INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

Bell & Howell Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

UNIVERSITY OF CALIFORNIA

RIVERSIDE

Wavelet Spectral Density Estimation of Continuous-Time Stationary Processes under Random Sampling

A Dissertation submitted in partial satisfaction of the requirements for the degree of

Doctor of Philosophy

in

Applied Statistics

by

Mark Eugene Lehr

March 2002

Dissertation Committee:

Dr. Keh-Shin Lii, Chairman

Dr. Christopher Roberts

Dr. Jie Chen

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

UMI Number: 3042849

Copyright 2002 by
Lehr, Mark Eugene

All rights reserved.

UMI Microform 3042849
Copyright 2002 by ProQuest Information and Learning Company.
All rights reserved. This microform edition is protected against
unauthorized copying under Title 17, United States Code.

ProQuest Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Copyright by

Mark Eugene Lehr

2002

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

The Dissertation of Mark Eugene Lehr is approved:

Christopher A. Robertson

Committee Chairman

University of California, Riverside

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

ACKNOWLEDGEMENTS

To name but a few, here are those that I would like to thank for being there with a comment or two.

Brandon – My Son *Thank you for staying with me your senior year, I would not have traded that time for anything in the world.*

Tiffany – My Daughter *I have shown you the way; it is up to you to determine where to go from here.*

Daniel Rowe – Office Mate *It makes it much easier when someone else is going through the same trials and tribulations. We certainly had our share! Now that I am the Godfather of your daughter, we can't help but stay in touch!*

Linda Penas – Bookie and Computer Girl *My random number generator was never able to achieve a 50% success rate on football picks.*

Harold – My Dad *"What is your dissertation topic?" It is a technique that cleans up noisy signals from randomly sampled data.*

Peggy – My Mom *"Get in there and get it done!" Well, it's done and I can finally say that it has all been worth the time and trouble when looking forward but never backward.*

Dr. Lii – Chairman & Friend *I always enjoyed our lunches together, the papers we wrote, the late nights, and all your advice.*

Janet – My Wife *New job, new house, new pool, and aren't we very lucky!* ☺

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

DEDICATIONS

Cody - My Grandfather & Evelyn – My Grandmother

Though you no longer are here with me you'll always be in my thoughts and my heart.

These are poems that were placed on the walkway through the Humanities Building at

UCR. I think you would appreciate them!

TIME SPENT

Take time to find out the nature of the place,
Its quirks of wind and weather,
Its well-tried ways,
Since each spot differs in what it bears or rejects.
So nature shifts for herself and trees of all species flourish in
Woodland,
Orchard,
Sacred Grove.

VIRGIL GEORGICS

MOTHEROOT

Creation often needs two hearts.
One to root and one to flower.
One to sustain in time of drought,
and hold fast against winds of pain.
The fragile bloom that in the glory of its hour,
affirms a heart unsung, unseen.

MARILOU AWIAKTA

SELU

And especially to Brandon and Tiffany! ☺

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

ABSTRACT OF THE DISSERTATION

Wavelet Spectral Density Estimation of Continuous-Time Stationary Processes under Random Sampling

by

Mark Eugene Lehr

Doctor of Philosophy, Graduate Program in Applied Statistics

University of California, Riverside, March 2002

Professor Keh-Shin Lii, Chairman

It has become increasingly accepted that wavelet based estimation techniques are generally better adapted to function estimates having large variations or, for lack of a better term, roughness. We consider a class of nonlinear wavelet estimators for the spectral density function of a zero-mean, stationary, not necessarily Gaussian continuous-time stochastic process, which is sampled at irregular intervals. A stationary point process is used to model the sampling method. The biases as well as the covariance properties of these alias-free estimators are investigated for their theoretical aspects. Simulation examples are presented to illustrate the salient features of the properties to be expected from such an analysis.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Contents Page

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

List of Figures and Tables

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

1 Introduction and Methodology

Data analysis is literally the most basic function performed by individuals no matter their age, educational background or employment classification. How else are decisions made? How else are conclusions reached? Of course, nothing is implied as to how well this task is being accomplished only that we all participate in the process. This decision process occurs with or without error from the day we are born and proceeds on a daily basis throughout our lives. Some approaches are more successful at revealing certain underlying fundamental principles then others and more importantly, the more revealing the analysis technique the more likely the correct conclusions are reached given the nature of the data.

The following treatise utilizes a technique that has been around in a popular form for a little over a decade and applies it to a problem that is related to randomly sampled time series data. It is not meant to represent a definitive, once and for all approach to the analysis of this type of problem but does present the theoretical foundation for utilizing such a technique and develops the rationale for why it can be used as a very important tool in analyzing information having the stated structure.

1.1 Introduction

Time Series data refers to any recorded process that occurs over time. As an example, one of the more famous time series is Sunspot data recorded over many years [Herzberg, 1985]. The basic idea is to investigate the possibility of recurrent patterns

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

or the applicability of theorized models that explain the deviations. This would apply equally as well to Stock Market fluctuations, unemployment data, and any business cycles, just to name a few. If these patterns actually existed and could be determined, predictions could be made that utilize the underlying information. Corresponding forecasts could be taken advantage of in any number of ways.

Normally, time series data is analyzed based upon the fact that the time interval between samples is constant. However, there are a number of real world instances where this is not possible or might not be entirely desirable. Consider data dropouts where data has been lost. In addition, data might not be recorded regularly and may seem to be taken at random, if not in practice then in actuality.

A natural way to analyze data, specifically time series data, is to transform it to an orthogonal system whereby the same information is simply represented in a different format. Spectral analysis represents one of these transformations of time series data to the frequency domain. It has been around for well over 200 years in physics, and most recently in the communication industry where it is all too familiar to every electrical engineer. There are advantages to analyzing data in the domain in which it exists. Fourier transformations are the natural way to look at data when it has periodic tendencies. Another type of transformation is the wavelet. Wavelets have their own advantages because they form a very general and quick procedure for deducing important differential and stepwise characteristics.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

In this paper we will first present the relevance of the above type of data and then provide an overview of the techniques utilized in analyzing this information. The following sections discuss in general terms, the major considerations of each technique. In Section 2 we will present some historical and brief mathematical aspects of these analyses. A simple synthesis will be suggested, combining these approaches to investigate a problem of estimating spectral densities from randomly sampled time series data. Section 3 will develop the mathematical basis and theory of this technique. Section 4 will apply the procedure to data of a sufficient nature so that we can see it's utility. Last but not least Section 5 will present were to go from here now that the foundation has been laid for future investigations. The appendices include some preliminary background, definitions, propositions, and lemmas which are helpful in understanding the previous sections.

1.2 The Applicability of Fourier Analysis

An extremely large class of problems falls under the Fourier or Spectral analysis methods. With the advent of computers and algorithms that compute the Fourier transform in $O(N\log_2 N)$, N represents the number of sampled data points, it has become commonplace to expect any time series study to include this type of foraging. The analysis utilizes trigonometric functions (i.e. sines, cosines) and has grown to encompass mathematical rigor in the areas of estimation, convergence criteria, existence and uniqueness.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

The Fourier transform and its cousin the Laplace transform represent linear decompositions and are widely used in science and engineering. To give an idea of just how pervasive and diverse these techniques have become, it is commonplace to see Fourier transforms used in boundary value problems, optics, quantum mechanics, image compression, image restoration, probability theory, economics, and the list goes on and on. As a matter of fact, it has become widely accepted that this is not just a technique for analysis but represents a physical phenomenon of nature and the world we live in. Hence, it is relevant to analyze data on the basis of where its functional characteristics reside or what function space the physical process exists.

1.3 The Applicability of Random Sampling

It is pertinent to mention briefly some of the inherent physical constraints to Fourier analysis. The sampled data is normally recorded at fixed intervals. This makes the recording process rather easy to mechanize and automate. However, there are mathematical as well as physical limitations present when we use such a procedure.

First, it is well known that a signal can be recreated exactly without error from a recorded sequenced set that is sampled at twice the rate of its highest frequency. This rate is referred to as the Nyquist rate or Nyquist frequency. The Sampling Theorem from which the Nyquist rate is derived is a major tool that has remarkable consequences.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Having stated this remarkable result there is also a downside. Over-sampling data can only help in reconstruction. However, under-sampling, meaning that the Nyquist frequency has not been reached, causes complications. This is known as aliasing. The higher frequency energy is not lost just placed into other frequency bins. The overall result is that we cannot reconstruct the signal. Errors are present due to this ambiguity and the farther away we are from the Nyquist frequency; the more energy will leak into the surrounding domain.

There is one technique that will mitigate the sampling limitation. Instead of sampling at uniform rates we can randomly sample according to any number of random schemes. Mechanizing this scheme with modern electronics is not an implementation problem. Multiple transducers can be utilized to create infinitely small separations in the sampled data. There are of course time constants associated with the transducers, but sampling can be tagged well ahead of the actual reading so this is not an implementation constraint.

1.4 The Applicability of Wavelet Analysis

One of many important results that could be obtained from spectral density estimation is the identification of peaks in the frequency band, which are closely related to fundamental periodicities in the time domain. A number of factors may obscure the identification of these peaks such as the sampling process, the estimation process, the signal to noise ratio, or even the proximity of the peaks or depressions to other

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

fundamental frequencies. Typical kernel estimators can smooth peak effects that may tend to obscure essential signal characteristics and their identification.

Wavelets, on the other hand, have been used effectively in estimating density functions limiting the associative smoothing effect penalty. This tendency is based on scale as the primary tool to de-noise signals of interest. All this will be emphasized in Section 2, including the fact that stochastic sampling has inherent advantages in the area of aliasing. Stochastic sampling adds additional complexities to the computations but should be weighed against the benefits especially when equally spaced samples are not possible or desirable, as is the case with dropouts, or other possible random effects.

Wavelets are relatively new to the bag of tools that can be applied to retrieving information from data. They represent another physical way to look at data and the world in which it was created. Unlike Fourier, there is an infinite number of wavelet analyzing functions that can be chosen, however it is not clear that more than a popular few are worth implementing for any general technique. It is true, on the basis of matching filters that utility can be derived by choosing properties of the wavelet function to correspond to the signal itself but due to the spatial mismatch of signal and filter that normally occurs this does not appear to be an advantage. What makes wavelets relatively nice to investigate data with is the ability to calculate the transformation with even faster algorithms then afforded by the Fast Fourier Transform (FFT). The speed is O(N) so the analysis is quick to occur and has been shown to be optimal in a certain mean squared sense.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Currently, the fields utilizing the wavelet technique and those holding the most promise involve imaging processing, numerical analysis, functional estimation, compression, and coding. There is definitely a link between Fourier analysis and Wavelets. Both are orthogonal decompositions, and they have fast algorithms that can implement either. They are also linear in nature. Normally, Fourier techniques can be utilized to create any number of Wavelet analyzing functions. However, there are a number of distinctions as well. Localization with Wavelets is possible while power spectral densities are global. Periodic functions are best suited for Fourier analysis while discontinuous functions fall under the Wavelet umbrella. Since wavelets maybe formulated in the Fourier domain it is only right that wavelets play a role in improving Fourier estimation techniques.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

2 Literature Review

The basis for this thesis is rooted in sampling processes, frequency representations of time series data, and wavelet based density estimation. All these components will be synthesized to produce a coherent application to an analysis topic that is of interest in the application realm. In order to accomplish this discussion, an examination of the literature with respect to these three topics is a prelude to the derivations and simulations that follow.

2.1 Fourier Analysis

Baron Jean Baptiste Joseph Fourier(1768-1830) was a French Mathematician and Physicist. One of the most important and seminal works of the 19th-century was his development of the mathematical theory of heat conduction published as the Theorie Analytique de la Cahleur, 1823. In it, Fourier developed the foundations of the series named after him. It was applied to the solution of boundary-value problems utilizing partial differential equations in thermodynamics. More importantly, it brought to a close a fair amount of controversy that a function could be represented by a series involving sines and cosines of integral multiples of the variable.

The Fourier transform decomposes a function in sinusoids of different frequencies, which sum to the original waveform. The transform and its inverse are typically represented in one of 6 ways.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

$$
\begin{aligned}
f(\Omega) &= \int_{-\infty}^{\infty} e^{2\pi t\Omega} f(t)dt \quad &\Leftrightarrow \quad f(t) &= \int_{-\infty}^{\infty} e^{-2\pi t\Omega} f(\Omega)d\Omega \\
f(\Omega) &= \int_{-\infty}^{\infty} e^{-2\pi t\Omega} f(t)dt \quad &\Leftrightarrow \quad f(t) &= \int_{-\infty}^{\infty} e^{2\pi t\Omega} f(\Omega)d\Omega \\
f(\omega) &= \int_{-\infty}^{\infty} e^{-it\omega} f(t)dt \quad &\Leftrightarrow \quad f(t) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{it\omega} f(\omega)d\omega \\
f(\omega) &= \int_{-\infty}^{\infty} e^{it\omega} f(t)dt \quad &\Leftrightarrow \quad f(t) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-it\omega} f(\omega)d\omega \\
f(\omega) &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-it\omega} f(t)dt \quad &\Leftrightarrow \quad f(t) &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{it\omega} f(\omega)d\omega \\
f(\omega) &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{it\omega} f(t)dt \quad &\Leftrightarrow \quad f(t) &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-it\omega} f(\omega)d\omega
\end{aligned}
\tag{2.1.1}
$$

The t in the above equation is measured in seconds, Ω is a frequency represented in hertz or cycles per second and ω is a frequency represented in radians per second. The scale factor π is applied based upon the conversion $\omega=2\pi\Omega$ and the desire to have symmetry for the inverse transform. The exponent is a complex representation of trigonometric functions in the imaginary plane (e^{ix} = cos(x)+isin(x)). They are all equivalent transforms and since the frequency representation is symmetrical about the frequency domain axis then the positive axis need only be represented visually. Refer to Appendix B for elementary properties of the transform with respect to types of functions (real, imaginary, even, odd), scaling, shifting, and some important theorems relating to Convolution, Correlation, and Parseval's representation. Most of which will be utilized in numerous derivations throughout this paper.

Another important concept is the Sampling Theorem and since a large part of the proceeding sections rely on this concept it should be discussed in a little detail here to reinforce the utility of point processes and random sampling later.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Theorem 2.1 *Shannon's Reconstruction Theorem* Any time series signal $f(t)$ that belongs to the Complex set and, $t \in \Re$ consisting of frequency components bandwidth limited to the Nyquist Critical Frequency Ω_c can be exactly reconstructed from the corresponding sampled data sequence where Δ is the sampling interval in seconds and $\Omega_c = 1/2\Delta$ in Hertz or cycles per second.

It is helpful to derive, and simulate the results of the above Theorem so that understandings of the concepts are more intuitively comprehended. E.T. Whittaker in 1915 and Kotel'nikov in 1933 developed the sampling series that predated the Sampling Theorem. However, the importance to communication theory and Shannon's application of the tenets of the series in 1949 assured him of its permanent reference and is one of the central foundations of sampling theory. First let's start with the definition of a time domain Δ impulse function [Feuer, 1996]:

$$s(t,\Delta) = \Delta \sum_{k \in Z} \delta_{t,k\Delta} \,. \tag{2.1.1}$$

This is just a scaled impulse function where δ is the Dirac delta function and in the above case, when $t = k\Delta$ the unit impulse occurs. A sampled signal is represented upon utilizing the Δ impulse function with the continuous function thus

$$f_s(t,\Delta) = \Delta \sum_{k \in Z} f(t)\delta_{t,k\Delta} \,. \tag{2.1.2}$$

Now, we shall look at this function in the frequency domain and transform back to determine the reconstruction formula. The assumption will be made that the signal is

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

band limited, meaning the entire power of the signal falls below a critical frequency Ω_c which is half the sampling frequency Ω_s.

$$f_s(t,\Delta) \Leftrightarrow F_s(\Omega) = F(\Omega) for \quad \Omega < |\Omega_c = \Omega_s / 2| \qquad (2.1.3)$$

A band-limited frequency domain function can be defined as

$$B(\Omega) = \begin{cases} 1 & \Omega \le |\Omega_c| \\ 0 & \Omega > |\Omega_c| \end{cases}. \qquad (2.1.4)$$

With this band-limited signal we can take the convolution of the following equation

$$F(\Omega) = B(\Omega) F_s(\Omega)$$

to obtain

$$f(t) = \int_{\tau \in \Re} b(\tau) \Delta \sum_{k \in Z} f(t) \delta_{t-\tau, k\Delta} d\tau . \qquad (2.1.5)$$

Utilizing the Fourier transform of the step function $B(\Omega) \Leftrightarrow b(\tau)$ and including the Dirac properties with the integral gives the final reconstruction formula of

$$f(t) = \Delta \sum_{k \in Z} f(k\Delta) \frac{\sin(2\pi\Omega_c (t - k\Delta))}{\pi(t - k\Delta)}. \qquad (2.1.6)$$

This formula shall be applied to several examples that follow.

The quarter hertz signal (Figure 2.1.1) is sampled at twice the critical frequency. The reconstructed signal is exact at the sampled points but has errors when reconstructing the entire signal. This does not seem to mesh with the Sampling Theorem except from the standpoint that the signal has not been infinitely sampled. In other words, the reconstruction in the center has very little error compared to the

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

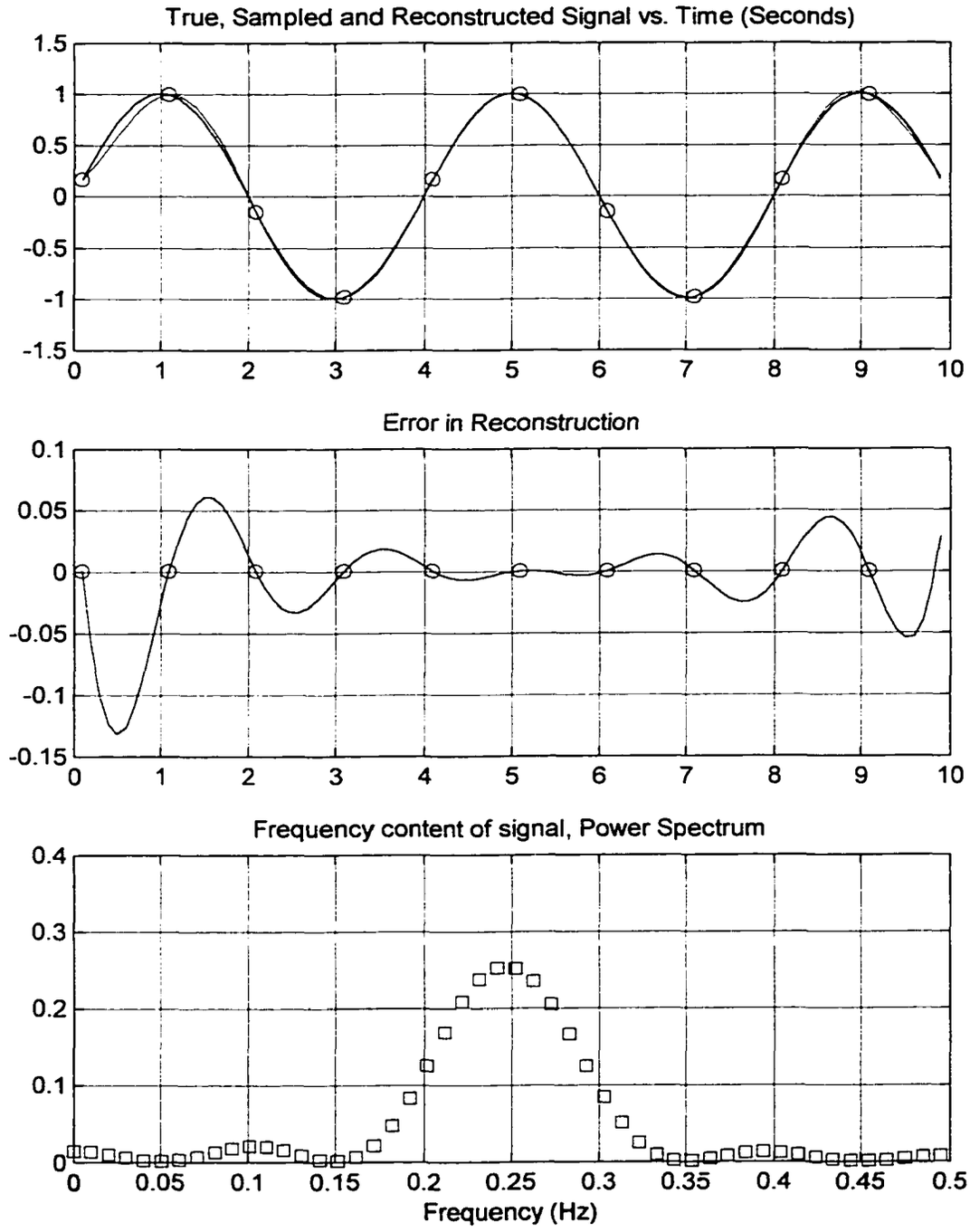


Figure 2.1.1 Quarter Hertz Over-Sampled Signal with Reconstruction

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

sampled endpoints. As the signal is sampled at the critical frequency there is also an implication that enough of the signal has been sampled.

The three quarter hertz signal (Figure 2.1.2) requires sampling at a minimum of 1.5 hertz but the sampling is done as before at 1 hertz. The construction is exact at the sampled points but the reconstructed function has larger errors. This is not just at the ends but is in error everywhere except at the sampled points. The frequency estimate of one quarter hertz is in error by a factor of 3. What this represents is *Aliasing*, which shows that the signal has been folded in the frequency domain back to the quarter hertz. The central point is that under-sampled signals are misinterpreted to erroneous frequency components and random sampling alleviates this condition [Lehr and Lii, 1997].

2.2 Wavelet Analysis

A description of wavelet properties, terminology, and analysis should start with the work of Stephane Mallet [Mallet, 1989], and Yves Meyer [Meyer, 1993]. Multiresolution signal analysis has its roots in perception with [Marr, 1983] and the way the human mind perceives objects at different scales from the perspective of visual acuity. It is also theorized as the way we initially perceive and learn about the world around us with scales in mind [Lehr & Lii, 1996]. This makes it easier to analyze and compare signals. It is also a possible reason why, as with the Fourier transform, that

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

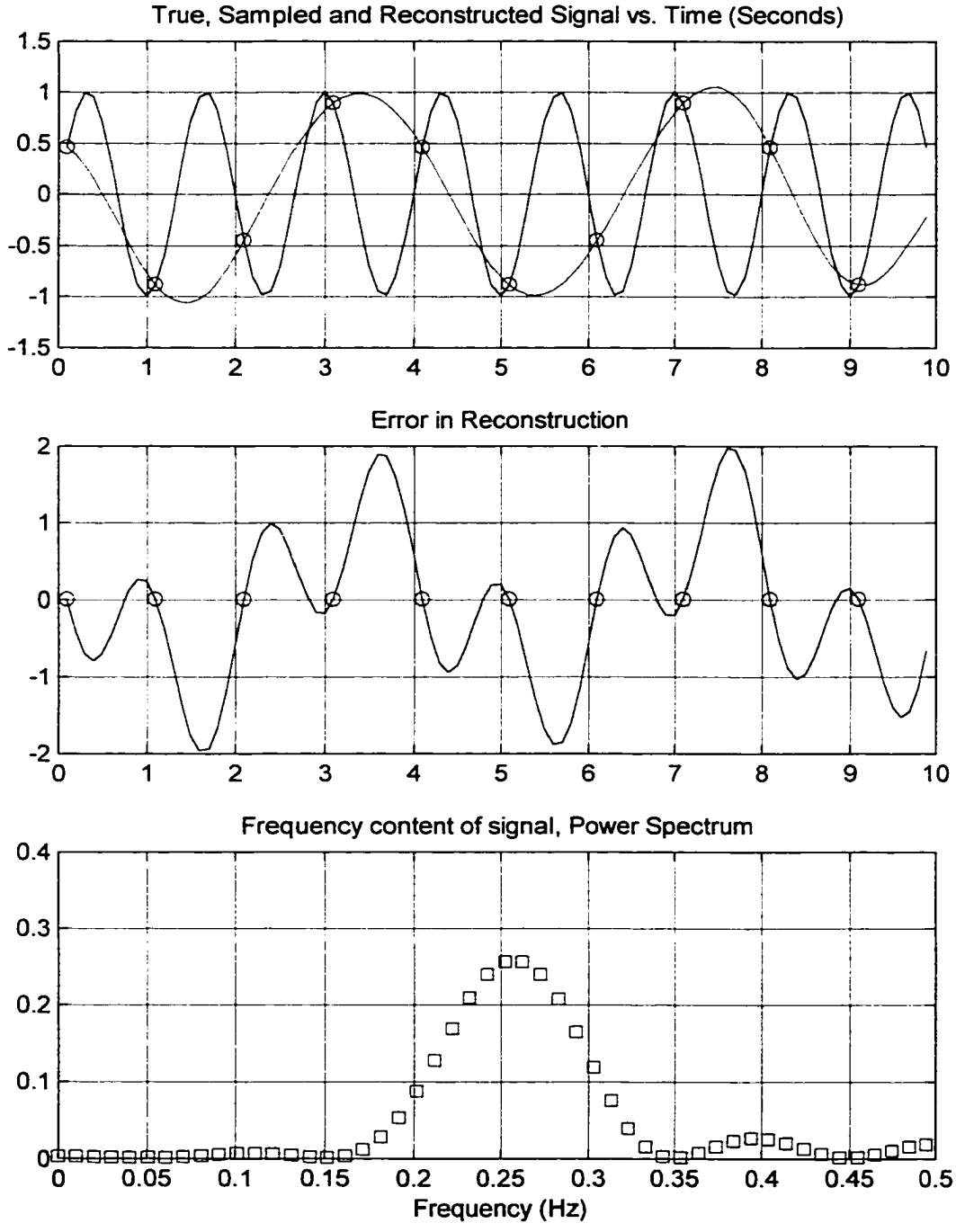


Figure 2.1.2 Three Quarter Hertz Under-Sampled Signal with Reconstruction

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

there is an inherent underlying property that makes wavelets much more than mathematical curiosities.

A mathematical description of multiresolution decomposition proceeds by expressing a function $f \in L^2$ as a limit of successive approximations [Akansu, 1992]. It is postulated that there exists a sequence of closed subspaces $\{V_m : m \in Z\}$ of $L^2(\Re)$ having the following properties

(1) Coarser subspaces are contained in finer subspaces.

$$\cdots V_2 \subset V_1 \subset V_0 \subset V_{-1} \subset V_{-2} \cdots \tag{2.2.1}$$

(2) The only object common to all subspaces is the null set.

$$\bigcap_{m \in Z} V_m = \{0\} \tag{2.2.2}$$

(3) Any signal can be approximated with arbitrary precision.

$$\bigcup_{m \in Z} V_m = L^2(\Re) \tag{2.2.3}$$

(4) There exists an orthonormal basis for the subspaces as well as an orthogonal complement

$$\begin{gathered} V_{m-1} = V_m \oplus W_m \\ V_m \perp W_m \end{gathered} \tag{2.2.4}$$

(5) Functions are related to subspaces by scaling

$$f(x) \in V_m \Leftrightarrow f(2x) \in V_{m-1} \tag{2.2.5}$$

for any function $f \in L^2(\Re)$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

The above 5 conditions provide the necessary framework for defining a basis function that can be scaled according to each of the subspaces. These functions are termed father wavelet $\phi(t) \in V_0$ and it's orthogonal complement the mother wavelet $\psi(t) \in W_0$. There are an infinite number of Wavelet basis possible. The first basis was recognized by Haar in 1911 and has his namesake.

2.3 Estimation Techniques for Sampled Time Series Data

Lii and Masry, [1994] developed spectral density estimates of randomly sampled data utilizing kernel smoothing. Asymptotic estimates for bias and covariance were derived and it was concluded that the bias was independent of the sampling point process. In addition there existed sampling point processes where the asymptotic variance was uniformly smaller that that of a Poisson sampling scheme for all spectral densities and frequencies. The advantage to this approach involved the alias-free nature of the resulting estimates.

Donoho, et.al. [1995] published a seminal work with respect to defining wavelet estimation techniques. They synthesize the current relevant approach to all the estimation techniques and conclude that wavelets are computationally practical and spatially adaptive. The shrinkage method is nearly minimax for a wide variety of loss functions and smoothness classes.

Neumann, [1995] considered nonlinear wavelet estimators of spectral densities with non-Gaussian conditions. The goal was to recognize peaks in the spectral density,

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

which are indications of periodicities. Whereas kernel estimators, and splines are an appropriate tool for estimating functions with a degree of smoothness, they are not necessarily the tool to utilize on less regular functions.

Hall, et. al. [1998] developed block thresholding rules for curve estimation using kernel and wavelet methods. They pointed out that wavelet methods adapt to erratic fluctuations in signals, have excellent mean squared error properties when used to estimate piecewise smooth functions, obtain the minimax convergence rates that are close to optimal over large function classes. The traditional linear estimators achieved relatively good performance for smooth functions.

2.4 Problem Synthesis

The synthesis of this investigation involves a combination of all the topics previously discussed. Namely, the derivation of estimation techniques, which utilize randomly sampling, Fourier transforms, and wavelets to determine the power spectral density and it's associated properties. This proposed estimate will be alias free due to the random sampling process and asymptotically consistent due to the wavelet implementation. This section presents the distributional characteristics of the wavelet coefficients and the process that will be used to reduce the variance of the estimate to achieve consistency.

