

A. HOLDSWORTH
A WAVELET PERSPECTIVE ON EPIDEMIOLOGICAL TIME SERIES

**WAVELET ANALYSIS OF TIME SERIES AND ITS APPLICATION TO
EPIDEMIOLOGICAL DATA**

By

AMBER HOLDSWORTH, BSc HONOURS

A thesis submitted to the School of Graduate Studies in partial fulfillment of the requirements for the
degree Master of Science.

McMaster University

MASTER OF SCIENCE (2008)
(Math)

McMaster University
Hamilton, Ontario

TITLE: Wavelet Analysis of Time Series and its Application to Epidemiological Data

AUTHOR: Amber Marie Holdsworth, BSc Honours (University of Calgary)

SUPERVISOR: Dr. Nicholas Kevlahan

NUMBER OF PAGES: ix, 75

ABSTRACT

Wavelets offer many unique tools for data analysis. The first part of this thesis is an exposition of wavelet transforms and many of the associated tools for the analysis of time series. Specifically, wavelet denoising, wavelet power spectra and the detection of singularities are examined in detail. The second half of the thesis consists of a wavelet perspective on epidemiological time series. Focus is placed on analysing the incidence of infection of measles in Ontario, Canada. Other incidence data sets are also considered: chicken pox, rubella, and whooping cough. We show that wavelet analysis can be used to evaluate mathematical models in epidemiology by testing them against observed data, as well as to characterize the fine scale structure of the data. With some serendipity, it is also shown that distinct data sets for the same disease are characterized by a similar multifractal signature.

Acknowledgments

First and foremost I would like to thank my supervisor, Dr. Nicholas Kevlahan, whose wisdom, encouragement, and professionalism added considerably to my graduate experience. For his tact, patience, and attention to detail, I cannot thank him enough. I would also like to thank Dr. David Earn for his insightful comments and advice and Brad Wagner for many informative discussions and the use of his stochastic model. Finally, I thank my family for their continual support. Especially Mike Nelles for his tireless efforts.

Contents

1	Introduction	1
2	The Wavelet Transform	5
2.1	The Fourier Transform	5
2.2	The Continuous Wavelet Transform	6
2.2.1	Implementing the Continuous Wavelet transform	8
2.3	The Discrete Wavelet Transform	8
2.3.1	Implementation of DWT	12
2.3.2	Higher Dimensions	13
2.4	Properties of Wavelet Basis	14
2.5	Biorthogonal Wavelets	17
2.5.1	The Lifting Scheme	18
2.6	Summary	20
3	Wavelet Analysis Tools	21
3.1	Reproducing the Fourier Energy Spectrum	21
3.2	Signal Regularity	23
3.2.1	Detecting/Measuring Singularities	25
3.2.2	Local Regularity: Maximum Modulus Method	28
3.2.3	Local Regularity: Wavelet Leaders	28
3.2.4	Cusp and Chirp type Singularities	29
3.3	Multifractal Analysis	31
3.3.1	Fractal Sets and Functions	31
3.3.2	WTMM Multifractal Formalism	33
3.3.3	Examples	34
3.3.4	Wavelet Denoising	39
3.4	Testing the Capabilities of Wavelet Thresholding	40

4	Applications in Epidemiology	49
4.1	Modeling in Epidemiology	49
4.1.1	Noise and Stochasticity	51
4.1.2	Seasonal Forcing	51
4.2	Power Spectra	52
4.3	Wavelet Analysis of Epidemiological Time Series	54
4.3.1	The Data	54
4.3.2	Denoising the Data	55
4.3.3	Measles Power Spectra	57
4.3.4	Singularity Spectra	60
5	Conclusions	69

