-dimensional (2D) space corresponding to $x, z > 0$ [14]. The wave traveling in the positive x direction with a speed $c = 1$ can be described without loss of generality by the band limited function Eq. (1) with the replacement of the argument x by $(x-t)/n$ in expressions (1) and (4) [14]. This new function $f_{red}(x,t)$ is superoscillatory near $x = t$ [14]. It represents a rigidly moving polychromatic packet with associated frequencies $\omega_m = k_m$ [14]. The function $f_{red}(x,t)$can be approximated by the following expression: $f_{red.app}(x,t) = \exp[ig(x-t)]\exp\left(\frac{(x-t)^2}{2X^2}\right); X = \sqrt{\frac{n}{g^2-1}}$ [14]. The function $f_{red.app}(x,t)$ represents superoscillations over the interval $-2X < x < 2X$ for $t = 0$ [14]. In order

to capture the superoscillations, the region near $x = 0$ is selected with a Gaussian window of width L which is opened and closed with a Gaussian switching function over an interval near $t = 0$ [14]. The window must faithfully transmit the red light including the superoscillaitons [14].

For the sake of definiteness, we consider the time-dependent superoscillating signal of the type Eq. (1) assuming that:

$$x = \frac{\omega_0 t}{n}; g = \frac{\omega}{\omega_0} \tag{14}$$

Substituting relationships Eq. (14) into Eq. (1) we obtain:

$$f_n(\omega_0 t, \omega/\omega_0) = \left\{ \cos\left(\frac{\omega_0 t}{n}\right) + i\frac{\omega}{\omega_0}\sin\left(\frac{\omega_0 t}{n}\right) \right\}^n \tag{15}$$

Expression (15) is the signal band limited by the frequency ω_0 with the superoscillation manifested by a single peak of a width $\Delta t = \omega^{-1}$ while the ratio ω/ω_0 may be arbitrary larger than unity. It should be noted that:

$$\lim_{n\to\infty} \mathrm{Re} f_n(\omega_0 t, \omega/\omega_0) = \cos(\omega t) \tag{16}$$

For finite n and under the first condition of Eq. (11) which now takes the form $\omega_0 t < \sqrt{n/\left[(\omega/\omega_0)^2 - 1\right]}$, expression (16) reduces to the following approximation:

$$f_n(\omega_0 t, \omega/\omega_0) \approx \cos(\omega t) \tag{17}$$

Equation (17) shows that the band limited signal Eq. (15) oscillates with a frequency ω higher than the band limit ω_0 for the arbitrary long time depending on n.

3. The possible applications of superoscillations

Optical superoscillations can be used in the subwavelength imaging [15]. This super-resolution technology is based on a superoscillatory lens (SOL) which represents a nanostructured mask [15]. SOL illuminated with a coherent light source creates a focus at a distance which is larger than the near-field of the mask [15]. Indeed, the ability to focus beyond the diffraction limit is related to the superoscillation, since the band-limited functions in such a case oscillate faster than their highest Fourier components [11]. Superoscillatory binary masks do not use evanescent waves and focus at distances tens of wavelengths away from the mask [15]. The superoscillation-based imaging has the following advantages with respect to other technologies: (1) it is non-invasive which allows to place the object at a substantial distance from SOL; (2) it can operate at the wide range of wavelengths from X-rays to microwaves; and (3) the resolution of the SOL can be improved by refining the design, increasing the size of the

superoscillatory mask and by increasing the dynamic range of the light detection [15]. SOL can be also used for the creation of sub-diffraction-limit optical needles [16]. An optical needle could be created by converting the central region of the SOL into an opaque area forming a shadow, and changing the diameter of the blocking region without varying the rest of SOL [16]. The possible applications of the sub-diffraction-limit optical needles are the far-field super-resolution microscopy and nanofabrication [16].

The possible applications of superoscillations for data compressions have been discussed [8]. However, the superoscillations are unstable in a way that tiny perturbations of a band-limited superoscillating function can induce very high-frequency components [8]. For this reason, the practical use of the superoscillations in imperfect communication channels is difficult [8].

4. The fundamental properties of CWT and DWT

There exist different types of a wavelet transform: CWT, discrete wavelet transform (DWT) [6, 7, 17], multi-wavelets [17, 18], and complex wavelets [19]. We applied these types of wavelets to the problems related to the signal processing in optical communication systems [20–23]. We have found that CWT is the most appropriate for the analysis of superoscillations.

In this section, we consider some fundamental features of CWT. Unlike the Fourier transform and STFT, the CWT is characterized by the time and frequency selectivity [6, 7]. It can localize events both in time and in frequency in the entire time-frequency plane [6, 7]. That is why CWT is unique mathematical tool for the investigation of the superoscillations where the time-frequency analysis in different regions of the spectrum is necessary as it is mentioned earlier [2, 6]:

The CWT $W(a,b)$ of any square integrable function $f(t)$ with respect to a wavelet $\psi(t)$ is defined as follows [7]:

$$W(a,b) \equiv \int_{-\infty}^{\infty} f(t) \frac{1}{\sqrt{|a|}} \psi^* \left(\frac{t-b}{a} \right) dt \tag{18}$$

Here a, b are real, the asterisk denotes complex conjugation, the energy signals $f(t), \psi(t) \subset L^2(R)$, $L^2(R)$ is the set of square integrable functions such that $\int_{-\infty}^{\infty} |\psi(t)|^2 dt < \infty$. The real- or complex-value continuous-time function $\psi(t)$ is called the mother wavelet. It satisfies the following condition that $\int_{-\infty}^{\infty} \psi(t) dt = 0$ [7].

CWT $W(a,b)$ is a function of two variables: (1) the scale or dilation variable a determines the amount of time scaling or dilation and (2) the translation or time shift variable b represents the shift of $\psi_{a,0}(t)$ by an amount b along the time axis and indicates the location of the wavelet window along it [6, 7]. The inverse scaling parameter $1/a$ is a measure of frequency [6].

Defining

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}}\psi\left(\frac{t-b}{a}\right) \tag{19}$$

and substituting expression (19) into Eq. (18) we obtain [7]:

$$W(a,b) \equiv \int_{-\infty}^{\infty} f(t)\psi^*_{a,b}(t)dt \tag{20}$$

The energy conservation law for the mother wavelet has the form for all values of *a, b* [7].

$$\int_{-\infty}^{\infty} \left|\psi_{a,b}(t)\right|^2 dt = \int_{-\infty}^{\infty} |\psi(t)|^2 dt \tag{21}$$

Consider some typical mother wavelets [6, 7]. The Haar wavelet is a piecewise continuous function. It has the form:

$$\psi(t) = \begin{cases} 1, 0 \le t < 1/2 \\ -1, 1/2 \le t < 1 \\ 0, otherwise \end{cases} \tag{22}$$

The Mexican hat wavelet is obtained by taking the second derivative of the negative Gaussian function $-\exp(-t^2)/2$ [7]. It is given by [7]:

$$\psi(t) = (1 - 2t^2)\exp(-t^2) \tag{23}$$

The time dependence of the Mexican hat wavelet is shown in **Figure 1**.

The Morlet wavelet represents a sinusoidal function modulated by a Gaussian function given by [7]:

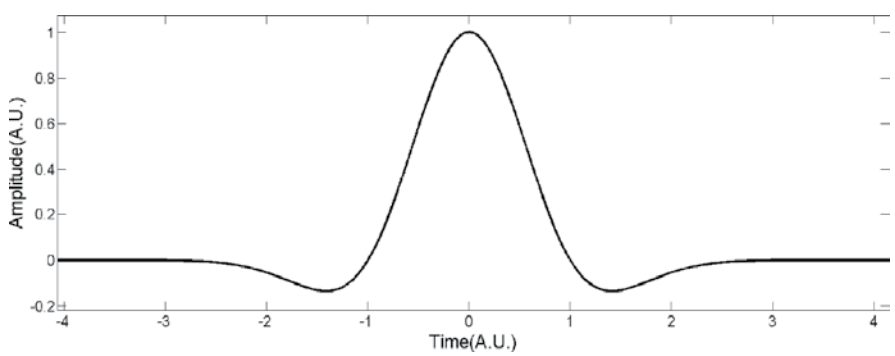


Figure 1. The Mexican hat wavelet.

$$\psi(t) = \exp(-t^2)\cos\left(\pi t\sqrt{\frac{2}{\ln 2}}\right) \tag{24}$$

The time dependence of the Morlet wavelet is shown in **Figure 2**. It is a wavelet of an infinite duration, but most of the energy in this wavelet is confined to a finite interval [7].

CWT can be used in pattern detection and classification [6, 7]. Indeed, taking into account the definition of the inner product $\langle x(t), y(t)\rangle$ of two finite energy signals as:

$$\langle x(t), y(t)\rangle = \int_{-\infty}^{\infty} x(t)y^*(t)dt \tag{25}$$

one can say that CWT is a collection of the inner products of a signal $f(t)$ and the translated and dilated mother wavelets $\psi_{a,b}(t)$ for all a, b: $W(a,b) = \left\langle f(t), \psi_{a,b}(t)\right\rangle$ [7]. CWT can be also considered as the cross-correlation at lag b between $f(t)$ and the mother wavelet dilated to scale factor a [7]. Comparing the definition of the cross-correlation function:

$$R_{x,y}(\tau) \equiv \langle x(t), y(t-\tau)\rangle = \int x(t)y^*(t-\tau)dt \tag{26}$$

and CWT expression (20) we can write [7]:

$$W(a,b) = \left\langle f(t), \psi_{a,0}(t-b)\right\rangle = R_{f,\psi_{a,0}}(b) \tag{27}$$

The CWT is characterized by the time selectivity or the so-called windowing effect because the segment of $f(t)$ influencing the value of $W(a,b)$ for any (a,b) coincides with the interval over which $\psi_{a,b}(t)$ has the bulk of its energy [7]. The CWT frequency selectivity can be described by its representation as a collection of linear, time-invariant filters with impulse responses which are dilations of the mother wavelet reflected about the time axis [7]. Indeed, using the definition of the convolution of the input signal $x(t)$ and the system impulse response $h(t)$:

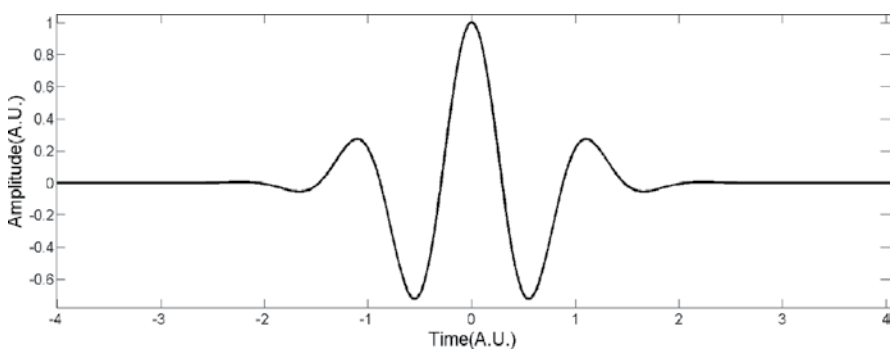


Figure 2. The Morlet wavelet.

$$h(t)*x(t) \equiv \int_{-\infty}^{\infty} h(\tau)x(t-\tau)d\tau \tag{28}$$

we can write for the CWT:

$$W(a,b) = f(b)*\psi^*_{a,0}(-b) \tag{29}$$

Consequently, for any given a, CWT $W(a,b)$ is the output of a filter with the impulse response $\psi^*_{a,0}(-b)$ and input $f(b)$ [7]. There exists a continuum of filters characterized by the scale factor a as a parameter [7]. We define the Fourier transform $\Psi(\omega)$ of the mother wavelet $\psi(t)$:

$$\Psi(\omega) = F\{\psi(t)\} \equiv \int_{-\infty}^{\infty} \psi(t)\exp(-i\omega t)dt \tag{30}$$

Then, the corresponding Q factor is determined as $Q = \omega_c/\Delta\omega$, where ω_c is center frequency of the Fourier transform Eq. (30) and $\Delta\omega$ is the 3-dB bandwidth defined as the difference between the two frequencies on either side of the peak at which $|\Psi(\omega)|^2$ is exactly half its peak value $|\Psi(\omega)|^2_{\max}$ [7]. The Q factor is invariant with respect to the wavelet dilation, since $F\{\psi(t/a)\} = |a|\Psi(a\omega)$. The center frequency ω_c of $|\Psi(\omega)|^2$ for any a is at $1/|a|$ times the center frequency of the mother wavelet $\psi(t)$, and its 3-dB bandwidth is $1/|a|$ times the 3-dB bandwidth of the mother wavelet $\psi(t)$ which gives the same value of the Q factor as the one mentioned above [7]. Hence the continuum of filters mentioned above is a set of constant Q bandpass filters which results in the frequency selectivity of the CWT [7]. For large values of a, the corresponding filter has a frequency response with a low center frequency ω_0, and the corresponding CWT $W(a,b)$ captures the frequency content of the signal $f(t)$ around this low frequency [7]. The bandpass filter shifts to higher frequencies region with the decrease of a in such a way that the CWT $W(a,b)$ at small scales contains information about $f(t)$ at the higher end of its frequency spectrum [7]. The time and frequency resolution of the CWT $W(a,b)$ are based on the duration and bandwidth of the mother wavelet $\psi(t)$, respectively. The first moments t_c and ω_c of the mother wavelet $\psi(t)$ and its Fourier transform $\Psi(\omega)$, respectively, are given by [7]:

$$t_c \equiv \frac{\int_{-\infty}^{\infty} t|\psi(t)|^2dt}{\int_{-\infty}^{\infty} |\psi(t)|^2dt}; \omega_c = \frac{\int_{-\infty}^{\infty} \omega|\Psi(\omega)|^2d\omega}{\int_{-\infty}^{\infty} |\Psi(\omega)|^2d\omega} \tag{31}$$

Expressions (31) provide the location of the center of $\psi(t)$ and $\Psi(\omega)$ along the time and frequency axes, respectively [6, 7]. A measure of the wavelet duration Δt, or the root mean square (RMS) duration, and the RMS bandwidth of the wavelet $\Delta\omega$ are given by, [6, 7], respectively:

$$\Delta t = \sqrt{\frac{\int_{-\infty}^{\infty} (t - t_c)^2 |\psi(t)|^2 dt}{\int_{-\infty}^{\infty} |\psi(t)|^2 dt}}; \Delta\omega = \sqrt{\frac{\int_{-\infty}^{\infty} (\omega - \omega_c)^2 |\Psi(\omega)|^2 d\omega}{\int_{-\infty}^{\infty} |\Psi(\omega)|^2 d\omega}} \tag{32}$$

Expressions (31) and (32) can be used only for the mother wavelet $\psi(t)$ and its Fourier transform $\Psi(\omega)$ rapidly decaying in time and frequency, respectively, since the integrals in the numerators of these expressions should have finite values [7]. For the RMS duration Δt_ψ and bandwidth $\Delta\omega_\psi$ of the mother wavelet $\psi(t)$, the RMS duration $\Delta t_\psi(a)$ and bandwidth $\Delta\omega_\psi(a)$ of its dilation $\psi_{a,0}(t)$ are given by [7]:

$$\Delta t_\psi(a) \equiv |a|\Delta t_\psi; \Delta\omega_\psi(a) = \Delta\omega_\psi/|a| \tag{33}$$

Combining expressions (33) we obtain:

$$\Delta t_\psi(a)\Delta\omega_\psi(a) = \Delta t_\psi \Delta\omega_\psi = c_\psi = const \tag{34}$$

It has been shown that the smallest time-bandwidth product is equal to 1/2., and condition Eq. (34) takes the form [6, 7]:

$$\Delta t_\psi(a)\Delta\omega_\psi(a) \geq \frac{1}{2} \tag{35}$$

Equation (34) shows that the product of the wavelet duration and bandwidth is invariant to dilation. For small values of *a*, the CWT is characterized by good time resolution and poor frequency resolution because the RMS duration of the dilated wavelet is small while the RMS bandwidth of the dilated wavelet is large [7]. For large values of *a*, the time resolution of the CWT is poor, and its frequency resolution is good. The CWT provides better frequency resolution for the low-frequency region of the spectrum and poorer frequency resolution for the high-frequency region of the spectrum [7]. It can be shown that the translation parameter *b* does not influence the mother wavelet duration and bandwidth [7].

The inverse CWT can be evaluated under the following sufficient condition for the mother wavelet Fourier transform $\Psi(\omega)$ [7]:

$$\int_{-\infty}^{\infty} \frac{|\Psi(\omega)|^2}{|\omega|} d\omega \equiv C; 0 < C < \infty \tag{36}$$

Then the inverse CWT has the form [7]:

$$f(t) = \frac{1}{C} \int_{a=-\infty}^{\infty} \int_{b=-\infty}^{\infty} \frac{1}{|a|^2} W(a,b)\psi_{a,b}(t) da db \tag{37}$$

The variable time-frequency resolution is an important property of the CWT which permits to use CWT for the analysis of the signals consisting of the slowly varying low-frequency components and the rapidly varying high-frequency components [7]. For this reason, the CWT is a unique tool for the study of the superoscillating signals described in Section 2.

Suppose that the dilation parameter a and the translation parameter b are discrete and take a form $a = 2^k, b = 2^k l$, where k and l are integers [6, 7, 17]. Then, Eq. (37) takes the form [7]:

$$f(t) = \sum_{k=-\infty}^{k=\infty} \sum_{l=-\infty}^{l=\infty} d(k,l) 2^{-k/2} \psi\left(2^{-k}t - l\right) \tag{38}$$

The two-dimensional sequence $d(k,l)$ is defined as DWT of $f(t)$ [7]. The values of DWT $d(k,l)$ are related to the values of CWT $W(a,b)$ Eq. (18) at $a = 2^k, b = 2^k l$ [7]. Then DWT $W_{kl}(a,b)$ takes the form [6, 7]:

$$W_{kl}(a,b) = \int_{-\infty}^{\infty} f(t) \frac{1}{2^{k/2}} \psi^*\left(2^{-k}t - l\right) dt \tag{39}$$

Comparison of CWT and DWT shows that the signal $f(t)$ in the both cases is expressed in terms of dilations and translations of a single mother wavelet [6]. DWT is used in the multiresolution analysis (MRA) which is based on a hierarchy of approximations to functions in N various subspaces $W_{N-1}, W_{N-2}, \ldots, W_1$ of a linear vector space $V_N = W_{N-1} \oplus W_{N-2} \oplus \ldots, W_1 \oplus V_1$ [6]. In general case, the wavelet $\psi(t)$ providing the DWT corresponding to the MRA must satisfy the following conditions [7]:

$$\int_{-\infty}^{\infty} \psi(t) dt = 0; \int_{-\infty}^{\infty} |\psi(t)|^2 dt = 1; \langle \psi(t), \psi(t-n) \rangle = \delta(n); \langle \psi(t), \phi(t-n) \rangle = 0 \tag{40}$$

Here $\delta(n) = 1, n = 0$ and $\delta(n) = 0, n \neq 0$ and $\phi(t)$ is the scaling function $\phi(t)$ satisfying the following conditions [7]:

$$\int_{-\infty}^{\infty} \phi(t) dt = 1; \int_{-\infty}^{\infty} \left|\phi(t)\right|^2 dt = 1; \langle \phi(t), \phi(t-n) \rangle = \delta(n) \tag{41}$$

The scaling function $\phi(t)$ and the wavelet function $\psi(t)$ are defined by the following equations, respectively [7]:

$$\phi(t) = \sum_{n=-\infty}^{\infty} c(n) \phi(2t-n); \psi(t) = \sum_{n=-\infty}^{\infty} d(n) \phi(2t-n), n = 0, \pm 1, \pm 2, \ldots \tag{42}$$

where $c(n), d(n)$ are sequences of scalars. It is seen that the scaling function $\phi(t)$ is determined by its own dyadic dilation and translation. For this reason, the equation for $\phi(t)$ is called a dilation Equation [6, 7, 17]. It can be shown that the DWT is equivalent to filtering a signal by a band of filters with nonoverlapping bandwidths differing by a factor of 2 [17].

http://dx.doi.org/10.5772/intechopen.76333

5. The applications of wavelet transforms

The different types of WT are widely used in different areas of mathematics and engineering [17]. The number of scientific books and articles concerning wavelet transforms (WT) applications is enormous and hardly observable. In this section, we briefly review some typical applications of wavelet transforms in optical communication systems and signal processing. Wavelet methods may complement the Fourier techniques due to their following specific features mentioned above [17]. Wavelets are functions of two parameters which represent the dilation and translation while the Fourier transform is characterized by the dilation only. In the case of wavelets, the width of the window through which the signal is observed is varying as a function of location. For a wavelet method, the window function in the time-frequency plane is nonuniform being a function of both time and frequency.

Wavelet transforms as a mathematical tool can be successfully used in the electromagnetic problems and signal processing applications [6, 7, 17–24]. Wavelet based signal processing represents a useful technique for the compression of certain classes of data demonstrating isolated band-limited properties [17]. Wavelets may be used as basis functions for the solution of Maxwell's equations in the integral or differential form [17]. Signal denoising process can be implemented by using wavelets with a smaller computational complexity as compared to the Fourier technique [17].

Wavelets can be successfully applied to signal and image processing including noise reduction, signal and image compression, signature identification, target detection, and interference suppression [6].

Wavelet packet transform (WPT) can be used in optical communications [20, 24]. WPTs are the generalization of wavelet transforms where the orthogonal basis functions are wavelet packets instead of ordinary wavelets [24]. Discrete WPT (DWPT) is used in the coherent optical orthogonal frequency division multiplexing (CO-OFDM) systems [24]. The detailed analysis of CO-OFDM communication systems can be found in [20, 24] and references therein. In a WPT-OFDM system, each channel occupies a wavelet packet, that is, a subcarrier in wavelet domain [24]. Inverse DWPT (IDWPT) is used at the transmitter which reconstructs the time domain signal from wavelet packets [24]. DWPT are used at the receiver in order to decompose the time domain signal into different wavelet packets by means of successive low-pass and high-pass filtering in the time domain [24].