Let us assume that we have samples $\{X_t\} \sim WN(0,\sigma^2)$. Using the Fourier transform to calculate the periodogram gives $\{I \sim (\sigma^2,(1 \wedge 2)\sigma^4)\}$. The periodogram is

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

exponentially distributed with the above mean and variance as well as being uncorrelated. The 2 applies to the variance at frequencies of 0 or π. We will leave out this detail to approximate the wavelet coefficient properties in general terms. The resultant spectral density estimate would be a constant random noise sequence across the frequency distribution with the mean and variance properties noted above. As the number of samples increases the variance will not diminish. Transforming the periodogram in the wavelet domain will distribute the mean and variance of the noise much differently then in the Fourier domain. The periodogram can be transformed into wavelet vectors represented by the following

$$\left\{\underline{\bar{\hat{f}}}_0, \underline{\hat{f}}_0, \underline{\hat{f}}_1, \underline{\hat{f}}_2, \cdots, \underline{\hat{f}}_J, \cdots, \underline{\hat{f}}_{J+m}\right\} \tag{2.4.1}$$

and defined in detail shortly. The first term $\underline{\bar{\hat{f}}}_0$ is the low-pass filtered component at scale 0 and each successive term is a high-pass filtered component at differing scales. This represents a multiresolution decomposition as defined by Mallet [Mallet, 1989]. To state simply, the first term contains the mean components of the signal at scale while the rest contain derivative or differencing information at each of the scales

The following table depicts the number of components in each of the vectors and their distributional qualities when transformed in the wavelet domain.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Table 2.4.1 Distributional Characteristics of Wavelet components associated with the transformed Periodogram

Vector Representation	Distribution of each component in the Vector (*mean*, var*iance*)	Number of components
$\underline{\hat{f}}$ -> the original Periodogram	(σ^2,σ^4)	$n=2^{J+m+1}$
$\underline{\hat{f}}_{J+m}$ ->wavelet vector at scale J+m	$(0,\sigma^4)$	n/2
...	...	...
$\underline{\hat{f}}_{J}$ ->wavelet vector at scale J	$(0,\sigma^4)$	$n/2^{m+1}$
...	...	...
$\underline{\hat{f}}_{2}$ ->wavelet vector at scale 2	$(0,\sigma^4)$	$n/2^{J+m-1}$
$\underline{\hat{f}}_{1}$ ->wavelet vector at scale 1	$(0,\sigma^4)$	$n/2^{J+m}$
$\underline{\hat{f}}_{0}$ ->wavelet vector at scale 0	$(0,\sigma^4)$	$n/2^{J+m+1}$
$\underline{\bar{\hat{f}}}_{0}$ ->mean wavelet vector at scale 0	$(2^{(J+m+1)/2}\sigma^2,\sigma^4)$	$n/2^{J+m+1}$

The wavelet transform for this illustration is implemented with a series of filters that have coefficients h_k, l_{1-k} with properties that are termed quadrature mirror filters.

$$h_k = (-1)^k l_{1-k}. \tag{2.4.2}$$

The filters that transform to the wavelet domain are termed analyzing filters and the filters used to return to the frequency domain are termed the synthesis filters. The sequence $\{L:l_k,k\in Z\}$ represents the low-pass filter coefficients and similarly the sequence {H} represent the high-pass filters. The analysis {L, H} filters and the synthesis filters {L*, H*} have the properties that L=L* and H=H* which is why they are termed quadrature mirror filters. Quadrature refers to orthogonality and mirror refers to equality. Applying the filters produces the following distributional properties for (mean,variance)

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

$$\bar{\hat{f}}_{J+m,k} = \sum_n l_{2k-n} \hat{f}_n \sim (\sqrt{2}\sigma^2, \sigma^4),$$

$$\hat{f}_{J+m,k} = \sum_n h_{2k-n} \hat{f}_n \sim (0, \sigma^4).$$

This is a simply a linear combination of (σ^2, σ^4) uncorrelated random variables so the distribution of each component is easily calculated. The low-pass filter coefficients sum to $\sqrt{2} = \sum_n l_{2k-n}$ while the high-pass filter coefficients sum to $0 = \sum_n h_{2k-n}$, hence the means of the distributions. The sum of squares of both coefficients sum to $1 = \sum_n l_{2k-n}^2 = \sum_n h_{2k-n}^2$ which similarly determines the variance. Note: [Daubechie, 1992] uses this type of scaling while others scale the coefficients by 1 or by a half. It is also important to note that this is independent of the type of wavelet utilized. We can again apply the same process for each space in the multiresolution analysis with the next scale

$$\bar{\hat{f}}_{J+m-1,k} = \sum_n l_{2k-n} \bar{\hat{f}}_{J+m,k} \sim (2\sigma^2, \sigma^4), \qquad (2.4.3)$$

$$\hat{f}_{J+m-1,k} = \sum_n h_{2k-n} \bar{\hat{f}}_{J+m,k} \sim (0, \sigma^4). \qquad (2.4.4)$$

The pattern is established for each multiresolution space such that each of the high-pass filtered components has exactly the same distributional characteristics as the lower. However, the successively low-passed filtered components increase their mean by a scale factor of $\sqrt{2}$ having the same variance regardless of the multiresolution space. Therefore, the generalization can be made from equation (2.4.3&4)

$$\bar{\hat{f}}_0 = \sum_n l_{2k-n} \bar{\hat{f}}_1 \sim (2^{(J+m+1)/2}\sigma^2, \sigma^4),$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

$$\hat{f}_0 = \sum_n h_{2k-n}\bar{\hat{f}}_0 \sim (0,\sigma^4)$$

the mean increases by the scale factor of the low-pass filtered component and the variance remains the same.

The purpose of utilizing the wavelet technique is to reduce the variance. This is accomplished by thresholding as discussed in Section 2.3. We can achieve any amount of variance reduction by understanding the maximum theoretical amount that can be achieved. Simply stated the variance can be halved successively by thresholding each scale. To see what happens with the reverse synthesis process we can approximate what happens when thresholding is applied. To reverse the process the same filter coefficients are utilized with

$$\bar{\hat{f}}_{0,n} = \sum_k (l_{2k-n}\bar{\hat{f}}_{1,k} + h_{2k-n}\hat{f}_{1,k}). \qquad (2.4.5)$$

To perfectly reconstruct the signal Equation 2.4.5 is applied to all the unaltered coefficients. This again is a linear combination so it is easy to calculate the mean. However, if we threshold and reduce or eliminate the effect of the high pass filtered terms (2.4.5) becomes $\bar{\hat{f}}_{0,n} = \sum_k l_{2k-n}\bar{\hat{f}}_{1,k}$. In general terms the mean will be reduced by a factor of $1/\sqrt{2}$ whereas the variance will be halved with this procedure. Each scale that is similarly thresholded will yield an additional 50% reduction in variance.

Figure 2.4.1 illustrates the effect of thresholding and its inherent variance reduction capabilities. A sequence of 2^{10} random variables are shifted and scaled from a beta(0.1,0.1) distribution to have the approximate characteristic distribution IID(0,1)

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

for simplicity sake. The first panel illustrates the histogram of one realization. The second panel contains the periodogram of this random sequence. Its distributional qualities are approximately IID(1,1) which is as predicted. The third panel portrays three separate curves. The first is the theoretical mean of the noise distribution with zero variance. The second curve contains the wavelet thresholded variance reduction of the data contained in the second panel. Eight levels or 2^8 coefficients were utilized in the threshold process. The reduction in variance is predicted to be 97%, which has been achieved. We should not expect to see more than $(1-2^m)*100$ per cent reduction in the variance when m levels are utilized in the thresholding process. A comparison was made with a super smoother technique [Friedman, 1984] that has kernel smoothing properties. Both achieved very reasonable results and agree with the overall theoretical mean of the noise process.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

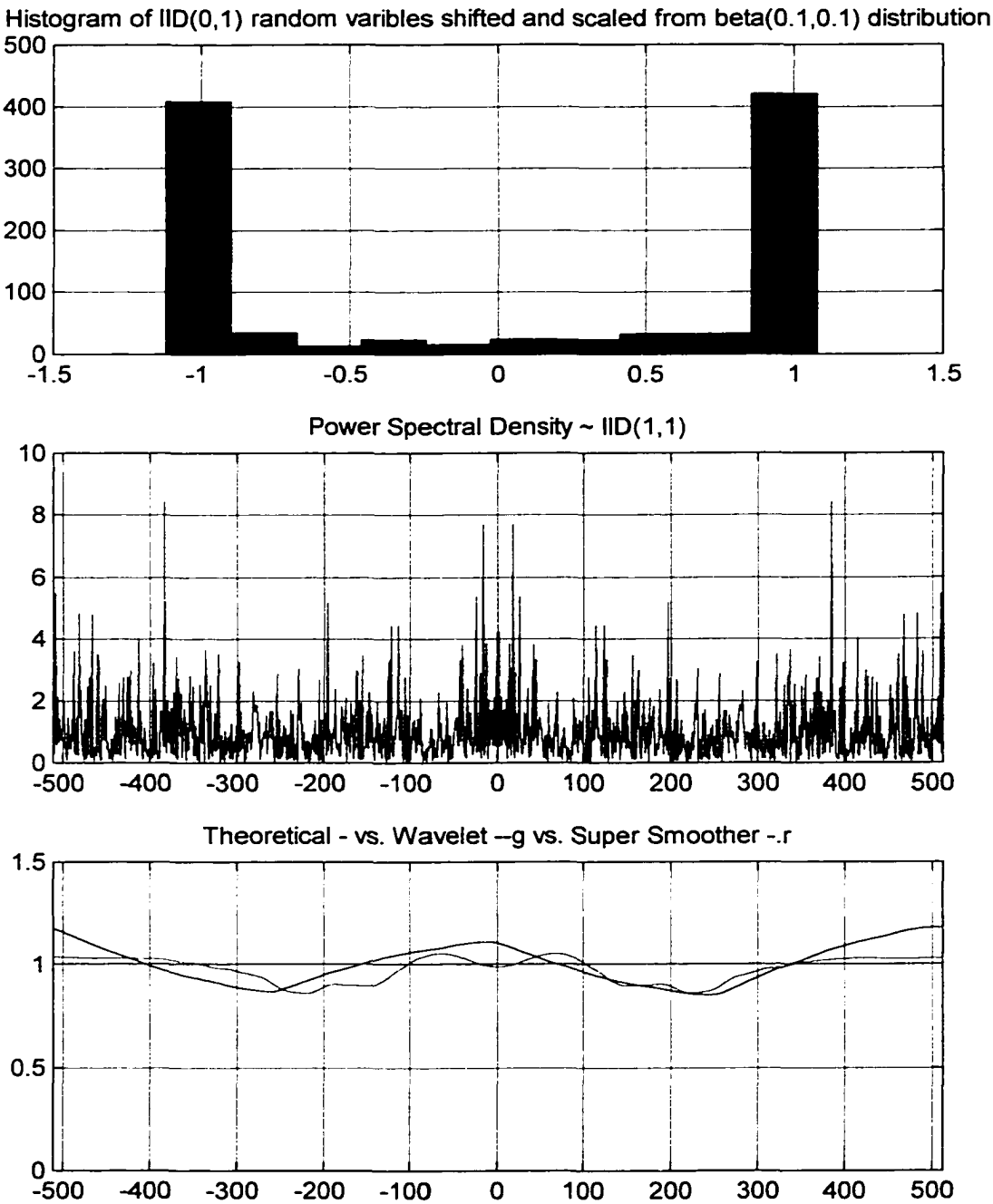


Figure 2.4.1 Variance Reduction using Wavelets

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

3 Theoretical Derivations of the Estimates

The following discussion starts with the development of an intuitive estimate of the spectral density obtained from random sampling. The estimate defined in Section 3.1 has desirable asymptotic properties with respect to the bias. The bias tends to zero as the sample size grows. However, the variance of the estimate does not improve as the sample size increases, hence the inconsistency of this estimate. We then contrast this result with the desirable theoretical properties of kernel estimates. These estimators have been applied to stationary processes under random sampling schemes and do achieve consistent results.

Wavelet estimation techniques are one possible solution to obtaining a consistent estimate without the inherent smoothing effects of kernel estimates. The intuitive estimate is transformed into the wavelet domain resulting in a new estimate for the spectral density. In sections 3.2 and 3.4 the statistical properties of the wavelet coefficients are determined for the first and second order cumulants. Naturally, this new representation of the previous intuitive estimate with all linear terms in the expansion has exactly the same properties as the previous estimate. The desirable asymptotic properties in bias remain as well as the undesirable inconsistency.

In order to achieve a consistent result, we shall introduce a small bias to the estimate in the form of a truncation to the series as outlined in Section 3.3. This effect balances the bias and variance resulting in a consistent estimate depicted in Section 3.5.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

3.1 The Intuitive Estimate

To set the stage for terminology and notation we borrow from Lii and Masry, [1994] by letting $X = \{X(t), t \in R\}$ be a zero mean stationary process with finite second-order moments, continuous covariance function $R_X(t) \in L_1$ and spectral density function $f_X = (f_X(\lambda), \lambda \in R)$. The sampling process is a stochastic point process $\{\tau_k\}_{k \in Z}$ that is stationary, orderly, and independent of X, with finite second-order moments [Daley and Jones, 1972]. With this in mind, let $N(\ )$ represent the counting process associated with the sampling scheme $\{\tau_k\}_{k \in Z}$ and $\beta = cum(N(0,1])$ be the mean intensity of the point process.

If the sampled process is defined as $Z(B) = \sum_{\tau_k \in t} X(\tau_k)$ where ($B \in \{B\}$, the Borel set) then the increment process Z has finite second-order moments. In differential notation we can write $dN(t) = N((0, t+dt]) - N((0,t])$ and $dZ(t) = X(t)dN(t)$ which upon evaluation gives $cum(dZ(t)) = cum(X(t)dN(t)) = 0$ due to the independence of the signal and the sampling technique. The covariance density c_N for distinct t_j's is defined as

$$c_N(t_2 - t_1)dt_1 dt_2 \equiv cum(dN(t_1), dN(t_2)),$$

and the Fourier transform of the covariance density becomes

$$\eta(\lambda) = (2\pi)^{-1} \int_R e^{-iu\lambda} c_N(u)du, \qquad c_N \in L_1 .$$

The point process statistical properties are

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

$$cum[N((t,t+dt])] = \beta dt,$$

$$C_N(du)dt \equiv cum[N((t,t+dt]), N((t+u,t+u+du])].$$

C_N is termed the reduced covariance measure, which is a σ-finite measure on the Borel set with an atom at the origin $C_N(\{0\}) = \beta$. We assume that C_N is absolutely continuous, outside of the origin, with the measure and the density related by

$$C_N(B) = \beta\delta_0(B) + \int_B c_N(u)du, \qquad B \in \mathrm{B}.$$

With the preceding definitions of all the relevant terms, the spectral density of the incremental process Z is

$$\begin{aligned} f_Z(\lambda) \equiv\ & \beta^2 f_X(\lambda) + (2\pi)^{-1}\beta R_X(0) \\ & + \int_R f_X(\lambda - u)\eta(\lambda)du. \end{aligned} \tag{3.1.1}$$

As above for the counting process we can define a covariance measure of the increment process $C_Z = C_Z^{(2)}$ with

$$dC_Z(u)dt \equiv cum[dZ(t), dZ(t+u)].$$

A few comments about the densities are important at this point. The covariance density c_N is zero when dealing with a Poisson sampling process. It should also be noted that the spectral density of the increment process is bounded, uniformly continuous but $f_Z(\lambda) \notin L_1$ in general.

The underlying function of interest $f_X(\lambda)$ is the focus of our investigation, which will require the rearrangement of equation (3.1.1) so that it portrays the term, which will be estimated by $\hat{f}_X = \left(\hat{f}_X(\lambda), \lambda \in R\right)$. Refer to Appendix B for the procedure

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

utilized to invert first order integral equations using Fourier techniques. Two more terms are defined with this implementation in mind by letting

$$\gamma(u) \equiv \frac{c_N(u)}{\beta^2 + c_N(u)} \tag{3.1.2}$$

and assuming its integrability

$$\gamma(u) \in L_1 \quad and \quad \Gamma(\lambda) = (2\pi)^{-1} \int_R e^{-iu\lambda} \gamma(u) du \in L_1.$$

We now obtain $f_X(\lambda)$ in the process with

$$f_X(\lambda) = \frac{1}{\beta^2} \left\{ f_Z(\lambda) - \frac{\beta R_X(0)}{2\pi} - \int_{-\infty}^{\infty} \Gamma(\lambda - u) \left[f_Z(u) - \frac{\beta R_X(0)}{2\pi} \right] du \right\}. \tag{3.1.3}$$

By Lii and Masry, [1994] is also established that a sufficient condition for the sampling process $\{\tau_k\}$ to be alias free is that the denominator of equation (3.1.2) be greater than zero $\beta^2 + c_N(u) > 0$ a.e.

A Natural Estimator

To find the spectral density of X from the discrete data $\{X(\tau_k), \tau_k\}_{k=1}^{N(T)}$ where N(T) is the number of samples in [0,T] we utilize the following sampled equations that are estimates of the components to $f_X(\lambda)$. The estimate for the covariance at lag 0 is

$$\hat{R}_X(0) \equiv (\beta T)^{-1} \int_0^T X^2(t) dt. \tag{3.1.4}$$

The periodogram for the incremental process Z is defined as

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

$$\hat{f}_Z(\lambda) \equiv (2\pi T)^{-1}\left|\int_0^T e^{-iu\lambda} X(u)du\right|^2. \tag{3.1.5}$$

With these functions we can now estimate the spectral density of X utilizing terms that are associated with the sampling process with

$$\begin{aligned} \hat{f}_P(\lambda) &= \beta^{-2}\left\{\hat{f}_Z(\lambda) - (2\pi)^{-1}\beta\hat{R}_X(0)\right\}, \\ \hat{f}_{NP}(\lambda) &= -\int_R \Gamma(\lambda-u)\hat{f}_P(u)du, \end{aligned} \tag{3.1.6}$$

which can be characterized as having Poisson f_p and Non-Poisson f_{np} like components equating to

$$\hat{f}_X(\lambda) = \hat{f}_P(\lambda) + \hat{f}_{NP}(\lambda). \tag{3.1.7}$$

Of course, the convolution component for the Non-Poisson condition is $\hat{f}_{NP}(\lambda) = 0$ when dealing with Poisson sampling schemes.

Bias Characteristics

Now we examine estimates to determine their statistical attributes. The correlation estimate at lag 0 utilizes Fubini's Theorem and the independence of the sampling method with the process itself thereby giving

$$\begin{aligned} cum\left(\hat{R}_X(0)\right) &= (\beta T)^{-1} cum\left(\int_R cum\left(X^2\right)cum(dN(t))\right) \\ &= (\beta T)^{-1}\int_R cum\left(X^2\right)cum(dN(t)) \end{aligned}$$

By previous definition since

$$\begin{aligned} cum\left(X^2(t)\right) &= R_X(0), \\ cum(dN(t)) &= \beta dt, \end{aligned}$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

then we obtain the following unbiased estimate of the correlation of X at lag 0 with

$$\begin{aligned} cum\left(\hat{R}_X(0)\right) &= \frac{1}{\beta T}\int_0^T R_X(0)\beta dt \\ &= R_X(0). \end{aligned} \tag{3.1.8}$$

The Fourier integral transform of the sample is defined as

$$d_{Z,T}(\lambda) = \int_0^T e^{-i\lambda t} X(t) dN(t). \tag{3.1.9}$$

The estimate of the spectral density of Z utilizing the above and (3.1.5) becomes

$$\begin{aligned} cum\left(\hat{f}_Z(\lambda)\right) &= cum\left(\frac{1}{2\pi T}\left|\int_0^T e^{-i\lambda t} X(t) dN(t)\right|^2\right) \\ &= \frac{1}{2\pi T} cum\left(d_{Z,T}(\lambda), d_{Z,T}(-\lambda)\right). \end{aligned}$$

Note: the following equations that are numbered starting with the letter "A" are referred to as assumptions made in deriving solutions for the estimates. Making use of Theorem 4.1 of Brillinger, [1972]

$$\int_{-\infty}^{\infty}(1+|u|)d|C_Z^2(u)| < \infty, \tag{A3.1.1}$$

gives

$$cum\left\{d_{Z,T}(\lambda), d_{Z,T}(u)\right\} = 2\pi D_T(\lambda+u) f_Z(\lambda) + O(1).$$

The continuous Dirichlet kernel D_T is defined as

$$D_T(\lambda) \equiv \int_0^T e^{-i\lambda t} dt = \frac{1-e^{-i\lambda T}}{i\lambda} \tag{3.1.10}$$

and the O(1) term is uniform in λ. Hence,

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

$$
\begin{aligned}
cum\left(\hat{f}_Z(\lambda)\right) &= \frac{1}{2\pi T} cum\left(d_{Z,T}(\lambda), d_{Z,T}(-\lambda)\right) \\
&= \frac{D_T(0) f_Z(\lambda) + O(1)}{T}.
\end{aligned}
$$

By definition, and using L'Hopitals Rule then

$$
D_T(\lambda) \xrightarrow{\lambda \to 0} \left. \frac{d\left(1 - e^{-i\lambda T}\right)}{d(i\lambda)} \right|_{\lambda=0} = \left. \frac{iTe^{-i\lambda T}}{i} \right|_{\lambda=0} = T.
$$

The resultant expectation becomes

$$
cum\left(\hat{f}_Z(\lambda)\right) = f_Z(\lambda) + O\left(\frac{1}{T}\right). \tag{3.1.11}
$$

This is asymptotically unbiased where the O(1/T) term is uniform in λ. For the Poisson sampling case the estimate becomes

$$
\begin{aligned}
cum\left(\hat{f}_P(\lambda)\right) &= cum\left(\frac{1}{\beta^2}\left\{\hat{f}_Z(\lambda) - \frac{\beta \hat{R}_X(0)}{2\pi}\right\}\right) \\
&= \frac{1}{\beta^2}\left\{cum\left(\hat{f}_Z(\lambda)\right) - \frac{\beta cum\left(\hat{R}_X(0)\right)}{2\pi}\right\} \\
&= \frac{1}{\beta^2}\left\{f_Z(\lambda) - \frac{\beta R_X(0)}{2\pi}\right\} + \frac{1}{\beta^2} O\left(\frac{1}{T}\right) \\
&= f_P(\lambda) + O\left(\frac{1}{T}\right),
\end{aligned} \tag{3.1.12}
$$

which is an asymptotically unbiased estimate where the order of magnitude term is uniform in λ. Similarly the non-Poisson term can be estimated using the previous result with

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

$$
\begin{aligned}
cum\left(\hat{f}_{NP}(\lambda)\right) &= cum\left(-\int_R \Gamma(\lambda-u)\hat{f}_Z(u)du\right) \\
&= -\int_R \Gamma(\lambda-u)cum\left(\hat{f}_Z(u)\right)du \\
&= -\int_R \Gamma(\lambda-u)f_Z(u)du - \int_R \Gamma(\lambda-u)O\left(\frac{1}{T}\right)du.
\end{aligned}
$$

Since $\Gamma \in L_1$ and the order term is uniform in λ then we have an asymptotically unbiased estimate for the non-Poisson component with

$$
cum\left(\hat{f}_{NP}(\lambda)\right) = f_{NP}(\lambda)+O\left(\frac{1}{T}\right). \tag{3.1.13}
$$

Putting this together for the spectral density of X results in

$$
\begin{aligned}
cum\left(\hat{f}_X(\lambda)\right) &= cum\left(\hat{f}_P(\lambda)+\hat{f}_{NP}(\lambda)\right) \\
&= f_P(\lambda)+f_{NP}(\lambda)+O\left(\frac{1}{T}\right) \\
&= f_X(\lambda)+O\left(\frac{1}{T}\right).
\end{aligned} \tag{3.1.14}
$$

This is an asymptotically unbiased estimate of the spectral density of X. Additionally, we can see that it is asymptotically independent of the sampling process.

Covariance Characteristics

For the covariance estimates the following assumptions [Lii and Masry,1994] are generalized for higher order cumulants with respect to the signal $\{X(t),-\infty < t < \infty\}$ and the sampling process $\{\tau_k\}_{k=-\infty}^{\infty}$. As before, all assumptions are numbered with the letter A at the beginning to distinguish them from referenced equations. The following

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

assumptions refer to an integer $K \geq 2$ for generality with respect to moments of the cumulants but the current utilization in this section never exceeds K=4.

Integrability of the signals cummulants

$$\int_{R^{k-1}} \left(1 + |u_j|\right) \left| c_X^{(k)}(u_1, \ldots, u_{k-1}) \right| du_1 \ldots du_{k-1} < \infty \qquad \text{(A3.1.2)}$$

for j = 1,..., k-1; k = 2, ..., K, where $c_X^{(k)}$ is the kth order cumulant of X. Incidentally $c_X^{(2)}(u) = R_X(u)$. In particular,

$$c_X^{(4)}(0, u, u) \in L_\infty \cap L_1 . \qquad \text{(A3.1.3)}$$

Similarly for the sampling process

$$\int_{R^{k-1}} \left(1 + |u_j|\right) \left| c_N^{(k)}(u_1, \ldots, u_{k-1}) \right| du_1 \ldots du_{k-1} < \infty \qquad \text{(A3.1.4)}$$

for j = 1,..., k-1; k = 2, ...K , and $c_N^{(k)}$ is the kth order cumulant defined by

$$cum\{dN(t_1), \ldots, dN(t_k)\} = c_N^{(k)}(t_2 - t_1, \ldots, t_k - t_1) dt_1 \cdots dt_k$$

for distinct $t_{j's}$. Note that $c_N^{(2)}(u) = c_N(u)$.

The preceding assumptions require that $E|X(t)|^k < \infty$ as well as $E|dN(t)|^k < \infty$ for the existence of the cumulants $\{c_X^{(k)}, c_N^{(k)}, k = 2, \ldots K\}$. Next we define the cumulants $C_Z^{(k)}(u_1, \ldots u_{k-1})$ of the increment process Z by

$$cum\{dZ(t_1), \ldots, dZ(t_k)\} = dC_Z^{(k)}(t_2 - t_1, \ldots, t_k - t_1) dt_1 \qquad \text{(3.1.15)}$$

and $C_Z^{(k)}$ is of bounded variation over finite cubes. Then under assumption (A3.1.2) and (A3.1.4) we have ([Brillinger, 72] p. 485).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

$$\int_{R^{k-1}} \left(1+|u_j|\right) d\left|C_Z^{(k)}(u_1,\ldots,u_{k-1})\right| < \infty, \quad j=1,\ldots k; k=2,\ldots,K. \tag{3.1.16}$$

The kth-order cumulant spectrum is defined by

$$f_Z^{(k)}(\lambda_1,\ldots,\lambda_{k-1}) \equiv \frac{1}{(2\pi)^{k-1}} \int_{R^{k-1}} e^{-i\sum_{j=1}^{k-1}\lambda_j} dC_Z^{(k)}(u_1,\ldots,u_{k-1}) \tag{3.1.17}$$

and we note that this is bounded, uniformly continuous but not integrable in general, and $f_Z^{(2)}(\lambda) = f_Z(\lambda)$. Similarly, we utilize ([Brillinger, 72] Theorem 4.1) as we did with A3.1.1 to obtain

$$cum\left(d_{Z,T}(\lambda_1),\ldots d_{Z,T}(\lambda_k)\right) = (2\pi)^{k-1} D_T\left(\sum_{j=1}^{k}\lambda_j\right) f_Z^{(k)}(\lambda_1,\ldots,\lambda_{k-1}) + O(1), \quad k=2,\ldots,K. \tag{3.1.18}$$

The O(1) term is uniform in the λ's.

We can rewrite the spectral density of X by first considering the inverse Fourier transform of

$$\gamma(u) = \int_R e^{i\lambda u}\Gamma(\lambda)d\lambda .$$

At u = 0 we obtain the equality $\gamma(0) = \int_R \Gamma(\lambda)d\lambda = \int_R \Gamma(\lambda_0 - \lambda)d\lambda$. This can be utilized to rearrange the spectral estimation function from Equation 3.1.3

$$\hat{f}_X(\lambda) = \frac{1}{\beta^2}\left\{\hat{f}_Z(\lambda) - \frac{\beta\hat{R}_X(0)}{2\pi} - \int_{-\infty}^{\infty}\Gamma(\lambda-u)\left[\hat{f}_Z(u) - \frac{\beta\hat{R}_X(0)}{2\pi}\right]du\right\}$$

to the more convenient form

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

$$\hat{f}_X(\lambda)=\frac{1}{\beta^2}\int_R(\delta(\lambda-u)-\Gamma(\lambda-u))\hat{f}_Z(u)du+\frac{\gamma(0)-1}{2\pi\beta}\hat{R}_X(0) \qquad (3.1.19)$$

for utilization in the covariance calculations.

To simplify the notation let $Q(\lambda)=\delta(\lambda)-\Gamma(\lambda)$ which leads to the spectral density estimate definition as

$$\hat{f}_X(\lambda) = \frac{1}{\beta^2}\int_R Q(\lambda-u)\hat{f}_Z(u)du+\frac{\gamma(0)-1}{2\pi\beta}\hat{R}_X(0). \qquad (3.1.20)$$

The covariance of the estimated spectral density of X becomes

$$\begin{aligned} cum\left(\hat{f}_X(\lambda_1),\hat{f}_X(\lambda_2)\right) &= (2\pi\beta)^{-2}[1-\gamma(0)]^2 Var\left(\hat{R}_X(0)\right) \\ &\quad +\beta^{-4}\int_{R^2} Q(\lambda_1-u_1)Q(\lambda_2-u_2)cum\left(\hat{f}_Z(u_1),\hat{f}_Z(u_2)\right)du_1du_2 \\ &\quad +(2\pi\beta)^{-1}\int_R Q(\lambda_1-u)cum\left(\hat{f}_Z(u),\hat{R}_X(0)\right)du \\ &\quad +(2\pi\beta)^{-1}\int_R Q(\lambda_2-u)cum\left(\hat{f}_Z(u),\hat{R}_X(0)\right)du, \\ &= T_1+T_2+T_3+T_4. \end{aligned} \qquad (3.1.21)$$

The magnitude of T_1 may first be calculated by

$$\begin{aligned} Var\left(\hat{R}_X(0)\right) &= Var\left((\beta T)^{-1}\int_0^T X^2(t)dN(t)\right), \\ &= (\beta T)^{-2}\int_0^T\int_0^T \mathrm{cov}\left(X^2(t)dN(t),X^2(s)dN(s)\right). \end{aligned}$$

Since the signal and the sampling process are independent then

$$\begin{aligned} cum\left(\hat{R}_X(0),\hat{R}_X(0)\right) &= (\beta T)^{-2}\int_0^T\int_0^T cum\left(X^2(t)dN(t),X^2(s)dN(s)\right), \\ &= (\beta T)^{-2}\int_0^T\int_0^T cum\left(X^2(t),X^2(s)\right)cum\left(dN(t),dN(s)\right). \end{aligned}$$

The quantity $E\left(X^2(t)X^2(s)\right)=cum\left(X^2(t),X^2(s)\right)+cum^2\left(X(t),X(t)\right)$. Using Mendel, [1995] we have the following identity

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

$$
\begin{aligned}
E\left(X^2(t)X^2(s)\right) &= cum(X(t),X(t),X(s),X(s)) \\
&+ cum(X(t),X(s))cum(X(t),X(s)) \\
&+ cum(X(t),X(s))cum(X(t),X(s)) \\
&+ cum^2(X(t),X(t)).
\end{aligned}
$$

This equates to

$$
\begin{aligned}
cum\left(X^2(t),X^2(s)\right) &= cum(X(t),X(t),X(s),X(s)) + 2cum^2(X(t),X(s)) \\
&= c_X^{(4)}(0,\tau,\tau) + 2R_X^2(\tau)
\end{aligned}
$$

where the variable $\tau = t - s$ and $R_X(\tau) = c_X(\tau) = c_X^{(2)}(\tau)$. With the change of variable

$$
\begin{aligned}
Var\left(\hat{R}_X(0)\right) &= (\beta T)^{-2}\int_0^T\int_0^T cum\left(X^2(t),X^2(s)\right)cum(dN(t),dN(s)) \\
&= \beta^{-2}T^{-1}\int_{-T}^{T}\left(1-|\tau|T^{-1}\right)\left\{2R_X^2(\tau)+c_X^{(4)}(0,\tau,\tau)\right\}\left\{\beta^2+\beta\delta(\tau)+c_N(\tau)\right\}d\tau \\
&\le \beta^{-2}T^{-1}\int_{-T}^{T}\left\{2R_X^2(\tau)+c_X^{(4)}(0,\tau,\tau)\right\}\left\{\beta^2+c_N(\tau)\right\}d\tau \\
&\quad + \beta^{-1}T^{-1}\left\{2R_X^2(0)+c_X^{(4)}(0,0,0)\right\} \\
&= O\left(T^{-1}\right).
\end{aligned}
\tag{3.1.22}
$$

The second and third row in the above equation utilizes the stationarity of the process, and the symmetry of the cumulants. The final result follows with the conditions that $R_X^2 \in L_1$ and $R_X^2 c_N \in L_1$ given $R_X \in L_\infty, c_N \in L_1$ and (A3.1.3). The order is

$$
\begin{aligned}
T_1 &= (2\pi\beta)^{-2}\left[1-\gamma(0)\right]Var\left(\hat{R}_X(0)\right) \\
&= (2\pi\beta)^{-2}\left[1-\gamma(0)\right]O\left(T^{-1}\right) \\
&= O\left(T^{-1}\right).
\end{aligned}
\tag{3.1.23}
$$

The second term T_2 is evaluated next by first examining the covariance inside the integral from (3.1.21)

$$
cum\left(\hat{f}_Z(u_1),\hat{f}_Z(u_2)\right) = (2\pi T)^{-1}cum\left(d_{Z,T}(u_1)d_{Z,T}(-u_1),d_{Z,T}(u_2)d_{Z,T}(-u_2)\right).
$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Applying Mendel, [1995] again

$$\begin{aligned} cum\left(d_{Z,T}(u_1)d_{Z,T}(-u_1),d_{Z,T}(u_2)d_{Z,T}(-u_2)\right) &= cum\left(d_{Z,T}(u_1),d_{Z,T}(-u_1),d_{Z,T}(u_2),d_{Z,T}(-u_2)\right) \\ &+ cum\left(d_{Z,T}(u_1),d_{Z,T}(u_2)\right)cum\left(d_{Z,T}(-u_1),d_{Z,T}(-u_2)\right) \\ &+ cum\left(d_{Z,T}(u_1),d_{Z,T}(-u_2)\right)cum\left(d_{Z,T}(-u_1),d_{Z,T}(u_2)\right). \end{aligned}$$

With (3.1.18) we can rewrite such that

$$\begin{aligned} cum\left(\hat{f}_Z(u_1),\hat{f}_Z(u_2)\right) = (2\pi T)^{-2} \quad & \left\{\left[(2\pi)^3 D_T(0) f_Z^{(4)}(u_1,-u_1,u_2)+O(1)\right]\right. \\ & +\left[2\pi D_T(u_1+u_2) f_Z^{(2)}(u_1)+O(1)\right] \\ & \times\left[2\pi D_T(-u_1-u_2) f_Z^{(2)}(-u_1)+O(1)\right] \\ & +\left[2\pi D_T(u_1-u_2) f_Z^{(2)}(u_1)+O(1)\right] \\ & \left.\times\left[2\pi D_T(-u_1+u_2) f_Z^{(2)}(-u_1)+O(1)\right]\right\}. \end{aligned}$$

(3.1.24)

We use the definition of the Fejer kernel $\Delta_T(\lambda) = T^{-1}|D_T(\lambda)|^2$ to simplify the results.

$$\begin{aligned} Cov\left(\hat{f}_Z(u_1),\hat{f}_Z(u_2)\right) = \quad & \left\{\left[2\pi T^{-1} f_Z^{(4)}(u_1,-u_1,u_2)\right]\right. \\ & +T^{-1}|f_Z(u_1)|^2\left[\Delta_T(u_1+u_2)+\Delta_T(-u_1+u_2)\right] \\ & +O(1)\frac{f_Z(u_1)}{2\pi T^2}\left[D_T(u_1+u_2)+D_T(u_1-u_2)\right] \\ & +O(1)\frac{f_Z(-u_1)}{2\pi T^2}\left[D_T(-u_1-u_2)+D_T(-u_1+u_2)\right] \\ & \left.+\frac{O(1)}{2\pi T^2}\right\}. \end{aligned}$$

(3.1.25)

Now we can look at each term in the covariance of the spectral density of Z to determine the order of magnitude of each component. We assume the 4th order spectrum is bounded $f_Z^{(4)} \in L_\infty$ from (3.1.17) then the order of this term is $O\left(T^{-1}\right)$ and similarly for the terms involving the Dirichlet kernel since even when evaluated at its

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

maximum the order is only $O(T^{-1})$. The term to be cautious of contains the Fejer kernel, which at 0 evaluates to T. Therefore the most significant term in T_2 contains

$$|f_Z(u_1)|^2 \times T^{-1}[\Delta_T(u_1+u_2)+\Delta_T(-u_1+u_2)]$$

which when evaluated leads to

$$cum(\hat{f}_Z(u_1),\hat{f}_Z(u_2)) = \begin{cases} O(T^{-1}) & |u_1| \neq |u_2| \\ |f_Z(u_1)|^2 + O(T^{-1}) & |u_1| = |u_2| \neq 0 \\ 2|f_Z(u_1)|^2 + O(T^{-1}) & |u_1| = |u_2| = 0 \end{cases}. \qquad (3.1.26)$$

And as

$$\lim_{T\to\infty} cum(\hat{f}_Z(u_1),\hat{f}_Z(u_2)) = |f_Z(u_1)|^2\{\delta_{u_1,u_2} + \delta_{u_1,-u_2}\} \qquad (3.1.27)$$

which implies that the covariance terms asymptotically tend to 0 and the variance terms are a function of the squared spectral density of the sampled process. To finish the second term we have to apply it to the integral with the results from (3.1.26). The function $Q(\lambda)=\delta(\lambda)-\Gamma(\lambda)$ is $Q \in L_1$. Realizing the integral has the following properties

$$\int_{R^2} Q(\lambda_1-u_1)Q(\lambda_2-u_2)du_1du_2 = \begin{cases} 1 & c_n = 0 \\ (1-\gamma(0))^2 & c_n \neq 0 \end{cases} \qquad (3.1.28)$$

the order remains the same as (3.1.26) and has asymptotic properties of (3.1.27).