List of Figures

2.1	The Mexican hat wavelet basis at the position, t_0 , and scale, a	7
2.2	a) Daubechies 2 wavelet with 1 vanishing moment. b) Daubechies 4 wavelet with 2 vanishing moments. c) Daubechies 6 wavelet with 3 vanishing moments. d) Daubechies 20 wavelet with 10 vanishing moments.	17
3.1	Paul and Perrier wavelets in Fourier space. Both wavelets have infinitely many vanishing moments, but Perrier's wavelet decreases faster at the origin in Fourier space	23
3.2	(a) A random phase signal constructed from an algebraically decreasing Fourier spectra. (b) The Paul wavelet is used to compute the power spectrum.	24
3.3	The Mexican Hat wavelet used to generate the wavelet scalogram for a signal with multiple singularities. The WC are amplified appropriately to the 256 bit color map. Lighter pigments correspond to higher values of wavelet coefficients.	26
3.4	A cusp singularity with Holder exponent $\alpha = 0.6$. WTMM (left) and WL (right). Recall: $\alpha = slope - 1/2$	30
3.5	A chirp type singularity with Holder exponent $\alpha = 0.6$, $\beta = 1$. WTMM (left) and WL (right). Recall: $\alpha = slope - 1/2$	30
3.6	The singularity spectrum (right) for $B_{0.7}$ is generated using the Mexican hat wavelet and 2^{14} points. The scaling exponent (left) is computed with $q \in [-5, 5]$	35
3.7	a) The Cantor function with $p_1 = 0.4$, $p_2 = 0.6$. 2^{20} points were used. b) Partition function. c) Scaling exponent. d) Singularity spectrum.	36
3.8	a) Log transformed scaling exponent for FBM, $H = 0.7$ (2^{20} points). b) Log transformed scaling exponent for the Cantor function with $p_1 = 0.4$, $p_2 = 0.6$ (2^{20} points). c) Derivative of the scaling exponent for both functions. The mean is subtracted for easy comparison. d) The behavior of $\tau'(q)$ for the Cantor function as compared with equation 3.31.	38
3.9	The convergence of the approximated spectrum to the theoretical spectrum as the number of points used increases.	39
3.10	Wavelet de-noising via Hard thresholding.	41

3.11	Wavelet de-noising via Visushrink. The recovered sine wave “shrinks” more for higher noise levels, σ	41
3.12	The results of signal denoising using the Haar wavelet are shown. The length of the input signal is $N = 2^{11}$	42
3.13	The affect of wavelet de-noising on the wavelet scalogram. a) Cantor function plus GWN with $\sigma = 0.01$. b) The Cantor function. c) The Cantor function recovered with hard thresholding. d) Cantor function recovered with soft thresholding. The scalograms are plotted within the cone of influence to eliminate edge effects.	43
3.14	The affect of wavelet de-noising on the wavelet scalogram. a) Cantor function plus GWN with $\sigma = 0.1$. b) The Cantor function. c) The Cantor function recovered with hard thresholding. d) Cantor function recovered with soft thresholding. The scalograms are plotted within the cone of influence to eliminate edge effects.	43
3.15	Pointwise error analysis for translation invariant de-noising of the Sine function. The slopes of the associated best fit lines, m are found in the legend. a) The known amplitude of the noise, σ , is used. b) σ is estimated.	44
3.16	Pointwise error analysis for translation invariant de-noising of the Cantor function. The slopes of the associated best fit lines, m are found in the legend. a) The known amplitude of the noise, σ , is used. b) σ is estimated.	45
3.17	The error in slope as a function of the SNR for a random phase signal with $E(k) \approx k^{-8}$	46
3.18	Both figures show an instance of de-noising the power spectrum for a random phase signal with algebraically decreasing spectra ($\sigma = 0.1$). Left: $E(k) \approx k^{-8}$. Right: $E(k) \approx k^{-\frac{5}{3}}$	47
4.1	Comparison of the sinusoidally forced stochastic model incidence with Ontario incidence.	55
4.2	Comparison of the term-time model incidence with Ontario incidence.	56
4.3	The singularity spectrum for Ontario measles incidence 1921-1989 plotted with spectra for the de-noised data. The Haar wavelet is used to de-noise the time series and the Mexican hat wavelet is used to plot the multifractal spectra.	57
4.4	De-noising of measles incidence data 1904-1989. On the left, the Haar wavelet is used to decompose the signal. The top graph shows the actual incidence data. Below, the noised removed via Soft thresholding is shown. On the bottom the noised removed via Hard thresholding is shown. The graphs on the right use the Daubechies 20 wavelet, the ones on the left use the Haar wavelet.	58