We proposed a novel hierarchical architecture of the $1Tb/s$ transmission system based on DWPT-OFDM in order to reduce the computational complexity of the digital signal processing (DSP) algorithms [20]. We separated the low bit rate and high bit rate signal channels in such a way that the low bit rate signals are processed in the electrical domain, while the high bit rate signals are processed optically [20]. We have shown theoretically that the performance of the WPT based CO-OFDM can be significantly improved by increasing the spectral efficiency (SE) of the system and mitigating the channel chromatic dispersion [20].

Recently, some novel applications of different types of wavelet transforms have been reported. CWT can be applied for the improvement of the time-delay estimation (TDE) method in the different-wavelength based inteferometric vibration sensor in a fiber link [25].

The maximal overlap DWPT (MODWPT) has been used for the real-time estimation of root mean square (RMS) power value, active power, reactive power, apparent power, and power factor in power electronic systems [26].

The time-reversal (TR) technique is used for the detection and localization of objects in microwave imaging [27]. TR technique is based on an assumption that in a lossless medium, for every wave component propagating away from a source point along a certain path there exists a corresponding time-reversed wave propagating along the same path back to the original point of the source [27]. This assumption is caused by the time invariance of the Maxwell's Equations [27]. TR can achieve super-resolution by using the multipath propagation [27]. However, TR in real media is deteriorated due to the dispersion and losses [27]. A compensation method based on CWT has been proposed which can overcome both the dispersion and attenuation of the electromagnetic wave propagating in a dispersive and lossy medium [27]. In this method, the adjustable-length windows are used in such a way that the long-time windows and short-time windows are applied at low and high frequencies, respectively [27]. Wavelets depend on both the time and frequency which results in the signal decomposition into different time and frequency components. The dispersion and attenuation of these components can be compensated by different filters. Unlike the short-time Fourier transform (STFT) method, the proposed CWT method can be applied in real-life scenarios, and its resolution is about three times higher than in other methods [27].

Online monitoring and control of power grid require the accurate and fast estimation of harmonics [28]. The WT has been widely used in the estimation of time-varying harmonics [28]. In particular, undecimated WPT (UWPT) is one of multiresolution techniques characterized by redundancy and time invariance which can be implemented by a set of filter banks [28]. Unlike DWPT, the UWPT does not perform downsampling on wavelet coefficients at each decomposition level preserving time-invariant property which permits the accurate estimation of the time-varying harmonics in one cycle of the fundamental frequency [28]. The comparison of the simulation results obtained by using the UWPT based method and the experimental results shows that the UWPT algorithm has better estimation accuracy for different types of signals [28].

We for the first time to our best knowledge applied CWT to the theoretical investigation of superoscillations which requires the dynamic time-frequency analysis of the strongly localized signals. CWT appeared to be a powerful mathematical tool for the identification of the superoscillation characteristic features.

6. The simulation results and discussion

We theoretically investigated the superoscillations of the signal defined by the real part of expression (15):

$$\mathrm{Re} f_n(\omega_0 t, \omega/\omega_0) = \mathrm{Re}\left\{\cos\left(\frac{\omega_0 t}{n}\right) + i\frac{\omega}{\omega_0}\sin\left(\frac{\omega_0 t}{n}\right)\right\}^n \tag{43}$$

The signal Eq. (38) is band limited by the maximum frequency $\omega_0/2\pi$ as it was mentioned above. Without the loss of generality, we have chosen the frequency $\omega_0/2\pi = 100KHz$, $n = 4$

and $n = 5$. We used the Mexican hat mother wavelet Eq. (23) and the Morlet mother wavelet Eq. (24), since their oscillating temporal behavior is similar to the behavior of the superoscillating signal Eq. (38).

We investigated the scalogram of the energy contours for the spectral component of the signal Eq. (38) at the highest frequency $f(\omega_0 t) = \cos(\omega_0 t)$ and used the results for the analysis of the superoscillation features.

The behavior of the component with the frequency $\omega_0/2\pi$ in the time domain (lower box) and its scalogram (the upper box) are shown in **Figure 3**. The pseudo-frequency $\omega_a/2\pi$ shown in the scalogarm is defined as follows:

$$\omega_a/2\pi = \omega_c/2\pi a \tag{44}$$

where ω_c is the mother wavelet central frequency defined by the second expression (31) [6, 7]. It is seen from **Figure 3** that the homogeneous scalogram in the time-pseudo-frequency plane is strictly periodic. The maximum energy at the scalogram corresponds to the pseudo-frequency $\omega_a/2\pi \approx 30KHz$.

The spectra, the temporal behavior, and the scalograms of the signal Eq. (38) for $n = 4$ and $n = 5$ are presented in **Figures 4–7**, respectively. The spectra shown in **Figures 4** and **6** are obtained by using the Fourier transform. For this reason, the superoscillations are absent in the spectra shown in **Figures 4** and **6**, and the highest frequency corresponding to the maximum spectral amplitude is $\pm 100KHz$ in the both cases.

The scalograms shown in the upper box of **Figures 5** and **7** are obtained by evaluating the CWT of the signal Eq. (38) with the Mexican hat mother wavelet Eq. (23).

The superoscillations with the time duration of about $\Delta T(n = 4) = 8\mu s$ can be identified in the lower box of **Figure 5**. These superoscillations correspond to the frequency of about $\omega_4/2\pi = 1/\Delta T(n = 4) = 125KHz > \omega_0/2\pi = 100KHz$. The energy contours of these super-oscillations in the time-pseudo-frequency plane shown in the upper box of **Figure 5** correspond to the pseudo-frequency of about $\omega_{a4}/2\pi = \omega_4/2\pi a \approx (50 - 60)KHz > \omega_a/2\pi \approx 30KHz$. Their

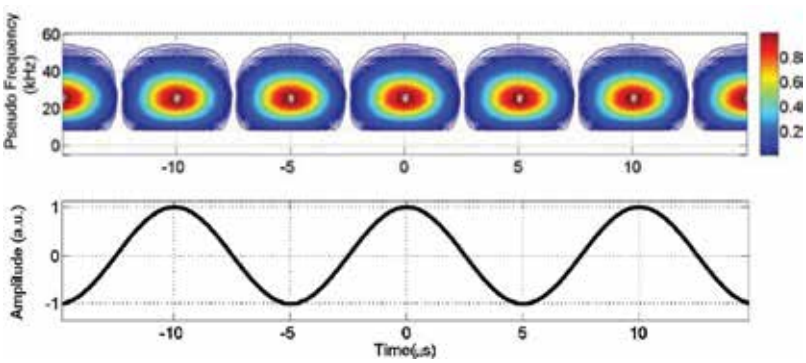


Figure 3. The time dependence (lower box) and scalogram (upper box) of the monochromatic sinusoidal signal $f(\omega_0 t) = \cos(\omega_0 t)$, $\omega_0/2\pi = 100KHz$.

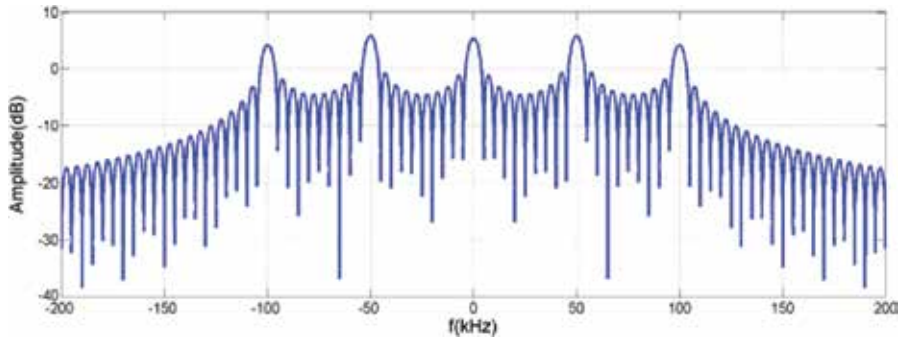


Figure 4. The spectrum of the signal $\mathrm{Re}f_n(\omega_0 t, \omega/\omega_0)$ for $\omega_0/2\pi = 100KHz$, $n = 4$.

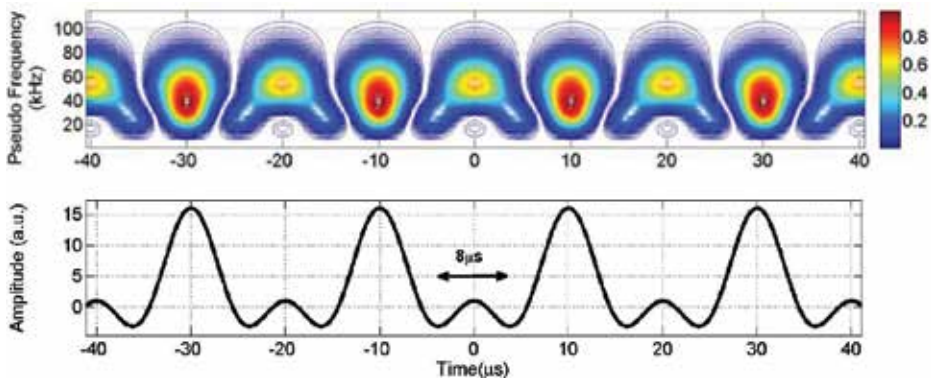


Figure 5. The time dependence (lower box) and scalogram (upper box) of the superoscillating signal $\mathrm{Re}f_n(\omega_0 t, \omega/\omega_0)$, $\omega_0/2\pi = 100KHz$, $n = 4$, the Mexican hat mother wavelet.

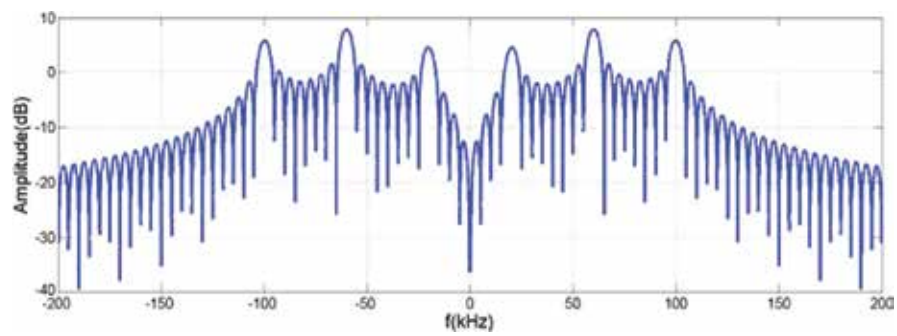


Figure 6. The spectrum of the signal $\mathrm{Re}f_n(\omega_0 t, \omega/\omega_0)$ for $\omega_0/2\pi = 100KHz$, $n = 5$.

maxima are strongly manifested at the time intervals with the center localized near $t = 0, \pm 20\mu s$ and the pseudo-frequencies higher than the ones corresponding to the highest pseudo-frequency shown in the upper box of **Figure 3**.

http://dx.doi.org/10.5772/intechopen.76333

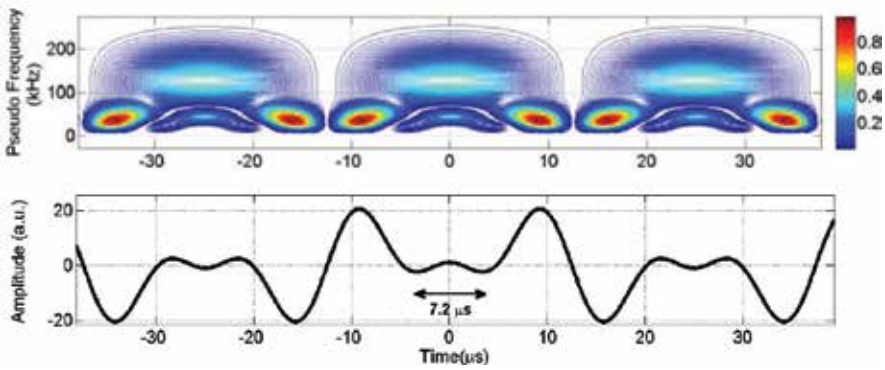


Figure 7. The time dependence (lower box) and scalogram (upper box) of the superoscillating signal $\mathrm{Re}f_n(\omega_0 t, \omega/\omega_0)$, $\omega_0/2\pi = 100KHz$, $n = 5$, the Mexican hat mother wavelet.

It is seen from the lower box of **Figure 7** that the superoscillations with the time duration of about $\Delta T(n=5) \approx 7.2\mu s < \Delta T(n=4)$ are identified for $n = 5$. They have the frequency $\omega_5/2\pi = 1/\Delta T$ $(n=5) \approx 139KHz > \omega_4/2\pi = 125KHz$.

The corresponding energy contours are identified in the scalogram (upper box of **Figure 7**) in the time intervals localized near $t = 0, \pm 25\mu s$. The maxima of the corresponding energy contours are localized at the pseudo frequency of about $\omega_{a5}/2\pi = \omega_5/2\pi a \approx (100-130)KHz$ $> \omega_a/2\pi \approx 30KHz$. Evidently, we can identify the higher frequency superoscillations by increasing n and using CWT.

The scalograms for different mother wavelets are also different. In order to compare the CWT results consider the application of the Morlet mother wavelet Eq. (24) for the superoscillating signal $\mathrm{Re}f_n(\omega_0 t, \omega/\omega_0)$, $\omega_0/2\pi = 100KHz$, $n = 5$.

Comparison of **Figures 7** and **8** shows that the spectral features of superoscillations are pronounced at the higher pseudo-frequencies of about $(250-300)GHz$, because the central frequencies of the

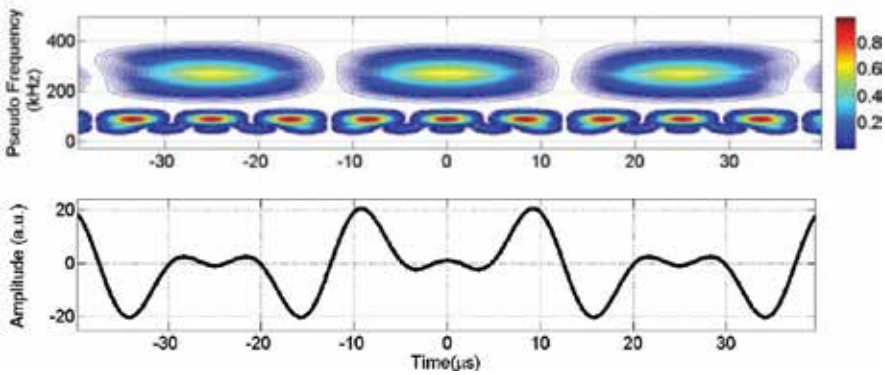


Figure 8. The time dependence (lower box) and scalogram (upper box) of the superoscillating signal $\mathrm{Re}f_n(\omega_0 t, \omega/\omega_0)$, $\omega_0/2\pi = 100KHz$, $n = 5$, the Morlet mother wavelet.

Mexican hat and Morlet mother wavelets are different. The real superoscillation frequencies can be obtained by pseudo-frequency multiplication by a scaling parameter a according to Eq. (39).

The theoretical results of the wavelet analysis clearly show that the superoscillations with the local frequency larger that the band limit of the signal can be identified by using CWT.

7. Conclusions

We for the first time to our best knowledge applied CWT for the theoretical analysis of superoscillations in the time and frequency domain. We discussed the basic properties of the superoscillating signals containing the components with the frequencies larger than the maximum frequency in the signal spectrum. We also considered some possible applications of superoscillations in optics and signal processing. The superoscillating components are extremely weak and short in the time domain. They cannot be identified by the Fourier transform since they require the time-frequency analysis. We discussed the fundamental properties of CWT and DWT and their typical applications. The CWT is a unique tool for the superoscillation studies because it provides the localization of the signal both in time and in the frequency domain. We used the Mexican hat and the Morlet mother wavelets for the CWT of the sinusoidal superoscillating signal because these mother wavelets are similar to the signal oscillations. The theoretical results clearly show that the superoscillation frequency, time duration, and energy contours can be identified by using the CWT of the corresponding signal. Generally, CWT with different mother wavelets can be used for the analysis of superoscillating signals with different structures.

Author details

Yossef Ben Ezra, Boris I. Lembrikov*, Moshe Schwartz and Segev Zarkovsky

*Address all correspondence to: borisle@hit.ac.il

Department of Electrical Engineering and Electronics, Holon Institute of Technology (HIT), Holon, Israel

References

[1] Katsav E, Schwartz M. Yield-optimized superoscillations. IEEE Transactions on Signal Processing. 2013;**61**:3113-3118. DOI: 10.1109/TSP.2013.2258018

[2] Ferreira PJSG, Kempf A. Superoscillations: Faster than the Nyquist rate. IEEE Transactions on Signal Processing. 2006;**54**:3732-3740. DOI: 10.1109/TSP.2006.877642

[3] Lee DG, Ferreira PJSG. Superoscillations of prescribed amplitude and derivative. IEEE Transactions on Signal Processing. 2014;**62**:3371-3378. DOI: 10.1109/TSP.2014.2326625

[4] Lee DG, Ferreira PJSG. Direct construction of superoscillations. IEEE Transactions on Signal Processing. 2014;**62**:3125-3134. DOI: 10.1109/TSP.2014.2321119

[5] Eliezer Y, Bahabad A. Super-transmission: The delivery of superoscillations through the absorbing resonance of a dielectric medium. Optics Express. 2014;**22**:31212-31226. DOI: 10.1364/OE.22.031212

[6] Goswami JC, Chan AK. Fundamentals of Wavelets. 2nd ed. New Jersey: Wiley; 2011. 359 p. ISBN: 9780470484135

[7] Rao RM, Bopardikar AS. Wavelet Transforms. 1st ed. California: Addison-Wesley; 1998. 310 p. ISBN: 0-201-63463-5

[8] Kempf A. Black holes, bandwidths and Beethoven. Journal of Mathematical Physics. 2000; **41**:2360-2374. DOI: 10.1063/1.533244

[9] Aharonov Y, Anandan J, Popescu S, Vaidman L. Superposition of time evolutions of a quantum system and a quantum time-translation machine. Physical Review Letters. 1990; **64**:2965-2968. DOI: 10.1103/PhysRevLett.64.2965

[10] Berry MV. Evanescent and real waves in quantum billiards and Gaussian beams. Journal of Physics A: Mathematical and General. 1994;**27**:L391-L398. DOI: 0305-4470/94/110391

[11] Berry MV, Popescu S. Evolution of quantum superoscillations and optical superresolution without evanescent waves. Journal of Physics A: Mathematical and General. 2006;**39**:6965-6977. DOI: 10.1088/0305-4470/39/22/011

[12] Katsav E, Perlsman E, Schwartz M. Yield statistics of interpolated superoscillations. Journal of Physics A: Mathematical and Theoretical. 2017;**50**:025001-1-025001-20. DOI: 10.1088/1751-8113/50/2/025001

[13] Berry MV. Suppression superoscillations by noise. Journal of Physics A: Mathematical and Theoretical. 2017;**50**:025003-1-025003-9. DOI: 10.1088/1751-8113/50/2/025003

[14] Berry MV, Fishman S. Escaping superoscillations. Journal of Physics A: Mathematical and Theoretical. 2018;**51**:025205-1-025205-14. DOI: 10.1088/1751-8121/aa9b50

[15] Rogers ETF, Lindberg J, Roy T, et al. A superoscillatory lens optical microscope for subwavelength imaging. Nature Materials. 2012;**11**:432-435. DOI: 10.1038/NMAT3280

[16] Diao J, Yuan W, Yu Y, Zhu Y, Wu Y. Controllable design of super-oscillatory planar lenses for sub-diffraction-limit optical needles. 2016;**24**:1924-1933. DOI: 10.1364/OE.24.001924

[17] Sarkar TK, Salazar-Palma M, Wicks MC. Wavelet Applications in Engineering Electromagnetics. 1st ed. Boston, USA: Artech House; 2002. 347 p. ISBN: 1-58053-267-5

[18] Cotronei M, Montefusco LB, Puccio L. Multiwavelet analysis and signal processing. IEEE Trans. Circuits and Systems-II: Analog and Digital Signal Processing. 1998;**45**:970-987. DOI: 1057-7130(98)04672-2

[19] Kingsbury N. Complex wavelets for shift invariant analysis and filtering of signals. Journal of Applied and Computational Analysis. 2001;**10**:234-253. DOI: 10.10016/acha.2000.0343

[20] Ben Ezra Y, Lembrikov BI, Zadok A, Halifa R, Brodeski D. All-optical signal processing for high spectral efficiency (SE) optical communication. In: Das N, editor. Optical Communication. Croatia: InTech; 2012. pp. 343-366 ISBN: 978-953-51-0784-2

[21] Ben Ezra Y, Lembrikov BI. Application of complex wavelet packet transform (CWPT) in coherent optical OFDM (CO-OFDM) communication systems. In: Das N, editor. Advances in Optical Communication. Croatia: InTech; 2014. pp. 145-168 ISBN: 978-953-51-4122-8

[22] Ben-Ezra Y, Dahan D, Zarkovsky S, Lembrikov BI. High spectral efficiency (SE) OFDM system based on multi-wavelet packets. In: Proceedings of 17th International Conference on Transparent Optical Networks (ICTON 2015); July 5–9, 2015; Budapest, Hungary, (Mo. C1.3). DOI: 978-1-4673-7879-6/15

[23] Ben Ezra Y, Zadok A, Califa R, Munk D, Lembrikov BI. All-optical wavelet based orthogonal frequency division multiplexing (OFDM) system based on silicon photonic integrated components. IET Optoelectronics. 2016;**10**:44-50. DOI: 10.1049/iet-opt.2015.0064

[24] Li A, Shieh W, Tucker RS. Wavelet transform-based OFDM for optical communications. Journal of Lightwave Technology. 2010;**28**:3519-3528. DOI: 10.1109/JLT.2010.2089673

[25] Ma C, Liu T, Liu K, et al. A continuous wavelet transform based time delay estimation method for long range fiber interferometric vibration sensor. Journal of Lightwave Technology. 2016;**34**:3785-3780. DOI: 10.1109/JLT.2016.2583938

[26] Alves KD, Costa FB, de Araujo Ribeiro RL, de Sousa Neto CM, de Oliveira Alves Rocha T. Real-time power measurement using the maximal overlap discrete wavelet-packet transform. IEEE Transactions on Industrial Electronics. 2017;**64**:3177-3187. DOI: 10.1109/TIE.2016.2637304

[27] Abduljabbar AM, Yavuz ME, Costen F, Himeno R, Yokota H. Continuous wavelet transform-based frequency dispersion compensation method for electromagnetic time-reversal imaging. IEEE Transactions on Antennas and Propagation. 2017;**65**:1321-1329. DOI: 10.1109/TAP.2016.2647594

[28] Tiwari VK, Umarikar AC, Jain T. Fast amplitude estimation of harmonics using undecimated wavelet packet transform and its hardware implementation. IEEE Transactions on Instrumentation and Measurement. 2018;**67**:65-77. DOI: 10.1109/TIM.2017.2759358

Chapter 11

Use of Wavelet Techniques in the Study of Seawater Flux Dynamics in Coastal Lakes

Edwin Muchebve, Yoshiyuki Nakamura and Hiroshi Kamiya

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/intechopen.75177

Abstract

Lakes Shinji and Nakaumi form a coupled brackish lake system on the Japan Sea coast of Japan, and seawater frequently intrudes into these lakes. The study analyzed the influence of river discharge, tide level, and meteorological forcing on seawater flux at Nakaura Watergate. Continuous wavelet transform (CWT), wavelet coherence (WTC), cross-wavelet transform (XWT), and partial wavelet coherence (PWC) were used to analyze seawater intrusion. CWT reveals the characteristics of seawater flux, river discharge, tide level, and meteorological forcing. WTC and XWT showed the correlation between forcing variables and seawater flux. PWC analyzed the impact of river discharge, tide level, and meteorological forcing variables on seawater flux after controlling the effect of all other forcing variables. The results showed that tide level has an impact on seawater flux over 0.5-day and 1-day cycle throughout the analysis period, suggesting that astronomical tides play an important role in the salinity intrusion processes. The river discharge characteristics and its influence on seawater flux were clearly revealed especially during high river flows. Sometimes, atmospheric pressure and wind velocity affect tide level, thereby driving salinity transport. The study reveals the power of wavelet analysis in examining nonlinear time series such as salinity intrusion processes.