The Cauchy-Scwartz inequality will be used to bound the last two terms of (3.1.21) with $Cov^2(a,b) \le Var(a)Var(b)$ which gives

$$Cov(\hat{f}_Z(\lambda),\hat{R}_x(0)) \le Var^{1/2}(\hat{R}_x(0))Var^{1/2}(\hat{f}_Z(\lambda)) \qquad (3.1.29)$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

This gives terms that are at most $O(T^{-1/2})$. Therefore, the most significant term remains the second based upon the fact that one part is not a function of the number of samples.

The above assumptions (A3.1.2)-(A3.1.4) with the preceding analysis produce the following bounded equation for the covariance of the estimates

$$cum(\hat{f}_X(\lambda_1), \hat{f}_X(\lambda_2)) = \begin{cases} O(T^{-1/2}) & |\lambda_1| \neq |\lambda_2| \\ \beta^{-4}|f_Z(\lambda_1)|^2 + O(T^{-1/2}) & |\lambda_1| = |\lambda_2| \neq 0 \\ 2\beta^{-4}|f_Z(\lambda_1)|^2 + O(T^{-1/2}) & |\lambda_1| = |\lambda_2| = 0 \end{cases} \qquad (3.1.30)$$

which, along with (3.1.14), identifies the result of this natural estimator as asymptotically unbiased but not a consistent estimator due to the variance. A way to obtain a consistent estimate follows by utilizing weighted sums and will be compared in the next sections with the derivation of wavelet estimators.

Kernel Estimates

The behavior of kernel estimators with respect to the conditions already stated in relation to spectral densities has been investigated in Lii and Masry, [1994] and is summarized here for comparison purposes. Given a weighting function that is real, even, and having the following properties

$$\begin{aligned} \int_R W(\lambda)d\lambda &= 1, \\ W &\in L_\infty \cap L_1. \end{aligned}$$

then the weight can be scaled according to a bandwidth parameter b_T such that $W_T(\lambda) = b_T^{-1}W(b_T^{-1}\lambda)$. As $T \to \infty$, $b_T \to 0$ we can convolve this weight with our previous estimates from equation (3.1.7) to obtain a new estimate of the form

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

$$\hat{f}_{Z,T}(\lambda) = \int_R W_T(\lambda - u)\hat{f}_Z(u)du. \tag{3.1.31}$$

With this new estimate we consider the revised estimate of $f_X(\lambda)$ given by

$$\hat{f}_{X,W}(\lambda) = \beta^{-2}\left[\hat{f}_{Z,T}(\lambda) - (2\pi)^{-1}\beta\hat{R}_X(0) - \int_R \Gamma(\lambda - u)\left[\hat{f}_{Z,T}(u) - (2\pi)^{-1}\beta\hat{R}_X(0)\right]du\right].$$

Bias of the Kernel Estimates

Using the same assumptions as before with those of the weight produces a very similar result for the bias

$$E\left(\hat{f}_{X,W}(\lambda)\right) = \int_R W_T(\lambda - u)f_X(\lambda)du + O\left(T^{-1}\right). \tag{3.1.32}$$

It follows that $E\left(\hat{f}_{X,W}(\lambda)\right) \xrightarrow{T\to\infty} f_X(\lambda)$ for all continuity points of $f_X(\lambda)$. And in particular should both the Fourier transform of the weighting function $w(t)$ and the spectral density $f_X(\lambda)$ be twice differentiable $\left\{w^{(2)}, f_X^{(2)}\right\}$, bounded and continuous the

$$bias\left[\hat{f}_{X,W}(\lambda)\right] = -2^{-1}b_T^2 w^{(2)}(0)f_X^{(2)}(\lambda) + o\left(b_T^2\right) + O\left(T^{-1}\right).$$

Thus this is independent of the statistics of the sampling process and in particular is not dependent upon the sampling rate β.

Consistency of the Kernel Estimate

Utilizing the same assumptions (A3.1.2)-(A3.1.4) from Lii and Masry, [1994] the covariance is obtained with

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

$$Tb_T Cov\left(\hat{f}_{X,W}(\lambda_1), \hat{f}_{X,W}(\lambda_2)\right) = \begin{aligned} & 2\pi\beta^{-4} f_Z^2(\lambda_1)\left[\delta_{\lambda_1,\lambda_2} + \delta_{\lambda_1,-\lambda_2}\right] \\ & \times \int_R W^2(x)dx + o(1) + O\left(\sqrt{b_T}\right) \end{aligned} \tag{3.1.33}$$

By letting $b_T \to 0$ and $Tb_T \to \infty$ as $T \to \infty$ then we have, with equation (3.1.32) a consistent estimate. Note: there is a great deal of similarity with respect to the weighting function in particular the bandwidth parameter utilized in obtaining consistent estimates and the wavelet scaling functions.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

3.2 First Order Cumulant Properties (Wavelet Coefficients)

We will start the derivation utilizing the conditions established in section 3.1 as well as using Assumptions A3.1.1 to A3.1.4. The same process is involved and we will simply add to this by recognizing a few more concepts that will help in deriving the statistical properties of the wavelet coefficients.

The purpose of using wavelet estimators has succinctly been stated in Donoho and Johnstone, [1995] which is the rationale for attempting to create and examine their characteristics in this framework. Substantial work in related areas to spectral density estimation under equally spaced sampling has been accomplished in Gao, [1993] and Neumann, [1996]. Poisson random sampling spectral density estimation has been previously done by Lehr and Lii, [1997]. We proceed with defining the wavelet expansion of our function in this setting with some assumptions on the wavelet basis functions. Let the wavelet expansion of the spectral density estimate defined from equation (3.1.7) be

$$\hat{f}_X(\lambda) = \sum_{k \in Z} \hat{\alpha}_k \phi_{l,k}(\lambda) + \sum_{j \geq l} \sum_{k \in Z} \hat{\alpha}_{j,k} \psi_{j,k}(\lambda), \{j, l; j \geq l\} \in Z \tag{3.2.1}$$

where the wavelet coefficient estimates are given by

$$\hat{\alpha}_k = \int_R \hat{f}_X(\lambda) \phi_{l,k}(\lambda) d\lambda, \tag{3.2.2}$$

$$\hat{\alpha}_{j,k} = \int_R \hat{f}_X(\lambda) \psi_{j,k}(\lambda) d\lambda. \tag{3.2.3}$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Note: The summations depicted in equation (3.2.1) begin at some fixed scale level (l). The functions $\{\phi,\psi\}$ are referred to as the father and mother wavelets, which generate an orthonormal basis set, scaled and offset in the following manner

$$\phi_{l,k}(\lambda)=2^{l/2}\phi(2^{l}\lambda-k), \tag{3.2.4}$$

$$\psi_{j,k}(\lambda)=2^{j/2}\psi(2^{j}\lambda-k). \tag{3.2.5}$$

Refer to section 2.2 for the scaling, shifting and multiresolution properties of this basis.

Our first assumption shall be that the wavelet functions are similar in nature to the spectral density, that is if $f_X(\lambda)\in C^m$ then

$$\begin{aligned}&\{\phi,\psi\}\in C^r, r\geq m,\\ &\int_R \psi(x)x^k dx=0, 0\leq k\leq r, where\\ &\int_R \phi(x)dx=1,\\ &\int_R \phi^2(x)dx=\int_R \psi^2(x)dx=1.\end{aligned} \tag{A3.2.1}$$

The existence and utilization of (A3.2.1) is presented in Belykin, et.al., [1991] which deals with the compression of large matrices and Neumann, [1996] with non-Gaussian spectral density estimation.

Father Wavelet Estimate

Taking the expectation and using the results from the previous section gives

$$\begin{aligned}cum(\hat{\alpha}_k) &= cum\left(\int_R \hat{f}_X(\lambda)\phi_{l,k}(\lambda)d\lambda\right)\\ &= \int_R cum\left(\hat{f}_X(\lambda)\right)\phi_{l,k}(\lambda)d\lambda\\ &= \int_R f_X(\lambda)\phi_{l,k}(\lambda)d\lambda+\int_R O(T^{-1})\phi_{l,k}(\lambda)d\lambda.\end{aligned}$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

We can use the definition of the wavelet coefficient for the left half of the integral and determine the order for the right half using simple identities from the Wavelet Basis Lemma in Appendix A. Since $\phi \in L_1$, uniform in frequency we have as a result with a change of variable the estimate for the father wavelet coefficient is

$$\begin{aligned} cum(\hat{\alpha}_k) &= \int_R f_X(\lambda)\phi_{l,k}(\lambda)d\lambda + 2^{l/2}\int_R O(T^{-1})\phi(2^l\lambda - k)d\lambda \\ &= \alpha_k + O(2^{-l/2}T^{-1}) \end{aligned} \tag{3.2.6}$$

This is an asymptotically unbiased estimate of the father wavelet coefficient if $2^{-l/2}T^{-1} \to 0 \quad as \quad T \to \infty$.

Mother Wavelet Estimate

Taking the expectation and using the results from the previous section for the mother wavelet gives

$$\begin{aligned} cum(\hat{\alpha}_{j,k}) &= cum\left(\int_R \hat{f}_X(\lambda)\psi_{j,k}(\lambda)d\lambda\right) \\ &= \int_R cum\left(\hat{f}_X(\lambda)\right)\psi_{j,k}(\lambda)d\lambda, \\ &= \int_R f_X(\lambda)\psi_{j,k}(\lambda)d\lambda + \int_R O(T^{-1})\psi_{j,k}(\lambda)d\lambda \end{aligned}$$

We can use the definition of the mother wavelet coefficient for the left half of the integral and then determine the order from the right half using simple concepts from the last section and a change of variable thereby obtaining the magnitude of the estimate

$$\begin{aligned} cum(\hat{\alpha}_{j,k}) &= \int_R f_X(\lambda)\psi_{j,k}(\lambda)d\lambda + 2^{j/2}\int_R O(T^{-1})\psi(2^j\lambda - k)d\lambda \\ &= \alpha_{j,k} + O(2^{-j/2}T^{-1}) \end{aligned} \tag{3.2.7}$$

This is an asymptotically unbiased estimate of the mother wavelet coefficient if $2^{-j/2}T^{-1} \to 0 \quad as \quad T \to \infty$.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

3.3 First Order Cumulant Properties (Spectral Density Estimate)

The previous section explored the statistical behavior of individual wavelet coefficients with the intention of determining the consistency of the spectral density estimate based upon a wavelet representation. However, equation (3.2.1) simply transforms the estimate into the wavelet domain. The norm is the same and therefore we have the same properties associated with (3.1.14 and 30) unless we utilize kernel techniques vis-à-vis (3.1.31 through 33). However, implementing a non-linear thresholding scheme discussed in section 2.3 will modify our intuitive estimate and develop a possible candidate for consistency due to the variance reduction examined in Section 2.4. Note: This modification will not change the statistical properties of the Wavelet coefficients already determined with equations (3.2.6) and (3.2.7).

Wavelet Shrinkage

Donoho, et.al., [1994, 1995], have developed substantial theory for non-parametric smoothing based on the principals of wavelet shrinkage. The rationale for shrinkage rests on:

1. Signal features are normally represented by a few wavelet coefficients,
2. Noise affects all wavelet coefficients, and
3. By shrinking wavelet coefficients towards zero the noise can be reduced while preserving desirable features of the original function.

Theoretical results show that non-linear shrinkage functions such as

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

$$\delta_c^{soft}(x) = \begin{cases} 0 & |x| \le c \\ \mathrm{sgn}(x)|x-c| & |x| > c, \end{cases}$$
$$\delta_c^{hard}(x) = \begin{cases} 0 & |x| \le c \\ x & |x| > c, \end{cases} \qquad (3.3.1)$$

when applied to wavelet coefficients thereby forming a new estimate $\hat{f}_{shrink}$ can almost achieve the minimax risk over a broad class of functions F

$$Risk(\hat{f}_{Shrink}, f) \approx \inf_{\hat{f}} \sup_{f \in F} Risk(\hat{f}, f)$$

The value of c is typically chosen proportional to the standard deviation σ_j of the wavelet coefficients at the largest scale j as $c = \sigma_j \lambda$. The constant of proportionality is based upon the works of Leadbetter, [1983] with $\lambda = \sqrt{2 \log T}$. The application of the shrinkage occurs by applying the wavelet transform of the periodogram, modifying some of the coefficient values by thresholding and then inverting the transform. This is depicted in the following diagram (Figure 3.3.1). The optimal result refers to the fact that $\hat{f}_{Shrink}$ gives nearly the best possible estimate of $\hat{f}$ in the mean squared sense making a minimum number of assumptions about the underlying nature of the spectral density.

For our particular problem we can define a scale J(T)+1 which is O(logT) whereupon we apply shrinkage to the wavelet coefficients. The purpose of choosing a scale that is a function of T is related to allowing a function $2^{J(T)}/T \to 0$ as $T \to \infty$ as

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

suggested in the previous section. We start this process by considering 3 functions. The original transformed spectral density function

$$f_X(\lambda) = \sum_{k \in Z} \alpha_k \phi_{l,k}(\lambda) + \sum_{j \geq l} \sum_{k \in Z} \alpha_{j,k} \psi_{j,k}(\lambda),$$

the thresholded function

$$f_{J(T)}(\lambda) = \sum_{k \in Z} \alpha_k \phi_{l,k}(\lambda) + \sum_{j=l}^{J(T)} \sum_{k \in Z} \alpha_{j,k} \psi_{j,k}(\lambda), \tag{3.3.2}$$

and the error term

$$\varepsilon_{J(T)}(\lambda) = \sum_{j > J(T)} \sum_{k \in Z} \alpha_{j,k} \psi_{j,k}(\lambda). \tag{3.3.3}$$

Using Parseval's Theorem we obtain

$$\|f_X\|^2 = \|f_{J(T)}\|^2 + \|\varepsilon_{J(T)}\|^2, \tag{3.3.4}$$

where

$$\|f_X\|^2 = \sum_{k \in Z} \alpha_k^2 + \sum_{j \geq l} \sum_{k \in Z} \alpha_{j,k}^2 ,$$

$$\|f_{J(T)}\|^2 = \sum_{k \in Z} \alpha_k^2 + \sum_{j \geq l}^{J(T)} \sum_{k \in Z} \alpha_{j,k}^2 , \text{ and}$$

$$\|\varepsilon_{J(T)}\|^2 = \sum_{j > J(T)} \sum_{k \in Z} \alpha_{j,k}^2 .$$

By definition, the energy of the signal is finite meaning $f_X \in L_2 \Rightarrow f_{J(T)}, \varepsilon_{J(T)} \in L_2$ since we assumed previously that $f_X \in L_1 \cap L_\infty$. We will additionally posit that the

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

spectral density function belongs to a Sobolev class, which formalizes an assumption made in Section 3.2 with (A3.2.1). The assumption is $f_X \in W_2^m(c)$ where

$$W_p^m(c) = \left\{ f : \int_R \left| f^{(m)}(\lambda) \right|^p d\lambda \le c^p \right\}. \tag{A3.3.1}$$

The value of the threshold $c = (\lambda \sigma_j | j > J(T))$ commonly found in the literature and discussed in Section 2 is not necessary to determine the order of magnitude of the error at this point so we let $c = \max(\alpha_{j,k} | j > J(T))$. This allows us to define a slightly modified estimate

$$\hat{f}_{J(T)} = \hat{f}_X(\lambda) - \sum_{j>J(T)} \sum_{k \in Z} \alpha_{j,k} \psi_{j,k}(\lambda). \tag{3.3.5}$$

The order of magnitude associated with this threshold is found by examining the terms at scale $J(T)+1$ for $\hat{\alpha}_{J(T)+1,k}$. Recalling that $cum(\hat{\alpha}_{j,k}) = \alpha_{j,k} + O(2^{-j/2} T^{-1})$ by equation (3.2.7). Now returning to the problem and limiting the sum as above

$$\hat{f}_{J(T)}(\lambda) = \sum_{k \in Z} \hat{\alpha}_k \phi_{l,k}(\lambda) + \sum_{j=l}^{J(T)} \sum_{k \in Z} \hat{\alpha}_{j,k} \psi_{j,k}(\lambda). \tag{3.3.6}$$

Using a Taylor series expansion with remainder $\Re(x)$ for our function gives

$$\begin{aligned} f(x) = f(\Delta) + f^{(1)}(\Delta)(x-\Delta) + f^{(2)}(\Delta)\frac{(x-\Delta)^2}{2!} + \cdots \\ + f^{(m)}(\Delta)\frac{(x-\Delta)^m}{m!} + (x-\Delta)^{m+1}\Re(x). \end{aligned}$$

Then consider the coefficient

$$\alpha_{j,k} = \langle f_X, \psi_{j,k} \rangle = \int_R f_X(x) \psi_{j,k}(x) dx$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

which can be expanded with our previous assumptions to

$$\alpha_{J,k} = \int_R f_X(\Delta)\psi_{J,k}(x)dx + \ldots + \int_R (x-\Delta)^{(m+1)}\Re(x)\psi_{J,k}(x)dx .$$

Now with

$$\begin{array}{lccccccc} & \psi_{J,k}(x) & = & 2^{J/2}\psi_{J,k}(2^J x - k) & & & & \\ let & x & = & 2^{-J}t & \Rightarrow & dx & = & 2^{-J}dt \\ and & \Delta & = & 2^{-J}k & \Rightarrow & f(\Delta) & = & f(2^{-J}k) \end{array}$$

which gives the expanded version of the Fourier coefficient to be

$$\begin{aligned} \alpha_{J,k} = \; & f_X(2^{-J}k)\int_R 2^{-J/2}\psi(t-k)dt + \\ & f_X^{(1)}(2^{-J}k)\int_R 2^{-J/2}\frac{(t-k)}{2^J}\psi(t-k)dt + \\ & \vdots \; + \\ & f_X^{(m)}(2^{-J}k)\int_R 2^{-J/2}\frac{(t-k)^m}{m!2^{Jm}}\psi(t-k)dt + \\ & \int_R 2^{-J/2}\frac{(t-k)^{m+1}}{2^{J(m+1)}}\Re\left(\frac{t}{2^J}\right)\psi(t-k)dt. \end{aligned}$$

Using assumption A3.2.1 gives

$$\int_R t^r\psi(t)dt = 0 \quad \forall \quad 0 \le r \le m$$

we have

$$\alpha_{J,k} = 2^{-J(m+3/2)}\int_R \Re\left(\frac{t}{2^J}\right)(t-k)^{m+1}\psi(t-k)dt .$$

In addition, if the remainder is bounded uniformly with λ with the m+1 moments of the mother wavelet

$$\int_R \Re(2^{-J}t)(t-k)^{m+1}\psi(t-k)dt = O(1) \qquad \text{(A3.3.2)}$$

which is a reasonable assumption then

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

$$\alpha_{j,k} \quad = \quad O\left(2^{-j(m+3/2)}\right).$$

We already know by Parseval's Theorem that all the sum of squares of the coefficients are bounded in totality and thereby at each scale j such that

$$\sum_{j>J(T)}\sum_{k\in Z}\alpha_{j,k}^2 = \sum_{j>J(T)} O\left(2^{-2j(m+3/2)}\right)$$

and the term with the largest order of magnitude occurs when $j = J(T)+1$. The conclusion reached at this point is that the order of the approximation given the Sobolev class previously defined of spectral density functions with matched wavelet basis is

$$\left\|\varepsilon_{J(T)}\right\|^2 \quad = \quad \sum_{j>J(T)}\sum_{k\in Z}\alpha_{j,k}^2 \quad = \quad O\left(2^{-2(J(T)(m+3/2)}\right) \qquad (3.3.7)$$

Now we can return to estimating the order of the error, given that we have the truncation error to the function

$$cum\left(\hat{f}_{J(T)}(\lambda)\right) = cum\left(\sum_{k\in Z}\hat{\alpha}_k\phi_{l,k}(\lambda) + \sum_{j=l}^{J(T)}\sum_{k\in Z}\hat{\alpha}_{j,k}\psi_{j,k}(\lambda)\right).$$

Using the previous results (3.2.6,.7) for the expectation of the wavelet coefficients and in conjunction that the error that are uniform in frequency then

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

$$
\begin{aligned}
cum\left(\hat{f}_{J(T)}(\lambda)\right) &= cum\left(\sum_{k\in Z}\hat{\alpha}_k\phi_{l,k}(\lambda)+\sum_{j=l}^{J(T)}\sum_{k\in Z}\hat{\alpha}_{j,k}\psi_{j,k}(\lambda)\right) \\
&= \sum_{k\in Z}cum(\hat{\alpha}_k)\phi_{l,k}(\lambda)+\sum_{j=l}^{J(T)}\sum_{k\in Z}cum(\hat{\alpha}_{j,k})\psi_{j,k}(\lambda) \\
&= \sum_{k\in Z}\alpha_k\phi_{l,k}(\lambda)+\sum_{j=l}^{J(T)}\sum_{k\in Z}\alpha_{j,k}\psi_{j,k}(\lambda)+ \\
&\quad \sum_{k\in Z}O\left(2^{-l/2}T^{-1}\right)\phi_{l,k}(\lambda)+\sum_{j=l}^{J(T)}\sum_{k\in Z}O\left(2^{-j/2}T^{-1}\right)\psi_{j,k}(\lambda)
\end{aligned}
$$

(3.3.8)

Using the Wavelet Basis Lemma (A.1.23 & 24) this can be reduced with our previous definitions simplifying to

$$
E\left(\hat{f}_{J(T)}(\lambda)\right)=f(\lambda)-\varepsilon_{J(T)}+O\left(T^{-1}\right)\sum_{k\in Z}\phi\left(2^{l/2}\lambda-k\right)+\sum_{j=l}^{J(T)}O\left(T^{-1}\right)\sum_{k\in Z}\psi\left(2^{j/2}\lambda-k\right). \tag{3.3.9}
$$

The magnitude of the estimate has been derived

$$
E\left(\hat{f}_{J(T)}(\lambda)\right)=f(\lambda)+O\left(2^{-J(T)(m+3/2)}\right)+O(((J(T)+2)-l)/T). \tag{3.3.10}
$$

Making a comparison with (3.1.12) we see that there is a term with O(1/T) which is what would be expected. The additional components of J(T) is associated with the number of scale summations. The added bias in equation (3.3.9) due to thresholding is small by comparison. The total bias of this approximation asymptotically tends to 0 as $J(T)/T \to 0$ and $T \to \infty$. In the next section we will examine the covariance properties.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

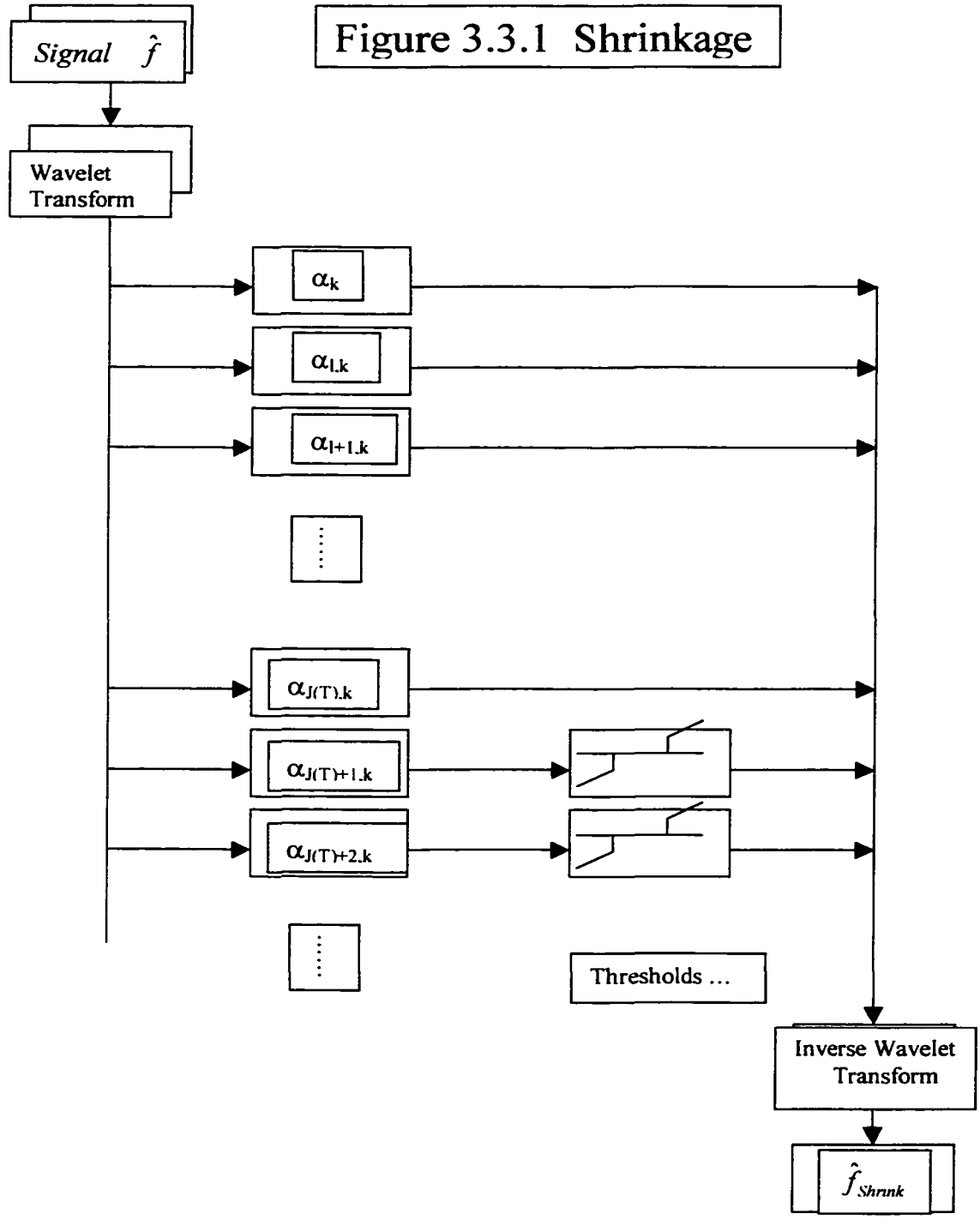


Figure 3.3.1 Shrinkage

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

3.4 Second Order Cumulant Properties (Wavelet Coefficients)

We developed a new estimate based on wavelets in the last section, which gives control over the number of coefficients in the wavelet expansion. Equation (3.3.10) showed that the bias would be small and can be controlled. In this section we calculate the second order cumulant properties of the coefficients. Once this is determined then the final analysis can be made as to the consistency of the new estimate defined in equation (3.3.6).