4.5	Probability density histograms of noise extracted from Ontario incidence data. The top two figures are associated with the Haar wavelet for soft (left) and hard(right) thresholding. The bottom two are associated with the Daubechies 20 wavelet.	59
4.6	The local wavelet power spectrum for Ontario measles incidence 1904-1989. The data is log transform, the mean is subtracted and the data is normalized by the variance, $1/\rho^2$, lag1 = 0, 95% confidence interval shown in thick contours. The Morlet wavelet is used.	59
4.7	Local wavelet power spectra for the stochastic SEIR models. For the term-time forced model (left) $\beta = 0.1$ (a), 0.25 (c). For the sinusoidally forced model, (right) $\beta = 0.05$ (b), 0.08 (d) are shown. The data is log transform, the mean is subtracted and the data is normalized by the variance, $1/\rho^2$, lag1 = 0.9, 95 % confidence interval shown in thick contours. The analyzing wavelet is the Morlet wavelet.	60
4.8	Singularity spectra of measles data compared with stochastic data. The line $x = 0$ is plotted to demarcate the smooth ($\alpha \geq 0$) and non-smooth ($\alpha < 0$) regions. The line $x = 1/2$ is plotted to divide regions of persistent and anti-persistent behaviors. The Mexican Hat wavelet was used to compute all singularity spectra.	61
4.9	Scaling exponents for Ontario measles incidence of infection data and stochastic model incidence of infection data. Non-linear behavior of the scaling functions indicates that all of the data exhibits a multifractal structure.	62
4.10	Singularity spectra of the time-term stochastic model for different amplitudes of seasonal forcing compared with ON data. The Mexican Hat wavelet is used to plot the Multifractal spectra.	63
4.11	The effect of adding Gaussian white noise to the models on the singularity spectra for the stochastic SEIR models. a) $\sigma = 1000$, b) $\sigma = 2500$. Figure 4.8 shows the singularity spectra for the models with no noise added.	64
4.12	Left: Singularity spectra of incidence of infection for several diseases. Right: corresponding scaling exponents.	65
4.13	Top: Singularity spectra for incidence of chicken pox. Bottom: Singularity spectra for incidence of measles. The similarities between the various Chicken pox data sets are contrasted with their differences from the measles data sets.	67
4.14	Left: Scaling exponent for chicken pox incidence. Right: Scaling exponent for measles incidence.	68

Chapter 1

Introduction

Developed in the 1980s, thanks to contributions from Meyer, Morlet, Grossman and others, wavelet transforms provide a range of tools for data analysis and data compression [8]. The concepts that lead to the development of wavelets may be traced back much further than the 1980s. In fact, wavelet transforms can be thought of as a natural extension of Fourier transforms.

The idea that functions or signals may be better understood by transforming them to convenient representations is central to wavelet analysis. Joseph Fourier capitalized on this idea in the early 1800s with the introduction of the Fourier transform. This transform is utilized in many fields of study for its ability to unveil the frequency content of a signal. Wavelets should not be thought of as replacing Fourier transforms, but as a new scientific language which offers a unique description of the analyzed data. Fourier transforms provide a description in terms of the frequency content alone, while wavelets provide information about both frequency and time scale.

Fourier transforms provide precise information about how much of each frequency a signal contains, but give no information about where these frequencies occur. In a sense, the time information is hidden, or masked, by the transform. After all, reconstruction of the time series would not be possible if the time information were lost completely. This lack of time resolution can lead to insurmountable errors [19]. Imagine recording a signal over a period of several hours. If the last few minutes of the recording are in error, then this error will propagate through the entire Fourier transform. Local information about the signal becomes global information in the Fourier transform.

The Windowed Fourier transform is an early attempt to establish localization in time. The idea of this transform is to analyze smaller segments of the signal with the Fourier transform to obtain some time localization. A window of constant length is chosen and “slides” across the signal. If the window is chosen too narrow, low frequency components of the signal become virtually imperceptible. If the window is chosen too large, time resolution is lost. The need for a

transform that allowed for variable window lengths became apparent.