Keywords: seawater flux, coastal lakes, wavelet analysis, wavelet coherence

1. Introduction

The wavelet transform has been used for process understanding since the early 1980s. It originated in geophysics and was significantly developed, both theoretically and application-oriented, in the signal processing and mathematics community [1]. Wavelets enable linear and

optimal extraction of information from a time series of any length [2]. The continuous wavelet transform (CWT) provides redundant information by mapping a time series into a function of time and frequency. The discrete wavelet transform (DWT) computes the transform for discrete values for time and frequency [3]. Hence, DWT is simple, easy to implement, and has low computational requirements. CWT requires a high computational time; however, it allows a larger freedom in wavelets selection than DWT. Redundant information in pictures from the CWT makes it easier to interpret results from the analysis of dynamic time series data. For an analysis where the main purpose is to reveal patterns or hidden information and data compression is not of concern, then redundancy provided by CWT may be useful [3]. Generally, CWTs are useful for dynamical analyses, while DWTs are better for information compression [2, 3].

This study utilized the CWT rather than DWT. The CWT may use an arbitrary number of *daughter wavelets* built from *mother wavelet* to match salinity oscillation periods, as needed for the optimal extraction of information concerning astronomical and meteorological forcing of the salinity intrusion. The *daughter wavelets* will be complete but not orthogonal. A DWT can be complete and orthogonal, given that it is built from functions with geometrically spaced periods; however, it may not have frequency flexibility necessary for salinity intrusion analysis [2].

Various studies have used wavelet analysis to analyze nonstationary time series water quality, meteorological, and hydrological data. Torrence and Compo provided a practical guide to wavelet analysis and also analyzed time series of the El Niño-Southern Oscillation (ENSO) [4]. Liu et al. ratified the bias problem in the estimation of the wavelet power spectrum and applied wavelet analysis to Niño-3 SST data [5]. Wavelet analysis was also used to study water-quality parameters [6, 7]. Zhang et al. [8], Sovi et al. [9], and Somoza et al. [10] used wavelet analysis to characterize water level variation. A number of studies have analyzed tidal variation and its influence on rivers using wavelet analysis [2, 11–13]. Ideião et al. studied the variability of the total monthly rainfall with the aid of CWT [4, 14]. The cross-wavelet transform (XWT) and wavelet coherence (WTC) techniques were used to analyze geophysical time series, for example, the effects of tidal range and river discharge on the salinity intrusion [15, 16], the impacts of Arctic Oscillation index and ENSO on the Baltic sea ice [17, 18], the relative humidity, and the shortwave radiation dataset [19]. Partial wavelet coherence (PWC) and multiple wavelet coherence (MWC) were used to study the impact of ENSO on the variability of tropical cyclones [20]. Wavelet analysis has also been applied in economics field [3, 21].

This study used CWT to analyze the period characteristics of tide level, river discharge, meteorological forcing variables, and seawater flux. It also quantified the relationships between river discharge, tide and meteorological forcing variables, and salinity intrusion, using WTC, XWT, and PWC. Several studies examined the effects of river discharge, tidal range, and meteorological forcing on salinity intrusion [15, 16]. Meteorologically induced sea surface variation (MISSV) and large periodic river discharge are considered to be effective water exchange mechanism between Lakes Nakaumi and Shinji, and the Japan Sea [22]. Though tidal amplitude on the Japan Sea is small, astronomical tides appear to be an effective water exchange mechanism [16]. The study of the influence of external forces on salinity

intrusion is difficult; the same applies to the development and application of seawater intrusion countermeasures. However, understanding the dynamic characteristics of forcing mechanisms and their influence on seawater intrusion enable the development and implementation of appropriate mitigation measures, for example, the regulation of river flow or the operation of flood control gates.

The previous wavelet analysis of salinity intrusion in the Sakai Channel was studied for one summer season [16]. Since salinity intrusion varies with season and time, the assessment of its dynamics requires a long-term analysis. Therefore, in order to understand the nonlinear characteristics of salinity intrusion in Sakai Channel in other seasons, this study conducted long-term wavelet analysis.

2. Materials and methods

2.1. Study area

Lakes Shinji and Nakaumi form a coupled brackish lake system in the western part of Japan (**Figure 1**). Lake Shinji has an average depth of 4.5 m, a surface area of 80 km^2, and a volume of 0.366 km^3. Lake Nakaumi has an average depth of 5.4 m, a surface area of 86.2 km^2, and a volume of 0.47 km^3. The Ohashi River (7.0 km long) connects the two lakes and the Sakai Channel (7.5 km long) connects Lake Nakaumi to the Japan Sea. The Hii River at the west-end of Lake Shinji supplies the lake system with most of its fresh water. Lake Shinji is a mesohaline lake with an average salinity between 1 and 6 PSU. Lake Nakaumi has a strongly differentiated two-layer system; the salinity of the surface water is 14–20 PSU and that of the bottom layer is 25–30 PSU. Hence, these brackish lakes are stably stratified due to salinity (density) differences, and density gradients have a large impact on water movement in this system [23].

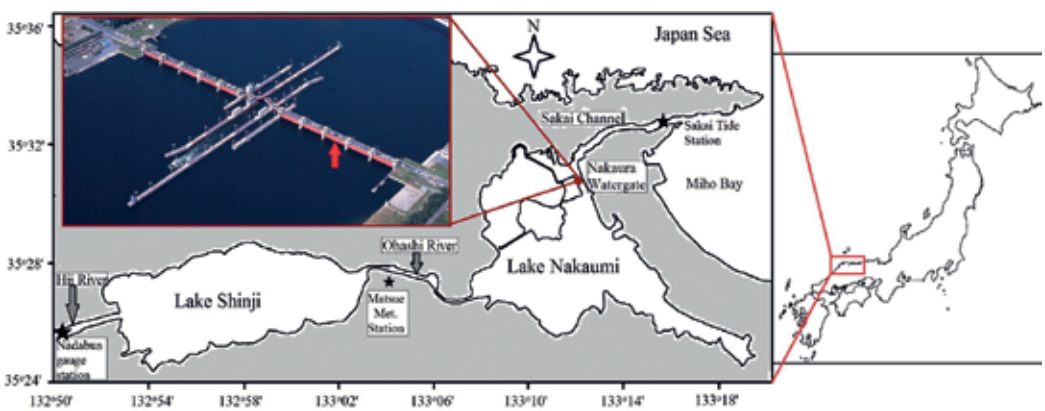


Figure 1. Location of Lakes Shinji and Nakaumi, and Nakaura Watergate monitoring station (insert). The red arrow indicates the third eastern-side floodgate, the location of sampling, and measurement equipment. Also shown are meteorological and hydrological monitoring stations (source: [16]).

2.2. Observations

Salinity data used in the study were collected at Nakaura Watergate monitoring station shown in **Figure 1**. Nakaura Watergate (width 414 m, depth 6.8 m) had five floodgates in the east and in the west (each 32 m long), and three floodgates at the center. On the western pile of the third east-side floodgate (indicated by a red arrow, **Figure 1**), submerged water pumps were installed for water sampling at 1, 2, 4, and 6 m from the bottom. The water was pumped to acrylic boxes in the floodgate administration building, where water temperature, electrical conductivity, and dissolved oxygen were measured every 30 min using custom-made sensors (Alec Electronics Co., Ltd.). Salinity was calculated from electrical conductivity.

Continuous measurements of salinity over a period of 6 years (February 1998 to March 2004) are available, although there are periods with missing data. Instantaneous seawater flux for the period January 2001 to October 2003 was used in this study. The salinity data were averaged to a 1-h interval to match the intervals of the other data used in the analysis. One-hour interval meteorological data (atmospheric pressure, wind speed, and direction) were collected at Matsue Meteorological Station (available on Japan Meteorological Agency website, http://www.jma.go.jp/jma/index.html). The wind was treated as a mathematical vector, and the mathematical convention for the direction was used, that is, wind direction was converted from "meteorological direction" to "math direction." The wind vector was resolved into its u (wx) and v (wy) components. Wind from the west was denoted as positive u and from the south a positive v. Tidal data used were recorded at Mihonoseki tide gauge station, and river discharge was recorded in Hii River at Nadabun gauging station (available on Japan's water information system website, http://www1.river.go.jp/).

2.3. Methodology

2.3.1. Salinity transport

Instantaneous advective salt transport (M_s, kg $m^{-1}s^{-1}$) per unit width of a section, normal to the longitudinal flow of the channel, is given by the following expression [24, 25]:

$$M_s = \int_0^h \rho VS dz = \overline{\rho VS} \cdot h \tag{1}$$

where ρ is the density, V is the longitudinal velocity component, and S is the longitudinal salinity. The upper bar denotes averaging over the total depth of the water column, h.

The study lakes are shallow. Pressure variation in shallow lakes is negligible. Therefore, the density of water was calculated using the following approximate density formula neglecting pressure [26]:

$$\rho = 999.83 + 0.808S - 0.0708(1 + 0.068T)T - 0.003(1 - 0.012T)(35 - S)T \tag{2}$$

where T is the temperature in °C and S is the salinity in PSU.

http://dx.doi.org/10.5772/intechopen.75177

2.3.2. Continuous wavelet transform (CWT)

CWT decomposes a time series into a time-frequency space and determines both the dominant modes of variability and their variation with time [4]. The wavelet is applied as a bandpass filter to the time, stretching it in time by varying its scale(s) and normalizing it to have a unit energy [8, 18].

A wavelet $\psi(t)$ is a function that oscillates around the t-axis and loses strength as it moves away from the center, behaving like a small wave [4]. Beginning with a mother wavelet ψ, a family of "daughter wavelet" $\psi(\tau, s)$ is computed by scaling and translating ψ [4]:

$$\psi_{\tau,s}(t) = \frac{1}{\sqrt{|s|}}\psi\left(\frac{t-\tau}{s}\right); s, \tau \in \mathbb{R}; s \neq 0 \tag{3}$$

where $\psi(t)$ is the mother wavelet, $\psi_{\tau,s}(t)$ is the daughter wavelet, t is a nondimensional "time" parameter, s is a scaling or a dilation factor that controls the width of the wavelet, and τ is a translation parameter controlling the location of the wavelet.

This study used the Morlet wavelet, which consists of a plane wave modulated by a Gaussian or in other words a complex exponential function multiplied by a Gaussian window. Hence, it represents the best compromise between frequency and time localization. A complex wavelet is essential for this study, as it yields a complex transform, with information on both the amplitude and the phase, crucial to study the synchronization of oscillations between different time series [3, 27]. Morlet wavelet is defined as [4]

$$\psi_0(t) = \pi^{-1/4} e^{i\omega_0 t} e^{-t^2/2} \tag{4}$$

where $\psi_0(t)$ is the Morlet wavelet, ω_0 is the fundamental frequency, which gives the number of oscillations within the wavelet itself.

CWT of a time series $x(t) \in L^2(\mathbb{R})$ with respect to the wavelet ψ is a function of two variables, $W_{x;\psi}(\tau, s)$ [3]:

$$W_{x;\psi}(\tau, s) = \int_{-\infty}^{\infty} x(t) \frac{1}{\sqrt{|s|}} \psi^*\left(\frac{t-\tau}{s}\right) dt \tag{5}$$

where $W_{x;\psi}(\tau, s)$ is the CWT of a time series $x(t)$, $L^2(\mathbb{R})$ denotes the set of square integrable functions, that is, the set of functions defined on the real line and satisfying $\int_{-\infty}^{\infty} |x(t)|^2 dt < \infty$, ψ^* is the complex conjugation of ψ.

2.3.3. Cross-wavelet transform (XWT)

The cross-wavelet transform (XWT), a multiscale signal analytical technique, combines the wavelet transform and cross-spectrum analysis. XWT analyzes multiple time-frequencies of

two time series from multiple time scale points, thereby exposing regions with a common high power, and further reveals information about the phase relationship in time-frequency space, hence determining correlations [15, 18].

For two time series x_n and y_n, their cross-wavelet transform is given by [4, 15]

$$W_n^{XY}(s) = W_n^X(s)W_n^{Y*}(s) \tag{6}$$

where $W_n^X(s)$ and $W_n^Y(s)$ are wavelet transforms of x_n and y_n, respectively, and * denotes complex conjugation. $\left|W_n^{XY}(s)\right|$ is the cross-wavelet power. If two time series have background power spectra P_k^X and P_k^Y, then their theoretical distribution of the cross-wavelet power is given by [4, 18].

$$D\left(\frac{\left|W_n^X(s)W_n^{Y*}(s)\right|}{\sigma_X\sigma_Y} < p\right) = \frac{Z_v(p)}{v}\sqrt{P_k^X P_k^Y} \tag{7}$$

where σ_X and σ_Y are the respective standard deviations, $Z_v(p)$ is the confidence level associated with the probability p for a probability distribution function (*pdf*) defined by the square root of the product of two chi-squared (χ^2) distributions.

2.3.4. Wavelet coherence (WTC)

Wavelet coherence (WTC) between two CWTs can find significant coherence even though the common power is low and show how confidence levels against red noise backgrounds are calculated. This can be thought of as a local correlation between two time series in the time-frequency space. It finds locally phase-locked behavior. The significance level of the WTC is determined using Monte Carlo methods [15]

$$R_n^2(s) = \frac{\left|S\left(s^{-1}W_n^{XY}(s)\right)\right|^2}{S\left(s^{-1}\left|W_n^X(s)\right|^2\right)\cdot S\left(s^{-1}\left|W_n^Y(s)\right|^2\right)} \tag{8}$$

where S is a smoothing operator.

2.3.5. Partial wavelet coherence

CWT is increasingly being used in the analysis of marine sciences time series data. However, most of the CWT analyses have been limited to univariate and bivariate analyses, that is, the wavelet power spectrum, the wavelet coherency, and the wavelet phase difference [3]. Wavelet analysis tools have already been extended to allow for multivariate analyses [3, 20]. PWC and PPD are the examples of recent wavelet analysis techniques. The PWC technique is similar to partial correlation and it identifies the resulting wavelet coherence between two time series after eliminating the influence of their common dependence [20]. The applicability of PWC to geophysics was demonstrated during the study of the "stand-alone" relationship between the

"ratio of number of typhoons to number of tropical cyclones" and "large-scale atmospheric factors" after removing the effect of El Nino-Southern Oscillation (ENSO) [20].

The squared multiple wavelet coherence ($R^2_{1(23\ldots p)}$) between the series x_1 and all the other series $x_2,\ldots, x_p$ is given by the following formula [3]:

$$R^2_{1(23\ldots p)} = 1 - \frac{C^d}{C^d_{11}} \tag{9}$$

where C denotes the $p \times p$ matrix of all the complex wavelet coherencies ϱ_{ij}, that is, $C = \left(\varrho_{ij}\right)_{i,j=1}^{p}$, $C^d = \det C$.

The complex partial wavelet coherence ($\varrho_{1j.q_j}$) of x_1 and $x_j (2 \le j \le p)$ allowing for all the other series is given by [3]

$$\varrho_{1j.q_j} = -\frac{C^d_{j1}}{\sqrt{C^d_{11}}\sqrt{C^d_{jj}}} \tag{10}$$

The partial wavelet coherence ($R_{1j.q_j}$) of x_1 and x_j allowing for all the other series is defined as the absolute value of Eq. (10), that is,

$$R_{1j.q_j} = \frac{\left|C^d_{j1}\right|}{\sqrt{C^d_{11}C^d_{jj}}} \tag{11}$$

and the squared partial wavelet coherence of x_1 and x_j allowing for all the other series is simply the square of $R_{1j.q_j}$.

The partial phase delay (phase difference) of x_1 and x_j given all the other series is defined as the angle of $\varrho_{1j.q_j}$. A complex wavelet function contains information about both the amplitude and the phase, making it suitable to capture oscillatory behavior. Complex partial wavelet coherence, $\varrho_{1j.q_j}$ considered can be separated into its real part, $\Re\left(\varrho_{1j.q_j}\right)$, and imaginary part, $\Im\left(\varrho_{1j.q_j}\right)$, or in its amplitude, $\left|\varrho_{1j.q_j}\right|$, and phase angle $\phi_{1j.q_j}$. The phase difference, $\phi_{1j.q_j}$, is given as follows [3]:

$$\phi_{1j.q_j} = Arctan\left(\frac{\Im\left(\varrho_{1j.q_j}\right)}{\Re\left(\varrho_{1j.q_j}\right)}\right) \tag{12}$$

A phase difference of zero indicates that the time series moves together at the specified time-frequency; if $\phi_{x_1x_j} \epsilon (0, 90)$, the series moves in a phase and the time series of x_1 leads x_j; if $\phi_{x_1x_j} \epsilon (-90, 0)$, then x_j leads x_1. A phase difference of 180 (or -180) indicates an antiphase

relationship; if $\phi_{x_1 x_j} \epsilon (90, 180)$, then x_j leads x_1; if $\phi_{x_1 x_j} \epsilon (-180, -90)$, then x_1 leads x_j (see **Figure 2**). Phase difference can be converted into instantaneous time lag between two time series by dividing the phase difference, $\phi_{1j.q_j}$ by the angular frequency corresponding to the scale s, $\omega(s)$.

2.3.6. Significance tests

It is important to assess the statistical significance of the wavelet, cross-wavelet power and the wavelet coherence. The assessment of the statistical significance levels and confidence intervals against red noise backgrounds was done using direct Monte Carlo simulations.

2.3.7. Wavelet packages and parameters used

Ng and Kwok provided the software for CWT, WTC, and XWT, which is available at http://www.cityu.edu.hk/gcacic/wavelet. It is used in collaboration with the software provided by Grinsted, which is available at http://www.glaciology.net/wavelet-coherence. The software for PWC was provided by Aguiar-Conraria and Soares and is available at http://sites.google.com/site/aguiarconraria/joanasoares-wavelets. Some of the parameters used in the analysis are as follows: the mother wavelet function—Morlet; the sampling time (dt)—1 h; the spacing between discrete scales (dj)—0.5 h; the level of significance—5%; and the number of Monte Carlo simulations used to assess statistical significance—1000. Default values were taken for other parameters.

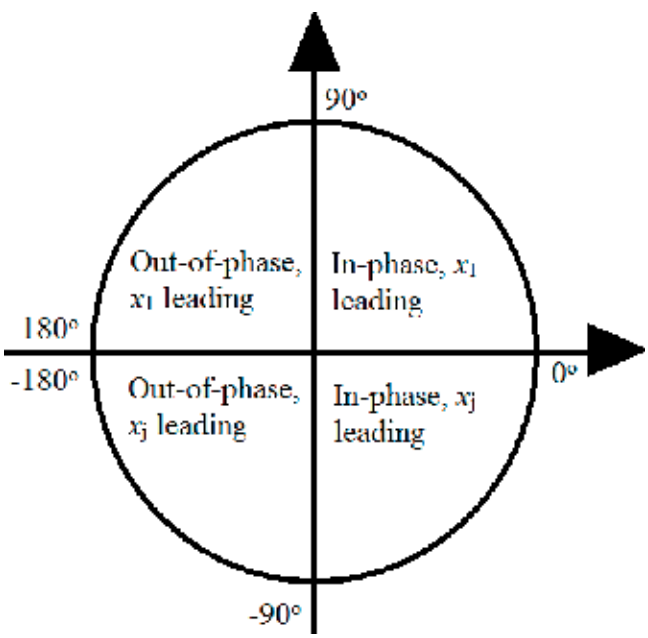


Figure 2. Phase difference circle.

http://dx.doi.org/10.5772/intechopen.75177

3. Results and discussion

3.1. Analysis of period characteristics

3.1.1. Variability of seawater flux

Figure 3 shows the seawater flux per unit width at Nakaura Watergate and its CWT coefficient chart. Positive values of the time series indicate seawater flux toward the Japan Sea and negative toward Lake Nakaumi. The CWT coefficient chart for seawater flux has stable period characteristics, with high power oscillations in the 12-h and 1-day period band throughout the analysis period. Both the red color and the black contour indicate that cycles are strong and statistically significant at 95% confidence level (hereinafter statistically significant).