First, the covariance of the wavelet coefficient estimates must be derived. The starting point begins by recognizing there will be three different types of second order cumulants to investigate which are related to the scale and location of the coefficients

$$cum(\hat{\alpha}_k, \hat{\alpha}_{k'}), \quad cum(\hat{\alpha}_k, \hat{\alpha}_{j'k'}), \quad and \quad cum(\hat{\alpha}_{j,k}, \hat{\alpha}_{j'k'})$$

However, the derivation for any of the above will follow the same structure and the solutions will be very similar. We will show the solution containing only the mother wavelets coefficients to limit a fair amount of duplication. The covariance of the mother wavelet term is by definition

$$cum(\hat{\alpha}_{j,k}, \hat{\alpha}_{j'k'}) = \int_{R^2} cum(\hat{f}_x(\lambda_1), \hat{f}_x(\lambda_2)) \psi_{j,k}(\lambda_1) \psi_{j'k'}(\lambda_2) d\lambda_1 d\lambda_2 . \qquad (3.4.1)$$

Using (3.1.21) we can rewrite (3.4.1) in the following way

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

$$
\begin{aligned}
cum(\hat{\alpha}_{j,k},\hat{\alpha}_{j',k'}) &= \int_{R^2} [\,(2\pi\beta)^{-2}[1-\gamma(0)]^2 Var(\hat{R}_X(0)) \\
&\quad + \beta^{-4}\int_{R^2} Q(\lambda_1-u_1)Q(\lambda_2-u_2)cum(\hat{f}_Z(u_1),\hat{f}_Z(u_2))du_1du_2 \\
&\quad + (2\pi\beta)^{-1}\int_R Q(\lambda_1-u)cum(\hat{f}_Z(u),\hat{R}_X(0))du \\
&\quad + (2\pi\beta)^{-1}\int_R Q(\lambda_2-u)cum(\hat{f}_Z(u),\hat{R}_X(0))du\,]\psi_{j,k}(\lambda_1)\psi_{j',k'}(\lambda_2)d\lambda_1d\lambda_2, \\
&= T_1+T_2+T_3+T_4.
\end{aligned}
\tag{3.4.2}
$$

We will follow the procedure used in Section 3.1 with the first term in the above equation

$$
T_1 = \int_{R^2}(2\pi\beta)^{-2}[1-\gamma(0)]^2 Var(\hat{R}_X(0))\psi_{j,k}(\lambda_1)\psi_{j',k'}(\lambda_2)d\lambda_1d\lambda_2. \tag{3.4.3}
$$

We know from equation (3.1.23) that this term before the integration was $O(T^{-1})$ and when integrated with the mother wavelet as above will be exactly $O(\delta_{j,j'}\delta_{k,k'}/T)$ using Appendix A.1.7.

The fourth term is the most important because of it's magnitude which we have seen before in Section 3.1 and integrated here as

$$
T_4 = \beta^{-4}\int_{R^2} Q(\lambda_1-u_1)Q(\lambda_2-u_2)\psi_{j,k}(\lambda_1)\psi_{j',k'}(\lambda_2)Cum(\hat{f}_Z(u_1),\hat{f}_Z(u_2))du_1du_2d\lambda_1d\lambda_2. \tag{3.4.4}
$$

We develop the following equations utilizing the function Q(x) as in (3.1.15)

$$
\psi^Q_{j,k}(u) \equiv \int_R Q(\lambda-u)\psi_{j,k}(\lambda)d\lambda = \psi_{j,k}(\lambda) - \int_R \Gamma(\lambda-u)\psi_{j,k}(\lambda)d\lambda \quad \text{and} \tag{3.4.5}
$$

$$
\phi^Q_{l,k}(u) \equiv \int_R Q(\lambda-u)\phi_{l,k}(\lambda)d\lambda = \phi_{l,k}(\lambda) - \int_R \Gamma(\lambda-u)\phi_{l,k}(\lambda)d\lambda. \tag{3.4.6}
$$

Since $\phi_{l,k}\psi_{j,k} \in L_\infty \cap L_1$, and $\Gamma \in L_1$ then $\phi^Q_{l,k}, \psi^Q_{j,k} \in L_\infty \cap L_1$. The functions are uniformly bounded in frequency with

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

$$\psi_{j,k}^{Q}(\lambda)=\psi_{j,k}(\lambda)+O\left(2^{-j/2}\right), \tag{3.4.7}$$

$$\phi_{l,k}^{Q}(\lambda)=\phi_{l,k}(\lambda)+O\left(2^{-l/2}\right). \tag{3.4.8}$$

Utilizing equations (3.4.5 and 6) gives the following

$$T_4=\beta^{-4}\int_{R^2}\psi_{j,k}^{Q}(\lambda_1)\psi_{j',k'}^{Q}(\lambda_2)cum\left(\hat{f}_Z(\lambda_1),\hat{f}_Z(\lambda_2)\right)d\lambda_1 d\lambda_2$$

which in expanded form is

$$\begin{aligned}T_4 &= \beta^{-4}\int_{R^2}\psi_{j,k}^{Q}(u_1)\psi_{j',k'}^{Q}(u_2)\Big[\Big[2\pi T^{-1}f_Z^{(4)}(u_1,-u_1,u_2)\Big]\\ &\quad+T^{-1}|f_Z(u_1)|^2\left[\Delta_T(u_1+u_2)+\Delta_T(-u_1+u_2)\right]\\ &\quad+O(1)\frac{f_Z(u_1)}{2\pi T^2}\left[D_T(u_1+u_2)+D_T(u_1-u_2)\right]\\ &\quad+O(1)\frac{f_Z(-u_1)}{2\pi T^2}\left[D_T(-u_1-u_2)+D_T(-u_1+u_2)\right]\\ &\quad+O(1)(2\pi T^2)^{-1}\Big]du_1du_2\\ &= T_{4,2}+T_{4,2}+T_{4,2}+T_{4,2}+T_{4,2}.\end{aligned} \tag{3.4.9}$$

The first term $T_{4,1}$ of this expansion is $O\left(2^{-(j+j')/2}T^{-1}\right)$ since $f_Z^{(4)}(u_1,-u_1,u_2)\in L_1$. The third and fourth term $T_{4,3}$ and $T_{4,4}$ are of the same order of magnitude since the spectral density f_Z and D_T/T are uniformly bounded. The fifth term is obviously of smaller magnitude compared to all others discussed so far which leaves the second term $T_{4,2}$ to be dealt with. Utilizing the Fejer Kernel Lemma found in Appendix A.2

$$\int_R\psi_{j,k}^{Q}(\lambda)\Delta_T(\lambda)d\lambda=2\pi\psi_{j,k}^{Q}(0)+o\left(2^{j/2}\right) \tag{3.4.10}$$

will result in

$$\frac{2\pi}{\beta^4 T}\int_R|f_Z(\lambda)|^2\left(\psi_{j',k'}^{Q}(\lambda)+\psi_{j',k'}^{Q}(-\lambda)\right)\psi_{j,k}^{Q}(\lambda)d\lambda+\frac{o\left(2^{j/2}\right)}{\beta^4 T}\int_R|f_Z(\lambda)|^2\psi_{j,k}^{Q}(\lambda)d\lambda$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

(3.4.11)

As before, with the cross-variance terms T_2, T_3 we can use the Cauchy -Schwartz inequality.

The final result leads to the three 2nd order cumulant estimates

$$cum(\hat{\alpha}_k, \hat{\alpha}_{k'}) = 2\pi\beta^{-4}T^{-1}\int_R |f_Z(\lambda)|^2 \times (\phi^Q_{l,k}(\lambda) + \phi^Q_{l,k}(-\lambda))\phi^Q_{l,k'}(\lambda)d\lambda + o(T^{-1}) + O(2^{-l}T^{-1}), \quad (3.4.12)$$

$$cum(\hat{\alpha}_k, \hat{\alpha}_{j'k'}) = 2\pi\beta^{-4}T^{-1}\int_R |f_Z(\lambda)|^2 \times (\psi^Q_{j',k'}(\lambda) + \psi^Q_{j',k'}(-\lambda))\phi^Q_{l,k}(\lambda)d\lambda + o(2^{(l-j')/2}T^{-1}) + O(2^{-(j'+l)/2}T^{-1}) \quad (3.4.13)$$

$$cum(\hat{\alpha}_{j,k}, \hat{\alpha}_{j'k'}) = 2\pi\beta^{-4}T^{-1}\int_R |f_Z(\lambda)|^2 \times (\psi^Q_{j',k'}(\lambda) + \psi^Q_{j',k'}(-\lambda))\psi^Q_{j,k}(\lambda)d\lambda + o(2^{(j-j')/2}T^{-1}) + O(2^{-(j'+j)/2}T^{-1}) \quad (3.4.14)$$

These are consistent estimates of the wavelet coefficients if we let $2^j T^{-1} \to 0$ as $T \to \infty$ with $0 \le l \le j \le j'$ without loss of generality. The first term in each of the 3 equations are uniformly bounded across λ with $f_z \in L_\infty$.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

3.5 Second Order Cumulant Properties (Spectral Density Estimate)

In this section we calculate the statistical properties of the estimate defined in Section 3.3 related to it's second order cumulant. The conditions from Section 3.1 apply as well as the assumptions from Sections 3.1 through 3.4. Recalling the form of the estimate from (3.3.2)

$$\hat{f}_{J(T)}(\lambda)=\sum_{k\in Z}\hat{\alpha}_k\phi_{l,k}(\lambda)+\sum_{j>l}^{J(T)}\sum_{k\in Z}\hat{\alpha}_{j,k}\psi_{j,k}(\lambda)$$

The covariance of this estimate is

$$\begin{aligned} cum\left(\hat{f}_{J(T)}(\lambda),\hat{f}_{J(T)}(\omega)\right) &= \sum_{k\in Z}\sum_{k'\in Z}\phi_{l,k}(\lambda)\phi_{l,k'}(\omega)cum(\hat{\alpha}_k,\hat{\alpha}_{k'}) \\ &\quad+\sum_{j=l}^{J(T)}\sum_{k\in Z}\sum_{k'\in Z}\psi_{j,k}(\lambda)\phi_{l,k'}(\omega)cum(\hat{\alpha}_{j,k},\hat{\alpha}_{k'}) \\ &\quad+\sum_{j'=l}^{J(T)}\sum_{k'\in Z}\sum_{k\in Z}\psi_{j',k'}(\lambda)\phi_{l,k}(\lambda)cum(\hat{\alpha}_{j',k'},\hat{\alpha}_k) \\ &\quad+\sum_{j=l}^{J(T)}\sum_{j'=l}^{J(T)}\sum_{k\in Z}\sum_{k'\in Z}\psi_{j,k}(\lambda)\psi_{j',k'}(\omega)cum(\hat{\alpha}_{j,k},\hat{\alpha}_{j',k'}) \\ &= T_1+T_2+T_3+T_4. \end{aligned} \tag{3.5.1}$$

We observe that the 2nd and 3rd term of this expansion are duplicates so there are only 3 unique terms.

First recall the covariance terms from equations (3.4.12, 13, and 14). We know from equations (3.4.7, and 8) that the mother and father wavelet are significant terms and utilizing A.1.19 through 21 bounds the integral of these components. Therefore, all the individual cumulant terms are of $O(T^{-1})$ or smaller. Now to sum all the terms in the expansion with

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

$$
\begin{aligned}
cum\left(\hat{f}_{J(T)}(\lambda),\hat{f}_{J(T)}(\omega)\right) &= \sum_{k\in Z}\sum_{k'\in Z}\phi_{l,k}(\lambda)\phi_{l,k'}(\omega)\begin{pmatrix}O(T^{-1})+\\ o(T^{-1})+O(2^{-l}T^{-1})\end{pmatrix}\\
&+\sum_{j=l}^{J(T)}\sum_{k\in Z}\sum_{k'\in Z}\psi_{j,k}(\lambda)\phi_{l,k'}(\omega)\begin{pmatrix}O(T^{-1})+\\ o(2^{(l-j)/2}T^{-1})+\\ O(2^{-(j+l)/2}T^{-1})\end{pmatrix}\\
&+\sum_{j'=l}^{J(T)}\sum_{k'\in Z}\sum_{k\in Z}\psi_{j',k'}(\lambda)\phi_{l,k}(\lambda)\begin{pmatrix}O(T^{-1})+\\ o(2^{(l-j')/2}T^{-1})+\\ O(2^{-(j'+l)/2}T^{-1})\end{pmatrix}\\
&+\sum_{j=l}^{J(T)}\sum_{j'=l}^{J(T)}\sum_{k\in Z}\sum_{k'\in Z}\psi_{j,k}(\lambda)\psi_{j',k'}(\omega)\begin{pmatrix}O(T^{-1})+\\ o(2^{(j-j')/2}T^{-1})+\\ O(2^{-(j'+j)/2}T^{-1})\end{pmatrix}.
\end{aligned}
\tag{3.5.2}
$$

Utilizing A.1.24 and A.1.25 further simplifies to

$$
\begin{aligned}
cum\left(\hat{f}_{J(T)}(\lambda),\hat{f}_{J(T)}(\omega)\right) &= \begin{pmatrix}O(2^{l}T^{-1})+\\ o(2^{l}T^{-1})+O(T^{-1})\end{pmatrix}\\
&+2\sum_{j=l}^{J(T)}\begin{pmatrix}O(2^{(j+l)/2}T^{-1})+\\ o(2^{l}T^{-1})+\\ O(T^{-1})\end{pmatrix}\\
&+\sum_{j=l}^{J(T)}\sum_{j'=l}^{J(T)}\begin{pmatrix}O(2^{(j+j')/2}T^{-1})+\\ o(2^{j}T^{-1})+\\ O(T^{-1})\end{pmatrix}.
\end{aligned}
\tag{3.5.3}
$$

Therefore collecting significant terms of the covariance estimate and defining a common element $J'(T)=J(T)+1-l$ gives

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

$$cum\left(\hat{f}_{J(T)}(\lambda),\hat{f}_{J(T)}(\omega)\right)=O\left(2^{J'(T)}T^{-1}\right)+O\left(J'(T)^2T^{-1}\right)+o\left(J'(T)2^{J'(T)}T^{-1}\right).$$

(3.5.4)

The 2nd order cummulant asymptotically tends to 0 as $2^{J(T)}T^{-1}\to 0$ and $T\to\infty$.

Summarizing

Theorem 3.5.1 Given the conditions of the process and the sampling scheme described in Section 3.1 with assumptions A3.1.1, A3.2.1, and A3.3.1,.2

$$E\left(\hat{f}_{J(T)}(\lambda)\right)=f_X(\lambda)+O\left(2^{-J(T)(m+3/2)}\right)+O\left((J'(T)+1)T^{-1}\right)$$

The bias of this approximation asymptotically tends to 0 as $J(T)/T\to 0$ and $T\to\infty$.

Theorem 3.5.2 In addition to the conditions and assumptions of Theorem 3.5.1 we add assumptions A3.1.2,.3,.4 to obtain

$$\begin{aligned} cum\left(\hat{f}_{J(T)}(\lambda),\hat{f}_{J(T)}(\omega)\right) &= \frac{2\pi}{\beta^4 T}\int_R |f_Z(u)|^2 C(\lambda,\omega,u)du \\ &\quad + o\left(J'(T)2^{J'(T)}T^{-1}\right) \\ &\quad + O\left((J'(T))^2T^{-1}\right) \end{aligned}$$

$C(\lambda,\omega,u)\in L_1=O\left(2^{J'(T)}\right)$ and the 2nd order cummulant asymptotically tends to 0 as $2^{J(T)}T^{-1}\to 0$ and $T\to\infty$.

Theorem 3.5.3 The estimate $\hat{f}_{J(T)}(\lambda)$ is a consistent estimate of $f_X(\lambda)$ given Theorems 3.5.1, 3.5.2 and by letting $2^{J(T)}=T^{\alpha},0<\alpha<1$ will achieve a rate of convergence $O\left(T^{\frac{-2m}{1+2m}}\right)$.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

4 Simulated Studies

The following section was generated by simulating randomly sampled time series data. This was done in order to compare methods that estimate the power spectral density and to illustrate the relevant aspects of the analysis.

Several programming languages and statistical packages were used to prepare this section. The applications and libraries of significance are found in Splus, Splus-Wavelets, Matlab, and the Matlab-Toolbox. These packages have specific capabilities associated with vectorizing equations, which made the process of simulating the functions at least tractable. The vectorization in Matlab made it possible to develop and display some of the asymptotic features found in the following sections without spending an inordinate amount of time waiting for the results.

All the figures are located at the end of this section. There are a number of subplots per figure and the reference letter of the subplot termed a panel follows a left to right, top to bottom orientation. Using this ordering convention, if a figure has 3 subplots then panel A will be the uppermost plot and panel C will be at the bottom of the page with panel B in the middle. If there are 6 subplots on the page with 3 rows and 2 columns then panel A will be the upper left plot and panel F will be the lower right plot.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

4.1 Scale and Signal Energy

Given a function f(x), it is possible to decompose the function into wavelet components, which represent the signal strength across scale. This is analogous to transforming a time domain signal into the frequency domain. By examining the periodogram we can determine the amount of energy that exists in the signal within a given bandwidth. Unlike the frequency domain, which is represented globally by the periodogram, the wavelet representation provides localization capability. The scale however is directly analogous to the frequency (i.e. the smaller the scale, the higher the frequency of the signal at this specific point in the domain).

To represent the energy contained in a signal, it is useful to define a common engineering term as the SNR. The SNR represents the Signal to Noise Ratio defined as

$$SNR_{dB} = 20\log\left(\frac{\mu}{\sigma}\right)$$

The resultant logarithmic term is coined as dB standing for decibel. The original meaning of this term is buried in the past but it has a lot to do with vocal transmission/communication and the fact that hearing is a logarithmic function. The mean (μ) of the signal is used in this case with the standard deviation (σ) of the noise. This is by no means the only definition of SNR but it is not uncommon. Sometimes the peak signal is used which might be a better definition if one is to discern peak signal strength.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Now to investigate variations in SNR and its implications associated with scale and signal compression. Initially, we can hypothesize a constant added to white noise that has a relatively large amount of signal as compared to the noise. Figure 4.1 shows a 20 dB SNR realization. We take roughly 512 samples from an $IID(\mu = 10, \sigma = 1)$ and transform into the wavelet domain. The following 2 equations represent this transformation. The first equation is in component form and the second is in vector form.

$$\begin{aligned} f(x) &= \sum_{k \in R} \alpha_k \phi_k(x) + \sum_{j \geq l}^{J} \sum_{k \in R} \alpha_{j,k} \psi_{j,k}(x) \\ \underline{f} &= \underline{s}_J + \sum_{j \geq l}^{J} \underline{d}_j \end{aligned}$$

The scale spans $[l, J]$, which is a different ordering presented in Section 3 and utilizes a slightly different definition of the wavelets than what we have presented before by

$$\phi_k(\lambda) = 2^{-J/2} \phi\left(2^{-J}\lambda - k\right),$$

$$\psi_{j,k}(\lambda) = 2^{-j/2} \psi\left(2^{-j}\lambda - k\right).$$

We can determine the amount of signal at each scale by utilizing a dot chart with the energy at scale defined in the following way

$$E_j^d = \frac{1}{E} \sum_{k=1}^{n/2^j} d_{j,k}^2 \qquad j = 1,2,\ldots J, \qquad E_J^s = \frac{1}{E} \sum_{k=1}^{n/2^J} s_{J,k}^2, \qquad E = E_J^s + \sum_{j=1}^{J} E_j^d.$$

And, we can compare the energy of the original signal with that of the ordered wavelet coefficients with

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

$$E_{DWT}(n) = \frac{1}{E}\sum_{i=1}^{n} \alpha_{(i)}^2 \qquad i = 1,2,....N,$$

$$E_{Data}(n) = \frac{1}{E}\sum_{i=1}^{n} x_i^2 \qquad i = 1,2,....N.$$

All the above definitions can be found in the Splus-Wavelet package. The father and mother wavelet utilized in the analysis in this section are the symmlet. Returning back to Figure 4.1, we point out that the total number of coefficients is 512 which is equivalent to the number of samples simulated and the number of coefficients at each scale are

$$\begin{aligned} \underline{d}_1 &\Rightarrow 256 \\ \underline{d}_2 &\Rightarrow 128 \\ \underline{d}_3 &\Rightarrow 64 \\ \underline{d}_4 &\Rightarrow 32 \\ \underline{d}_5 &\Rightarrow 16 \\ \underline{d}_6 &\Rightarrow 8 \\ \underline{s}_6 &\Rightarrow 8 \end{aligned}$$

Panel A simply plots the signal and it is relatively easy to see that the mean is 10 and the standard deviation is about 1. Panel B exhibits the dot-chart and points out that almost all of the energy is contained in roughly 8 coefficients of $\underline{s}_6$. A minor amount is located in $\underline{d}_1$. This is not surprising since the s coefficients represent the low pass filter sum of the energy which contain the mean of the entire vector whereas the mean of the d vectors are zero. The signal to noise is such that the relative energy in the differencing coefficients d represents most of the noise. Panel C depicts the energy

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

plots which shows that most of the signal is contained in 8 coefficients which from panel B could only be $\underline{s}_6$. We can literally throw away all other coefficients and approximate the original signal very accurately with just s.

The next 2 figures 4.2, and 4.3 show similar information as previously described in the first figure. We can see that the signal to noise ratio is decreased to 0 then to -20 db. As we decrease the signal, the energy is transferred from the s coefficients to the d coefficients such that the energy represents 50% in each for figure 4.2. We could effectively take the previous 8 coefficients to approximate the signal and the results would be a significant decrease in noise as outlined in Section 2.4.

The last figure 4.3 is meant to point out that white noise transformed to the wavelet domain remains white. With a negative 20 db we see that when pure noise is passed to the wavelet transformation we would expect to see pure noise as an output. There are no compressive capabilities, meaning that the signal is spread equally across all coefficients. The same variance reduction could be accomplished with this signal as previously discussed.

4.2 The ARMA(2,2) process

The next figures examine the spectrum of a time series process composed of an ARMA(2,2) superpositioned with Gaussian White Noise. The process has been randomly sampled f_Z and reconstructed f_X. It is used extensively in this dissertation to examine relevant properties. The signal has peaks and troughs in the frequency

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

domain as well as near constant areas across the bandwidth. We let $\{\varepsilon_t, Z_t\} \in WN(0,1)$. The sequence is defined by

$$Y_t + a_1 Y_{t-1} + a_2 Y_{t-2} = b_0 \varepsilon_t + b_1 \varepsilon_{t-1} + b_2 \varepsilon_{t-2}$$

then white noise was added to obtain

$$X_t = Y_t + c_0 Z_t .$$

We let $\{a_1, a_2\} = \{0.2, 0.9\}$, $\{b_0, b_1, b_2\} = \{1,0,1\}$, and c = 0.5. The function is found in Neumann, [1996] and was utilized here since it contained peaks in the power spectrum as well as relatively flat areas.

Figures 4.5 through 4.7 examine the energy contained in theoretical functions of the power spectrum as well as a single realization. The theoretical functions show that most of the energy can be contained in a few number of coefficients somewhat similar to the 20db SNR previously discussed. It is also apparent that the vectors $\underline{d}_1, \underline{d}_2$ contain no apparent signal energy while $\underline{d}_3, \underline{d}_4$ contain some of the signal. The realizations on the other hand are more similar to the 0db SNR figure. Naturally, their energy is spread across the spectrum.

Normally we would not know the information discussed in the previous paragraph but it does remind us that when comparing the merits of one technique to another that certain results well depend on the underlying signal involved in the analysis.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Figures 4.8 and 4.9 show f_X, and an estimate $\hat{f}_X$ with their corresponding components in each of the wavelet vectors. This is what one realization looks like before and after thresholding. The largest 8 coefficients in $\underline{d}_1, \underline{d}_2, \underline{d}_3, \underline{d}_4$ were kept to show the smoothing effect as well as the peaks that remain distinct and sharp.

4.3 Comparative Simulations

Figures 4.10 to 4.25 show a multitude of data and results. Two figures will be discussed in detail then the results will be summarized in the next section. Basically, kernel techniques will be contrasted with wavelet techniques. The kernel technique will utilize 4 bandwidths, which are 1, 2, 3,and 4% of the span. The wavelet techniques will utilize 4 scales. The first technique will incorporate truncation for 1 through 4 scales. The next wavelet technique will incorporate thresholding across 1 through 4 scales. The last technique will compare block thresholds for 1% of the span across 1 through 4 scales. In all these, uniform and random sampling results will be contrasted with each other.

Figure 4.10 has 3 subplots. Panel A shows the theoretical power spectral density for uniformly and randomly sampled data. The point process utilized is Poisson. These theoretical curves will be utilized in calculating the mean square error and plotted against the estimate of the functions for each of the 4 techniques contrasted in this section. Panel B compares the mean squared estimation error for 3 curves spanning 10 to 50 simulations of the power spectral density. The largest error depicted represents

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

the simple average of the simulation runs. The mean squared error is calculated for the simple average. The curve with the circle o data points represents the 1% kernel smoothed estimate that is then averaged for the number of sample runs. The next curve with + data points represents the 2% kernel smoothed estimate. Panel C contains the same information as panel B but for Randomly spaced data.

Figure 4.11 contains 6 subplots that represent 50 realizations that have been averaged and compared to the theoretical results. Panels A, C, E correspond to 3 points from panel B in Figure 4.10. Similarly, panels B, D, F correspond to 3 points from Panel C in Figure 4.10. All the plots that follow are defined in a exactly the same manner with the technique varying across scale.

4.4 Relative Merits

We do not show an aliased signal in this section. The discussion and derivations in Sections 1, 2 and 3 point out that uniform sampling is susceptible to aliasing. Random sampling mitigates this problem. The results do support the theoretical developments in this paper. The bias and variance asymptotically tend to zero the larger the sample size.

It is not surprising that kernel techniques have slightly better integrated MSE. However, they tend to oversmooth the peaks whereas the wavelet techniques do better. If specific areas of the density namely peak effects are of interest then wavelets will tend to be a better analyzing tool thereby agreeing with the referenced literature.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