In the 1970s, petrochemical engineer Jean Morlet was faced with the problem of analyzing a signal that had high frequency components on a shorter time span than low frequency components. He wished to increase time resolution for the high frequency components and maintain a reasonable frequency resolution for the low frequency components. To accomplish this, Morlet developed a new kind of transform based on the Fourier transform. Instead of fixing the window length and filling the window with oscillations of different frequencies, Morlet kept the number of oscillations in the window fixed and varied the length of the window. The functions he generated had the incredible advantage that the high frequency functions were narrow while the low frequency functions were not [8]. Morlet called these wavelets “wavelets of constant shape” [19] because the overall shape of the wavelet remained the same despite stretching or compressing.

Mathematical physicist, Alex Grossman, realized that Morlet’s transform was very similar to methods Grossman himself had developed in the field of quantum mechanics. Together, they came up with a formula for the exact reconstruction of the signal from the transform. Mathematician Yves Meyer also became interested in wavelet research when he discovered that the reconstruction formula of Grossman and Morlet was actually a rediscovery of a formula developed in the 1960s by Calderon [8]. Through the joint efforts of researchers in physics, engineering, and mathematics, the theory of wavelet analysis found its roots.

The process of studying coarse to fine resolutions of a signal is referred to as multiresolution analysis. Mallat’s efforts to develop wavelet analysis in a multiresolution based framework lead to an efficient algorithm for computing the transform [23]. At the age of twenty-three Mallat contacted Meyer with the idea that orthogonal wavelet transforms would be a versatile tool in computer vision. With the help of Meyer, Mallat wrote a magnificent paper that would help to unify wavelet theory and establish a practical framework for wavelet construction. He showed how a *scaling function* could be used for fast computations of the transform. His fast $O(N)$ algorithm used truncated versions of infinite orthogonal wavelets.

The next major contribution to wavelet theory would come from Daubechies. She is credited with the introduction of compactly supported orthogonal wavelet basis. This improved both computational efficiency and accuracy. The importance of the contributions of Mallat and Daubechies cannot be overstated. The nuances of their work will become clearer in the next chapter.

This contagious affinity for wavelet analysis continues to inspire researchers. Bi-orthogonal wavelets and the lifting scheme have gained a lot of attention in recent years. Wim Sweldens can be credited with much of the work in this area. The great thing about wavelets constructed with the lifting scheme is that they are independent of the Fourier transform. This means that they can be used in settings where the Fourier transform is inappropriate. A more detailed exposition of biorthogonal wavelets will be given in the next chapter.

Three terms are commonly used without explanation in the literature: scale, resolution, oc-

tave. Scale is a term used to describe how changing the size of the wavelet will change the size of the signal components examined. Coarser scales highlight low frequency components, while finer scales highlight lower frequencies. Resolution is meant to refer to the number of wavelets being used or, in other words, the number of times the signal is sampled. The word octave refers to changes in the resolution of the wavelet. Increasing wavelet frequency on an interval, decreases the size or scale of the wavelet. Higher frequency components are encoded by a higher resolution of wavelets.

Table 1.1 lists and summarizes several applications of the wavelet transform. This list is by no means exhaustive. Adaptive meshing and data compression are not explored here, but represent two of the most fruitful applications of wavelet analysis. The reader is referred to an intriguing historical perspective on wavelets given by Barbara Burke Hubbard for a more comprehensive review [19].

Many areas of medical imaging had long utilized Fourier transforms until the advantages of wavelets became clear. Small recording/imaging errors do not corrupt the entire wavelet transform because of their excellent time localization properties. Medical imaging can be quite costly. Which is part of the reason wavelets have gained such high acclaim in this field. Applications include high resolution electrocardiographs, brain resonance images and MRIs [19].

As an example of the current use of wavelet technology the Joint Photographic Experts Group (JPEG) introduced a new image standard called JPEG 2000 [33]. JPEG 2000 uses the discrete wavelet transform to compress images. This standard has several advantages over the old standard such as reducing visual artifacts, and offering a multiresolution representation.