3.1.2. Variability of tide level

Figure 4 shows time series plot and CWT coefficient chart for the tide level. The high-power tide level oscillations have statistically significant periods of 12 h and 1 day. This implies considerable power spreads throughout the semi-diurnal and diurnal bands throughout the analysis period. The oscillations indicate spring-neap tidal variations since they appear twice a month. Also observed is a relatively strong statistically significant, though not regular, 2–6-day period cycle that occurs mainly in winter (December to March). Tide level and atmospheric pressure are negatively correlated. The time series of both shows that as the atmospheric pressure increases, the tide level decreases and vice versa (**Figures 4** and **5**).

3.1.3. Variability of sea level atmospheric pressure

The CWT coefficient chart for the atmospheric pressure (**Figure 5**) shows continuous statistically significant high power 64-day period cycles from April 2002. There are also 128-day to 1-year period cycles throughout the analysis period. There are some discontinuous and irregular high power oscillations in the 1-day and 2–32-day period cycles.

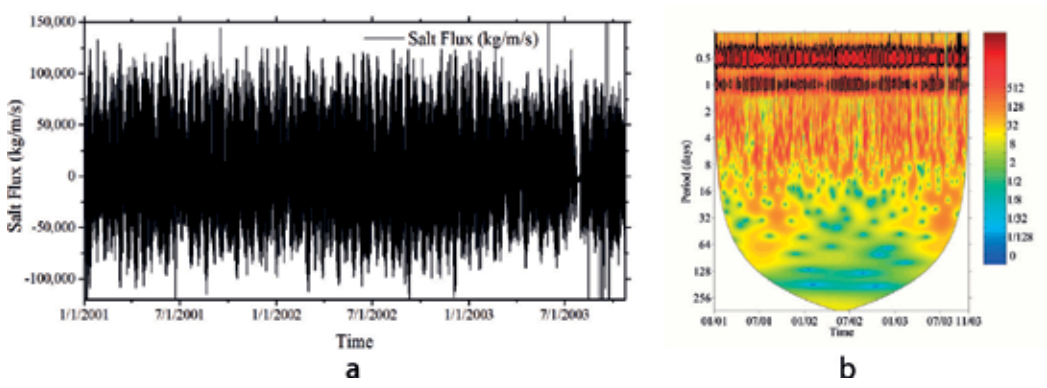


Figure 3. (a) The time series of salinity flux (salt flux (kg/m/s)) and (b) its wavelet power spectrum for the Jan 2001 to Oct 2003 period. Period is in days. The red color designates high power oscillations whilst blue is low power oscillations. The *black* contour designates 95% confidence level, using *red* noise as the background spectrum. *White* regions on either end indicate the "cone of influence" where edge effects become important.

3.1.4. Variability of river discharge

Figure 6 shows the time series plot and CWT coefficient chart for the total river discharge in Hii River. Wavelet coefficients acutely vary from the highest to the lowest, indicating an unstable river discharge. The chart also shows a distinct character that has long vertical peaks like a raindrop, which indicate that the period of oscillation varies from high to low almost

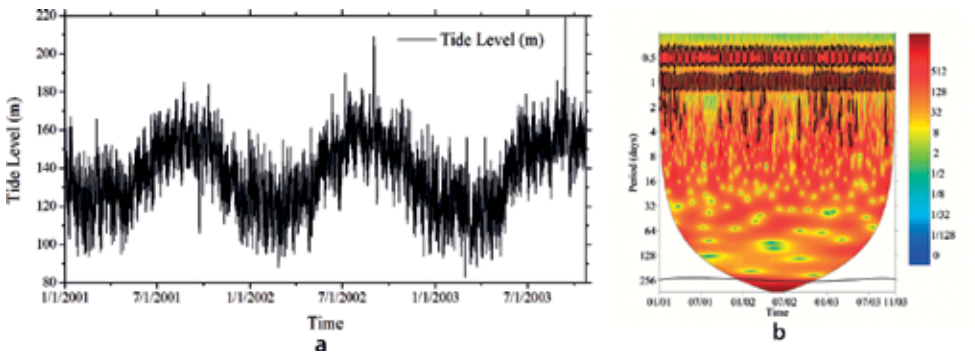


Figure 4. Tide level—time-series plot (a) and time-series wavelet power spectrum (b) for the Jan 2001 to Oct 2003 period.

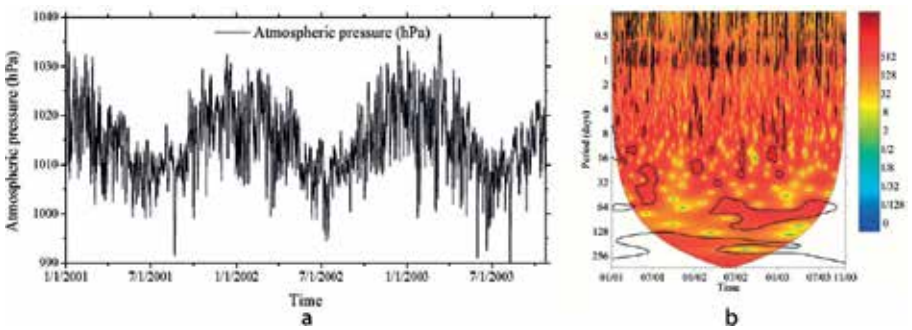


Figure 5. Atmospheric pressure–time-series plot (a) and time-series wavelet power spectrum (b) for the Jan 2001 to Oct 2003 period.

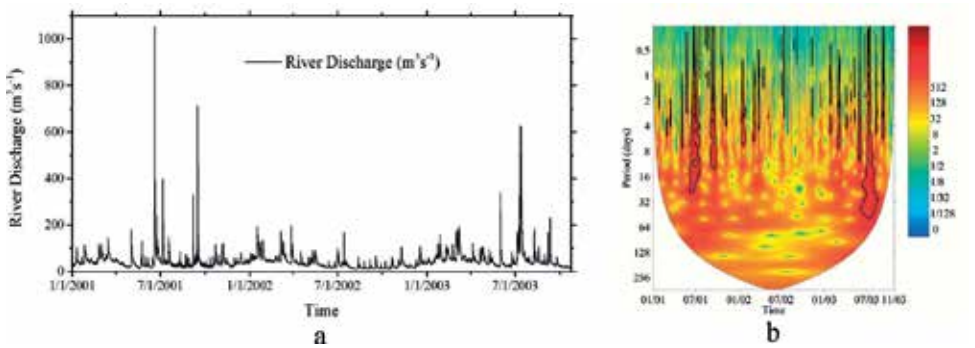


Figure 6. River discharge—time-series plot (a) and time-series wavelet power spectrum (b) for the Jan 2001 to Oct 2003 period.

http://dx.doi.org/10.5772/intechopen.75177

instantaneously. Each peak of energy corresponds to a high river discharge. From the time series plot and CWT coefficient chart, it is evident that the highest river flow rate occurs in June and July. However, there were little rains in the summer of 2002 compared to that of 2001 and 2003. The CWT coefficient chart did show the river discharge in the summer of 2002 as significant; however, the wavelet analysis of the periods June–September 2002 indicated high energy in June and July [16]. The river discharge during June–July 2002 was dwarfed by that of June–July 2001 and 2003 and hence the absence of high energy on the CWT coefficient chart.

3.1.5. Variability of wind velocity

Figure 7 shows the CWT coefficient chart for the wind velocity vectors. There are observations of fluctuating medium power in the 0.5-day period band for the North–South (*wy*) wind velocity component throughout the analysis period. The strong and statistically significant oscillations are in the 2–14-day period band throughout the analysis period. The East–West (*wx*) wind velocity component has discontinuous and irregular high power oscillations in the 2–16-day period band. There are continuous medium-power oscillations in the 32-day period band and 64-day to 1-year period band.

3.2. Analysis of dynamic relationships

3.2.1. Correlation between the tide level and the seawater flux

The WTC and XWT of the tide level and the seawater flux are shown in **Figure 8**, which displays that significant power sections appear continuously throughout the analysis period. This indicates

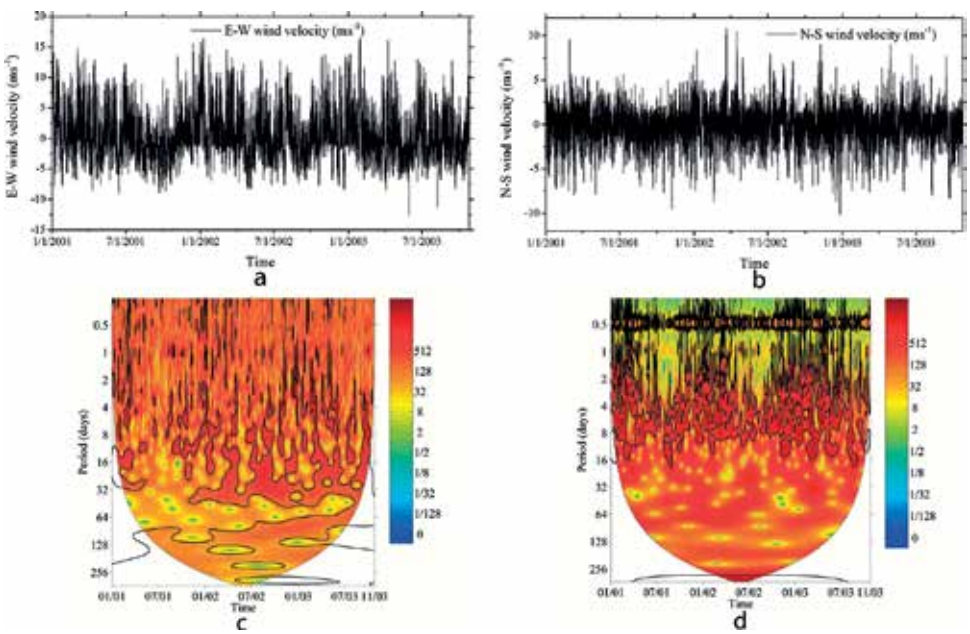


Figure 7. Wind velocity (*wx* (a & c) and *wy* (b & d))—time-series plot (a & b) and time-series wavelet power spectrum (c & d) for the Jan 2001 to Oct 2003 period.

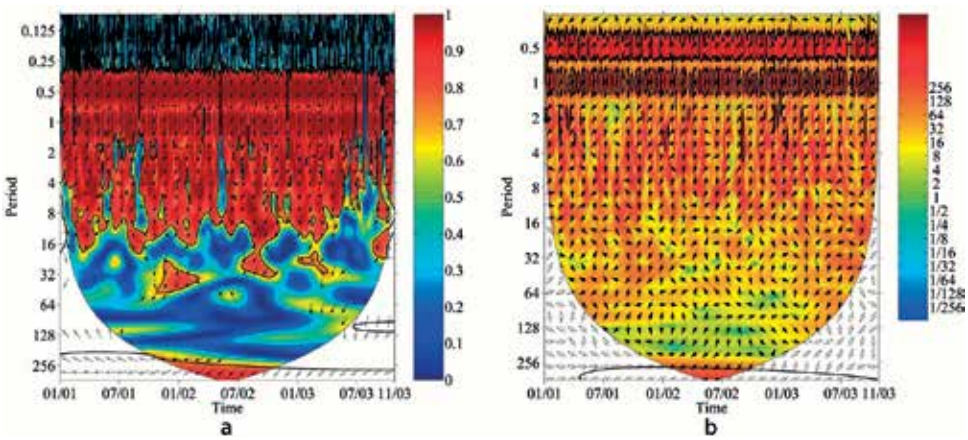


Figure 8. The WTC (a) and XWT (b) of the tide level and the salinity flux for the Jan 2001 to Oct 2003 period. The arrows represent phase difference. The arrow pointing up and to the right—in-phase relationship with the tide level leading. The arrow pointing down and to the left—out-of-phase relationship with the tide level leading.

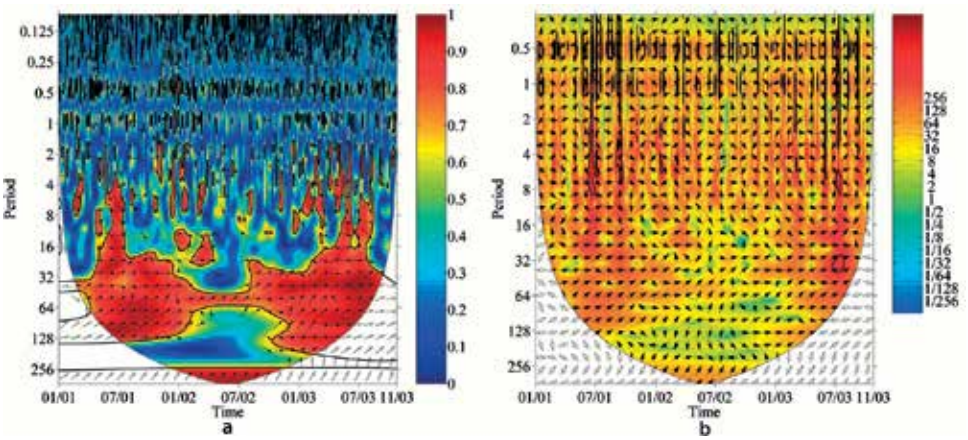


Figure 9. The WTC (a) and XWT (b) of river discharge and salinity flux for the Jan 2001 to Oct 2003 period.

that the influence of the tide level on the seawater flux is strong. Both WTC and XWT show significant power sections in the semi-diurnal and diurnal periods. WTC also shows almost continuous coherence between the tide level and the seawater flux in the 2–16-day period band. XWT does not show much common power in the 2–16-day period band. This indicates that the tide level influences the seawater flux mainly in the 0.5-day and 1-day period band. Both WTC and XWT show that in the 0.5-day and 1-day period band, the tide level and the seawater flux have an antiphase relationship with tides leading (the arrow pointing down and to the left). That is, a rise in the tide level leads to an increase in the negative seawater flux (seawater flux into Lake Nakaumi is denoted as the negative flux in this study).

3.2.2. Correlation between the river discharge and the seawater flux

Extensive significant power sections show the influence of river discharge on salinity, **Figure 9**. The WTC shows continuous coherence in the 16–128-day period band and discontinuous

http://dx.doi.org/10.5772/intechopen.75177

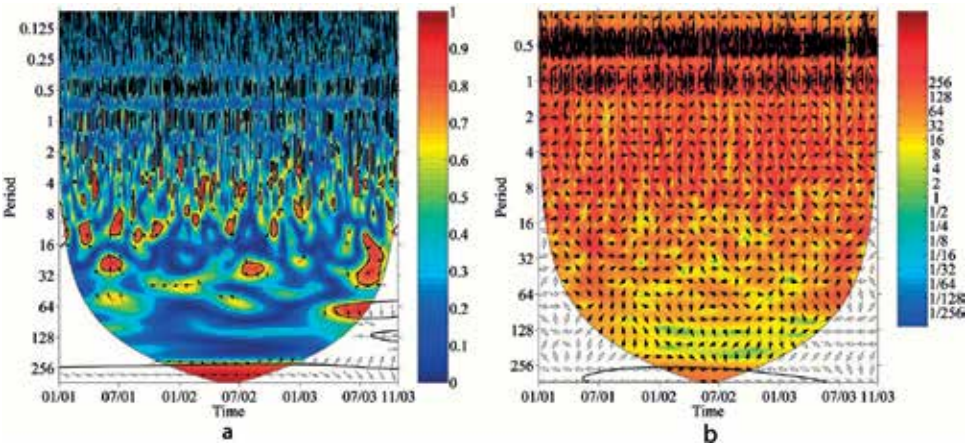


Figure 10. The WTC (a) and XWT (b) of the atmospheric pressure and the salinity flux for the Jan 2001 to Oct 2003 period.

coherence in the 2–16 period band. However, the XWT shows the occasional appearance of the significant power sections with irregular intervals especially during the summer period. The power section appears around June, July, September, and January. The influence of the river discharge on the seawater flux failed to pass the significance test at 5% level in other months and the summer of 2002.

3.2.3. Correlation between the atmospheric pressure and the seawater flux

The XWT and WTC of the atmospheric pressure and the seawater flux shown in **Figure 10** occasionally display extensive significant power sections in the 0.5-, 1–, and 256-day to 1-year period band, which stands out throughout analysis period, testing the existence of the correlation between the atmospheric pressure and the seawater flux. The WTC shows occasional correlation in the 2–16-day period band.

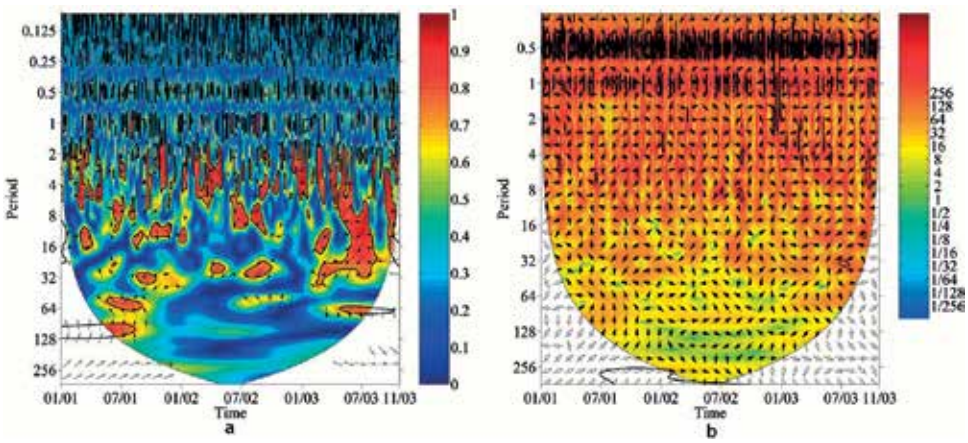


Figure 11. The WTC (a) and XWT (b) for the *wx* wind velocity component and salinity flux for the Jan 2001 to Oct 2003 period.

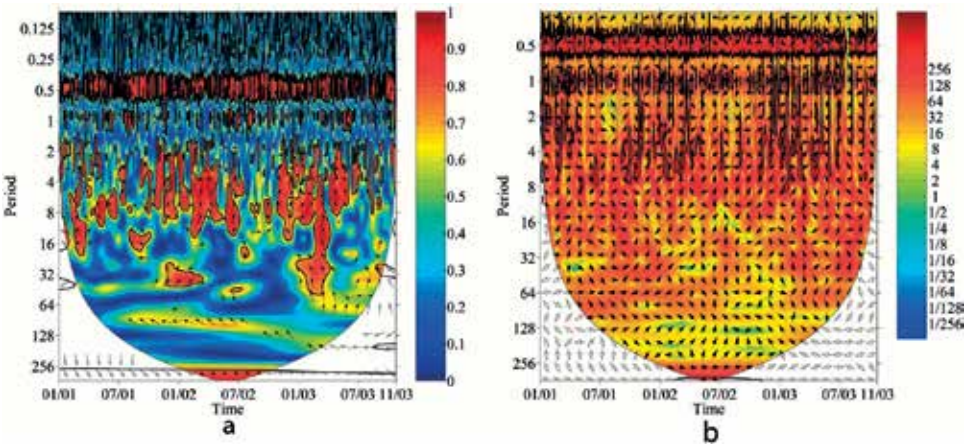


Figure 12. The WTC (a) and XWT (b) for the *wy* wind velocity component and salinity flux for the Jan 2001 to Oct 2003 period.

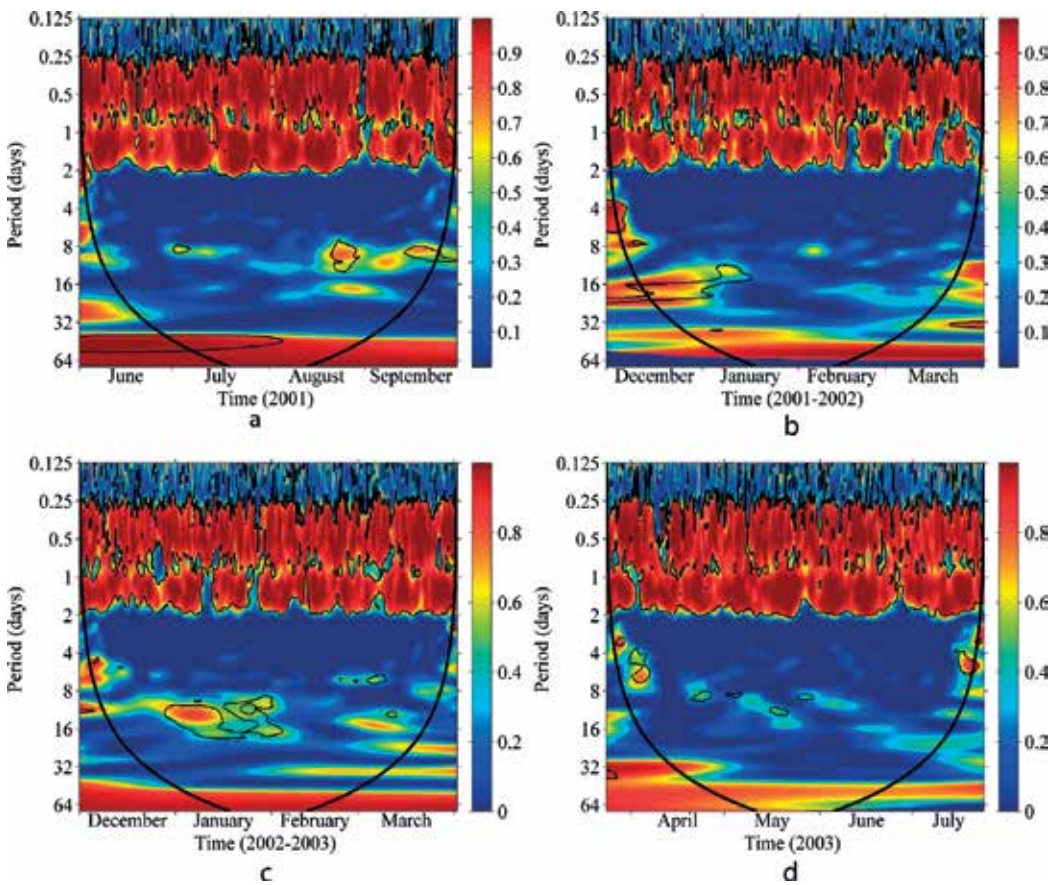


Figure 13. Partial wavelet coherence of observed tides versus salinity flux for the period (a) Jun-Sep 2001, (b) Dec 2001 - Mar 2002, (c) Dec 2002 - Mar 2003, (d) Apr-Jul 2003. The thick black line indicates the cone of influence that delimits the region not influenced by edge effects.