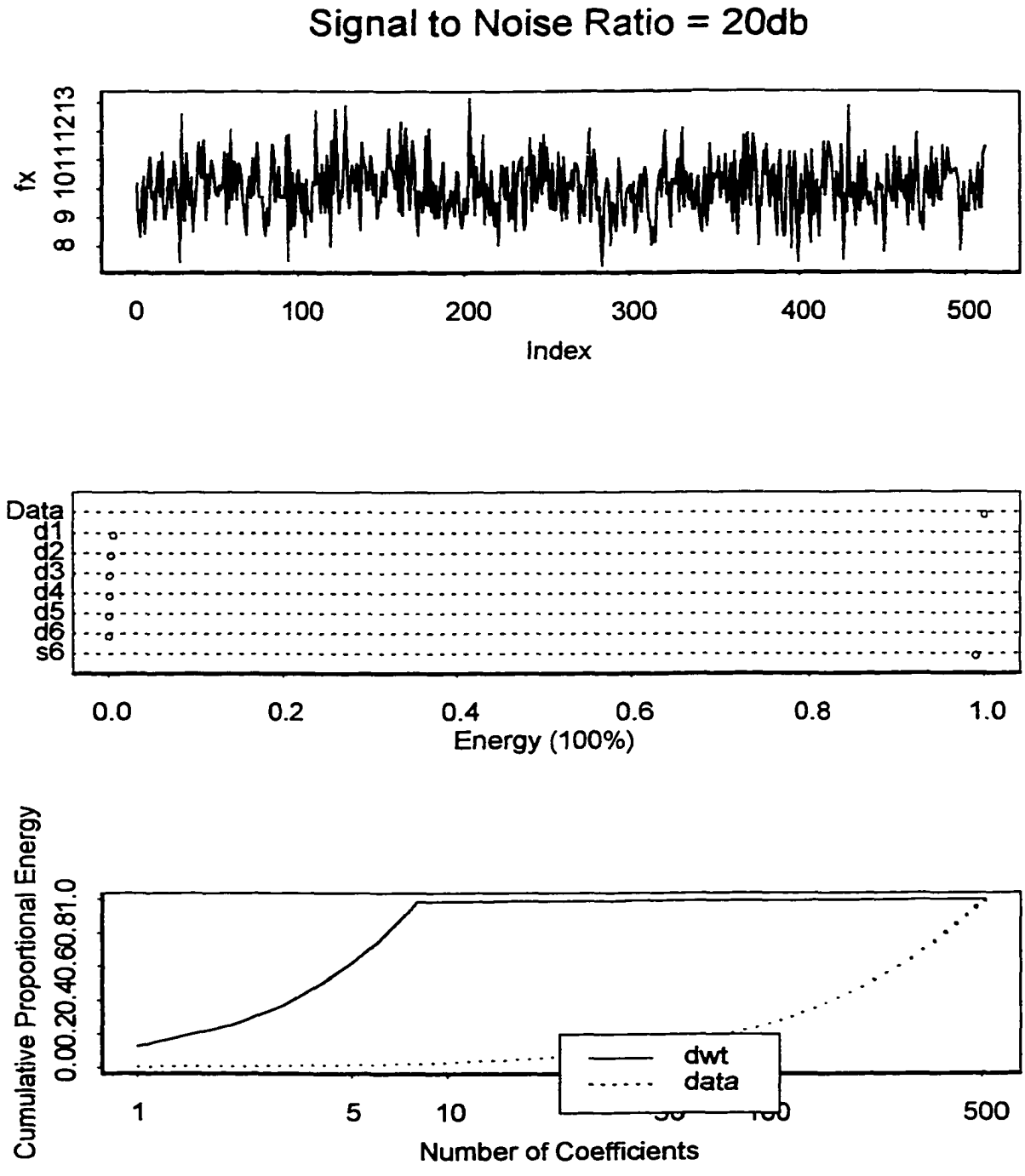


Figure 4.1 Energy Distribution of Large Signal to Noise Realization

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

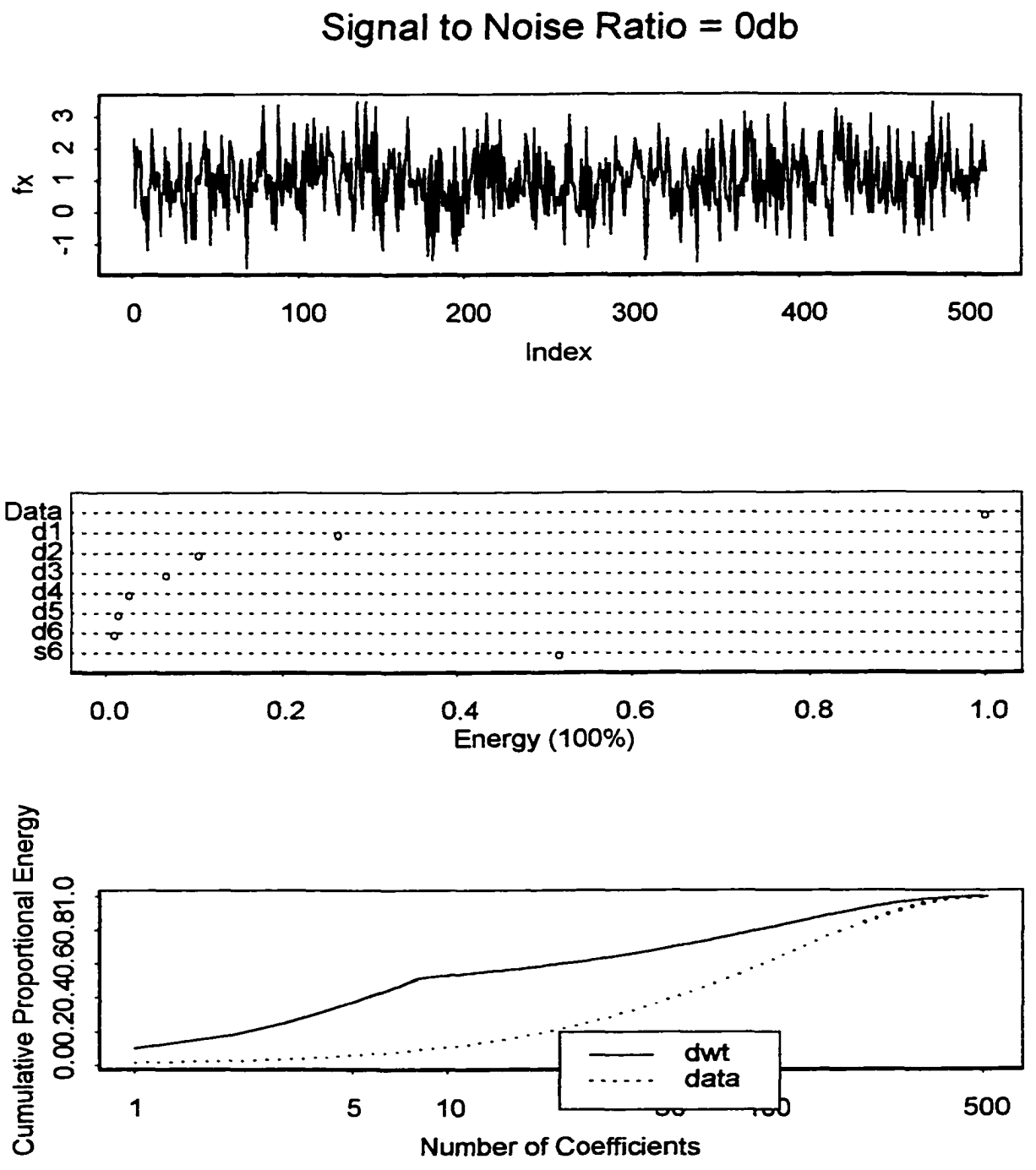


Figure 4.2 Energy Distribution of equal Signal to Noise Realization

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

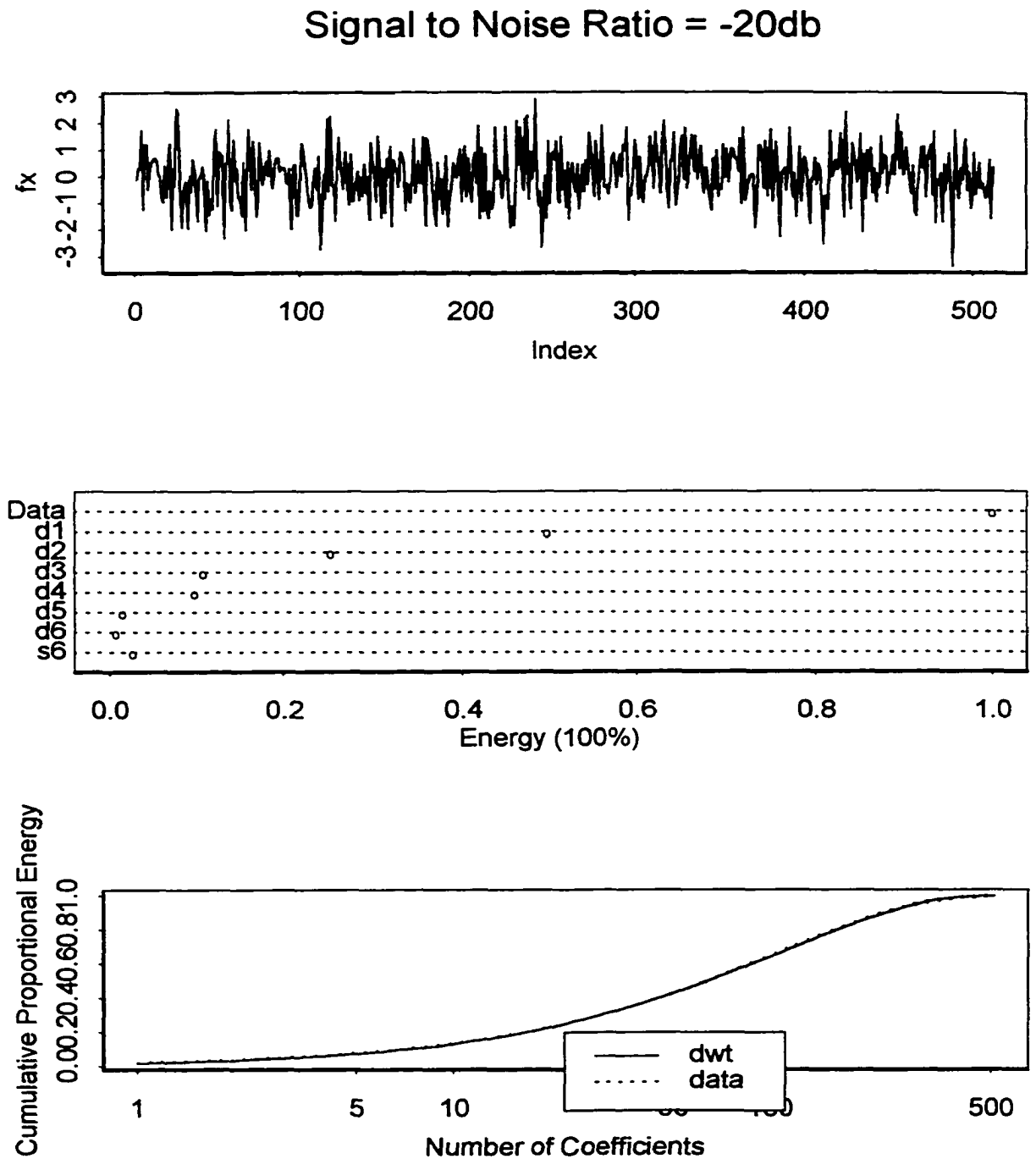


Figure 4.3 Energy Distribution of Large Noise to Signal Realization

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

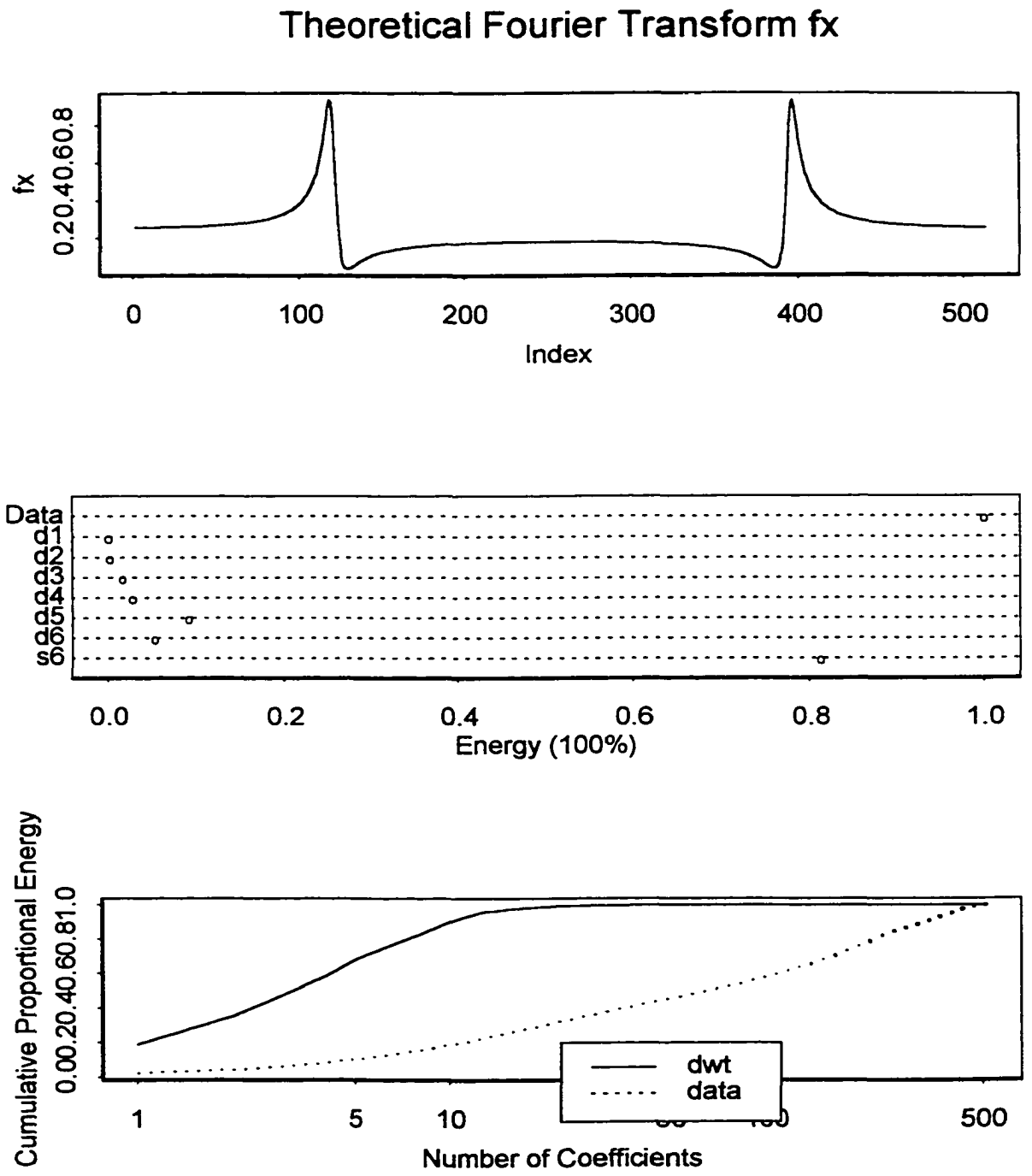


Figure 4.4 Energy Distribution of the Theoretical Spectral Density f_X

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

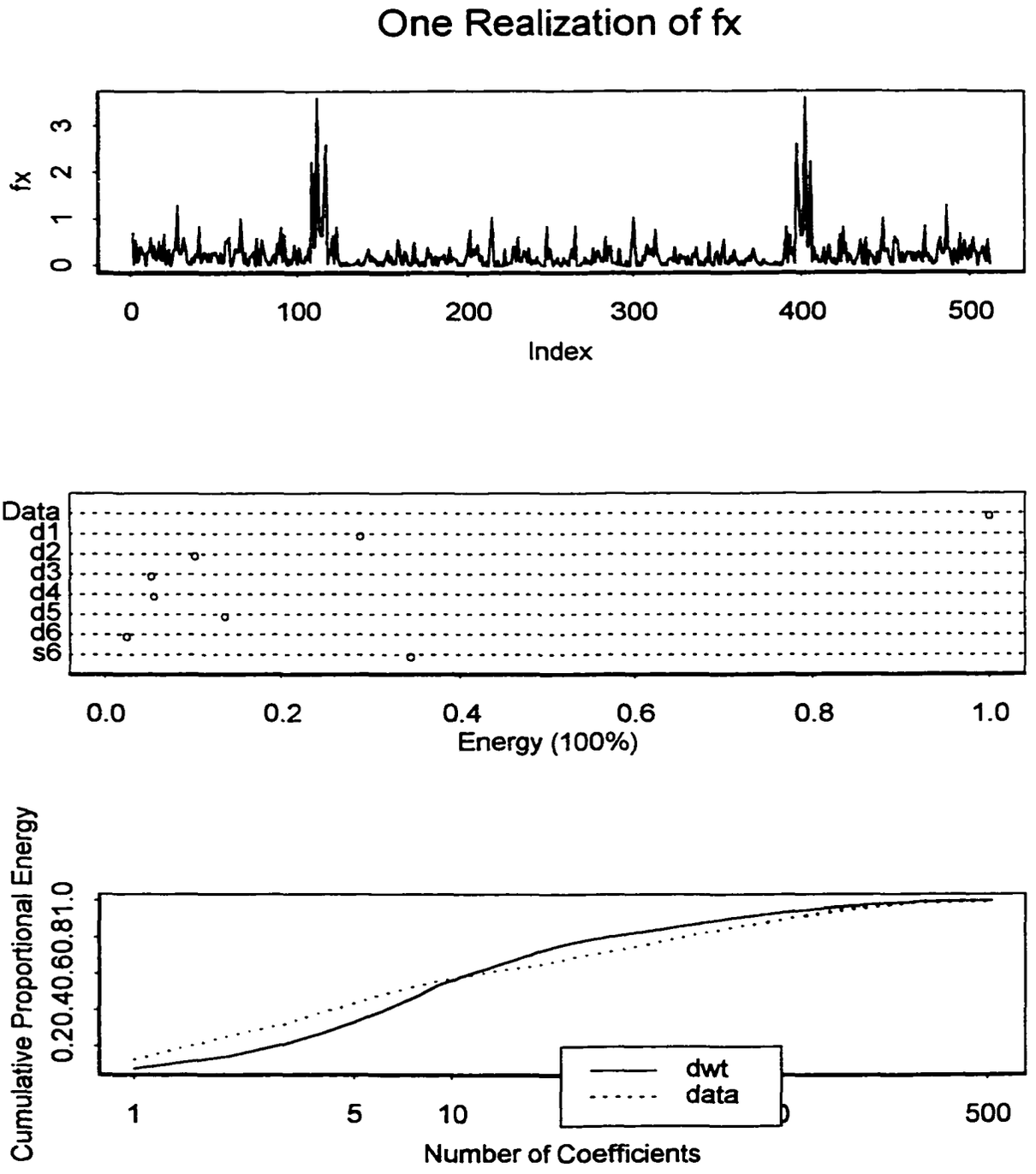


Figure 4.5 Energy Distribution of one Realization of the Spectral Density f_X

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

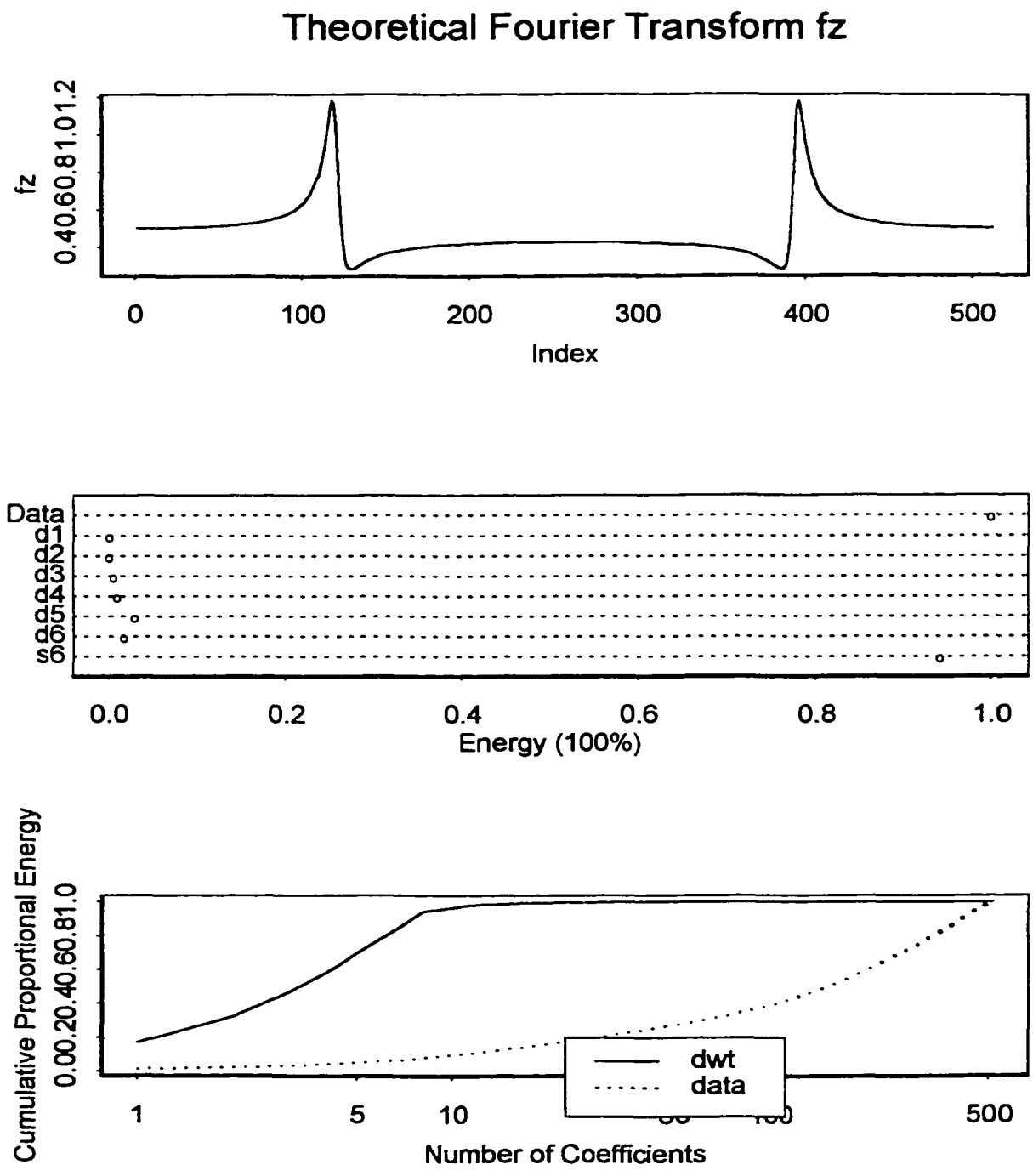


Figure 4.6 Energy Distribution of the Theoretical Spectral Density f_z

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

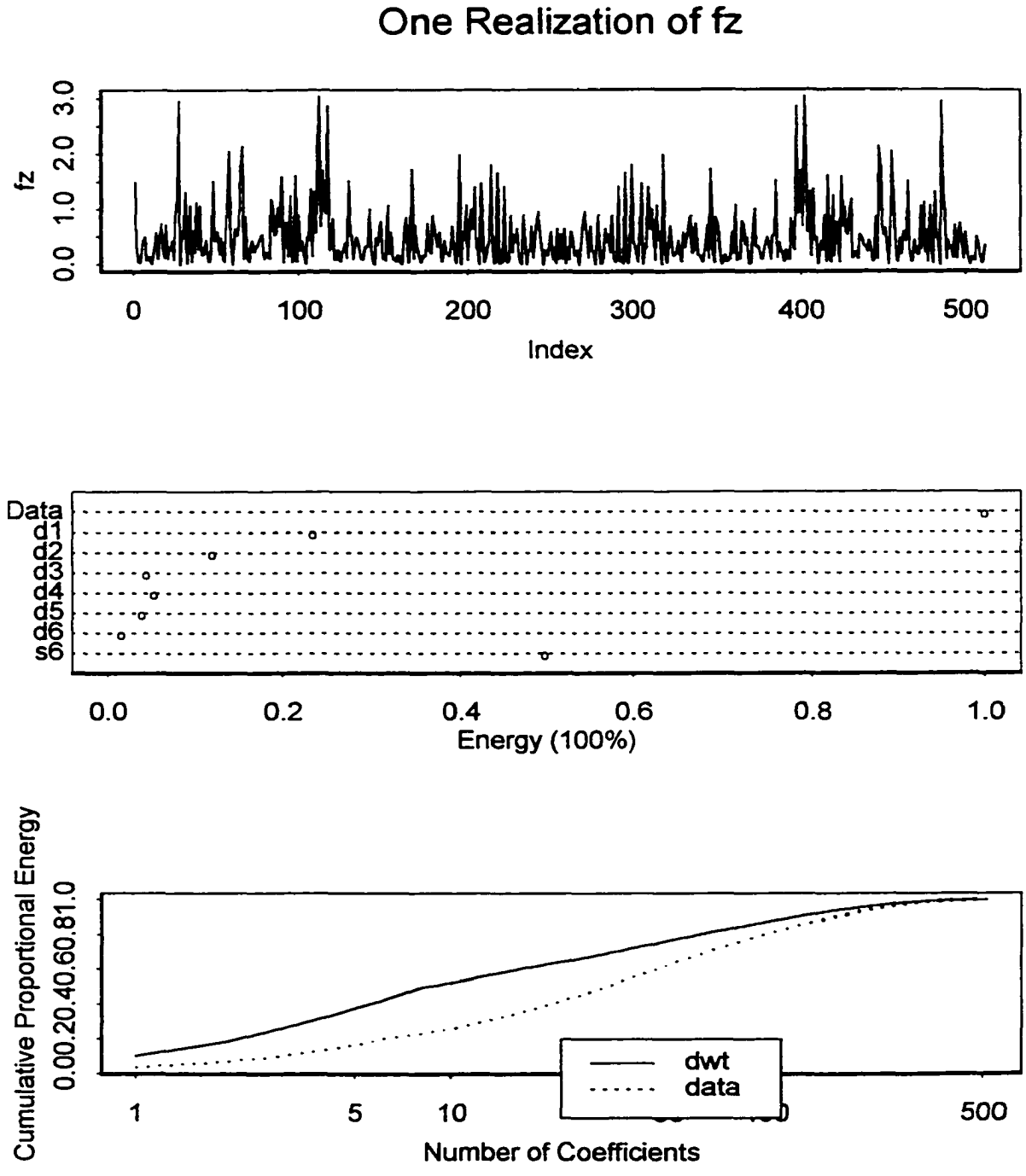


Figure 4.7 Energy Distribution of one Realization of the Spectral Density f_Z

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

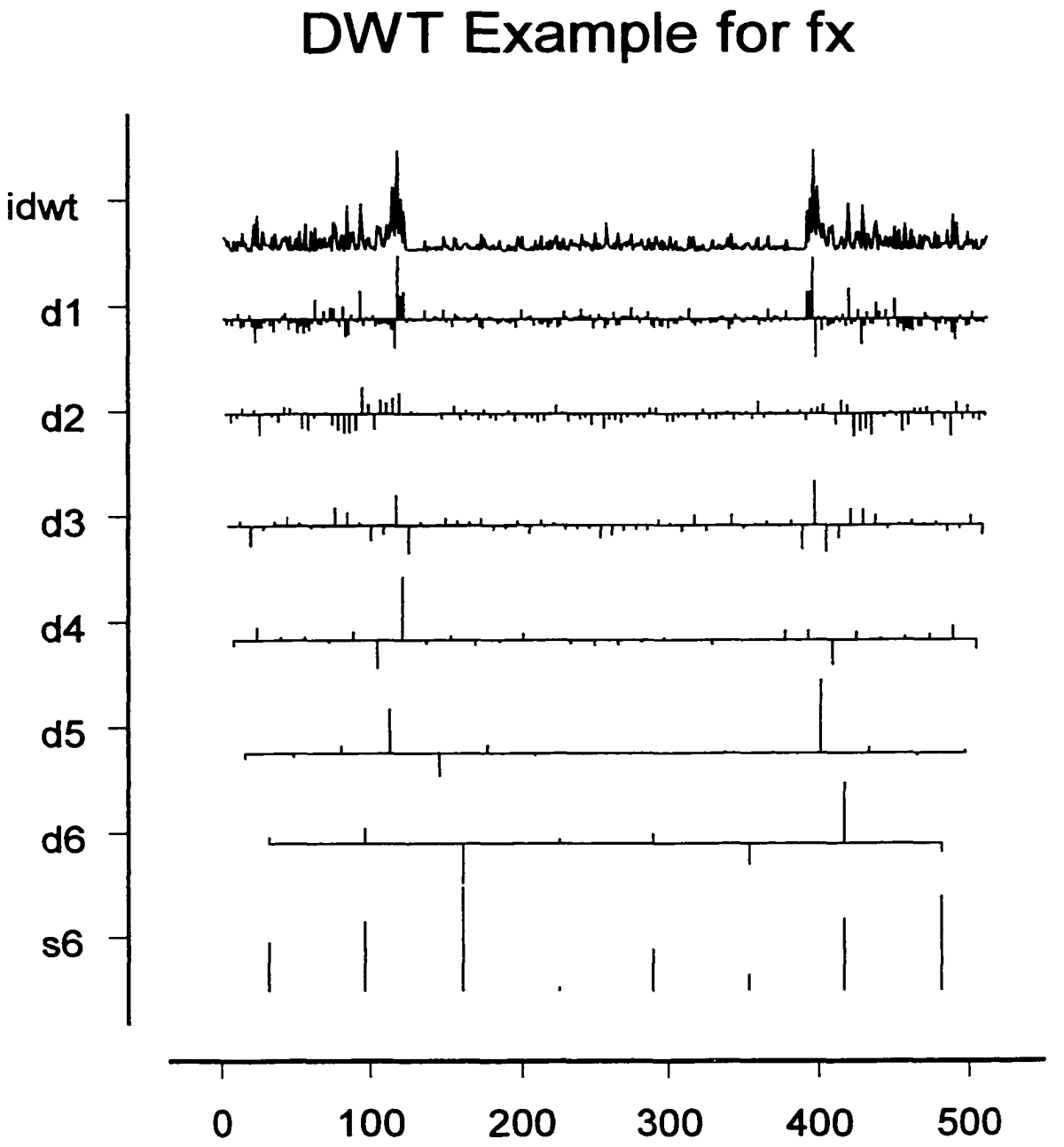


Figure 4.8 Spectral Density f_X with all Wavelet Coefficients

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

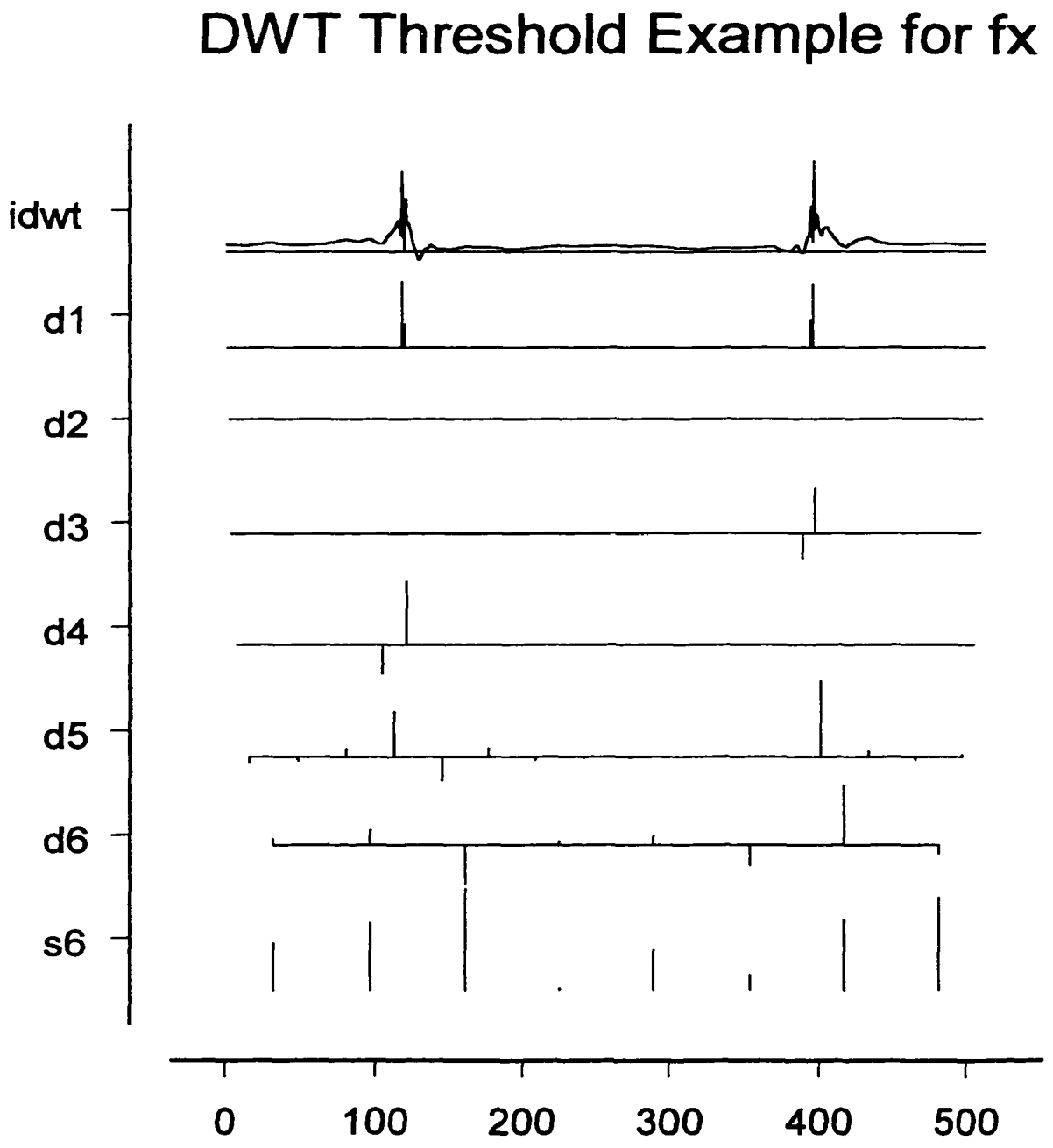
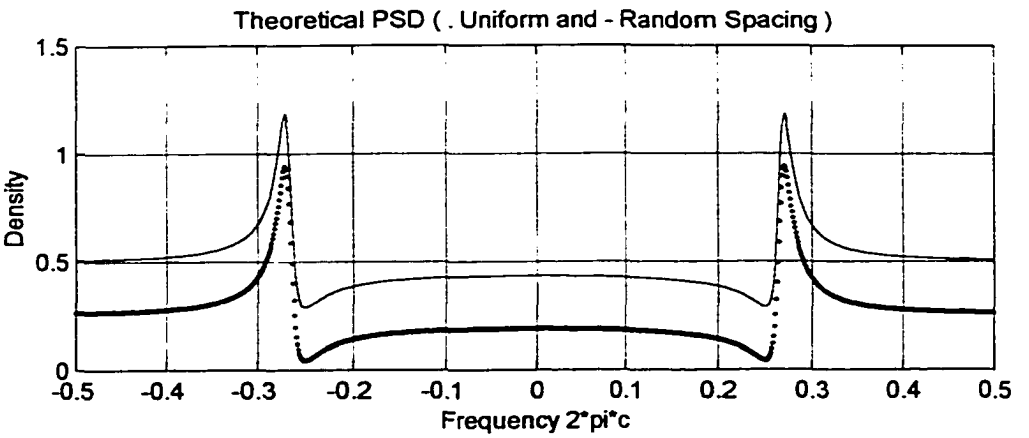
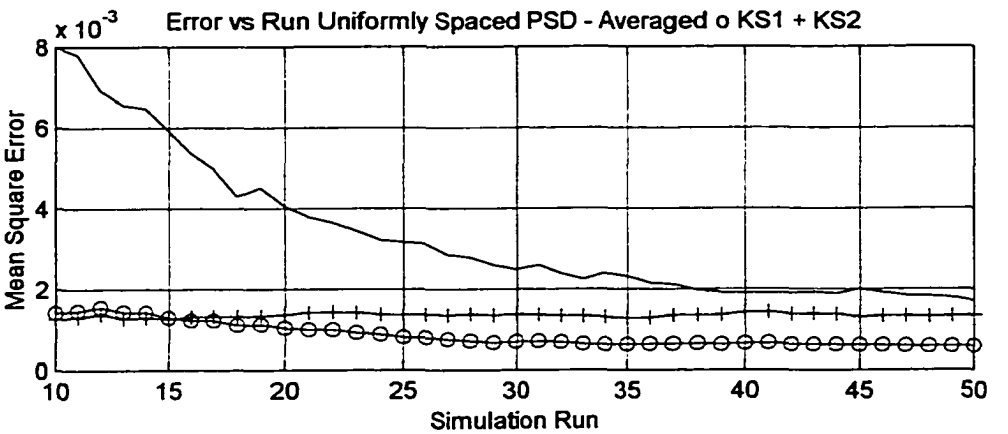


Figure 4.9 Reconstructed Spectral Density f_x with few Wavelet Coefficients

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.