3.2.4. Correlation between the wind velocity and the seawater flux

The XWT and WTC for the wind velocity vectors and the seawater flux are shown in **Figures 11** and **12**. East–West (*wx*) wind velocity vector also influences seawater flux. The WTC indicates significant discontinuous and irregular power sections on the periods 2–32 days throughout the statistical intervals as shown in **Figure 11**. The XWT shows some discontinuous and irregular significant power sections in the 0.5- and 1-day period band, testifying that sometimes a correlation exists between East–West wind velocity component and seawater flux.

The XWT and WTC for North–South (*wy*) wind velocity vector display continuous extensive significant power sections, and their center focuses on the period of 0.5 day (**Figure 12**). The significant power sections also appear at irregular intervals with varying periods in the 2–16-day period band. Powerful influence of wind speed is consistent in the 0.5 day throughout the analysis period.

3.3. Analysis of dynamic characteristics using partial wavelet coherence

Figures 13–17 shows PWC, the relationship, in the time-frequency domain, between seawater flux and each of the forcing variables, after eliminating the effect of other variables. Two summer

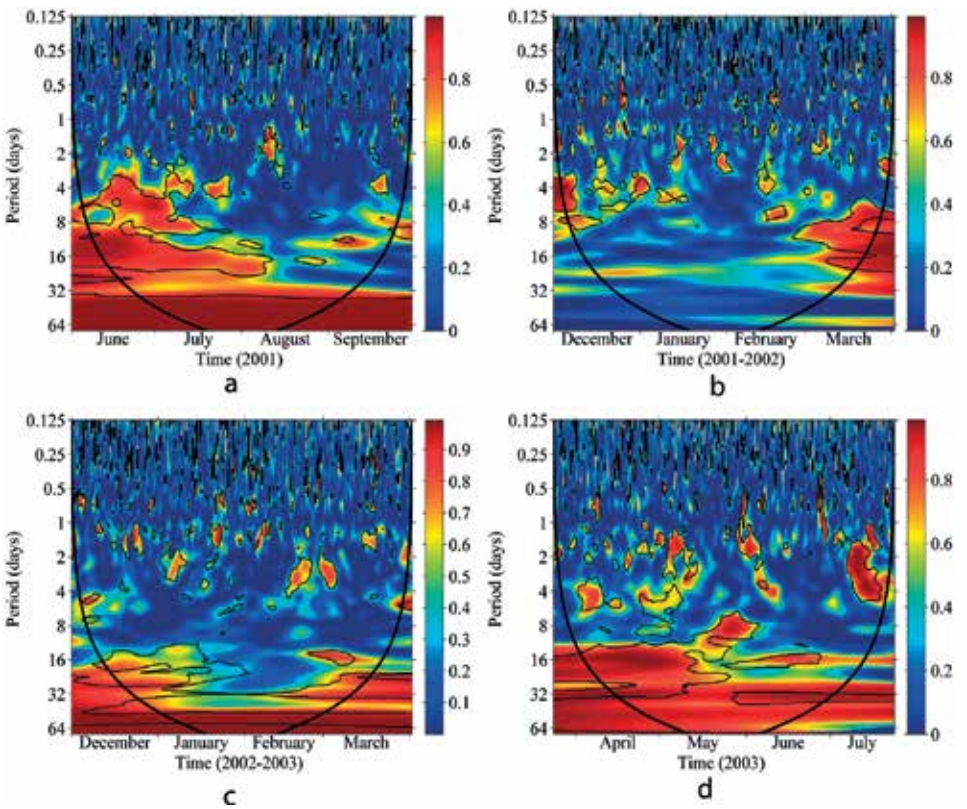


Figure 14. Partial wavelet coherence of river discharge versus salinity flux for the period (a) Jun-Sep 2001, (b) Dec 2001 - Mar 2002, (c) Dec 2002 - Mar 2003, (d) Apr-Jul 2003.

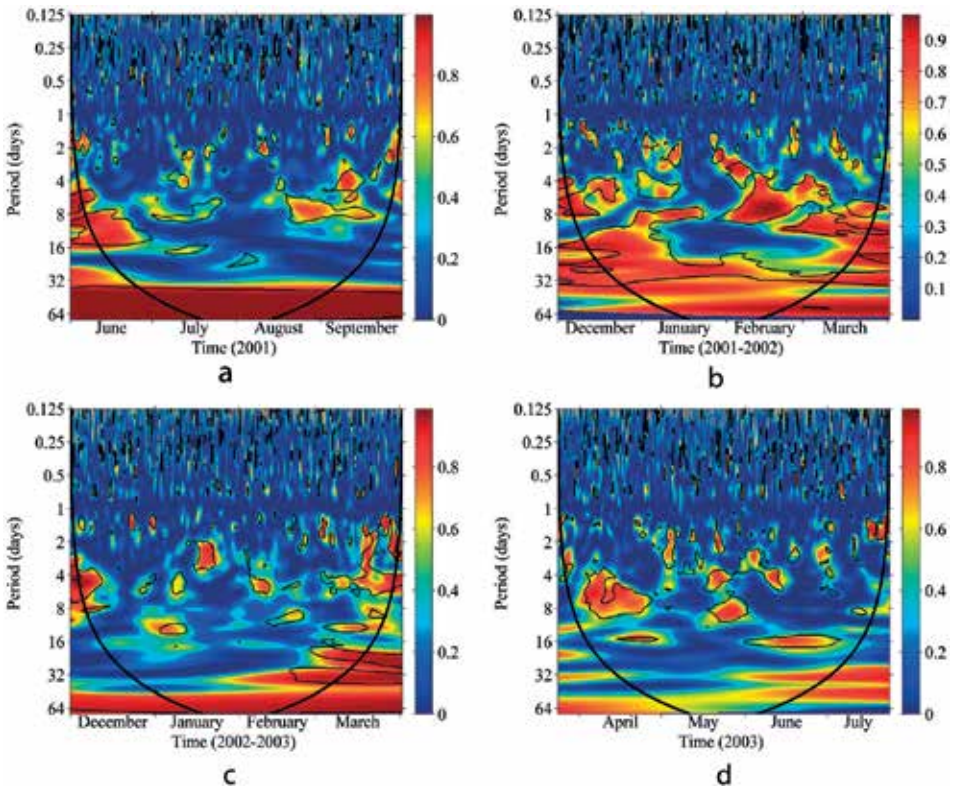


Figure 15. Partial wavelet coherence of atmospheric pressure versus salinity flux for the period (a) Jun-Sep 2001, (b) Dec 2001 - Mar 2002, (c) Dec 2002 - Mar 2003, (d) Apr-Jul 2003.

seasons and two winter seasons were analyzed separately in order to visualize the relationships that might otherwise be lost in a long-term analysis. The analysis of 2002 summer season was done before and will be compared with the current analysis [16].

3.3.1. PWC between the tide level and the seawater flux

Figure 13 shows extensive statistically significant coherence at the 5% level in all the seasons, indicating the existence of relationship between observed tides and salinity transport. The center of power sections focuses on periods 0.5 and 1 day. Tides have a positive impact on seawater flux over the periods 0.5 and 1 day throughout the year. This study reinforces previous conclusion that short-term salinity transport is highly influenced by tides [16].

3.3.2. PWC between the river discharge and the seawater flux

PWC between the river discharge and the seawater flux, after controlling for other forcing variables, shows statistically significant in-phase relationship in the 3–16-day period in all the seasons analyzed (**Figure 14)**. The 2001 summer (June and July) and spring/summer 2003 (April–June) show a significant continuous coherence between the river discharge and the seawater flux in the 16-day period band. In July 2003, though the coherence in the 16-day band

http://dx.doi.org/10.5772/intechopen.75177

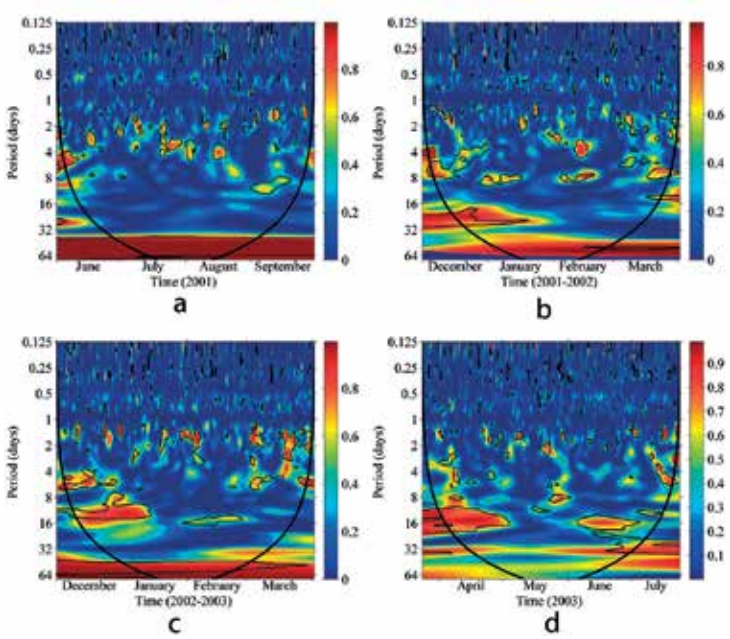


Figure 16. Partial wavelet coherence of East–West (*wx*) wind velocity component versus salinity flux for the period (a) Jun-Sep 2001, (b) Dec 2001 - Mar 2002, (c) Dec 2002 - Mar 2003, (d) Apr-Jul 2003.

is high, it is not statistically significant. The winter periods also showed some significant correlation between the river discharge and the seawater flux. 2001–2002 winters have coherences, which are discontinuous and occur at irregular intervals. March 2002 exhibits a significant continuous coherence in the 4–24-day period band. 2002–2003 winters have statistically significant coherences that occur continuously in the 16–32-day period band. This shows the existence of the relationship between the river discharge and the seawater flux, which generally coincides with an increased river discharge. The effect of increased river flows due to rain, typhoon events, and water releases upstream is clearly shown.

3.3.3. PWC between the atmospheric pressure and the seawater flux

During summer, a discontinuous and irregular statistically significant relationship between the atmospheric pressure and the seawater flux exists in the 4–16-day period band, indicating the existence of an on-and-off correlation between the tide level and the seawater flux (**Figure 15**). During winter, a continuous relationship exists in the 16–32-day period band. WTC and XWT show that the atmospheric pressure's influence on the flux of seawater is not stable, implying that it is short-lived and has a weak influence on seawater flux (**Figure 10**). However, PWC shows that the atmospheric pressure sometimes influences the seawater flux.

3.3.4. PWC between the wind velocity and the seawater flux

Figures 16 and **17** show partial wavelet coherence between the wind velocity and the seawater flux. An unstable relationship between wind vectors and seawater flux is exhibited. The East–West (*wx*) wind velocity component and seawater flux have a discontinuous and irregular

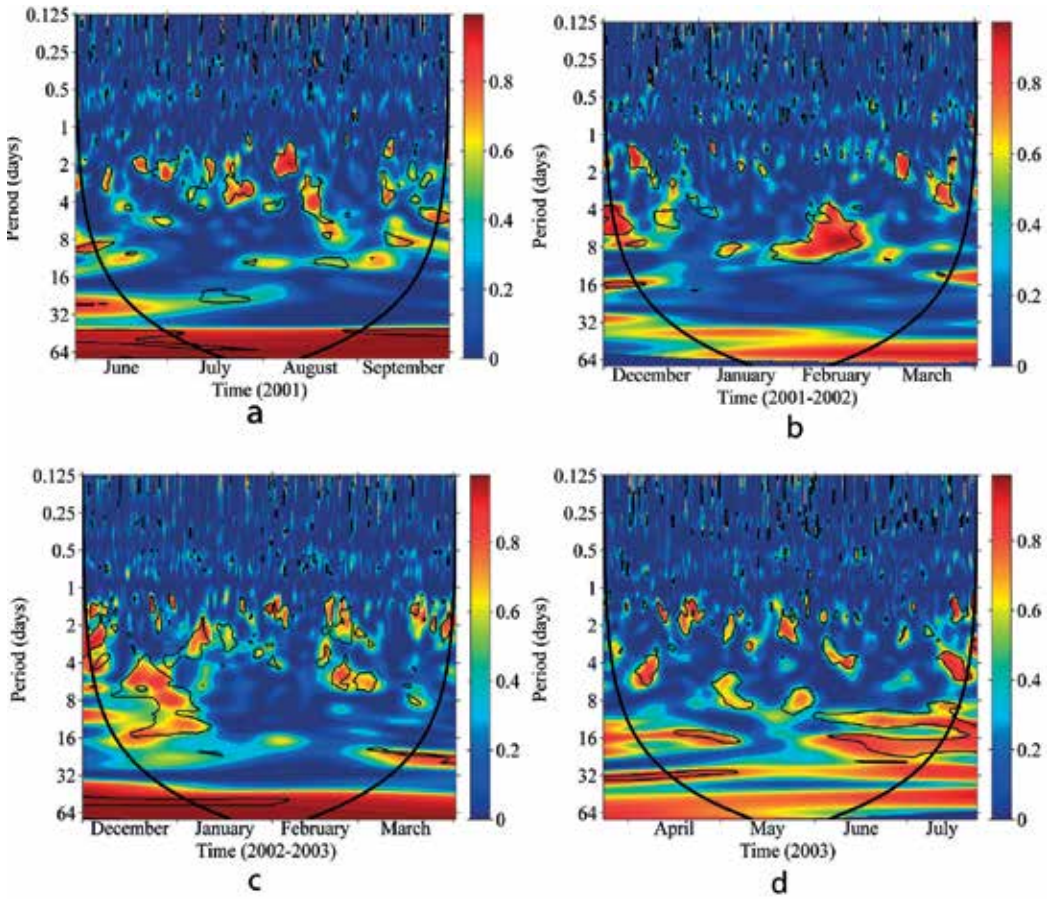


Figure 17. Partial wavelet coherence of North–South (*wy*) wind velocity component versus salinity flux for the period (a) Jun-Sep 2001, (b) Dec 2001 - Mar 2002, (c) Dec 2002 - Mar 2003, (d) Apr-Jul 2003).

relationship, mainly in the periods between 2 and 8 days. A continuous month-long relationship exists in the 8–16-day period band. The statistically significant month-long power sections for the North–South (*wy*) wind velocity component exist for 2001–2002 winters, 2002–2003 winters, and 2003 summer. The statistically significant correlation between the North–South (*wy*) wind velocity component and seawater flux exists mainly in the 2–16-day period band. Short-term oscillations are irregular and short-lived.

4. Conclusion(s)

This study explored the usefulness of continuous wavelet analysis in the investigation of salinity intrusion. The study summarized CWT, WTC, XWT, and PWC approaches and applied them in the analysis of the impact of forcing variables on the seawater flux in Sakai Channel. The study

revealed fundamental characteristics in the variation of forcing parameters and seawater flux, as well as their interactions. The only constraint in this study was a high computation time due to 1000 Monte Carlo simulation runs.

The CWT results show that the seawater flux and the tide level have regular oscillations in the 12-h and 1-day period band, indicating that the influence of astronomical tides is dominant. River discharge from the Hii River does not exhibit any periodical variations due to the irregularity of precipitation and the controlled release from upstream reservoirs. Atmospheric pressure exhibits a continuous high power (lasting over a month) with a period range from 16-day to 1-year. East–West (*wx*) and North–South (*wy*) wind velocity components show irregular oscillations with periods between 2 and 16 days.

WTC, XWT, and PWC revealed the influence of tide level, river discharge, atmospheric pressure, and wind velocity on seawater flux. WTC, XWT, and PWC showed that tides are consistently influential on the seawater flux in the 0.5- and 1-day period band. River discharge influenced seawater flux after heavy rains or water releases from upstream reservoirs. Atmospheric pressure and wind velocity occasionally influence seawater flux at Nakaura Watergate and may have an indirect influence on salinity transport through their effect on sea surface elevation. High drops of atmospheric pressure occasionally resulted in an increased tide level. This study reiterated the importance of tides in the transport of seawater in and out of Lakes Shinji and Nakaumi.

To conclude, the wavelet analysis of seawater intrusion studies proved useful. Wavelet coherence is helpful in the study of relationships between two time series. Partial wavelet coherence reveals the relationship between two time series after removing the effect of other time series. This is very useful when a dependent variable is under the influence of two or more variables. Wavelet analysis performs spectral analysis in frequency-time domain, revealing time-varying relationships across frequencies.

Acknowledgements

Shimane Prefectural Institute of Public Health and Environmental Science provided the salinity, water temperature, and current velocity data used in this study.

Author details

Edwin Muchebve[1]*, Yoshiyuki Nakamura[1] and Hiroshi Kamiya[2]

*Address all correspondence to: muchebve-edwin-tg@ynu.jp

1 Graduate School of Urban Innovation, Yokohama National University, Yokohama, Japan

2 Research Centre for Coastal Lagoon Environments, Shimane University, Matsue, Japan

References

[1] Kumar P, Foufoula-Georgiou E. Wavelet analysis for geophysical applications. Reviews of Geophysics. 1997;**35**(4):385

[2] Flinchem EP, Jay DA. An introduction to wavelet transform tidal analysis methods. Estuarine, Coastal and Shelf Science. 2000 Aug;**51**(2):177-200

[3] Aguiar-Conraria L, Soares MJ. The continuous wavelet transform: Moving beyond uni- and bivariate analysis. Journal of Economic Surveys. 2014;**28**(2):344-375

[4] Torrence C, Compo GP. A practical guide to wavelet analysis. Bulletin of the American Meteorological Society. 1998 Jan;**79**(1):61-78

[5] Liu Y, San Liang X, Weisberg RH. Rectification of the bias in the wavelet power spectrum. Journal of Atmospheric and Oceanic Technology. 2007 Dec;**24**(12):2093-2102

[6] Parmar KS, Bhardwaj R. Wavelet and statistical analysis of river water quality parameters. Applied Mathematics and Computation. 2013 Jun;**219**(20):10172-10182

[7] Shaharuddin M, Zaharim A, Nor MJM, Karim OA, Sopian K. Using wavelet transform on suspended particulate matter (PM10) and meteorological time series. In: Proceedings of the 12th WSEAS International Conference on Applied Mathematics. World Scientific and Engineering Academy and Society (WSEAS); 2007. pp. 179-183

[8] Zhang Q, Xu C-Y, Chen YD. Wavelet-based characterization of water level behaviors in the Pearl River estuary, China. Stochastic Environmental Research and Risk Assessment. 2009 Jan 13;**24**(1):81-92

[9] Sovi A, Poto K, Serši D, Kuspili N. Wavelet analysis of hydrological signals on an example of the River Sava. In: MIPRO, 2012 Proceedings of the 35th International Convention; Opatija, Croatia: IEEE; 2012. pp. 1223-1228

[10] Somoza RD, Pereira ES, Novo EML, Rennó CD. A water level relationship between consecutive gauge stations along Solimoes/Amazonas main channel: A wavelet approach. 2013 Oct;**178**:53-62

[11] Briciu A-E. Wavelet analysis of lunar semidiurnal tidal influence on selected inland rivers across the globe. Scientific Reports. 2014 Jan;**4**:4193

[12] Briciu A-E. Supplementary Information—Wavelet analysis of lunar semidiurnal tidal influence on selected inland rivers across the globe. Scientific Reports. 2014;**4**:4193

[13] Lim Y-H, Lye LM. Wavelet analysis of tide-affected low streamflows series. Journal of Data Science. 2004;**2**:149-163

[14] Ideião SMA, Santos CAG. Analysis of precipitation time series using the wavelet transform. Sociedade & Natureza. 2005 May;**1**(1):736-745

[15] Liu B, Yan S, Chen X, Lian Y, Xin Y. Wavelet analysis of the dynamic characteristics of saltwater intrusion—A case study in the Pearl River Estuary of China. Ocean and Coastal Management. 2014 Jul;**95**:81-92

[16] Muchebve E, Nakamura Y, Suzuki T, Kamiya H. Analysis of the dynamic characteristics of seawater intrusion using partial wavelet coherence: A case study at Nakaura Watergate, Japan. Stochastic Environmental Research and Risk Assessment. 2016 Dec 31;**30**(8):2143-2154

[17] Jevrejeva S, Moore JC, Grinsted A. Influence of the Arctic oscillation and El Niño-Southern oscillation (ENSO) on ice conditions in the Baltic Sea: The wavelet approach. Journal of Geophysical Research—Atmospheres. 2003 Nov 16;**108**(D21):4677

[18] Grinsted A, Moore JC, Jevrejeva S. Application of the cross wavelet transform and wavelet coherence to geophysical time series. Nonlinear Processes in Geophysics. 2004;**11**(5/6): 561-566

[19] Veleda D, Montagne R, Araujo M. Cross-wavelet bias corrected by normalizing scales. Journal of Atmospheric and Oceanic Technology. 2012;**29**(9):1401-1408

[20] Ng EKW, Chan JCL. Geophysical applications of partial wavelet coherence and multiple wavelet coherence. Journal of Atmospheric and Oceanic Technology. 2012;**29**(12):1845-1853

[21] Mandler M, Scharnagl M. Money growth and consumer price inflation in the euro area: A wavelet analysis. Discussion Paper. Deutsche Bundesbank; 2014. Report No.: 33/2014

[22] Nakata K, Horiguchi F, Yamamuro M. Model study of Lakes Shinji and Nakaumi—A coupled coastal lagoon system. Journal of Marine Systems. 2000 Oct;**26**(2):145-169

[23] Okuda S. Water movement and physical environments in brackish lakes—Shinjiko and Nakaumi. In: International Seminar on Restoration of Damaged Lagoon Environments; Matsue, Japan; 2004. pp. 1-6

[24] Vaz N, Lencarte Silva JD, Dias JM. Salt fluxes in a complex river mouth system of Portugal. PLoS One. 2012 Oct 10;**7**(10):e47349

[25] Moser GAO, Gianesella SMF, Barrera-Alba JJ, Bérgamo AL, Saldanha-Corrêa FMP, De Miranda LB, et al. Instantaneous transport of salt, nutrients, suspended matter and Chlorophyll-a in the tropical estuarine system of Santos. Brazilian Journal of Oceanography. 2005;**53**(3/4):115-127

[26] Klinger BA. Density of Seawater [Internet]. p. 4. Available from: http://mason.gmu.edu/~bklinger/seawater.pdf

[27] De Sousa RMD. Carbon Prices. Dynamic analysis of European and Californian markets [doctoral dissertation]. Maio: Universidade Nova de Lisboa; 2014

Chapter 12

Wavelets and LPG-PCA for Image Denoising

Mourad Talbi and Med Salim Bouhlel

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/intechopen.74453

Abstract

In this chapter, a new image denoising approach is proposed. It combines two image denoising techniques. The first one is based on a wavelet transform (WT), and the second one is a two-stage image denoising by PCA (principal component analysis) with LPG (local pixel grouping). In this proposed approach, we first apply the first technique to the noisy image in order to obtain the first estimation version of the clean image. Then, we estimate the noise-level from the noisy image. This estimate is obtained by applying the third technique of noise estimation from noisy images. The third step of the proposed approach consists in using the first estimation of the clean image, the noisy image, and the estimate of the noise-level as inputs of the second image denoising system (LPG-PCA). A comparative study of the proposed technique and the two others denoising technique (one is based on WT and and the second is based on LPG-PCA), is performed. This comparative study used a number of noisy images, and the obtained results from PSNR (peak signal-to-noise ratio) and SSIM (structural similarity) computations show that the proposed approach outperforms the two other denoising approaches (the first one is based on WT and the second one is based on LPG-PCA).

Keywords: image denoising, wavelet transform, noise estimation, LPG-PCA

1. Introduction

In the image acquisition process, the noise will be inevitably introduced so denoising is a necessary step for ameliorating the image quality. As a primary low-level image processing, noise suppression has been extensively studied, and numerous denoising approaches have been proposed, from the earlier frequency domain denoising approaches and smoothing filters [1] to the lately developed wavelet [2–11], curvelet- [12] and ridgelet- [13] based approaches, sparse representation [14] and K-SVD approaches [15], shape-adaptive transform [16], bilateral filtering [17, 18], nonlocal mean-based techniques [19, 20], and nonlocal collaborative filtering

[21]. With the quick development of modern digital imaging devices and their increasingly broad applications in our daily life, there are rising necessities of new denoising techniques for higher quality of image. The WT (wavelet transform) [22] proved its effectiveness in noise cancelation [2–11]. This transform decomposes the input signal into multiple scales which represent different time-frequency components of the original signal. At each scale, some operations, such as statistical modeling [4–6] and thresholding [2, 3], can be applied for canceling noise. Noise reduction is performed by transforming back the processed wavelet coefficients into spatial domain. Late development of WT-based denoising techniques includes ridgelet- and curvelet-based techniques [12, 13] for line structure conservation. Despite WT proved its effectiveness in denoising, it uses a fixed wavelet basis (with translation and dilation) for representing image. However, for natural images, a rich amount of different local structural patterns exists and therefore cannot be well represented by using just one fixed wavelet basis. Consequently, WT-based techniques can generate many visual artifacts in the denoising output. To overcome the problem of WT, in [23], Muresan and Parks proposed a spatially adaptive principal component analysis (PCA)-based denoising technique, which computes the locally fitted basis for transforming the image. In [15, 16], Elad and Aharon proposed K-SVD-based denoising approach and sparse and redundant representation by training a highly over-complete dictionary. In [16], Foi et al. applied a shape-adaptive discrete cosine transform (DCT) to the neighborhood, which can attain very sparse representation of the image and consequently lead to efficient denoising. All these approaches proved better denoising performance than classical WT-based denoising techniques. The NLM (nonlocal means) schemes used a very different philosophy from the above approaches in noise cancelation. The NLM idea can be traced back to [24], where the similar image pixels are averaged according to their intensity distance. Similar ideas were used in the bilateral filtering schemes [17, 18], where both the spatial and intensity similarities are exploited for pixel averaging. The NLM denoising framework was well established in [19]. In the image, each pixel is estimated as the weighted average of all the pixels and the weights are determined by the similarity between the pixels. This scheme was improved in [20], where the pair-wise hypothesis testing was used in the NLM estimation. Inspired from the success of NLM schemes, Dabov et al. [21] proposed a collaborative image denoising scheme by sparse 3D transform and patch matching. They look for similar blocks in the image by using block matching and grouped these blocks into a 3D cube. Then, a sparse 3D transform was applied to that cube, and noise was canceled by Wiener filtering in the transformed domain. The so-called BM3D approach attains remarkable denoising results, yet its implementation is a little complex. Lei Zhang et al. [25] have presented an efficient PCA-based denoising approach with local pixel grouping (LPG). PCA is a classical de-correlation technique in statistical signal processing, and it is pervasively used in dimensionality reduction and pattern, etc. [26]. The original dataset is transformed into PCA domain, and only the different most significant principal components are conserved. Consequently, trivial information and noise can be eliminated. In [23], a PCA-based scheme was proposed for image denoising by using a moving window for computing the local statistics, from which the local PCA transformation matrix was estimated. This technique applies PCA directly to the noisy image without data selection, and much residual noise and visual artifacts appear in the denoised image. In the LPG-PCA-based technique, Lei Zhang et al. [25] modeled a pixel and its nearest neighbors as a vector variable. The training samples of this variable are chosen by grouping the pixels with similar local spatial structures to the underlying one in the

local window. With such an LPG technique, the local statistics of the variables can be accurately calculated so that the image edge structures can be well conserved after shrinkage in the PCA domain for noise suppression. The LPG-PCA scheme proposed in [25] has two stages where the first stage yields an initial image estimation by eliminating the most of the noise and the second stage will further refine the first stage output. The two stages have the same procedures except for the noise-level parameter. Since the noise is significantly reduced in the first stage, the LPG precision will be much improved in the second stage so that the final denoised image is visually much better. When compared with WT which uses a fixed basis function for decomposing the noisy image, the proposed LPG-PCA approach is a spatially adaptive image representation so that it can better characterize the image local structures. When compared with BM3D and NLM approaches, the LPG-PCA-based technique proposed in [25] can use a relatively small local window to group the similar pixels for PCA training, yet it yields competitive results with state-of-the-art BM3D algorithm. In this paper we propose a new image denoising approach which combines the dual-tree discrete wavelet transform (DT-DWT)-based denoising approach [12] and the two-stage image denoising technique by PCA with local pixel grouping (LPG) [25]. To evaluate this proposed technique, we have compared it to the two techniques (the DT-DWT-based denoising technique [12] and LPG-PCA [25]). This comparison is based on PSNR and SSIM computation. In the rest of this paper, we first will deal with PCA. Then, we will be interested in DT-DWT [12]. After that we will deal with noise-level estimation from the noisy image proposed in [27, 28]. Then, we will present the two denoising techniques proposed in [12, 25]. After that we will detail the proposed image denoising technique, and finally we will give results and evaluation.

2. Principal component analysis (PCA)

Let

$$X = \begin{bmatrix} x_1^1 & x_1^2 & \dots & x_1^n \\ x_2^1 & x_2^2 & \dots & x_2^n \\ . & . & & . \\ x_n^1 & x_n^2 & & x_n^n \end{bmatrix} \tag{1}$$

$$X_i = \left[x_1^1 x_1^2 \dots x_1^n\right] \tag{2}$$

representing the sample vector of x_i. The mean value of X_i is computed as follows:

$$\mu_i = \frac{1}{n}\sum_{j=1}^{n} X_i(j). \tag{3}$$

And then, the sample vector is centralized as follows:

$$\overline{X}_i = X_i - \mu_i = \left[\overline{x}_i^1 \overline{x}_i^2 \dots \overline{x}_i^n\right] \tag{4}$$

with $\overline{x}_i^j = x_i^j - \mu_i$. Accordingly, the centralized matrix of X is expressed as follows:

$$\overline{X} = X_i - \mu_i = \left[X_i^T\, X_2^T \ldots \bar{x}_m^T\right]^T. \tag{5}$$

Finally, the covariance matrix of the centralized dataset is computed as follows:

$$\Omega = \frac{1}{n}\overline{X}\overline{X}^T. \tag{6}$$

The PCA aim consists in finding an orthonormal transformation matrix P in order to de-correlate $\overline{X}$, i.e., $\overline{Y} = P\overline{X}$, so that the matrix of covariance of $\overline{Y}$ is diagonal. Since the covariance matrix Ω is symmetrical; therefore, it can be expressed as follows:

$$\Omega = \phi\Lambda\phi^T \tag{7}$$

where $\Lambda = \text{diag}\{\lambda_1, \lambda_2, \ldots, \lambda_m\}$ is the diagonal eigenvalue matrix with $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_m$ and $\phi = \left[\phi_1\, \phi_2 \ldots \phi_m\right]$ represents the $m \times m$ orthonormal eigenvector matrix. The terms $\lambda_1, \lambda_2, \ldots, \lambda_m$ and $\phi_1, \phi_2, \ldots, \phi_m$are, respectively, the eigenvalues and the eigenvectors of Ω. By setting the matrix P as follow:

$$P = \phi^T, \tag{8}$$

$\overline{X}$ can be de-correlated, i.e., $\Lambda = \frac{1}{n}\frac{\overline{Y}}{\overline{Y}^T}$ and$\overline{Y} = P\overline{X}$. An interesting property of PCA is that it fully de-correlates the original dataset $\overline{X}$. In general, the signal energy will concentrate on a small subset of the PCA transformed dataset, whereas the noise energy will evenly spread over the whole dataset. Consequently, the noise and signal can be better distinguished in the domain of PCA.

3. LPG-PCA denoising algorithm

3.1. Modeling of spatially adaptive PCA denoising

In [25] and in previous literature, the noise v degrading the original image I is supposed to be white and additive with standard deviation σ and zero mean, and the noisy image,I_v, is expressed as follows:

$$I_v = I + v \tag{9}$$

Both noise v and image I are supposed to be uncorrelated. The purpose of image denoising consists in estimating the clean image I from I_v, and the estimate is denoted by $\widehat{I}$. The latter is expected to be as close as possible to the original image, I. Two quantities describe an image pixel. Those quantities are its intensity and the spatial location. However, the image local structure is represented as a set of neighboring pixels at different intensity levels. As most of the semantic information of an image is conveyed by its edge structures, edge conservation is highly required in denoising of this image. In [25], the pixel and its nearest neighbors were

modeled as a vector variable and perform denoising on the vector instead of the single pixel. According to **Figure 1**, for an underlying pixel to be denoised, Lei Zhang et al. [25] set a $K \times K$ window centered on it, and denoted by $x = [x_1, \ldots, x_m]^T$, $m = K^2$, the vector contains all components within the window.

As the observed image is the original image degraded by the noise, they denote the noisy vector of x by [25]:

$$x_v = x + v \tag{10}$$

x where $v = [v_1 \ldots v_m]^T$, $x_k^v = x_k + v_k, k = 1, \ldots, m$, and $x_v = \left[x_1^v \ldots x_m^v\right]$. For estimating x from the noisy vector, x_v, they are viewed as (both noiseless and noisy) vector variables so that one can use the statistical techniques such as PCA. For canceling the noise from the noisy vector x_v by using PCA, a set of training samples of x_v is needed so that the covariance matrix of x_v and therefore the PCA transformation matrix can be computed. For this aim, Lei Zhang et al. [25] have used an $L \times L$ $(L > K)$ training block centered on x_v in order to find the training samples, as illustrated in **Figure 1**. The simplest manner consists in taking the pixels in each possible $K \times K$ block within the $L \times L$ training block as the samples of noisy variable x_v. In this way, for each component x_k^v of x_v, there are in total $(L - K + 1)^2$ training samples. Though, there can be very different blocks from the given central $K \times K$ block in the $L \times L$ training window, taking all the $K \times K$ blocks as the training samples of $x_v$will lead to inaccurate estimation of the matrix of covariance of x_v, which subsequently leads to inaccurate estimation of the PCA transformation matrix and finally results in much residual noise. Consequently, selecting and grouping the training samples that are similar to the central $K \times K$ block are required before image denoising by applying the PCA transform.

3.2. Local pixel grouping (LPG)

Grouping the training samples similar to the central $K \times K$ block in the $L \times L$ training window is certainly a problem of classification, and therefore different grouping techniques such as

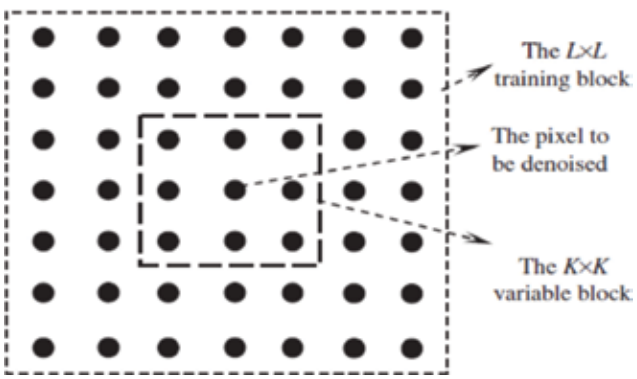


Figure 1. Illustration of the modeling of LPG-PCA-based denoising [25].

correlation based matching, block matching, K-means clustering, etc. can be used based on different criteria. The block matching-based technique may be the simplest but very efficient one, and it is used in [25] for LPG. There are totally $(L-K+1)^2$ possible training blocks of x_v in the $L \times L$ training window. We will denote $\overline{x}_0^v$ in the column sample vector which contains the pixels in the central $K \times K$ block, and denoted by $\overline{x}_i^v, i = 1, 2, \ldots, (L-K+1)^2 - 1$, the sample vectors correspond to the other blocks. Let $\overline{x}_i$ and $\overline{x}_0$ be, respectively, the associated noiseless sample vectors of $\overline{x}_i^v$ and $\overline{x}_0^v$. It can be simply computed that

$$e_i = \frac{1}{m}\sum_{k=1}^{m} \overline{x}_0^v(k) - \overline{x}_i^v(k)^2 \approx \frac{1}{m}\sum_{k=1}^{m} \overline{x}_0(k) - \overline{x}_0(k)^2 + 2\sigma^2 \tag{11}$$

In Eq. (11), the fact that noise v is white and uncorrelated with signal is used. With Eq. (11), if we have the following condition

$$e_i < T + 2\sigma^2 \tag{12}$$

where T designates a preset threshold, then we select $\overline{x}_i^v$ as a sample vector of x_v. Assume that n sample vectors of x_v are selected including the central vector $\overline{x}_0^v$. For the expression convenience, these sample vectors are denoted as $\overline{x}_0^v, \overline{x}_1^v, \ldots, \overline{x}_{n-1}^v$. The noiseless counterparts of those vectors are denoted as $\overline{x}_0, \overline{x}_1, \ldots, \overline{x}_{n-1}$, accordingly. Then, the training dataset for x_v is constituted by.

$$X_v = \left[\overline{x}_0^v, \overline{x}_1^v, \ldots, \overline{x}_{n-1}^v\right] \tag{13}$$

The noiseless counterpart of X_v is designated as $X = [\overline{x}_0, \overline{x}_1, \ldots, \overline{x}_{n-1}]$. To insure the existence of enough samples in calculating the PCA transformation matrix, $n$could not be too small. Practically speaking, it will be used in denoising at least $c \cdot m$ training samples of x_v, with $c = 8\tilde{}10$. That is to say that in case of $n < c \cdot m$, we will use the best $c \cdot m$-matched samples in PCA training. Often, the best $c \cdot m$-matched samples are robust for estimating the local statistics of image, and this operation makes the algorithm more stable for computing the PCA transformation matrix. The problem now is how to estimate from the noisy data X_v, the noiseless dataset X. Once this dataset X is estimated, the central block and therefore we can extract the central underlying pixel. Such procedure is applied to each pixel, and then the entire image I_v can be denoised. The LPG-PCA-based denoising is detailed in [25], and the denoising refinement in the second stage will be detailed in the next part of this paper.

3.3. Denoising refinement in the second stage

Most of the noise will be suppressed by employing the denoising procedures described in [25]. However, there is still much visually unpleasant residual noise in the denoised image. **Figure 2** shows an example of image denoising where (a) is the original image Cameraman, (b) the noisy version of it with $PSNR = 22.1$ dB and $\sigma = 20$, and (a) is the denoised image with $PSNR = 29.8$ dB by employing the LPG-PCA technique proposed in [25]. Despite the remarkable improvement of *PSNR*, one can still see much residual noise in the denoising output. There are mainly two reasons for the residual noise. First, because of the strong noise in the original dataset X_v, the covariance matrix $\Omega_{\overline{x}_v}$ is much noise degraded, which leads to

Figure 2. (a) Original image Cameraman, (b) corresponding noisy image ($PSNR = 22.1$ dB), (c) denoised image after the first round of the technique proposed in [25] ($PSNR = 29.8$ dB), and (d) denoised image after the second round of the proposed technique ($PSNR = 30.1$ dB) [25].

estimation bias of the PCA transformation matrix and therefore deteriorates the denoising performance; second, the strong noise in the original dataset will also lead to LPG errors, which therefore results in estimation bias of the covariance matrix $\Omega_{\overline{x}_v}$ or $\Omega\overline{x}$. Consequently, it is essential to further process the denoising output for a better image denoising. As the noise has been much canceled in the first round of LPG-PCA denoising, the LPG correctness and the estimation of $\Omega_{\overline{x}_v}$ or $\Omega\overline{x}$ can be much ameliorated with the denoised image. Consequently, the LPG-PCA denoising procedure for the second round for enhancing the denoising results.

According to this figure, we remark that the visual quality is much ameliorated after the second round of refinement. As shown in **Figure 3**, in the second round of LPG-PCA denoising technique [25], the noise-level should be updated.

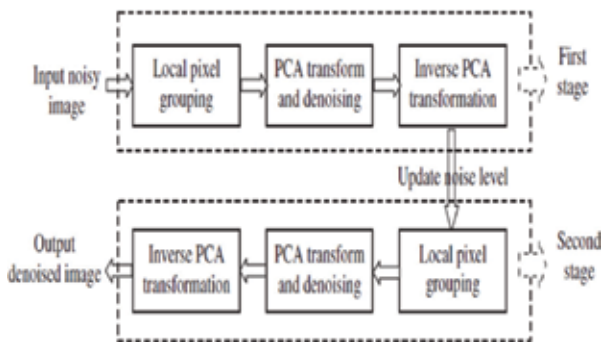


Figure 3. Flowchart of the two-stage LPG-PCA denoising technique proposed in [25].

Denote by $\widehat{I}$, the denoised version of the noisy image in the first stage. The $\widehat{I}$ can be expressed as

$$\widehat{I} = I + v_s \tag{14}$$

where v_s is the residual noise in the denoised image. The level estimation ofv_s is denoted by $\sigma_s = \sqrt{E\left[v_s^2\right]}$ and inputs it to the second round of LPG-PCA denoising algorithm. In [25], σ_s is estimated based on the difference between I_v and $\widehat{I}$. Let

$$\tilde{I} = I_v - \widehat{I} = v - v_s \tag{15}$$

We have $E\left[\tilde{I}^2\right] = E\left[v^2\right] + E\left[v_s^2\right] - 2E[v \cdot v_s] = \sigma^2 + \sigma_s^2 - 2E[v \cdot v_s]$. The v_s can be seen as the smoothed version of noise v, and it mainly contains the low-frequency component of v. Let $\tilde{v} = v - v_s$ be their difference, and $\tilde{v}$ mainly contains the high-frequency component of v. There is $E[v \cdot v_s] = E\left[\tilde{v} \cdot v_s\right] + E\left[v_s^2\right]$. Generally, compared to $E\left[v_s^2\right]$, $E[\tilde{v} \cdot v_s]$ is much smaller, and we can obtain the following approximation: $E[v \cdot v_s] \approx E\left[v_s^2\right] = \sigma_s^2$. Thus, from $E\left[\tilde{I}^2\right] = \sigma^2 + \sigma_s^2 - 2E(v \cdot v_s)$, we obtain

$$\sigma_s^2 \approx \sigma^2 - E\left[\tilde{I}^2\right] \tag{16}$$

In practice, v_s will include not only the residual noise but also the estimation error of noiseless image I. Consequently, in the implementation [25], of Lei Zhang et al. let

$$\sigma_s = C_s\sqrt{\sigma^2 - E\left[\tilde{I}^2\right]} \tag{17}$$

where C_s is a constant satisfying $C_s < 1$. In [25], Lei Zhang et al. found experimentally that setting C_s around 0.35 can lead to satisfying denoising results for most of the testing images. **Figure 2**d shows the denoised image ($PSNR = 30.1\ dB$) after the second round of the LPG-PCA denoising technique [25]. Although the *PSNR* is not too much ameliorated on this image, we can remark clearly that the visual quality is much ameliorated by efficiently eliminating the residual noise obtained from the first round of denoising.