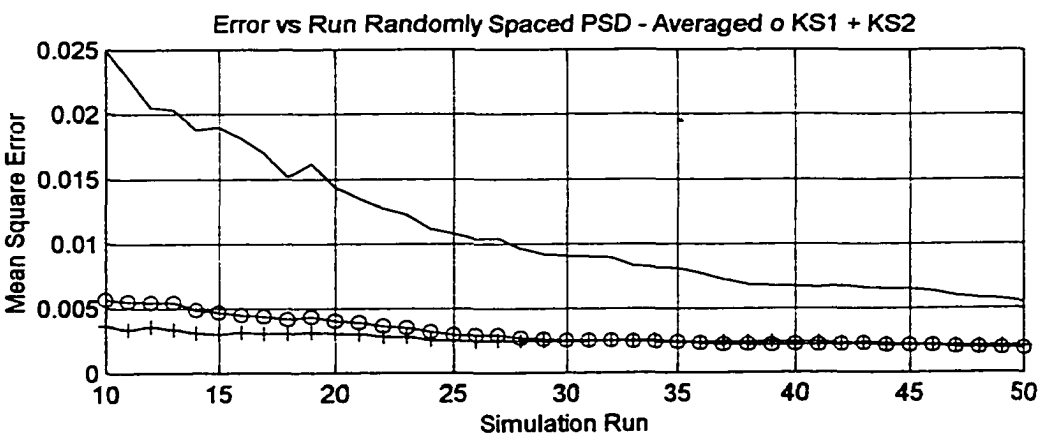


Figure 4.10 MSE of Kernel Smoothing – 1, 2 % Bandwidth

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

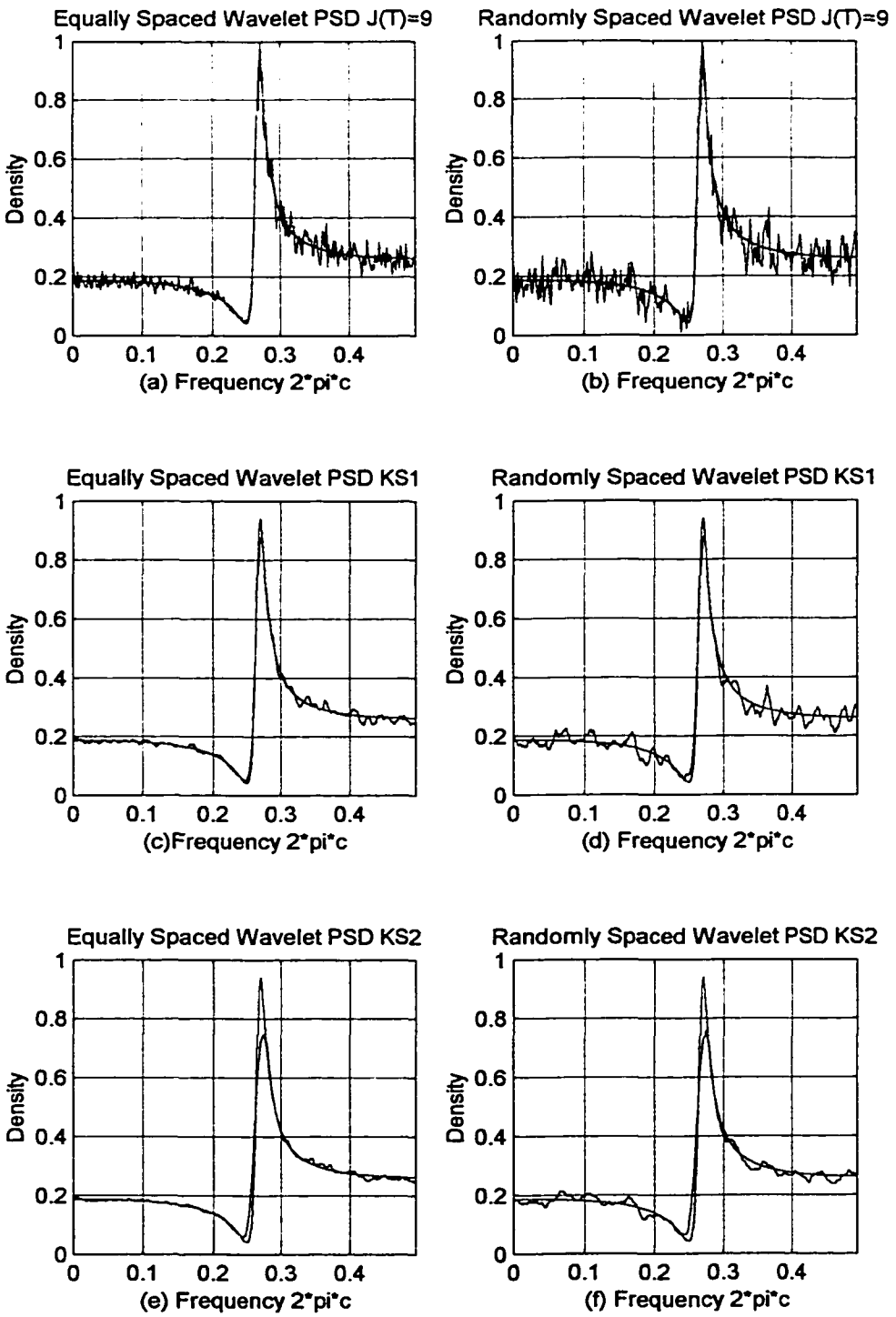
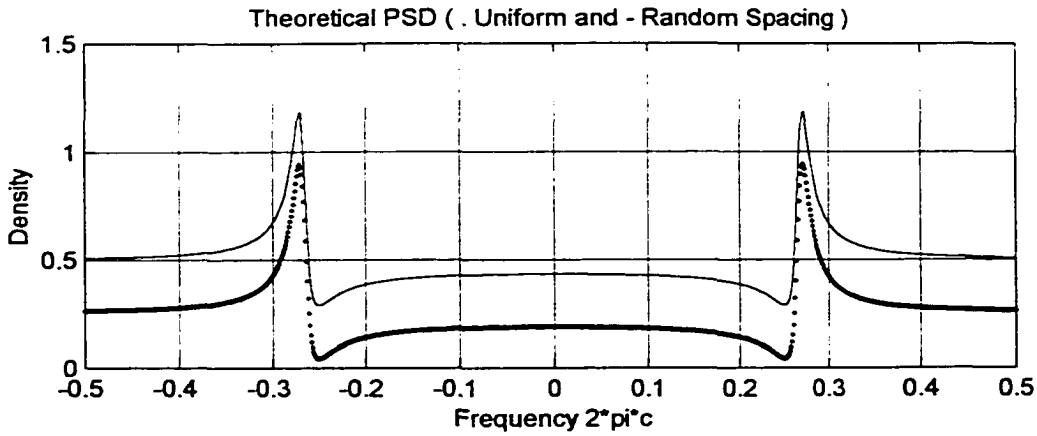
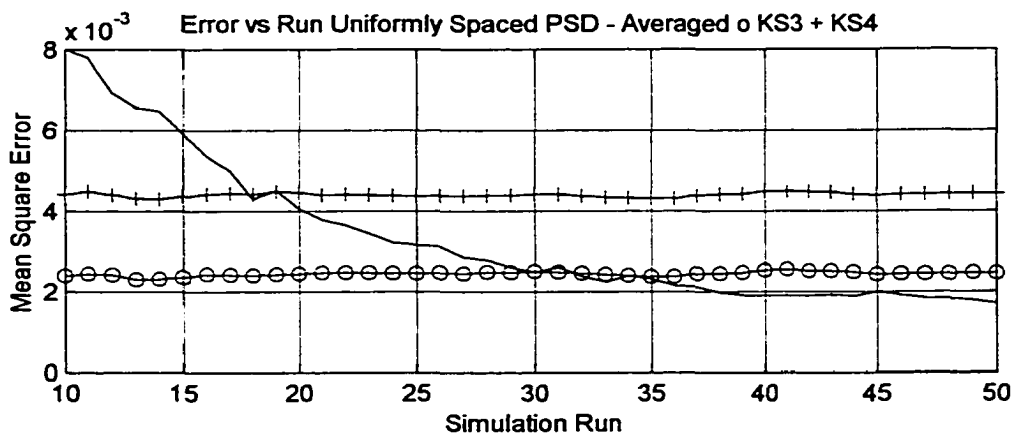


Figure 4.11 Reconstruction error with Kernel Smoothing – 1, 2 % Bandwidth

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.





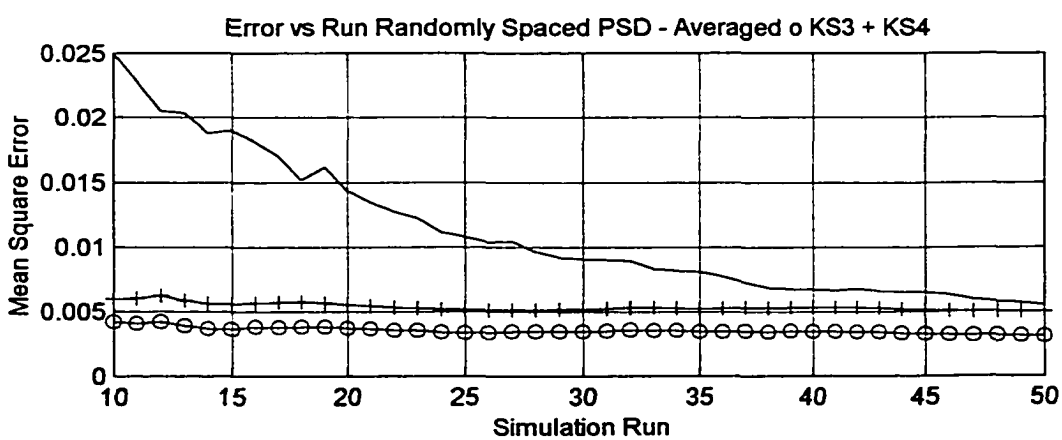


Figure 4.12 MSE of Kernel Smoothing – 3, 4 % Bandwidth

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

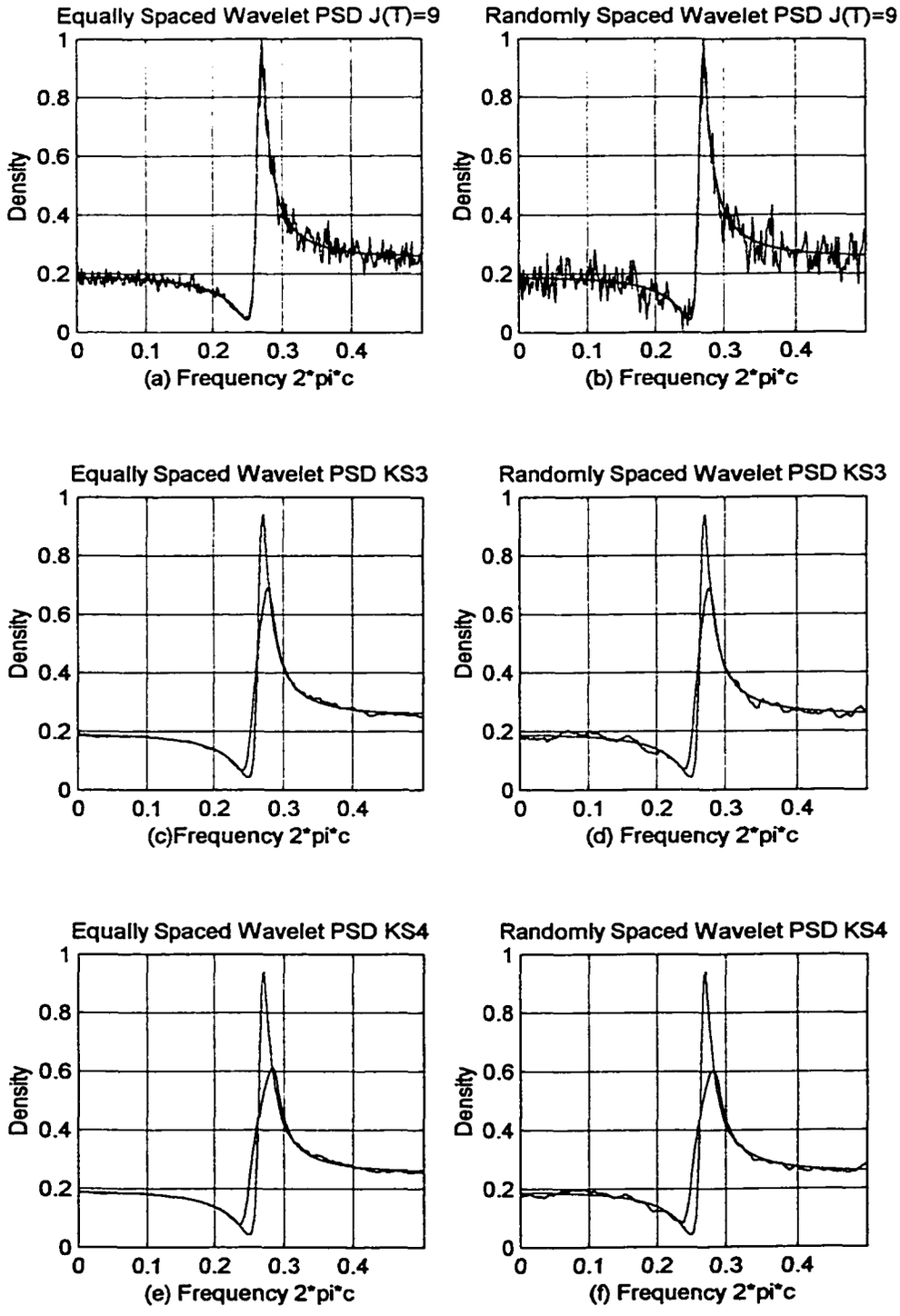
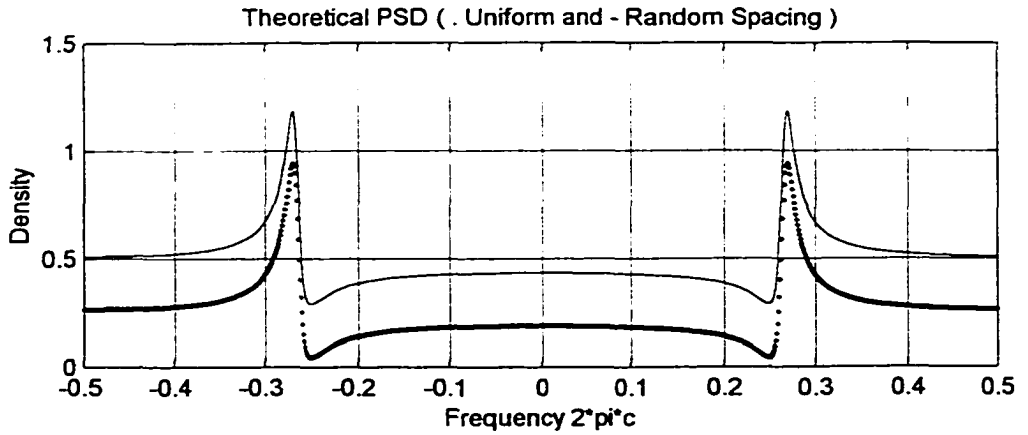
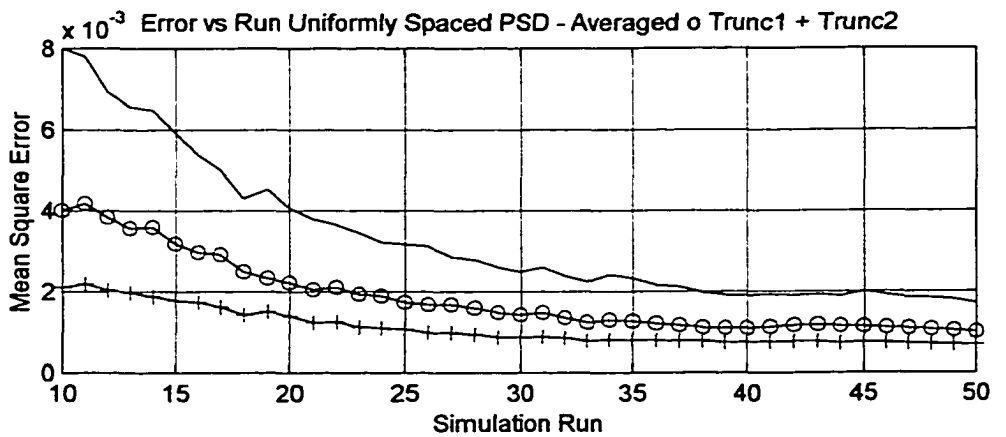


Figure 4.13 Reconstruction error with Kernel Smoothing – 3,4 % Bandwidth

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.





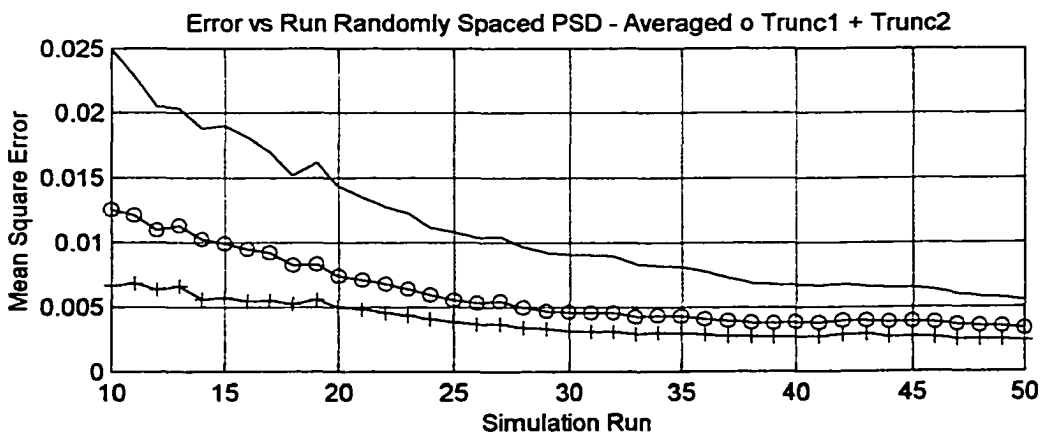


Figure 4.14 MSE of Truncated Scales 1, 2

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

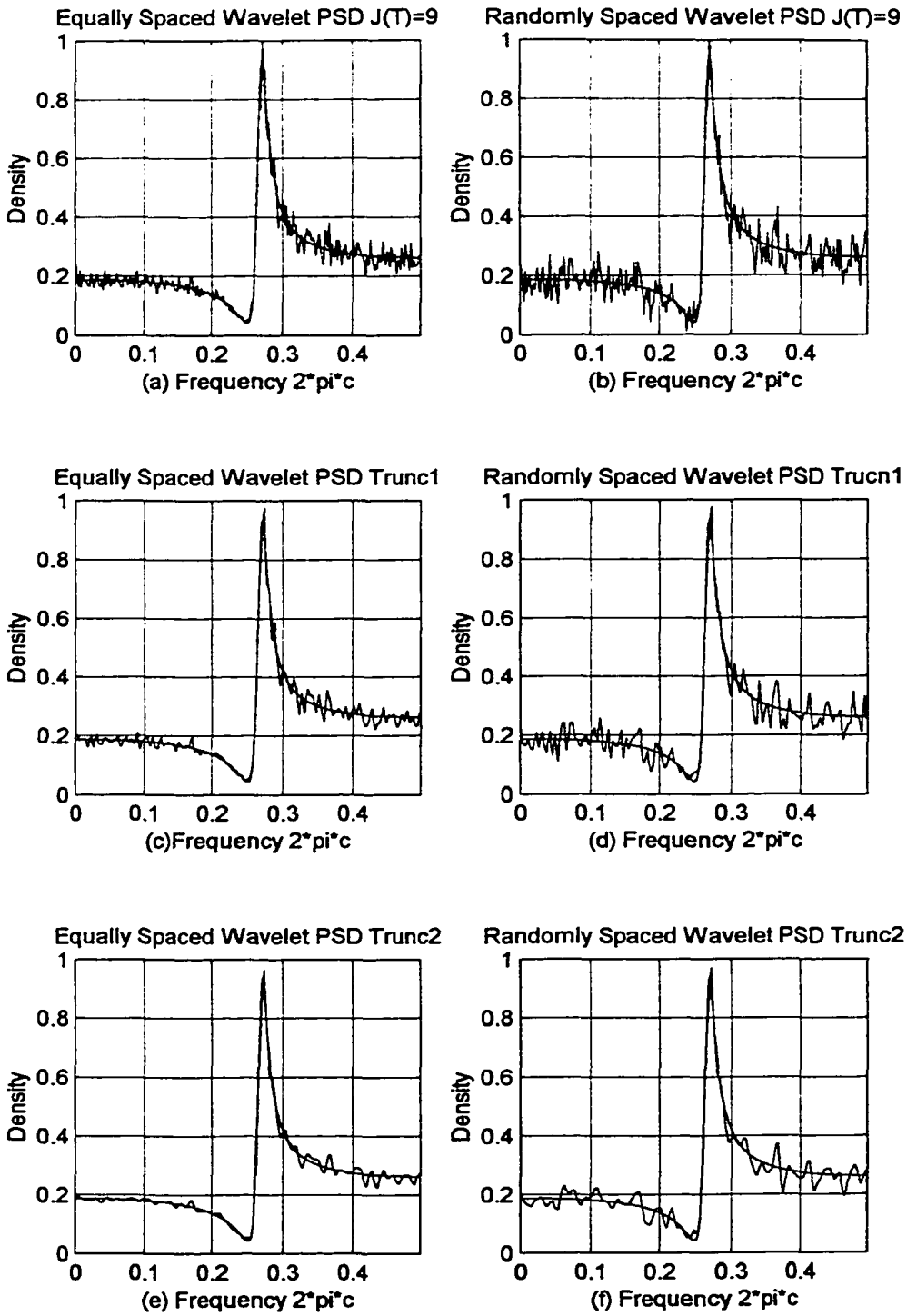
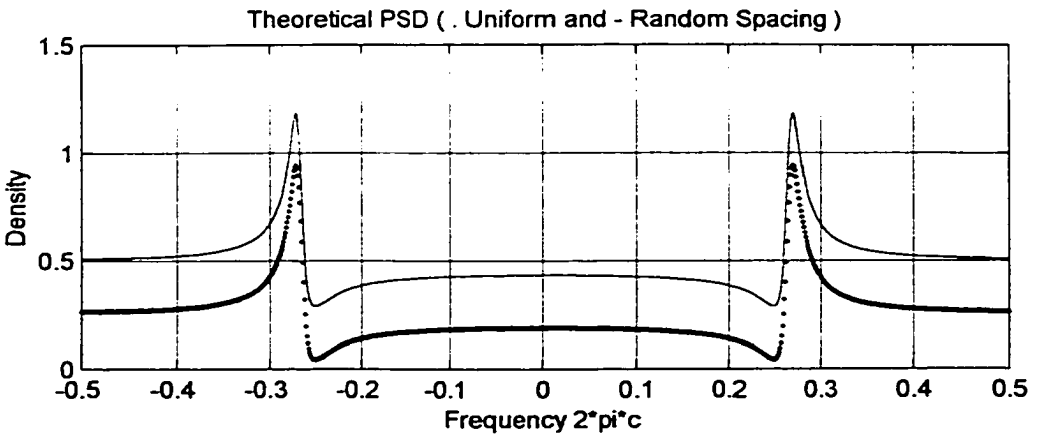
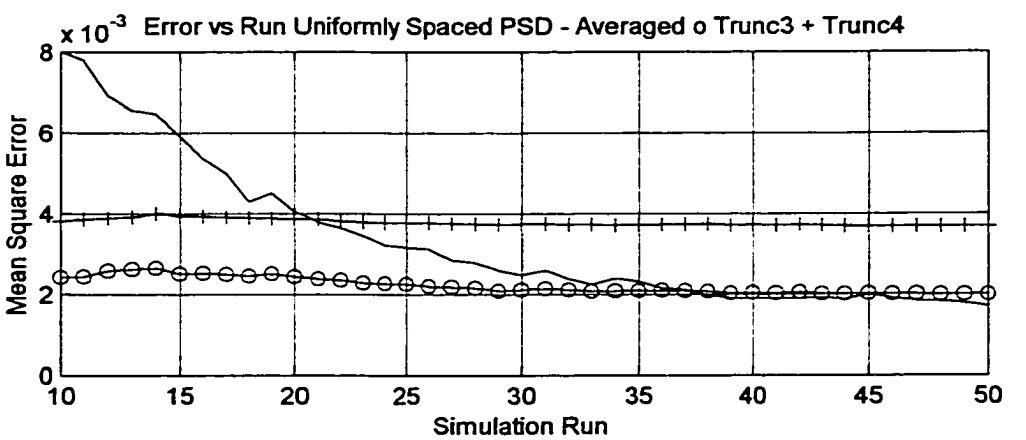


Figure 4.15 Reconstruction error under Truncated Scales 1, 2

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.





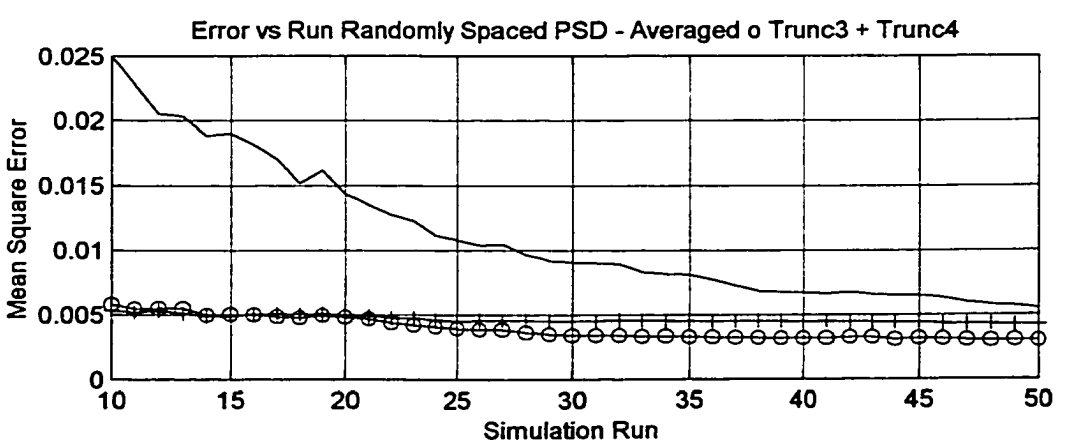


Figure 4.16 MSE of Truncated Scales 3, 4

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

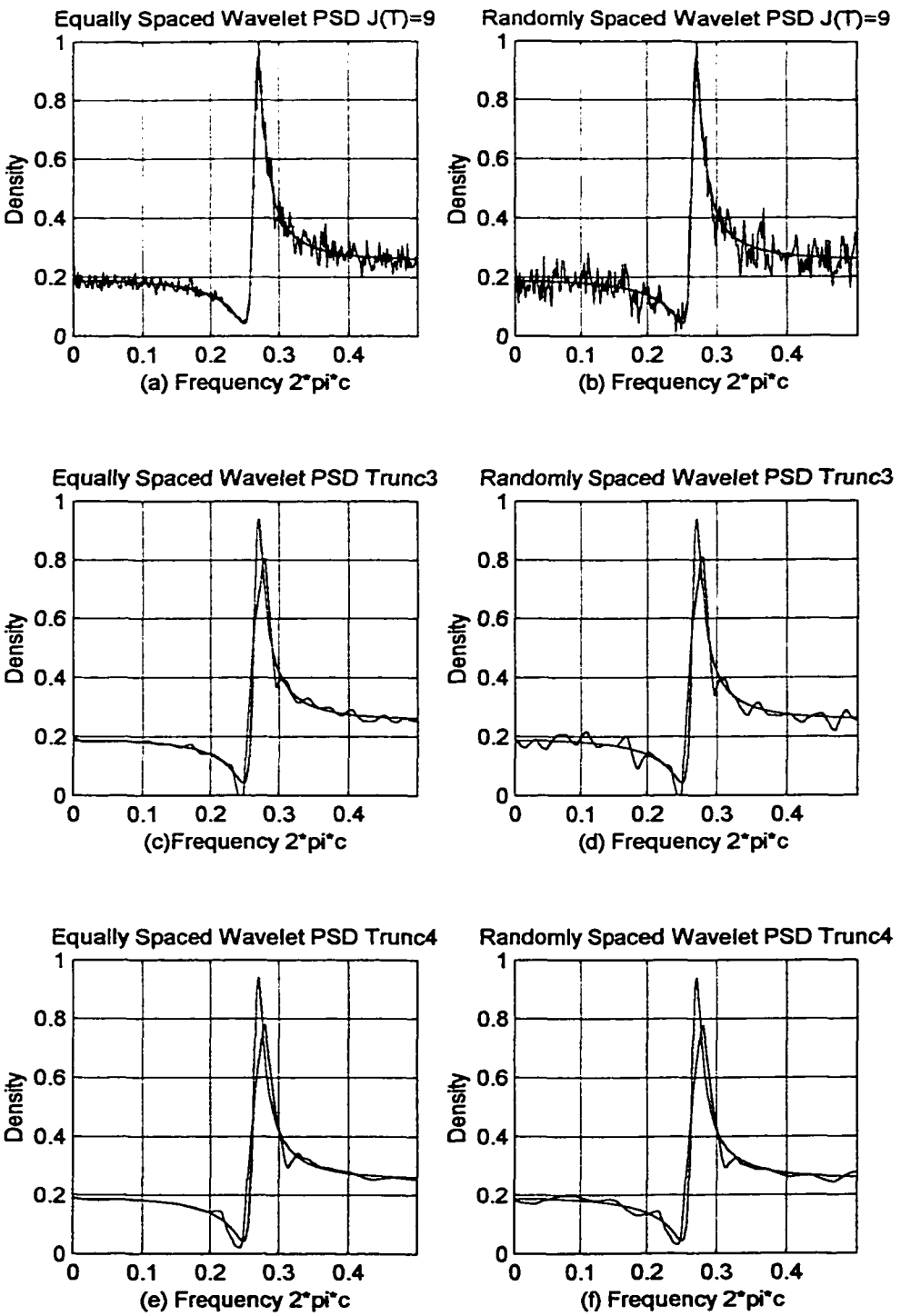


Figure 4.17 Reconstruction error under Truncated Scales 3, 4

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

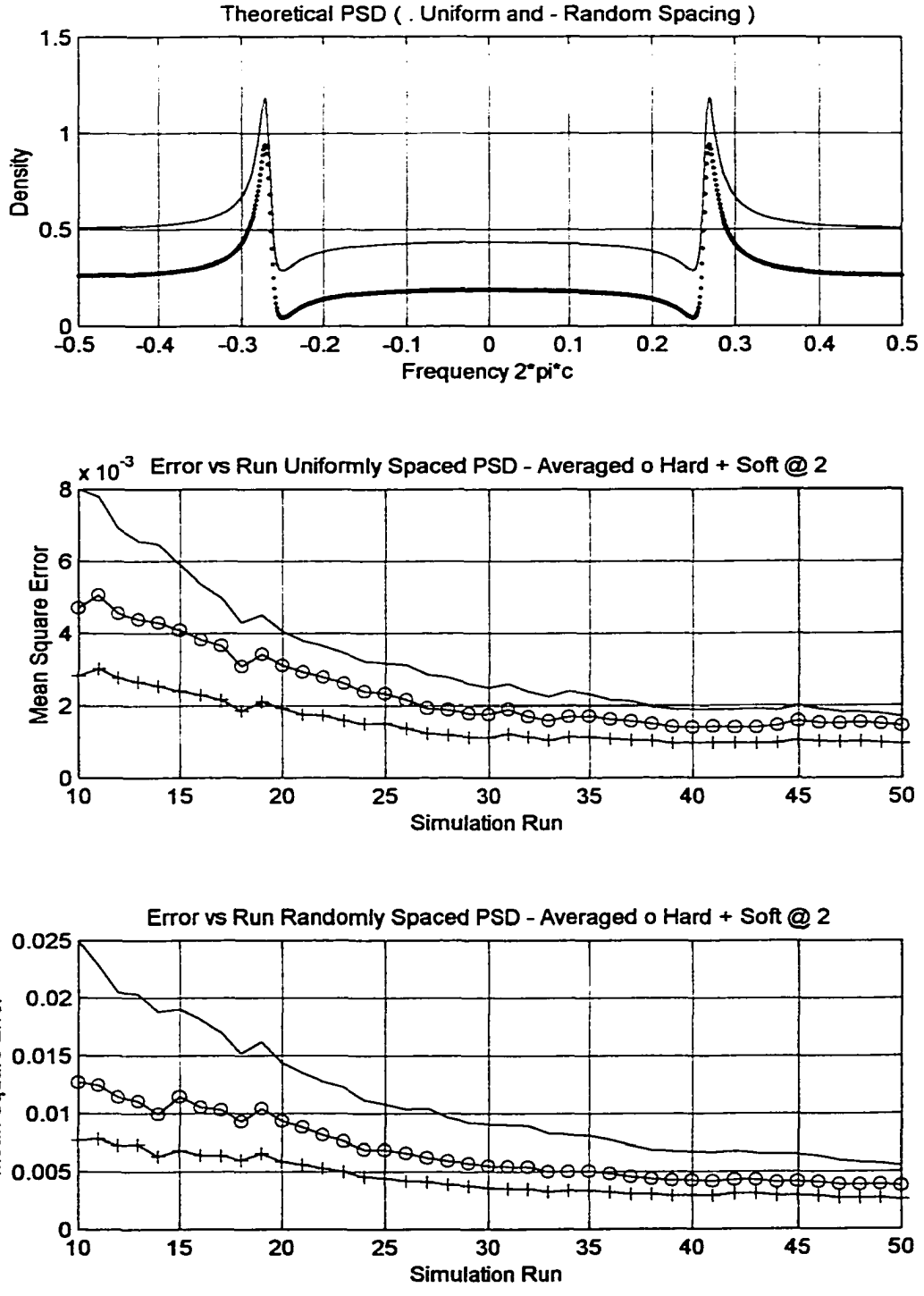


Figure 4.18 MSE of Hard and Soft Thresholding Scales 1, 2

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

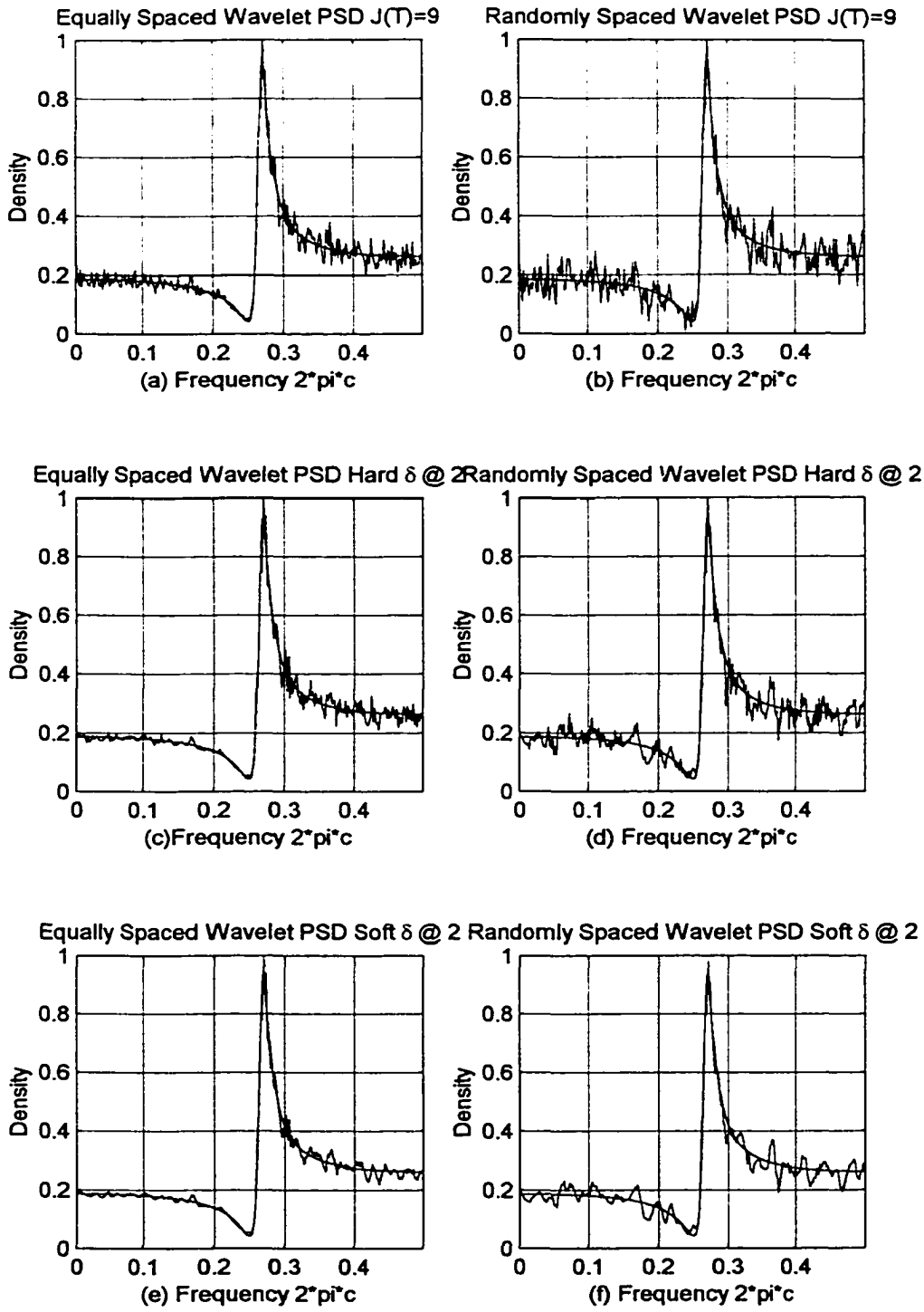
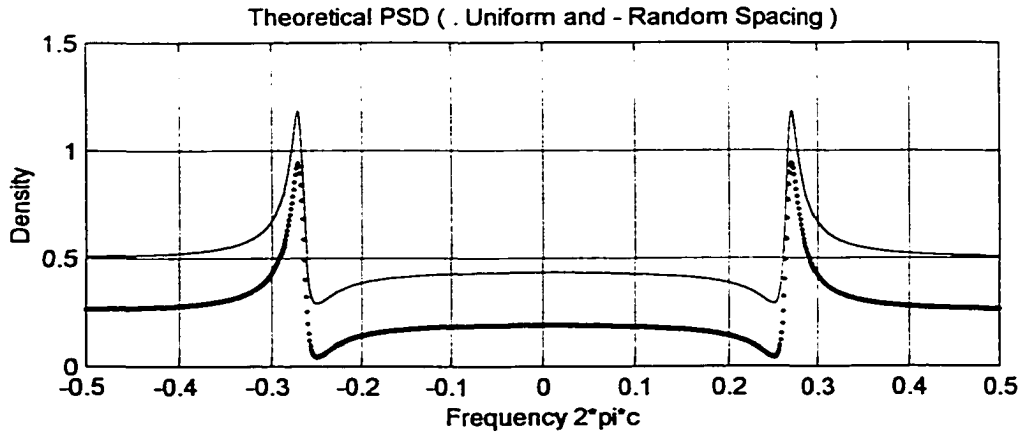
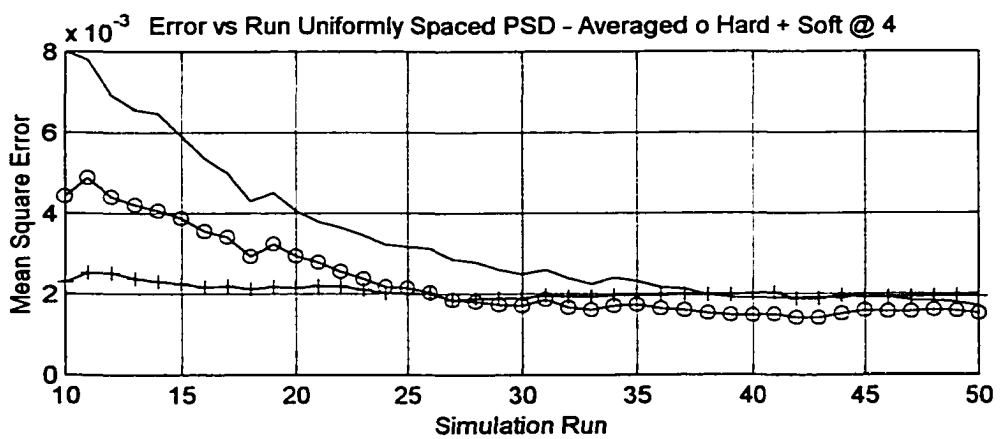


Figure 4.19 Reconstruction error under Hard and Soft Thresholding Scales 1, 2

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.





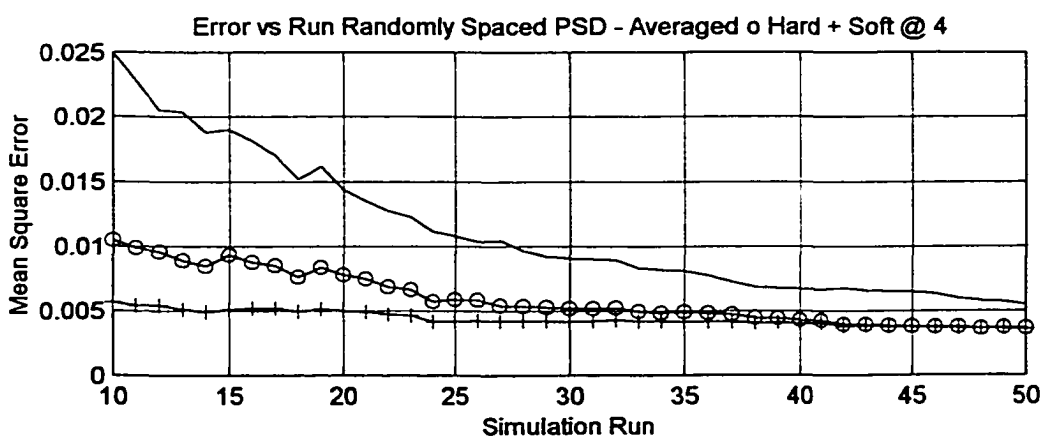


Figure 4.20 MSE of Hard and Soft Thresholding Scales 3, 4

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

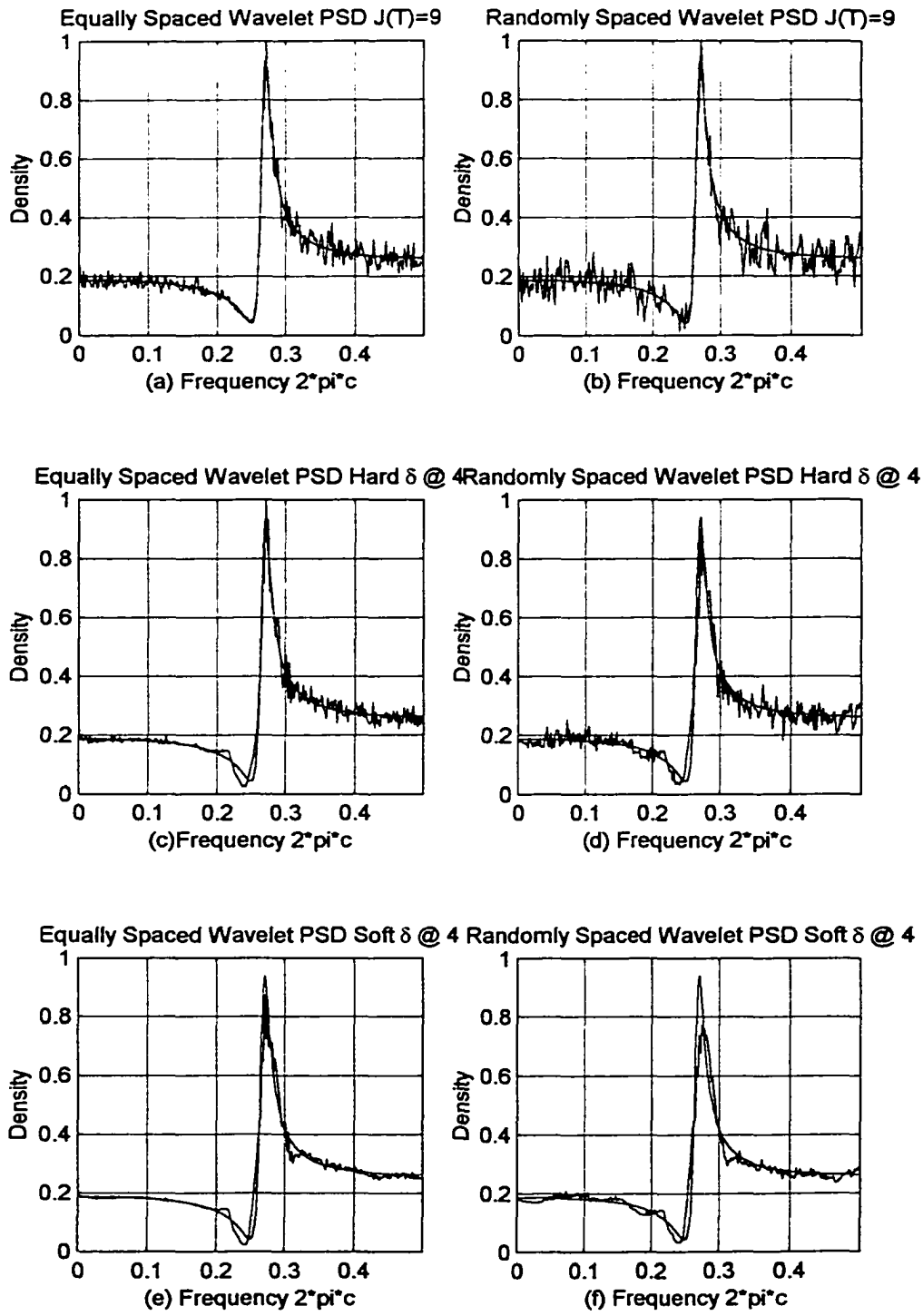


Figure 4.21 Reconstruction error under Hard and Soft Thresholding Scales 3, 4

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

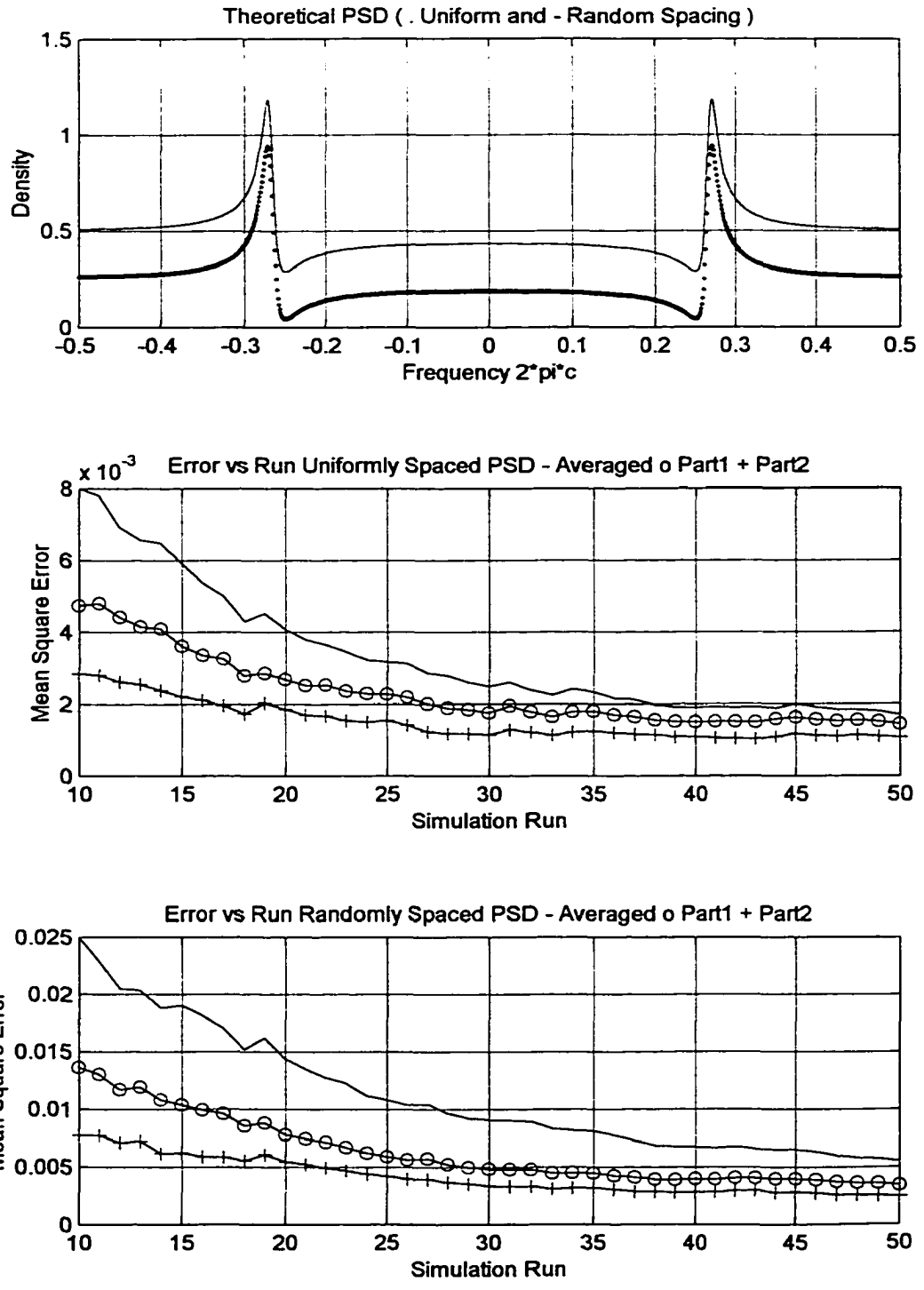


Figure 4.22 MSE of Block Partitions Scales 1, 2 Bandwidth 1%

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

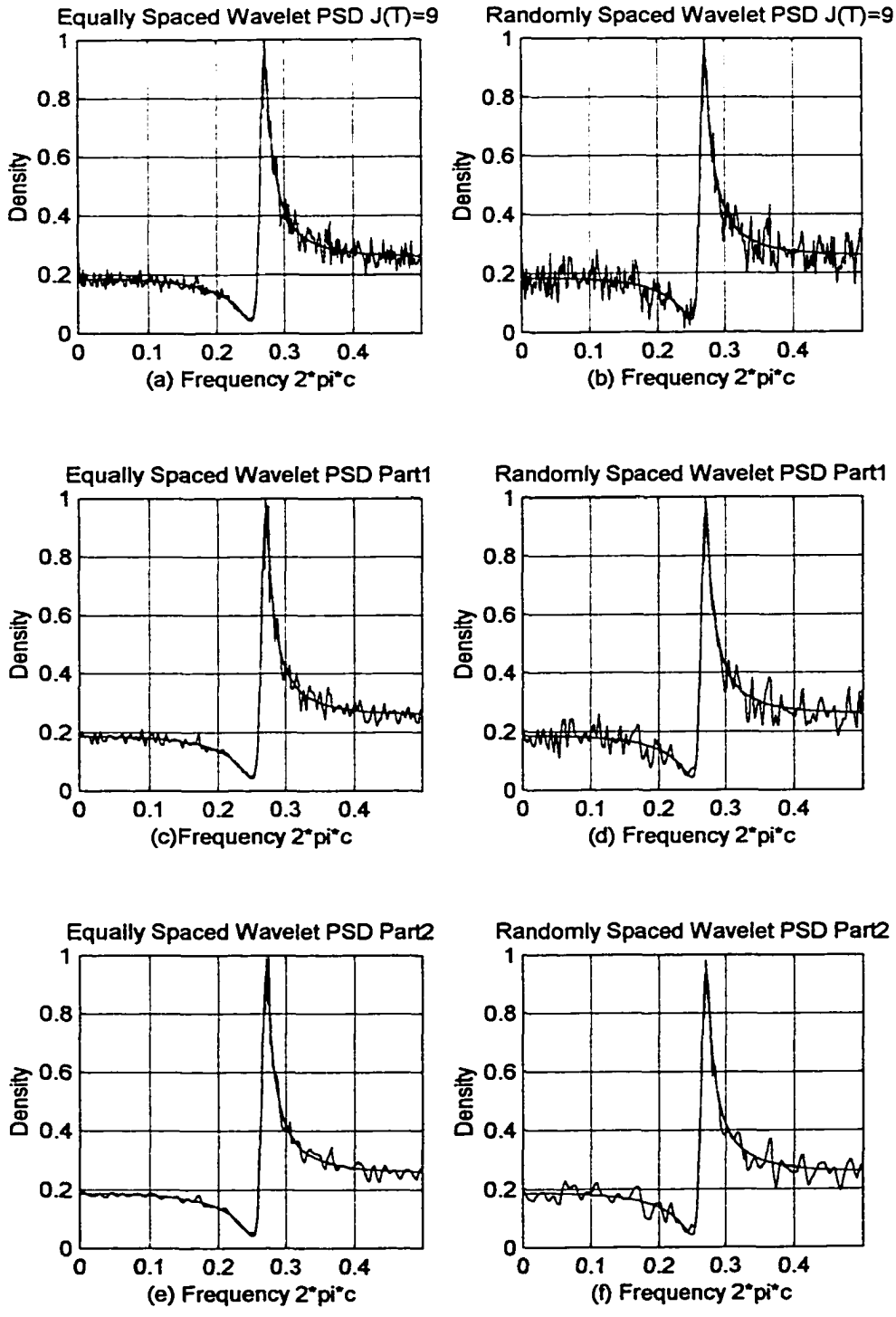
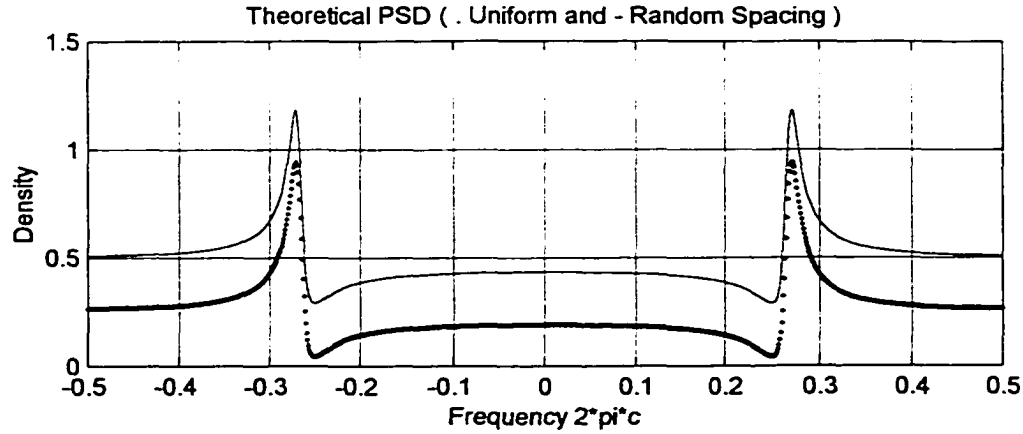
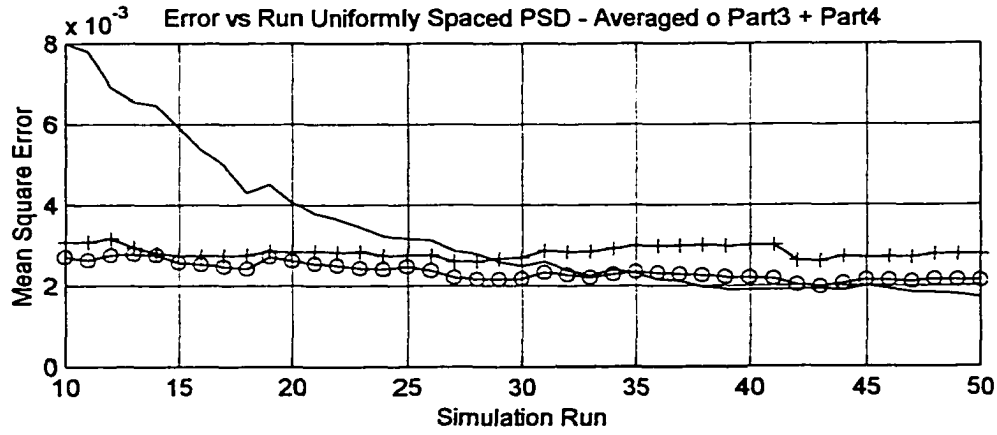


Figure 4.23 Reconstruction error under Block Partitions Scales 1, 2

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.





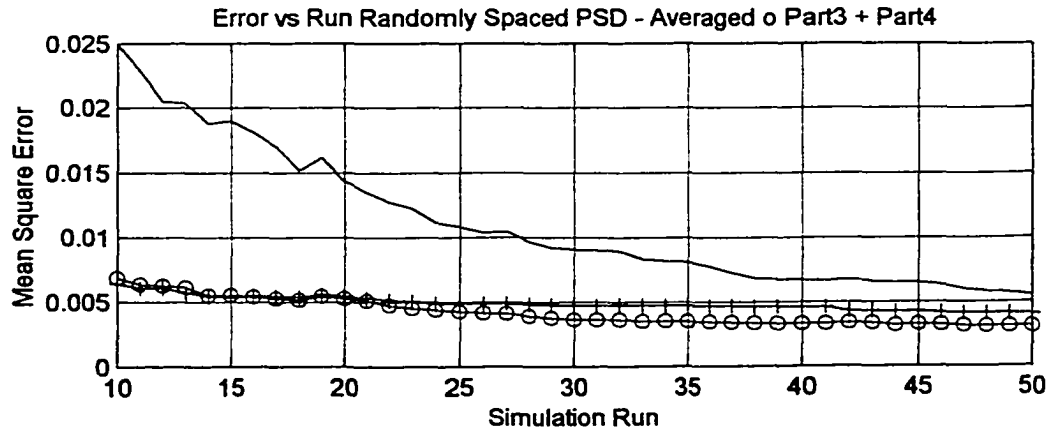


Figure 4.24 MSE of Block Partitions Scales 3, 4 Bandwidth 1%

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

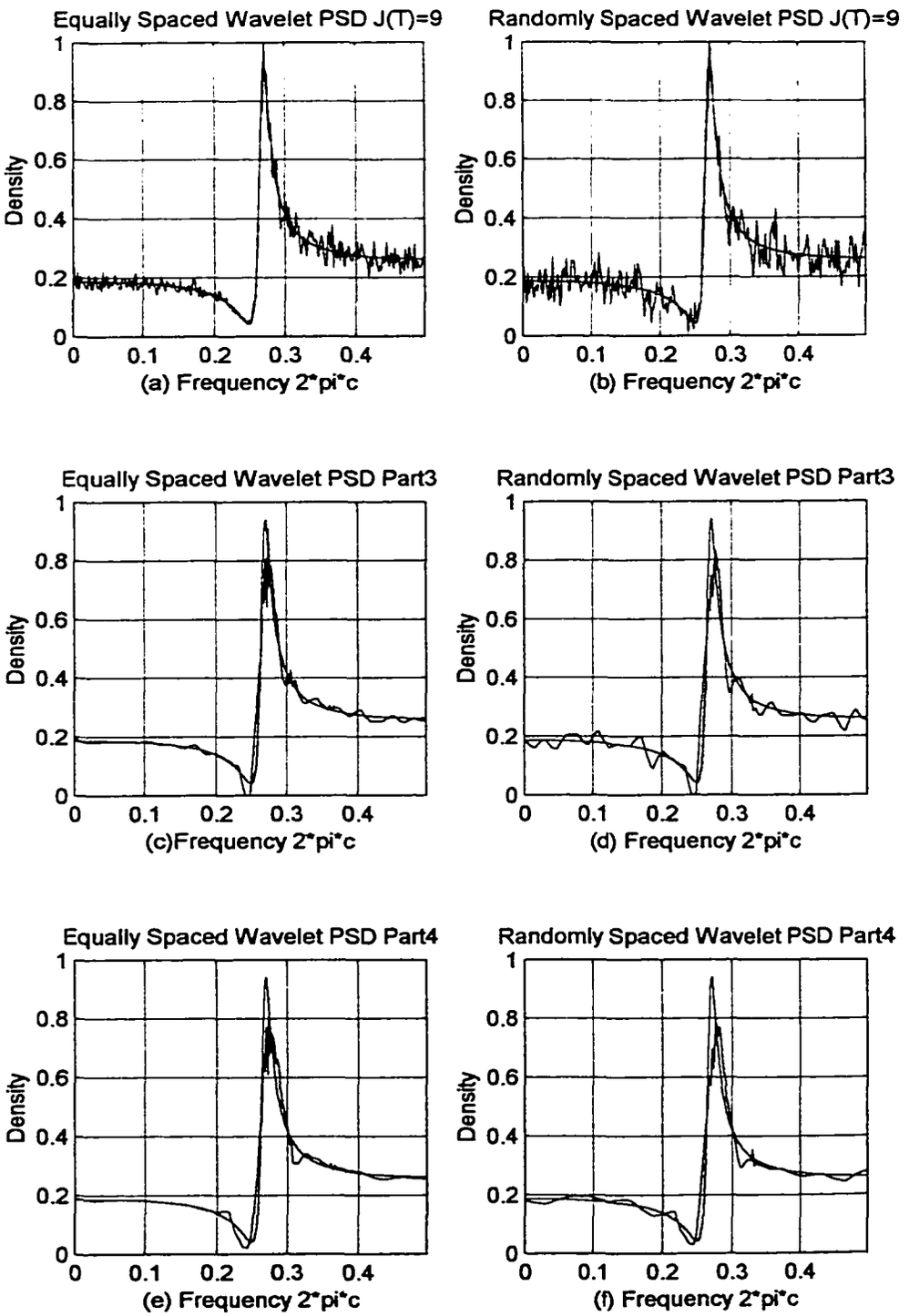


Figure 4.25 Reconstruction error under Block Partitions Scales

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

5 Conclusions and Future Research

Power spectral density estimates based on randomly sampled data provide superior estimates to that of uniformly sampled data when considering the effects of aliasing. The cost associated with random sampling is an increase in variance due to the sampling scheme, although the optimal rate of convergence is the same as uniform sampling. The estimator $\hat{f}_{J(T)}(\lambda)$ as defined in (3.3.6) has been shown to achieve the optimal rate of convergence for a Sobolev class (A3.3.1). Simulated power spectral densities based on randomly spaced time series data have been compared to kernel, and wavelet estimates that have been truncated, thresholded and block partitioned. Both the kernel and wavelet estimate are asymptotically consistent, but the wavelet estimates are able to enhance peak frequencies more effectively than kernel estimates.

It has been proposed [Hall et. al., 1998] that a larger functional class (specifically the Besov space) of data can be optimally analyzed if block thresholds are utilized for the wavelet coefficients by

$$\tilde{f}_{J(T)}(\lambda) = \sum_{k\in Z} \hat{\alpha}_k \phi_{l,k}(\lambda) + \sum_{j\geq l}^{J(T)} \sum_{k'\in Z} \left(\sum_{(k')} \hat{\alpha}_{j,k} \psi_{l,k}(\lambda) I\left(\overline{\alpha}_{j,k}{}^2 > \frac{c}{T} \right) \right).$$

This assumes that the coefficients within a certain region are all highly correlated or homogenous. However, the tradeoff is that a logarithmic penalty is mitigated with the block vs. single coefficient approach. An initial examination was done by simulation in section 4 and it does appear to hold promise. A theoretical analysis and simulation study could be performed to confirm this for randomly sampled power spectral

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

densities. In addition to block thresholding as proposed the form of the threshold could be adjusted based on local conditions instead of applied across the entire scale.