4. The proposed image denoising technique

As previously mentioned, in this chapter, a new image denoising technique is proposed. It combines two denoising approaches. The first one is a dual-tree discrete wavelet (DT-DWT)-based denoising method [12], and the second one is a two-stage image denoising by PCA with LPG [25]. This proposed technique consists at the first step in applying the first denoising approach [12] to the noisy image in order to obtain the first estimation of the clean image (the

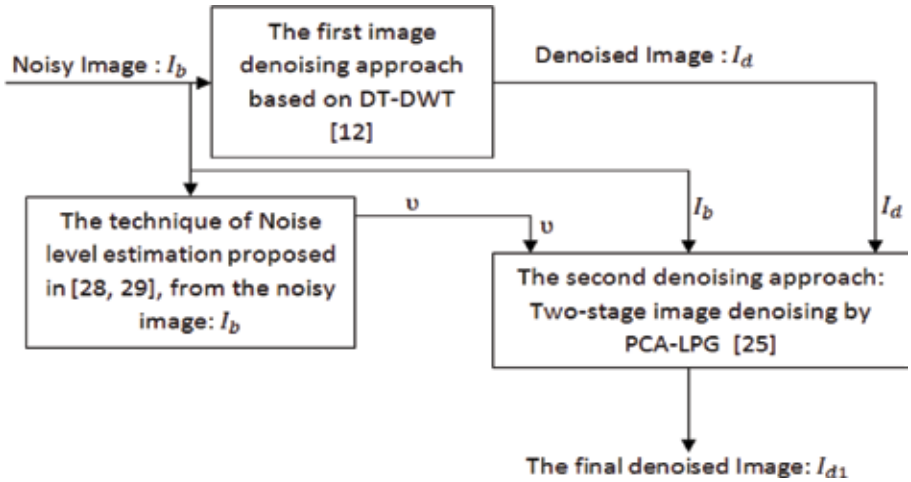


Figure 4. The block diagram of the proposed image denoising technique.

cleaned image). Then, we estimate the level of noise corrupting the clean image. The cleaned image, the noisy image, and the noise-level are used for applying the second approach which is two-stage image denoising by PCA with LPG [25]. **Figure 4** illustrates the block diagram of the proposed technique.

According to this figure, the first step of the proposed image denoising technique consists in applying the first denoising approach based on DT-DWT [12] to the noisy image, I_b, in order to obtain a first estimate of the clean image, I_d, and then estimates the noise-level, υ, from I_b. The noisy images I_b, I_d, and υ constitute the inputs of the second image denoising system proposed in [25, 27]. The output of this system and the overall proposed one are the final denoised image, I_{d1}. In the image denoising system (LPG-PCA denoising) proposed in [25, 27], Lei Zhang et al. have used the clean image, I, and the noise-level, υ, as the inputs of this system [27]. However, only the noisy image, I_b, is available, and for this raison, we have used in our proposed technique the denoising approach based on DT-DWT [12] in order to obtain a cleaned image, I_d, which is then used as a clean image, I. This clean image is one important input of the denoising system proposed by Lei Zhang et al. [27]. In the following two subsections, we will be interested in the first image denoising approach based on DT-DWT [12] and the technique of noise-level estimation proposed in [28, 29], from the noisy image, I_b.

5. The Hilbert transform

The Hilbert transform of a signal corresponds in Fourier plane to a filter with complex gain, $-i\,\mathrm{sign}(\gamma)$ [30]. This is corresponding to an impulse response $vp\left(\frac{1}{\pi t}\right)$ where vp is the principal value in Cauchy sense [30]. The analytic signal is then constructed as follows:

$$z(t) = x(t) + iH\{x(t)\} = x(t) + \frac{i}{\pi} vp \int_{-\infty}^{+\infty} \frac{x(s)}{t-s} ds \tag{18}$$

This analytic signal has only positive frequencies. The Hilbert transform of a real signal is also real. Instead of considering the Hilbert transform of the wavelet (which is defined through the associated filters), we can consider the Hilbert transform of the signal, and the analysis is performed with initial wavelet because we have $\left\langle f, H\psi_{a,t} \right\rangle = \left\langle Hf, \psi_{a,t} \right\rangle$ [30]. The latter equality is justified by the fact that the Hilbert transform is considered as a linear filter [30]. Therefore, we have the following scheme: let $X(n)$ be the signal to be analyzed with real wavelet by using the Mallat algorithm in order to obtain the wavelet coefficients, $d_1(j,k)$. Then, we analyze $HX(n)$ with the same wavelet, and we obtain the coefficients $d_2(j,k)$. Then, we construct the complex coefficients: $d_{complex} = d_1(j,k) + i\, d_2(j,k)$. As follows, the magnitude of those coefficients is named Hilbert magnitude. The drawbacks of this method are as follows: The support of the Hilbert transform of a wavelet having a compact support is infinite. There is a computing disadvantage because the cost of two wavelet transforms plus the Hilbert Transform. Theoretically speaking, it is possible to limit the drawback of the support of the Hilbert transform of the wavelet by using an approximate of the Hilbert transform. However, this approximation cannot be optimized for all scales [30]. One solution of this problem has been proposed by Kingsbury: the dual tree [30].

6. Dual-tree complex wavelet transform

The dual tree complex wavelet (DT-CWT) permits to make signal analysis by using two different trees of DWT, with filters selected in such manner to obtain approximately a signal decomposition using analytic wavelet [30]. **Figure 5** shows a tree of DT-CWT, using two different filter banks: h_1 and g_1 are high-pass filters of the first and second trees, and h_0 and g_0 are low-pass filters of the same two trees [30]. The first tree gives the coefficients of the real part, $d_r(j,k)$, and the second tree gives those of the imaginary part, $d_i(j,k)$. After that, we construct the complex coefficients $d_{\text{complex}}(j,k) = d_r(j,k) + i\ d_i(j,k)$. The magnitude of those coefficients is named dual-tree magnitude [30].

This Q-shift dual-tree complex wavelet transform (**Figure 5**) is in 1D. Synthesis of the filters adapted to this structure has been performed by many research works. Particularly, Kingsbury [30] proposed some filters named Q-shift. In [30], some filters are employed, and their utilization is equivalent to the signal analysis by wavelets illustrated in **Figure 6**.

We can see in this figure that the wavelet corresponding to the imaginary part tree is very near to the Hilbert transform of the wavelet corresponding to the real part tree [30]. Finally, the utilization of this structure requires an operation of prefiltering; it means that the filters used in the first step are not the same as those used in the next step. The advantages of this method compared to the simple Hilbert transform (Section 5) are [30]:

http://dx.doi.org/10.5772/intechopen.74453

- A lower computation cost (Two DWT),
- An approximate of the Hilbert transform, is optimized for each scale,
- The possibility of an exact reconstruction is preserved.

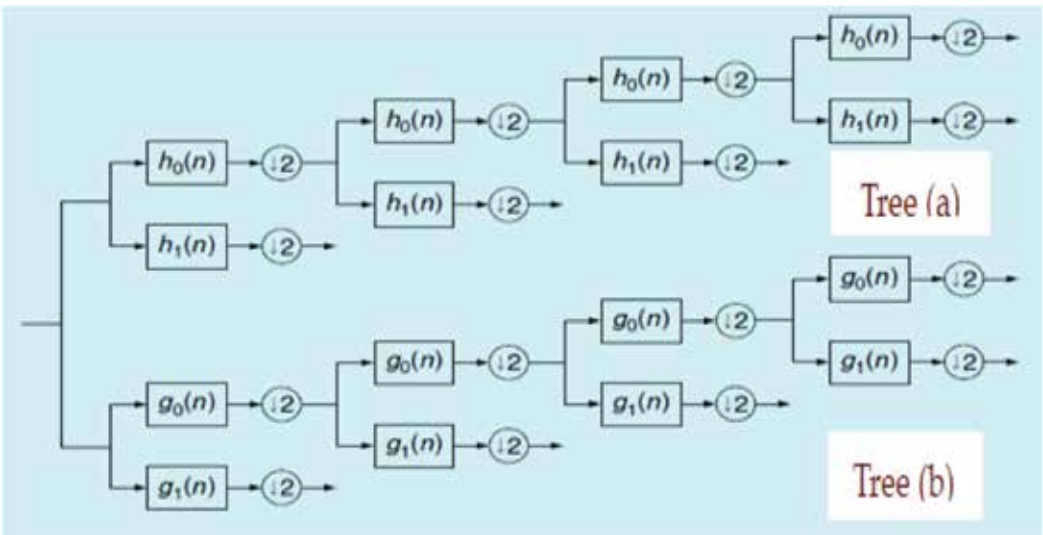


Figure 5. Dual tree of real filters for the Q-shift DT-CWT, giving real and imaginary parts of complex coefficients from tree (a) and tree (b), respectively [30].

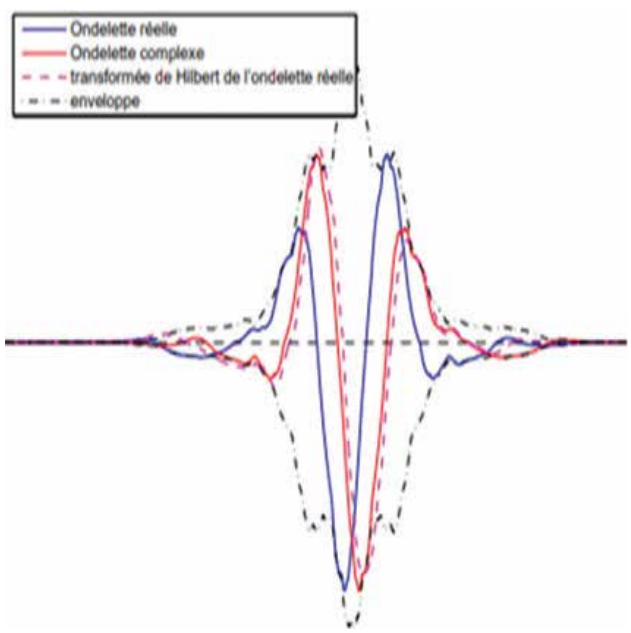


Figure 6. Q-shift wavelet obtained with filters Antonini [30].

The principal drawback of the DT-CWT is the non-possibility of the use of the well-known wavelets of the DWT (Daubechies wavelet, Spline, etc.) and therefore the non-possibility to choose the number of vanishing moments (all the Q-shift filter gives wavelets with two vanishing moments).

6.1. 2D DT-CWT

To explain how the DT-CWT produces oriented wavelets, consider the 2D wavelet $\psi(x,y) = \psi(x) \cdot \psi(y)$ associated with the row-column implementation of the wavelet transform, where $\psi(x)$ is a complex wavelet (approximately analytic) and is expressed as follows [31]:

$$\psi(x) = \psi_h(x) + i\,\psi_g(x). \tag{19}$$

Therefore, we obtain the following expression of $\psi(x,y)$**:**

$$\psi(x,y) = \left[\psi_h(x) + i\,\psi_g(x)\right]\left[\psi_h(y) + i\,\psi_g(y)\right] = \psi_h(x)\psi_h(y) - \psi_g(x)\psi_g(y) + i\left[\psi_g(x)\psi_h(y) + \psi_h(x)\psi_g(y)\right] \tag{20}$$

The following idealized diagram (**Figure 7**) illustrates the Fourier spectrum support of this complex wavelet [31].

Since the (approximately) 1D wavelet spectrum is supported on just one side of the frequency axis, the complex 2D wavelet ($\psi(x,y)$) spectrum is supported in just one quadrant of the 2D frequency plane. That is why the complex 2D wavelet, $\psi(x,y)$, is oriented. If the real part of this complex wavelet is taken, then the sum of two separable wavelets is obtained:

$$\text{Real Part}\,\{\psi(x,y)\} = \psi_h(x)\psi_h(y) - \psi_g(x)\psi_g(y). \tag{21}$$

Since the real function spectrum should be symmetric with respect to the origin, then the spectrum of this real wavelet is supported in two quadrants of the 2D frequency plane (**Figure 8**).

Unlike the real separable wavelet, the support of the spectrum of this real wavelet has not the checkerboard artefact and consequently this real wavelet (illustrated in the second panel of **Figure 11**), is oriented at -45°. It is deserving mentioning that this construction is depending on $\psi(x) = \psi_h(x) + i\,\psi_g(x)$ being (approximately) analytic or equivalently on $\psi_g(x)$ being

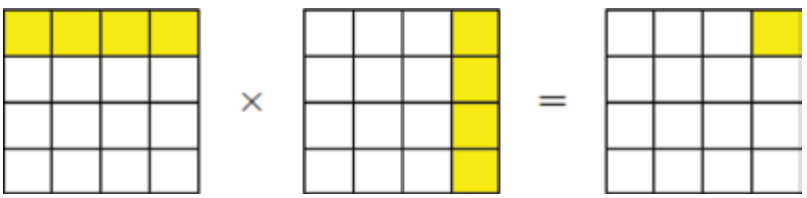


Figure 7. Idealized diagram illustrating the Fourier spectrum support of the complex wavelet, $\psi(x,y) = \psi(x) \cdot \psi(y)$ [31].

Figure 8. Idealized diagram illustrating the support of spectrum of this real wavelet, R*eal Part* $\{\psi(x,y)\}$ [31].

approximately the Hilbert transform of $\psi_h(x)$ **(**$\psi_g(x) \approx H\{\psi_h(x)\}$**)**. Note that $\psi_h(x)\psi_h(y)$ is the sub-band HH of a separable 2D real wavelet transform implemented employing the filters $\left\{h_0(n)^{0}, h_1(n)\right\}$. The term $\psi_g(x)\psi_g(y)$ is also the sub-band HH which is obtained from the application of a real separable wavelet transform. The latter is implemented by employing the filters $\left\{g_0(n)^{0}, g_1(n)\right\}$. To have a real 2D wavelet oriented at $+45^\circ$, we consider now the complex 2D wavelet $\psi_2(x,y) = \psi(x)\overline{\psi}(y)$ where $\overline{\psi}(y)$ is the complex conjugate of $\psi(y)$ and, as previously mentioned, $\psi(x)$ is approximately the analytic wavelet, $\psi_h(x) + i\,\psi_g(x)$. Therefore, we have

$$\begin{aligned}\psi_2(x,y) &= \left[\psi_h(x) + i\,\psi_g(x)\right]\left[\psi_h(y) - i\psi_g(y)\right] \\ &= \psi_h(x)\psi_h(y) + \psi_g(x)\psi_g(y) + i\left[\psi_g(x)\psi_h(y) + \psi_h(x)\psi_g(y)\right]\end{aligned} \tag{22}$$

The support in the 2D frequency plane of this complex wavelet spectrum is illustrated in **Figure 9**.

As above, the spectrum of the complex wavelet, $\psi_2(x,y)$, is supported in just one quadrant of the 2D frequency plane. If the real part of this complex wavelet is taken, then we have

$$\text{R}\textit{eal Part}\left\{\psi_2(x,y)\right\} = \psi_h(x)\psi_h(y) + \psi_g(x)\psi_g(y). \tag{23}$$

The spectrum of which is supported in two quadrants of the 2D frequency plane as illustrated in **Figure 10**.

Again, neither the wavelet nor the spectrum of this real wavelet has the spectrum of the checkerboard artifact. This real 2D wavelet is oriented at $+45^\circ$ as illustrated in the fifth panel of **Figure 11**. To have four more oriented real 2D wavelets, one can repeat this procedure on

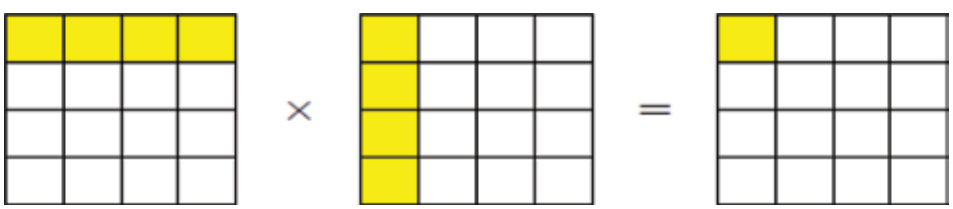


Figure 9. The idealized diagram in 2D frequency plane of the spectrum of this complex wavelet [31].

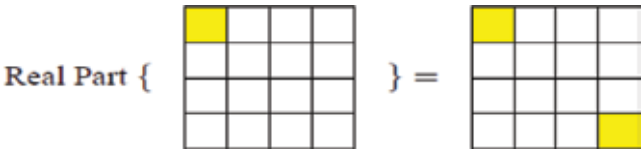


Figure 10. Idealized diagram in 2D frequency plane of the spectrum R*eal Part* $\{\psi_2(x,y)\}$ [31].

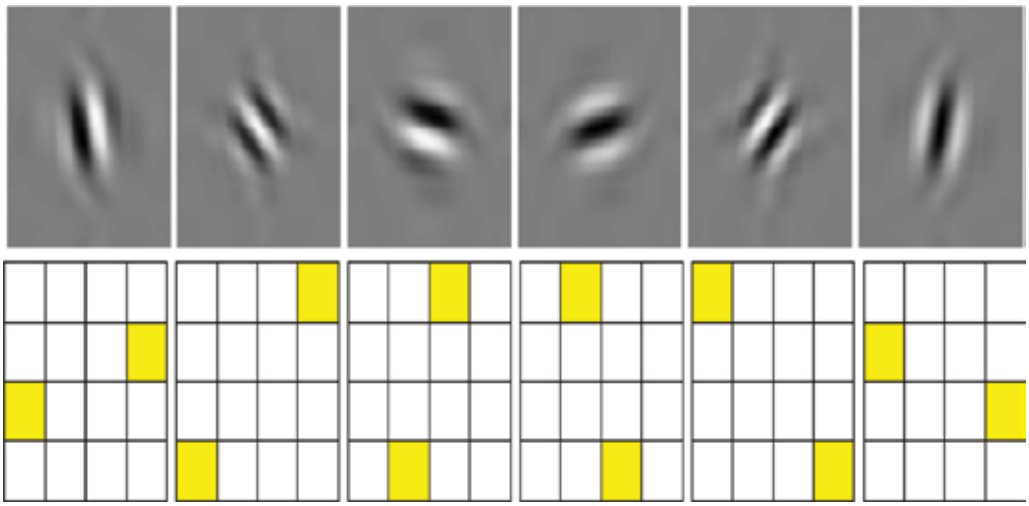


Figure 11. Typical wavelets associated with the real oriented 2D dual-tree wavelet transform. Top row illustrates the wavelets in the space domain: bottom row illustrates the (idealized) support of the Fourier spectrum of each wavelet in 2D frequency plane. The absence of the checkerboard phenomenon is observed in both frequency and spatial domains.

the complex wavelets expressed as follows: $\phi(x)\psi(y)$, $\psi(x)\phi(y)$, $\phi(x)\,\vec{\psi}\,(y)$, and $\psi(x)\overline{\phi}(y)$ where we have

$$\psi(x) = \psi_h(x) + i\,\psi_g(x) \tag{24}$$

$$\phi(x) = \phi_h(x) + i\,\phi_g(x) \tag{25}$$

By taking the real part of each of these wavelets, one can obtain four real oriented 2D wavelets. Moreover, the two already obtained in Eqs. (21) and (23). Precisely, we have six wavelets expressed as follows:

$$\psi_i(x,y) = \frac{1}{\sqrt{2}}\left(\psi_{1,i}(x,y) - \psi_{2,i}(x,y)\right) \tag{26}$$

$$\psi_{i+3}(x,y) = \frac{1}{\sqrt{2}}\left(\psi_{1,i}(x,y) + \psi_{2,i}(x,y)\right) \tag{27}$$

For $i = 1, 2, 3,$ the two separable 2-D wavelet bases are expressed as follow:

$$\psi_{1,1}(x,y) = \phi_h(x)\psi_h(y)^h, \psi_{2,1}(x,y) = \phi_g(x)\psi_g(y)^g, \tag{28}$$

$$\psi_{1,2}(x,y) = \psi_h(x)\phi_h(y)^h, \psi_{2,2}(x,y) = \psi_g(x)\phi_g(y)^g, \tag{29}$$

$$\psi_{1,3}(x,y) = \psi_h(x)\psi_h(y)^h, \psi_{2,3}(x,y) = \psi_g(x)\psi_g(y)^g, \tag{30}$$

The normalization factor $1/\sqrt{2}$ is used only so that the sum/difference operation constitutes an orthonormal operation. In **Figure 11** the six real oriented wavelets derived from a pair of typical wavelets satisfying $\psi_g(x) \approx H\{\psi_h(x)\}$ are illustrated. Compared to separable wavelets, these six non-separable wavelets succeed in isolating different orientations. Each of these six wavelets are aligned with a specific direction. Moreover, no checkerboard effect appears. In addition, they cover more distinct orientations than the separable wavelets obtained from the application of DWT. Moreover, since the sum/difference operation is orthonormal, the wavelet set is obtained from integer translates and dyadic dilations from a frame [31].

7. The technique of Noise-level estimation

In many image processing applications, the noise-level is an important parameter. For example, the performance of an image denoising technique can be much degraded due to the poor noise-level estimation. The most available denoising techniques simply supposed that the noise-level is known that largely prevents them from practical employment. Furthermore, even with the given true noise-level, those denoising techniques still cannot achieve the best performance, precisely for scenes with rich texture. Xinhao Liu et al. [28, 29] have proposed a technique of patch-based noise-level estimation, and they suggested that the noise-level parameter should be tuned according to the complexity of the scene. Their approach [28, 29] includes the process of selecting low-rank patches without high-frequency components from a single noisy image. Then, the noise-level was estimated from the selected patches employing principal component analysis. Because the exact noise-level does not always provide the best performance for non-blind denoising. Experiments prove that both the stability and precision are superior to the state-of-the-art noise-level estimation technique for different noise-levels and scenes.

8. Evaluation criteria

In this section, we will evaluate the three techniques which are the proposed image denoising techniques: the first image denoising approach based on DT-CWT [12] and the second denoising approach and the two-stage image denoising by principal component analysis with local pixel grouping [25]. This evaluation is based on the computation of PSNR and SSIM which are detailed in [32].

9. Results and discussion

In this work, we have applied the proposed image denoising technique, the first image denoising technique based on DT-CWT [12] and the second denoising technique and the two-stage image denoising by principal component analysis with local pixel grouping [25], on a number of digital images such as "House," "Lena," and "Cameraman." These images are degraded by additive white noise with different values of noise-level, σ. PSNR and SSIM values obtained from the application of the three mentioned techniques on the noisy images are listed in **Table 1**.

Technique	The first image denoising technique based on DT-DWT [12]	Two-stage image denoising by principal component analysis with local pixel grouping [25]: first stage	Two-stage image denoising by principal component analysis with local pixel grouping [25]: second stage	The proposed technique
House ($\sigma = 10$)	34.7138 (0.8778)	35.4 (0.9003)	35.6 (0.9012)	36.1223 (0.9130)
House ($\sigma = 20$)	31.6671 (0.8253)	31.8 (0.8084)	32.5 (0.8471)	33.0828 (0.8677)
House ($\sigma = 30$)	29.8494 (0.7877)	29.3 (0.7225)	30.4 (0.8185)	31.2095 (0.8393)
House ($\sigma = 40$)	28.5744 (0.8084)	27.3(0.6243)	28.9 (0.7902)	29.7344 (0.8084)
Lena ($\sigma = 10$)	33.6767(0.9170)	33.6 (0.9218)	33.7 (0.9243)	34.0765 (0.9271)
Lena ($\sigma = 20$)	30.0002 (0.8539)	29.5 (0.8346)	29.7 (0.8605)	30.5415 (0.8765)
Lena ($\sigma = 30$)	27.9859 (0.8016)	27.1 (0.7441)	27.6 (0.8066)	28.3595 (0.8292)
Lena ($\sigma = 40$)	26.6364 (0.7585)	25.4 (0.6597)	26.0 (0.7578)	26.8566 (0.7882)
Cameraman ($\sigma = 10$)	32.7481 (0.8989)	33.9 (0.9261)	34.1 (0.9356)	33.6141 (0.9241)
Cameraman ($\sigma = 20$)	28.9990 (0.8175)	29.8 (0.8320)	30.1 (0.8902)	29.7184 (0.8575)
Cameraman ($\sigma = 30$)	27.1022 (0.7641)	27.3 (0.7395)	27.8(0.8558)	27.8174 (0.8151)
Cameraman ($\sigma = 40$)	25.7866 (0.7241)	25.5 (0.6393)	26.2 (0.8211)	26.4954 (0.7826)
Monarch ($\sigma = 10$)	32.9907 (0.9369)	34.0 (0.9522)	34.2 (0.9594)	34.0698 (0.9553)
Monarch ($\sigma = 20$)	29.1114 (0.8811)	29.6 (0.8859)	30.0 (0.9202)	30.0384 (0.9145)
Monarch ($\sigma = 30$)	27.0058 (0.8346)	27.0 (0.8071)	27.4 (0.8769)	27.7209 (0.8735)
Monarch ($\sigma = 40$)	25.5973 (0.7950)	25.2 (0.7267)	25.9 (0.8378)	26.0832 (0.8293)
Peppers ($\sigma = 10$)	33.4942 (0.9056)	33.4 (0.8909)	33.3 (0.8943)	33.7904 (0.9189)
Peppers ($\sigma = 20$)	29.8124 (0.8424)	29.9 (0.8177)	30.1 (0.8413)	30.5252 (0.8743)
Peppers ($\sigma = 30$)	27.7810 (0.7924)	27.5 (0.7332)	27.9 (0.7973)	28.4765 (0.8356)
Peppers ($\sigma = 40$)	26.4045 (0.7507)	25.9 (0.6447)	26.7(0.7648)	26.9883 (0.8013)
Paint ($\sigma = 10$)	32.5488 (0.9165)	33.5 (0.9280)	33.6 (0.9311)	33.3567 (0.9276)
Paint ($\sigma = 20$)	28.5980 (0.8416)	26.8 (0.7467)	29.5 (0.8683)	29.4699 (0.8648)
Paint ($\sigma = 30$)	26.6067 (0.7817)	26.8 (0.7467)	27.2 (0.8088)	27.2540 (0.8077)
Paint ($\sigma = 40$)	25.2968 (0.7330)	25.0 (0.6590)	25.6 (0.7569)	25.6389 (0.7560)

Table 1. PSNR (dB) and SSIM results of the denoised images for the different techniques.

http://dx.doi.org/10.5772/intechopen.74453

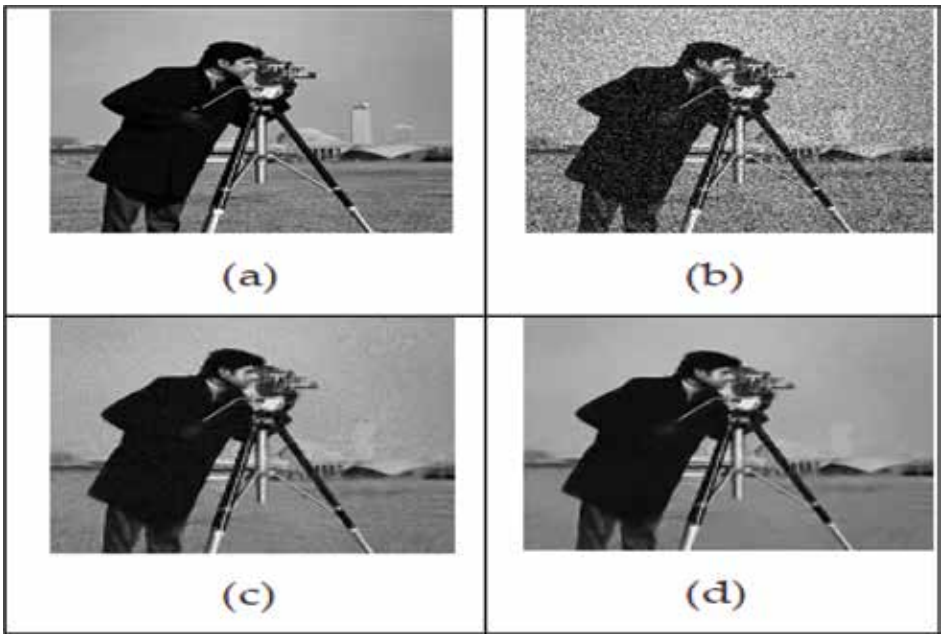


Figure 12. (a) Clean image (Cameraman.tif), (b) Noisy image with, (c) The denoised image by the proposed technique (the first stage) and denoised image by the proposed technique (the second stage).

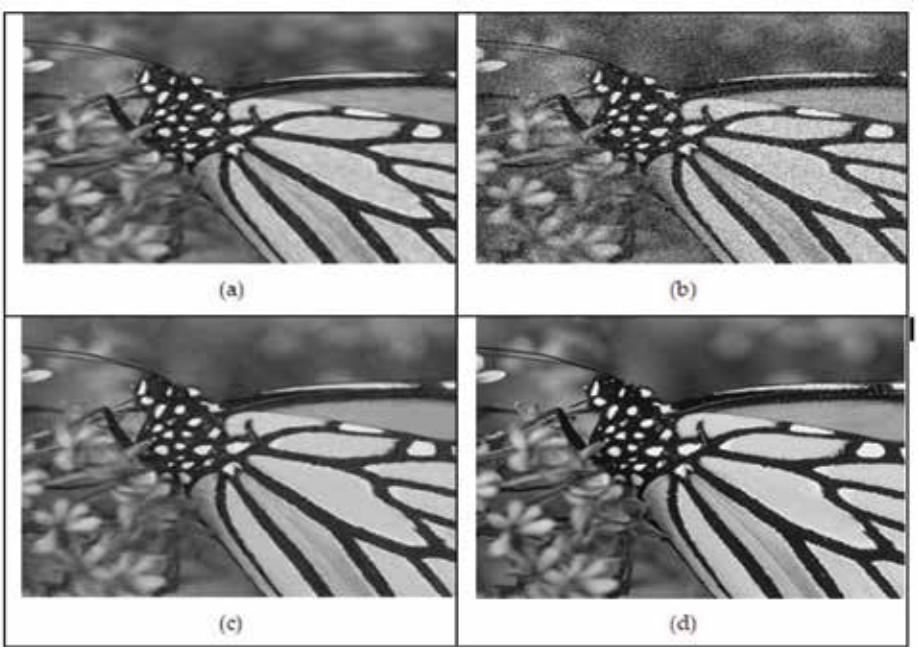


Figure 13. (a) Clean image (Monarch.tif), (b) Noisy image with, (c) The denoised image by the proposed technique (the first stage) and denoised image by the proposed technique (the second stage).

Figure 14. (a) Clean image (Lena.tif), (b) Noisy image with, (c) The denoised image by the proposed technique (the first stage) and denoised image by the proposed technique (the second stage).

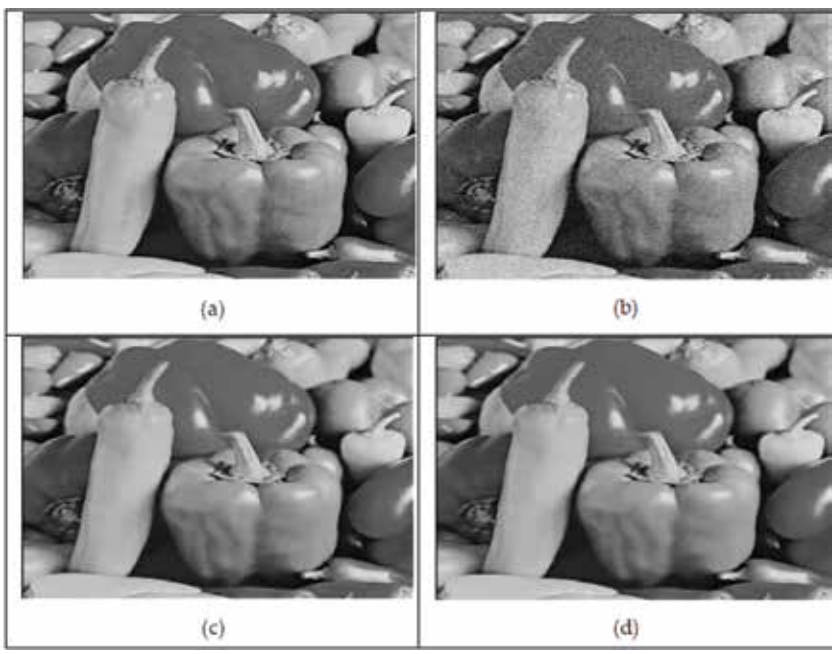


Figure 15. (a) Clean image (Peppers.tif), (b) Noisy image with, (c) The denoised image by the proposed technique (the first stage) and denoised image by the proposed technique (the second stage).

These obtained results (**Table 1**) show clearly that the proposed technique outperforms the denoising technique based on DT-CWT proposed in [12] and the denoising approach based on LPG-PCA [25]. **Figures 12–15** show four examples of image denoising using the proposed technique.

These figures show that the noise corrupting the original images is sufficiently suppressed. Moreover, the proposed technique permits to obtain denoised images with good perceptual quality. In each of these figures, the image (c) is obtained after the first denoising stage in the proposed technique. In this image (c), some noise is still existing, whereas it is considerably reduced into the image (d) obtained after the second denoising step. In the following subsection, we will give the results obtained by applying the proposed technique, the LPG-PCA-based denoising technique [25, 27] and the DT-DWT-based denoising one to a number of grayscale images. Those results are in terms of SNR and MSE and are listed in **Table 2**.

Technique	The first image denoising technique based on DT-DWT [12]	Two-stage image denoising by principal component analysis with local pixel grouping [25]	The proposed technique
House ($\sigma = 10$)	SNR = 78.00, MSE = 21.96	SNR = 79.41, MSE = 15.88	SNR = 79.44, MSE = 15.75
House ($\sigma = 20$)	SNR = 74.95, MSE = 44.29	SNR = 76.37, MSE = 31.97	SNR = 76.38, MSE = 31.92
House ($\sigma = 30$)	SNR = 73.14, MSE = 67.31	SNR = 74.50, MSE = 49.21	SNR = 74.50, MSE = 49.21
House ($\sigma = 40$)	SNR = 71.86, MSE = 90.28	SNR = 73.02, MSE = 69.16	SNR = 73.02, MSE = 69.14
Lena ($\sigma = 10$)	SNR = 74.67, MSE = 27.88	SNR = 75.28, MSE = 24.17	SNR = 75.28, MSE = 24.19
Lena ($\sigma = 20$)	SNR = 70.99, MSE = 65.02	SNR = 71.53, MSE = 57.40	SNR = 71.55, MSE = 57.19
Lena ($\sigma = 30$)	SNR = 68.97, MSE = 103.39	SNR = 69.35, MSE = 94.87	SNR = 69.37, MSE = 94.36
Lena ($\sigma = 40$)	SNR = 67.62, MSE = 141.07	SNR = 67.85, MSE = 134.09	SNR = 67.87, MSE = 133.35
Cameraman ($\sigma = 10$)	SNR = 75.33, MSE = 34.53	SNR = 76.19, MSE = 28.29	SNR = 76.23, MSE = 28.06
Cameraman ($\sigma = 20$)	SNR = 71.58, MSE = 81.88	SNR = 72.30, MSE = 69.38	SNR = 72.33, MSE = 68.80
Cameraman ($\sigma = 30$)	SNR = 69.68, MSE = 126.72	SNR = 70.39, MSE = 107.48	SNR = 70.45, MSE = 106.00
Cameraman ($\sigma = 40$)	SNR = 68.36, MSE = 171.56	SNR = 69.07, MSE = 145.72	SNR = 69.14, MSE = 143.51
Monarch ($\sigma = 10$)	SNR = 74.94, MSE = 32.65	SNR = 76.02, MSE = 25.47	SNR = 76.01, MSE = 25.55
Monarch ($\sigma = 20$)	SNR = 71.06, MSE = 79.78	SNR = 71.99, MSE = 64.45	SNR = 71.98, MSE = 64.53
Monarch ($\sigma = 30$)	SNR = 68.96, MSE = 129.56	SNR = 69.67, MSE = 109.89	SNR = 69.68, MSE = 109.62
Monarch ($\sigma = 40$)	SNR = 67.55, MSE = 179.20	SNR = 68.01, MSE = 161.05	SNR = 68.03, MSE = 160.25
Peppers ($\sigma = 10$)	SNR = 76.07, MSE = 29.08	SNR = 76.65, MSE = 25.43	SNR = 76.63, MSE = 25.56
Peppers ($\sigma = 20$)	SNR = 72.39, MSE = 67.89	SNR = 73.10, MSE = 57.61	SNR = 73.12, MSE = 57.43
Peppers ($\sigma = 30$)	SNR = 70.36, MSE = 108.38	SNR = 71.05, MSE = 92.34	SNR = 71.07, MSE = 92.02
Peppers ($\sigma = 40$)	SNR = 68.98, MSE = 148.80	SNR = 69.57, MSE = 130.09	SNR = 69.58, MSE = 129.58

Table 2. SNR (dB) and MSE results of the denoised images for the different techniques.

Those results show that the proposed technique outperforms the two other techniques (the LPG-PCA-based denoising technique [25, 27] and the DT-DWT-based denoising one [12]). In fact the proposed techniques are the highest values of SNR and lowest values of MSE.

10. Conclusion

In this chapter, a new image denoising technique is proposed. It combines two denoising approaches. The first one is a dual-tree discrete wavelet transform (DT-DWT)-based denoising technique, and the second one is a two-stage image denoising by principal component analysis with local pixel grouping (LPG-PCA). The first step of this proposed technique consists in applying the first approach to the noisy image in order to obtain a first estimate of the clean image. Then, we estimate the level of noise corrupting the original image. This estimation is performed by using a method of noise estimation from noisy images. The third step of the proposed technique consists in using this first clean image estimation, the noisy image, and this noise-level estimate as inputs of the second image denoising system (LPG-PCA-based image denoising) in order to obtain the final estimation of the clean image. A comparative study is performed between the proposed image denoising technique and two others denoising approaches where the first is based on DT-DWT and the second is based on LPG-PCA. This study is based on PSNR and SSIM computations, and the obtained results show that the proposed technique outperforms the two other denoising approaches. We also computed SNR (Signal to Noise Ratio) and MSE (Mean Square Error) and the obtained results also show that the proposed technique outperforms the others techniques.

Acknowledgements

We would like to thank all the people who contributed in some way to this work which was supported by the CRTEn (Center of Research and Technology of Energy) of Borj Cedria, Tunisia, and the Ministry of Higher Education and Scientific Research.

Author details

Mourad Talbi[1]* and Med Salim Bouhlel[2]

*Address all correspondence to: mouradtalbi196@yahoo.fr

1 Semi-Conductor, Nano-structures and Advanced Technologies Laboratory, Center of Researches and Technologies of Energy of Borj Cedria, Tunis, Tunisia

2 Sciences Electronics, Technologies of Information and Telecommunications (SETIT), Tunis, Tunisia

References

[1] Gonzalez RC, Woods RE. Digital Image Processing. 2nd ed. Englewood Cliffs, NJ: Prentice Hall; 2002

[2] Donoho DL. De-noising by soft thresholding. IEEE Transactions on Information Theory. 1995;**41**:613-627

[3] Coifman RR, Donoho DL. Translation-invariant de-noising. In: Antoniadis A, Oppenheim G, editors. Wavelet and Statistics. Berlin, Germany: Springer; 1995

[4] Mıhc-ak MK, Kozintsev I, Ramchandran K, Moulin P. Low-complexity image denoising based on statistical modeling of wavelet coefficients. IEEE Signal Processing Letters. 1999; **6**(12):300-303

[5] Chang SG, Yu B, Vetterli M. Spatially adaptive wavelet thresholding with context modeling for image denoising. IEEE Transaction on Image Processing. 2000;**9**(9):1522-1531

[6] Pizurica A, Philips W, Lamachieu I, Acheroy M. A joint inter- and intrascale statistical model for Bayesian wavelet based image denoising. IEEE Transaction on Image Processing. 2002; **11**(5):545-557

[7] Zhang L, Paul B, Wu X. Hybrid inter- and intra wavelet scale image restoration. Pattern Recognition. 2003;**36**(8):1737-1746

[8] Hou Z. Adaptive singular value decomposition in wavelet domain for image denoising. Pattern Recognition. 2003;**36**(8):1747-1763

[9] Portilla J, Strela V, Wainwright MJ, Simoncelli EP. Image denoising using scale mixtures of Gaussians in the wavelet domain. IEEE Transaction on Image Processing. 2003;**12**(11): 1338-1351

[10] Zhang L, Bao P, Wu X. Multiscale LMMSE-based image denoising with optimal wavelet selection. IEEE Transaction on Circuits and Systems for Video Technology. 2005;**15**(4):469-481

[11] Pizurica A, Philips W. Estimating the probability of the presence of a signal of interest in multiresolution single- and multiband image denoising. IEEE Transaction on Image Processing. 2006;**15**(3):654-665

[12] http://eeweb.poly.edu/iselesni/WaveletSoftware

[13] Chen GY, Ke'gl B. Image denoising with complex ridgelets. Pattern Recognition. 2007; **40**(2):578-585

[14] Elad M, Aharon M. Image denoising via sparse and redundant representations over learned dictionaries. IEEE Transaction on Image Processing. 2006;**15**(12):3736-3745

[15] Aharon M, Elad M, Bruckstein AM. The K-SVD: An algorithm for designing of overcomplete dictionaries for sparse representation. IEEE Transaction on Signal Processing. 2006;**54**(11):4311-4322

[16] Foi A, Katkovnik V, Egiazarian K. Pointwise shape-adaptive DCT for high-quality denoising and deblocking of grayscale and color images. IEEE Transaction on Image Processing. 2007; **16**(5)

[17] Tomasi C, Manduchi R. Bilateral filtering for gray and colour images. In: Proceedings of the 1998 IEEE International Conference on Computer Vision; Bombay, India; 1998. pp. 839–846

[18] Barash D. A fundamental relationship between bilateral filtering, adaptive smoothing, and the nonlinear diffusion equation. IEEE Transaction on Pattern Analysis and Machine Intelligence. 2002;**24**(6):844-847

[19] Buades A, Coll B, Morel JM. A review of image denoising algorithms, with a new one. Multiscale Modeling Simulation. 2005;**4**(2):490-530

[20] Kervrann C, Boulanger J. Optimal spatial adaptation for patch based image denoising. IEEE Transaction on Image Processing. 2006;**15**(10):2866-2878

[21] Dabov K, Foi A, Katkovnik V, Egiazarian K. Image denoising by sparse 3D transform-domain collaborative filtering. IEEE Transaction on Image Processing. 2007;**16**(8):2080-2095

[22] Mallat S. A Wavelet Tour of Signal Processing. New York: Academic Press; 1998

[23] Muresan DD, Parks TW. Adaptive principal components and image denoising. In: Proceedings of the 2003 International Conference on Image Processing. Vol. 1, September 14–17, 2003. pp. I101-I104

[24] Yaroslavsky LP. Digital Signal Processing—An Introduction. Berlin: Springer; 1985

[25] Zhang L, Dong W, Zhang D, Shi G. Two-stage image denoising by principal component analysis with local pixel grouping. Pattern Recognition. 2010;**43**:1531-1549

[26] Fukunaga K. Introduction to Statistical Pattern Recognition. 2nd ed. New York: Academic Press; 1991

[27] http://www4.comp.polyu.edu.hk/~cslzhang/LPG-PCA-denoising.htm

[28] Liu X, Tanaka M, Okutomi M. Noise-level estimation using weak textured patches of a single noisy image. In: IEEE International Conference on Image Processing (ICIP); 2012

[29] Liu X, Tanaka M, Okutomi M. Single-image noise-level estimation for blind denoising. IEEE Transactions on Image Processing. 2013;**22**(12):5226-5237

[30] Patrick L. Ondelettes complexes pour l'analyse des lois d'echelles. Master des Sciences de la Matière, Ecole Normale Supérieure de Lyon Université Claude Bernard Lyon 1; 2006

[31] Selesnick IW, Baraniuk RG, Kingsbury NG. The Dual-Tree Complex Wavelet Transform. IEEE Signal Processing Magazine. 2005;**22**(6):123-151

[32] Talbi M, Sira BF. Speech modulation for image watermarking. International Journal Multimedia and Image Processing (IJMIP). 2016;**6**(1/2):345-351

Edited by Sudhakar Radhakrishnan

This book is intended to attract the attention of practitioners and researchers in the academia and industry interested in challenging paradigms of wavelets and its application with an emphasis on the recent technological developments. All the chapters are well demonstrated by various researchers around the world covering the field of mathematics and applied engineering. This book highlights the current research in the usage of wavelets in different areas such as biomedical analysis, fringe-pattern analysis, image applications, network data transfer applications, and optical measurement techniques. The entire work available in the book is mainly focusing on researchers who can do quality research in the area of the usage of wavelets in related fields. Each chapter is an independent research, which will definitely motivate the young researchers to ponder on. These 12 chapters available in four sections will be an eye opener for all who are doing systematic research in these fields.

Published in London, UK

IntechOpen

ISBN 978-1-83881-536-3