Another avenue of investigation might utilize the density itself to help impose a constraint on the estimation problem by

$$\tilde{f}_{J(T)}(\lambda)=\sum_{k\in Z}\hat{\alpha}_k\phi_{l,k}(\lambda)+\sum_{j\geq l}^{J(T)}\sum_{k'\in Z}\left(\sum_{(k')}\hat{\alpha}_{j,k}\psi_{l,k}(\lambda)I\left(\overline{\alpha}_{j,k}{}^2>\frac{c}{T},\tilde{f}_{J(T)}>0\right)\right).$$

The density estimate is always positive and if by thresholding the resultant density becomes negative, then this would further narrow the decision on the threshold level or the size of the block partition. This becomes theoretically hard to analyze but certainly relatively easy to implement in a data driven algorithm.

Lastly, thresholding decisions are normally based on conditions associated with noise assumptions. Choosing the scales to threshold indirectly utilizes signal assumptions. It is possible to hypothesize directly a signal to noise ratio SNR as discussed in Section 4 and use this as the basis for a thresholding parameter. There is a large body of existing literature that can be utilized in the development of this approach and it seems a logical extension to the very general approach of wavelet density estimation.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Appendix A

This appendix contains 2 Lemma's used in Sections 3. All of the Lemma's presented and derived in this section relate to bounding components for the spectral density estimate.

A.1 Lemma – Wavelet Basis

Axioms and Identities (A.1.1 through A.1.7) will be stated but not derived which are related to an orthogonal multiresolution analysis (MRA) [Mallet, 1989 and Meyer, 1993]. Then identities by derivation (A.1.8 through A.1.30) will be stated followed by their derivation.

Axioms by Construction

The functions ϕ and ψ are denoted as the father and mother wavelet have the following properties

$$\int_R \phi(x)dx = 1, \tag{A.1.1}$$

$$\int_R \psi(x)dx = 0. \tag{A.1.2}$$

The functions ϕ and ψ can be scaled and shifted in the following manner

$$\phi_{l,k}(x) = 2^{l/2}\phi(2^l x - k), \tag{A.1.3}$$

$$\psi_{j,k}(x) = 2^{j/2}\psi(2^j x - k), \tag{A.1.4}$$

where $l, j, k \in Z$.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Identities by Construction

Given $\delta_{j,j'} = \begin{cases} 1 & j = j' \\ 0 & j \neq j' \end{cases}$ then

$$\int_R \phi_{l,k'}(x)\phi_{l,k}(x)dx = \delta_{k',k}, \tag{A.1.5}$$

$$\int_R \phi_{j,k}(x)\psi_{j',k'}(x)dx = 0, \tag{A.1.6}$$

$$\int_R \psi_{j,k}(x)\psi_{j',k'}(x)dx = \delta_{j,j'}\delta_{k,k'}. \tag{A.1.7}$$

Identities by Derivation

$$\int_R \phi_{l,k}(x)dx = 2^{-l/2} \tag{A.1.8}$$

$$\int_R \psi_{j,k}(x)dx = 0 \tag{A.1.9}$$

$$\int_R |\phi_{l,k}(x)|dx = O(2^{-l/2}) \tag{A.1.10}$$

$$\int_R |\psi_{j,k}(x)|dx = O(2^{-j/2}) \tag{A.1.11}$$

Let $\Delta \in \Re$ for equations A.1.12 through A.1.24.

$$\int_R |\phi_{l,k}(x+\Delta)\psi_{j,k}(x)|dx \leq 1 \tag{A.1.12}$$

$$\int_R |\phi_{l,k}(x+\Delta)\phi_{l,k'}(x)|dx \leq 1 \tag{A.1.13}$$

$$\int_R |\psi_{j,k}(x+\Delta)\psi_{j',k'}(x)|dx \leq 1 \tag{A.1.14}$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

$$\left\|\psi_{j',k'}(x+\Delta)\right\|^2 = \sum_{k\in Z}\left|\int_R \psi_{j',k'}(x+\Delta)\phi_{l,k}(x)dx\right|^2 + \sum_{j\geq l}\sum_{k\in Z}\left|\int_R \psi_{j',k'}(x+\Delta)\psi_{j,k}(x)dx\right|^2 \quad \text{(A.1.15)}$$

$$0 \leq \sum_{k\in Z}\left|\int_{x\in R} \psi_{j',k'}(x+\Delta)\phi_{l,k}(x)dx\right|^2 \leq \left\|\psi_{j',k'}(x+\Delta)\right\|^2 \quad \text{(A.1.16)}$$

$$0 \leq \sum_{k\in Z}\left|\int_R \psi_{j',k'}(x+\Delta)\psi_{j,k}(x)dx\right|^2 \leq \left\|\psi_{j',k'}(x+\Delta)\right\|^2 \quad \text{(A.1.17)}$$

$$0 \leq \sum_{j\leq l}\sum_{k\in Z}\left|\int_R \psi_{j',k'}(x+\Delta)\psi_{j,k}(x)dx\right|^2 \leq \left\|\psi_{j',k'}(x+\Delta)\right\|^2 \quad \text{(A.1.18)}$$

$$0 \leq \left|\int_R \psi_{j',k'}(x+\Delta)\psi_{j,k}(x)dx\right| \leq \int_R \psi_{j,k}^2(x)dx = 1 \quad \text{(A.1.19)}$$

$$\left\|\phi_{l,k'}(x+\Delta)\right\|^2 = \sum_{k\in Z}\left|\int_R \phi_{l,k'}(x+\Delta)\phi_{l,k}(x)dx\right|^2 + \sum_{j\geq l}\sum_{k\in Z}\left|\int_R \phi_{l,k'}(x+\Delta)\psi_{j,k}(x)dx\right|^2 \quad \text{(A.1.20)}$$

$$0 \leq \sum_{k\in Z}\left|\int_{x\in R} \phi_{l,k'}(x+\Delta)\phi_{l,k}(x)dx\right|^2 \leq \left\|\phi_{l,k'}(x+\Delta)\right\|^2 \quad \text{(A.1.21)}$$

$$0 \leq \sum_{k\in Z}\left|\int_R \phi_{l,k'}(x+\Delta)\psi_{j,k}(x)dx\right|^2 \leq \left\|\phi_{l,k'}(x+\Delta)\right\|^2 \quad \text{(A.1.22)}$$

$$0 \leq \sum_{j\leq l}\sum_{k\in Z}\left|\int_R \phi_{l,k'}(x+\Delta)\psi_{j,k}(x)dx\right|^2 \leq \left\|\phi_{l,k'}(x+\Delta)\right\|^2 \quad \text{(A.1.23)}$$

$$0 \leq \left|\int_R \phi_{l,k'}(x+\Delta)\phi_{l,k}(x)dx\right| \leq \int_R \phi_{l,k}^2(x)dx = 1 \quad \text{(A.1.24)}$$

$$\sum_{k\in Z}\left|\phi_{0,k}(x)\right| = O(1) \quad \text{(A.1.25)}$$

$$\sum_{k\in Z}\left|\phi_{l,k}(x)\right| = O\left(2^{l/2}\right) \quad \text{(A.1.26)}$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

$$\sum_{k\in Z}|\psi_{0,k}(x)| = O(1) \tag{A.1.27}$$

$$\sum_{k\in Z}|\psi_{J,k}(x)| = O(2^{J/2}). \tag{A.1.28}$$

$$\sum_{k\in Z}\phi^2{}_{l,k}(x) = O(2^{l}) \tag{A.1.29}$$

$$\sum_{k\in Z}\psi^2{}_{J,k}(x) = O(2^{J}). \tag{A.1.30}$$

Proof – The first derivations pertaining to A.1.8 through A.1.11 will use A.1.1 through A.1.4. With a change of variable we have

$$\begin{aligned}\int_R \phi_{l,k}(x)dx &= 2^{l/2}\int_R \phi(2^l x-k)dx \\ &= 2^{-l/2}\int_R \phi(z)dz \quad = \quad 2^{-l/2}.\end{aligned}$$

Similarly for the mother wavelet

$$\begin{aligned}\int_R \psi_{J,k}(x)dx &= 2^{J/2}\int_R \psi(2^J x-k)dx \\ &= 2^{-J/2}\int_R \psi(z)dz \quad = \quad 0.\end{aligned}$$

To find the order of magnitude of the absolute value of the mother and father wavelet we use the fact that they are $\{\phi,\psi\}\in L_1$. Then

$$\begin{aligned}\int_R |\phi_{l,k}(x)|dx &= 2^{l/2}\int_R |\phi(2^l x-k)|dx \\ &= 2^{-l/2}\int_R |\phi(z)|dz \quad = \quad O(2^{-l/2})\end{aligned}$$

Similarly for the mother wavelet

$$\begin{aligned}\int_R |\psi_{J,k}(x)|dx &= 2^{J/2}\int_R |\psi(2^J x-k)|dx \\ &= 2^{-J/2}\int_R |\psi(z)|dz \quad = \quad O(2^{-J/2})\end{aligned}$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

The next three derivations for equations A.1.12 through A.1.14 use the Buniakowsky-Schwarz inequality [Beckenbach and Bellman, 61] to obtain the magnitude of each integration with

$$\begin{aligned}
\int_R |\phi_{l,k}(x+\Delta)\psi_{j,k}(x)|dx &\le \left(\int_R |\phi_{l,k}(x+\Delta)|^2 dx \int_R |\psi_{j,k}(x)|^2 dx\right)^{1/2} = 1, \\
\int_R |\phi_{l,k}(x+\Delta)\phi_{l,k'}(x)|dx &\le \left(\int_R |\phi_{l,k}(x+\Delta)|^2 dx \int_R |\phi_{l,k'}(x)|^2 dx\right)^{1/2} = 1, \\
\int_R |\psi_{j,k}(x+\Delta)\psi_{j',k'}(x)|dx &\le \left(\int_R |\psi_{j,k}(x+\Delta)|^2 dx \int_R |\psi_{j',k'}(x)|^2 dx\right)^{1/2} = 1.
\end{aligned}$$

Expanding a function $f(x)=\psi_{j',k'}(x+\Delta)$ in the wavelet domain and then applying Parseval's Theorem leads to equations A.1.15 through A.1.18. The components of A.1.15 are individually stated in the next 3 equations. We start by expanding

$$f(x)=\sum_{k\in Z}\alpha_k\phi_{l,k}(x)+\sum_{l\le j\le\infty}\sum_{k\in Z}\alpha_{j,k}\psi_{j,k}(x).$$

Using Parseval

$$\left\|\psi_{j',k'}(x+\Delta)\right\|^2=\sum_{k\in Z}\alpha_k{}^2+\sum_{l\le j\le\infty}\sum_{k\in Z}\alpha_{j,k}{}^2.$$

Similarly, expanding a function $f(x)=\phi_{l,k'}(x+\Delta)$ in the wavelet domain and then using Parseval's Theorem leads to equations A.1.19 through A.1.24.

The last set of identities bounds the summation of the father and mother wavelets and their norms stated in equations A.1.25 to A.1.30. Since $\{\phi,\psi\}\in L_1\cap L_\infty$ we hypothesize a function g(x) such that $|\phi|,|\psi|<g\in L_1$ where g is montonically decreasing as $|x|\to\infty$. The summation of this function is bounded by

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

$$\sum_{m\in Z} g(\Delta+m) \le c + \int_R g(x)dx, \quad \Delta \in R.$$

Since $\sum_{m\in Z}|\phi(\Delta+m)| < \sum_{m\in Z} g(\Delta+m)$ and likewise $\sum_{m\in Z}|\psi(\Delta+m)| < \sum_{m\in Z} g(\Delta+m)$ then

$$\sum_{k\in Z}|\phi(x-k)| = O(1)$$
$$\sum_{k\in Z}|\psi(x-k)| = O(1).$$

The solution for the above at scale follows the same guidelines which gives

$$\sum_{k\in Z}|\phi_{J,k}(y_k)| = 2^{J/2}\sum_{k\in Z}|\phi(x_k)|.$$

Then for both the father and mother wavelet we have

$$\sum_{k\in Z}|\phi_{l,k}(x_k)| = O(2^{l/2}),$$
$$\sum_{k\in Z}|\psi_{J,k}(x_k)| = O(2^{J/2}),$$

The same approach utilized for A.1.27 and A.1.28 are applied for the sum of squares of the mother and father wavelet which completes the proof.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

A.2 Lemma - Fejer Kernel

Given the one form of the Fejer Kernel

$$\Delta_T(\lambda) = T^{-1}|D_T(\lambda)|^2 = \frac{T\sin^2\left(\frac{\lambda T}{2}\right)}{\left(\frac{\lambda T}{2}\right)^2},$$

functions $\psi^Q_{j,k}$ defined in (3.4.5), $\psi, \Gamma \in L_\infty \cap L_1$ then the following relationship is true

$$\int_R \psi^Q_{j,k}(\lambda - u)\Delta_T(u+z)du = 2\pi\psi^Q_{j,k}(\lambda+z) + o(2^{j/2}). \tag{A.2.1}$$

Proof: The integral of the Fejer Kernel is $\int_R \Delta_T(\lambda)d\lambda = 2\pi$. Let

$$M_T = \int_R \psi^Q_{j,k}(\lambda - u)\Delta_T(u+z)du - 2\pi\psi^Q_{j,k}(\lambda+z). \tag{A.2.2}$$

By change of variable and the utilization of the Fejer Kernel form previously stated we obtain the following

$$\begin{aligned} M_T &= \int_R \psi^Q_{j,k}(\lambda - u)\Delta_T(u+z)du - 2\pi\psi^Q_{j,k}(\lambda+z) \\ &= \int_R \psi^Q_{j,k}(\lambda + z - x)\Delta_T(x)dx - 2\pi\psi^Q_{j,k}(\lambda+z) \\ &= \int_R \{\psi^Q_{j,k}(\lambda + z - x) - \psi^Q_{j,k}(\lambda+z)\}\Delta_T(x)dx. \end{aligned}$$

Now we utilize another change of variable with u = xT such that

$$\Delta_T(x)dx = \frac{T\sin^2\left(\frac{xT}{2}\right)}{\left(\frac{xT}{2}\right)^2}dx \xrightarrow{u=xT} \Delta(u)du = \frac{\sin^2\left(\frac{u}{2}\right)}{\left(\frac{u}{2}\right)^2}du.$$

The quantity M_T transforms to

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

$$
\begin{aligned}
M_T &= \int_R \left\{\psi^Q_{J,k}(\lambda+z-x)-\psi^Q_{J,k}(\lambda+z)\right\}\Delta_T(x)dx, \\
&= \int_R \left\{\psi^Q_{J,k}\left(\lambda+z-\frac{u}{T}\right)-\psi^Q_{J,k}(\lambda+z)\right\}\Delta(u)du.
\end{aligned}
$$

It is useful at this time to define a term considering scale by

$$
\psi^Q(2^J x-k) \equiv \psi^Q_{J,k}(x)/2^{J/2} = \psi(2^J x-k)-\int_R \Gamma(x-u)\psi(2^J x-k)dx. \qquad \text{(A.2.3)}
$$

The final form of M_T to analyze is

$$
\begin{aligned}
M_T &= 2^{J/2}\int_R \left\{\psi^Q\left(2^J\left(\lambda+z-\frac{u}{T}\right)-k\right)-\psi^Q\left(2^J(\lambda+z)-k\right)\right\}\Delta(u)du \\
&= 2^{J/2}\int_R \left\{\psi^Q\left(y-\frac{2^J u}{T}\right)-\psi^Q(y)\right\}\Delta(u)du
\end{aligned}
\qquad \text{(A.2.4)}
$$

where $y = 2^J(\lambda+z)-k$. Let $w_\psi(x)$ be the modulus of continuity of ψ^Q then

$$
M_T \le 2^{J/2}\int_R w_\psi\left(\frac{|u|}{T/2^J}\right)\Delta(u)du. \qquad \text{(A.2.5)}
$$

The integrand in A.2.5 $\to 0$ as $T \to \infty$ and is bounded by a constant $\Delta \in L_1$. Hence by dominated convergence $2^{-J/2}M_T \to 0$ as $T \to \infty$.

COROLLARY LFK-A

Without loss of generality we can make j a function of T such that j=J(T) just so long as $2^{J(T)}/T \to 0$ as $T \to \infty$. Then the following statement is true with the above considerations from the preceding Lemma with

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

$$\int_R \psi^Q_{J(T),k}(\lambda - u)\Delta_T(u + z)du = 2\pi\psi^Q_{J(T),k}(\lambda + z) + o\left(2^{J(T)/2}\right). \qquad \text{(A.2.6)}$$

COROLLARY LFK-B

Referring to the Fejer Kernel Lemma, it is possible to replace $\psi^Q_{J,k}$ with any of the following functions $\Gamma, \psi_{J,k}, \phi_{l,k}$ utilizing the same assumptions and achieve the following results

$$\begin{array}{lll} \Gamma & \rightarrow & M_T = o(1) \\ \phi_{l,k} & \rightarrow & M_T = o\left(2^{l/2}\right) \\ \psi_{J,k} & \rightarrow & M_T = o\left(2^{J/2}\right) \end{array} \qquad \text{(A.2.7)}$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Appendix B -Fourier Transform

Definitions, Identities, Notation, and Formulations

The ***Fourier Transform*** and its ***Inverse Fourier Transform*** can be represented by

$$\begin{aligned} g(x) &= a\int_{-\infty}^{\infty} e^{-\iota xy} G(y)dy, \\ G(y) &= b\int_{-\infty}^{\infty} e^{\iota xy} g(x)dx, \end{aligned}$$

where **a*b=2π**. It is assumed that $g(x) \in L_1 \cap L_\infty$ and $g(x), g'(x)$ are piecewise continuous in every finite interval. The sign of the exponential is arbitrary and another common formulation of the two can be defined as

$$\begin{aligned} g(x) &= \int_{-\infty}^{\infty} e^{-2\pi\iota xY} G(Y)dy, \\ G(Y) &= \int_{-\infty}^{\infty} e^{2\pi\iota xY} g(x)dx. \end{aligned}$$

If we look closely at the definition and try to provide a physical meaning to the variables **x** and **y** then by giving **x** the dimension of time and **Y** - 1/time or cycles per second we have a natural transformation between the time and frequency domains. This means that **y** would refer to the angular frequency in radians per second. The constant **2π** is the dimensional transformation between the time and frequency domain. This constant in the inversion makes the units consistent such that

$$\begin{aligned} y \equiv 2\pi Y \equiv \frac{2\pi}{x}, or \\ \omega \equiv 2\pi f \equiv \frac{2\pi}{T}. \end{aligned}$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Conventions for this paper with respect to the ***Fourier Transform*** are

$$\begin{aligned} g(t) &= \int_{-\infty}^{\infty} e^{it\omega} G(\omega) d\omega, \\ G(\omega) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-it\omega} g(t) dt, \end{aligned}$$

where t refers to time and ω refers to radians/second. The symbol λ is used consistently to refer to the angular frequency in radians/second.

The ***Fourier Transform Pair*** is defined as

$$g(x) \Leftrightarrow G(y).$$

This by definition means that the Fourier Transform of **g(x)** is **G(y)**, or in other words

$$\begin{aligned} G(y) &= F\{g(x)\}, \\ g(x) &= F^{-1}\{G(y)\}. \end{aligned}$$

And subsequently an identity is possible by applying the inverse transform with

$$\begin{aligned} g(x) &= F^{-1}\{F\{g(x)\}\}, \\ G(y) &= F\{F^{-1}\{G(y)\}\} \end{aligned}$$

The ***Convolution Theorem*** and its Proof start with the convolution of two functions such as

$$g * h = \int_R g(u) h(x-u) du$$

It will be shown that

$$\begin{aligned} F\{g * h\} &= F\{g\} F\{h\}, \\ &= G(\omega) H(\omega). \end{aligned}$$

We have by definition of the Fourier transform

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

$$G(\omega)=\int_R e^{-i\omega u}g(u)du, H(\omega)=\int_R e^{-i\omega u}h(u)du,$$
$$G(\omega)H(\omega)=\int_R\int_R e^{-i\omega(u+v)}g(u)h(v)dudv.$$

Using a change of variable x = v + u and the concept of the Jacobian from advanced calculus, we know that

$$\frac{\partial(u,v)}{\partial(u,x)} = \begin{vmatrix} \frac{\partial u}{\partial u} & \frac{\partial u}{\partial x} \\ \frac{\partial v}{\partial u} & \frac{\partial v}{\partial x} \end{vmatrix} = \begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix} = 1$$

Thus by a change of variable we have

$$\begin{aligned} G(\omega)H(\omega) &= \frac{1}{2\pi}\int_R\int_R g(u)h(x-u)e^{-ixu}dudx, \\ &= \frac{1}{2\pi}\int_R e^{-ixu}\left(\int_R g(u)h(x-u)du\right)dx, \\ &= F\left\{\int_R g(u)h(x-u)du\right\}, \\ &= F(g*h), \end{aligned}$$

and equivalently

$$g*h=F^{-1}\{GH\}.$$

The ***Integral Equation*** involves the inversion of the integral by Fourier Transform techniques. Given the equation

$$g(x)=h(x)+\int_R h(u)r(x-u)du,$$

we need to solve for h(x) in terms of the other functions. By applying the Fourier Transform and simple algebra the result step by step is

$$\{g(x)=h(x)+h*r\}\Leftrightarrow\{G(\omega)=H(\omega)+H(\omega)R(\omega)\},$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

which assumes that the G,H and R transformations exist with the conditions already established then

$$\begin{aligned} H(\omega) &\equiv \frac{G(\omega)}{1+R(\omega)}, \\ &= G(\omega)-G(\omega)\frac{R(\omega)}{1+R(\omega)}, \\ &= G(\omega)-G(\omega)S(\omega). \end{aligned}$$

We can now perform the Inverse Fourier Transform to obtain the final result which is

$$h(x)=g(x)-\int_R g(u)s(x-u)du.$$

In order for this to be admissible we have by definition

$$\begin{aligned} s(x) &= F^{-1}\{S(\omega)\}, \\ &= F^{-1}\left\{\frac{R(\omega)}{1+R(\omega)}\right\}, \\ &= F^{-1}\left\{\frac{F\{r(x)\}}{1+F\{r(x)\}}\right\}. \end{aligned}$$

which places restrictions on the denominator.

The ***Correlation Theorem*** is very similar to the **Convolution Theorem** and may be obtained with the same approach.

$$\begin{aligned} Corr(g,h) &= \int_R g(u)h(u+\tau)du, \\ &\Leftrightarrow G(\omega)\overline{H(\omega)}. \end{aligned}$$

The bar denotes the complex conjugate. It follows immediately that the correlation of a function with itself is

$$Corr(h,h)=|H(\omega)|^2.$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

This is also known as the ***Wiener-Khinchin Theorem***. And finally we have the concept of the Total Power of a signal. If we define the Total Power of a signal as

$$Power = \int_R |h(t)|^2 dt$$

then we can use the following ***Identities/Properties*** to determine a similar quantity under the Fourier Transform.

$$\begin{aligned}
h(at) &\Leftrightarrow \frac{1}{|a|} H\left(\frac{f}{a}\right) \quad time-scaling \\
\frac{1}{|b|} h\left(\frac{t}{b}\right) &\Leftrightarrow H(bf) \quad frequency-scaling \\
h(t-t_0) &\Leftrightarrow H(f)e^{2\pi f t_0} \quad time-shifting \\
h(t)e^{-2\pi f_0 t} &\Leftrightarrow H(f-f_0) \quad frequency-shifting
\end{aligned}$$

In particular to prove the relationship with respect to power we need two relatively simple properties of the complex conjugate and convolution.

$$\begin{aligned}
\overline{g(t)} &\Leftrightarrow \overline{G(-f)}, \\
g(t)h(t) &\Leftrightarrow G(f) * H(f).
\end{aligned}$$

Using these two properties gives

$$\begin{aligned}
\overline{g(t)}h(t) &\Leftrightarrow \overline{G(-f)} * H(f), \\
\int_{-\infty}^{\infty} \overline{g(t)}h(t)e^{-2\pi f t} dt &= \int_{-\infty}^{\infty} \overline{G(-\lambda)}H(f-\lambda)d\lambda.
\end{aligned}$$

Now all we do is let the function h(t) = g(t) and set the frequency to 0 resulting in

$$\int_R |h(t)|^2 dt = \int_R |H(f)|^2 df,$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

which is also known as ***Parseval's Theorem***. This equates the total power of the signal to its frequency domain counterpart exhibiting the conservation of energy principal.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

References

Akansu, Ali N., and Richard A. Haddad, *Multiresolution Signal Decomposition*, Academic Press, 1992.

Anderson, T.W., *An Introduction to Multivariate Statistical Analysis*, John Wiley and Sons, Inc., New York, 1958.

Beckenbach, Edwin F., Bellman, Richard, *Inequalities*, Springer-Verlag, 1961

Beylkin, G., R. Coifman, and V. Rokhlin, Fast wavelet transforms and numerical algorithms, *Comm. Pure Appl. Math*, 1991.

Bloomfield, Peter, *Fourier Analysis of Time Series*, John Wiley and Sons, 1976.

Brockwell, Peter J., Richard A. Davis, *Time Series: Theory and Methods*, Second Edition, Springer Verlag, 1991.

Brillinger, D. R., The spectral analysis of stationary interval functions, *Proc. Sixth Berkeley Symposium Probability and Statistics*, pp. 483-513, 1972.

Brillinger, D. R., *Time Series Data Analysis and Theory*, Holt, Rinehart, and Winston Inc., pp. 19/93/247, 1975

Brillinger, D. R., Some Uses of Cumulants in Wavelet Analysis, *Nonparameteric Statistics*, Special Issue, NPS 001, 1995.

Cohen, Leon, *Time Frequency Analysis*, Prentice Hall, 1995.

Daley, D. J. and D. Vere-Jones, A summary of the theory of point processes, in P.A. W. Lewis, ed., *Stochastic Point Processes*, Wiley, New York, pp. 299-383, 1972.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Daubechie, Ingrid, Ten Lectures on Wavelets, *Society for Industrial and Applied Mathematics*, Philadelphia, PA, 1992.

Donoho, D. L. and I. M. Johnstone, Ideal spatial adaptation by wavelet shrinkage, *Biometrika*, 81,3, pp. 425-455, 1994.

Donoho, David l., Iain M. Johnstone, Gerard Kerkyacharian, and Dominique K. Picard, Wavelet Shrinkage: Asymptopia, *JRSS B*, 57, 2, 1995.

Feuer, Arie and Goodwin, Graham C., *Sampling in Digital Signal Processing and Control*, Birkhauser, 1996.

Friedman, J. H., A Variable Span Smoother. *Tech. Rep. No. 5, Laboratory for Computational Statistics*, Dept. of Statistics, Stanford Univ., California., 1984

Gao, Hung-Ye, *Wavelet Estimation of Spectral Densities in Time Series Analysis*, Ph.D. dissertation, University of Berkeley, 1993.

Gao, Hong-Ye, and Bruce Andrew, *S+Wavelets*, StatSci Division MathSoft Inc., 1994.

Gelb, Arthur (ed), *TASC: Applied Optimal Estimation*, MIT Press, pp. 204-207, eighth printing, 1984.

Hall, Peter, Gerard Kerkyacharian and Dominique Picard, Block Threshold Rules for Curve Estimation using Kernel and Wavelet Methods, *The Annals of Statistics*, vol. 26, no. 3, 922-942, 1998.

Hayes, Monson H., *Statistical Digital Signal Processing and Modeling*, John Wiley and Sons Inc., 1996.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Helstrom, C. W., *Statistical Theory of Signal Detection*, Pergamon Press, New York, 1960.

Herzberg, A. M., Data: *A Collection of Problems from Many fields for the Student and Research Worker*, Springer-Verlag, 1985.

Higgins, J.R., *Sampling Theory in Fourier and Signal Analysis*, Oxford Science Publications, 1996.

Hoel, Paul G., Sidney C. Port, and Charles J. Stone, Introduction to Probability Theory, Houghton Mifflin Company, 1971.

Hogg, Robert V., and Allen T. Craig, *Introduction to Mathematical Statistics*, Macmillan Publishing, 1978.

Lawson, J. L. and G. E. Uhlenbeck (eds.), Threshold Signals, *MIT Radiation Laboratory Series*, vol. 24, McGraw-Hill Book Company, New York, p. 177, 1950.

Lehr, Mark E., and Keh-Shin Lii, Template Basis Techniques to Pattern Recognition, *SPIE Proceedings*, Vol 2825, pp. 972-981, August 1996.

Lehr, Mark E., and Keh-Shin Lii, Automated Mine Identification using Wavelet Analyzing Functions, *Proceedings of the Technology and the Mine Problem Symposium*, Vol II, pp. 7/199-205, Nov 18-21, 1996.

Lehr, Mark E., and Keh-Shin Lii, Wavelet Spectral Density Estimation Under Irregular Sampling, *The Thrity-First Asilomar Conference on Signals, Systems and Computers*, November 2-5, 1997.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Lii, Keh-Shin and Elias Masry, Spectral estimation of continuous-time stationary processes from random sampling, *Stochastic Processes and their Applications* 52, pp. 39-64, 1994.

Mallet, Stephane, A theory for multiresolution signal decomposition: the wavelet representation, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(12), pp 2091-2110, 1989.

Masry, Elias, Alias-free sampling: an alternative conceptualization and its applications, *IEEE Trans. Inform. Theory*, IT-24, pp. 317-324, 1978.

Math Works, *MATLAB*, Prentice Hall, 2000.

Mendel, J.M., Tutorial on higher-order statistics (spectra) in signal processing and system theory: theoretical results and some applications, *IEEE vol. AC-16*, pp.847-869, 1991.

Mendel, Jerry M., *Lessons in Estimation Theory for Signal Processing, Communications, and Control*, Prentice Hall, 1995.

Meyer, Yves, *Wavelets: Algorithms and Applications*, SIAM, 1993.

Neumann, M. H., Spectral Density Estimation Via Non-Linear Wavelet Methods for Stationary Non-Gaussian Time Series, *Journal of Time Series Analysis*, Vol. 17, No. 6, 1996.

Nikias, C.L. and J. M. Mendel, Signal processing with higher-order spectra, *IEEE Signal Processing Magazine*, Vol. 10, pp. 10-37, 1993.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Nikias C. L. and A. Petropulu, *Higher-Order Spectral Analysis: A Nonlinear Signal Processing Framework*, Englewood Cliffs, NJ, Prentice Hall, 1993.

Papoulis, A., *Probability, random Variables, and Stochastic Processes*, 3rd edition, New York: McGraw-Hill, 1991.

Press, William H., Saul A. Teukolsky, William T. Vetterling, and Brian P. Flannery, *Numerical Recipes in C*, Cambridge University Press, Second Edition, pp. 576-584 and 602-606, 1992.

Rohatgi, V. K., *An Introduction to Probability Theory and Mathematical Statistics*, John Wiley and Sons, New York, pp. 612-640, 1976.

Skolnick, Merrill I., *Introduction to Radar Systems*, McGraw-Hill Book Company, Second Edition, 1980.

Strang, Gilbert, and Truong Nguyen, *Wavelets and Filter Banks*, Wellesley-Cambridge Press, 1996.

Swerling, P. Probability of Detection for Fluctuating Targets, *IRE Trans.*, vol. IT-6, pp. 269-308, April, 1960.

Triebel, Hans, *Theory of Function Spaces*, Monographs in Mathematics, Birkhauser, 1983

Vetterli, Martin, and Jelena Kovacevic, *Wavelets and Subband Coding*, Prentice Hall, 1995.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.