In this thesis, epidemiological time series are studied from a wavelet analysis point of view. Previous work in this area is scarce. Bauch, Grenfell and others used the wavelet power spectrum to analyze recurrence in epidemiological time series [5] [6]. Some of their results are reproduced in the last chapter and compared with spectra generated stochastically. Both wavelet power spectra and singularity spectra are examined as tools for determining the efficacy of stochastic models. Based on the literature review, we believe that this study is the first to use the language of multifractal analysis to provide a description of epidemiological time series.

Part of the inspiration for this work is the observation that so-called stochastic models exhibit fine scale fluctuations that could be indicative of the presence of noise, or an inherent fractal structure in the data. Wavelet de-noising is used to compare the numerical models to the observed incidence of infection. The following questions are investigated with regard to this fine scale structure: is this “noise” Gaussian or are these fluctuations a true feature of the dynamics? Does the data or model exhibit a multifractal structure?

The interesting questions alluded to here will be examined in the last chapter. Before any conclusions are drawn, a detailed description of wavelet analysis must be given. To this end, this thesis is comprised of two distinct sections. Chapters two and three develop the theory of

Table 1.1: WAVELET APPLICATIONS

DATA COMPRESSION:	Allows for efficient compression with minimal data loss due to low magnitude wavelet coefficients.
WAVELET DENOISING:	Used to remove noise from a signal. Efficient de-noising of signals with unknown smoothness. Applications in image processing.
METHODS FOR PDES:	Wavelet based numerical algorithms for generating an adaptive grid and Galerkin methods for solving PDEs. Applications to Geodesy, Climate Modeling, Weather Prediction.
MULTIFRACTAL ANALYSIS:	Compute the singularity spectrum for a given signal. Applications in Turbulence.
SPECTRAL ANALYSIS:	Offers time localized picture of the power spectrum. Applications to epidemiology, ecology, astronomy.
LOCAL/GLOBAL REGULARITY:	Determine the uniform and pointwise Lipschitz exponent for a singularity at a given point. Applications in Medicine.

wavelet transforms and present tools for the analysis of time series. Chapter four is devoted to the application of these techniques to incidence of infection time series.

Chapter 2

The Wavelet Transform

2.1 The Fourier Transform

The discovery that many periodic waveforms are well approximated by sums of sine and cosine functions lead to the development of the Fourier series. The Fourier transform (FT) is a generalization of the complex Fourier series obtained as the formal limit of the Fourier series as the period tends to infinity. Discrete Fourier transforms and fast Fourier transforms are used to approximate the FT using a discrete number of points. The inverse FT is used to restore the input signal.

FT offer a method of representing functions from the time (or spacial) domain in the frequency domain. The frequency domain is referred to as Fourier space. It is a function space with sine and cosine functions as a basis; $\{e^{-ikt}, k \in \mathbb{R}\}$. The Fourier space representation of a signal reveals the contribution of the sine and cosine functions at various frequencies.

Definition 2.1.1. 1. *The Fourier Transform for a function f which is integrable on \mathbb{R} is defined by:*

$$\hat{f}(k) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} f(t)e^{-ikt} dt \quad (2.1)$$

2. *The Discrete Fourier Transform (DFT):*

$$\hat{f}(k) = \sum_{l=0}^{N-1} f(t_l)e^{ikt_l} \quad l = 0, 1, 2, \dots, N-1. \quad (2.2)$$

The DFT transforms a signal of N complex numbers to a representation in Fourier space using N complex numbers. The signal is discretized by sampling at the points t_l .

The FT allows one to ascertain which frequencies are present in the signal, but gives no temporal information about the frequencies; it is insufficient for representing non-stationary signals.

A potential solution to this problem is offered by the *windowed Fourier transform* (WFT). This transform works by dividing the input signal, $f(t)$, into sections and analyzing the frequency component of each section separately. The WFT transform is said to be localized in time, however, the width of the window is fixed. Shifting a window of constant size across a signal results in a time-frequency representation with a constant resolution. A smaller scale basis is required to detect sharp changes in frequency, possibly due to discontinuities in the signal, while a larger scale basis is needed to uncover detailed frequency information about the time series. This motivates the need for the wavelet transform which allows for a multitude of basis functions that are not restricted by a fixed window length.

2.2 The Continuous Wavelet Transform

Wavelet transforms are a versatile tool for studying non-stationary time series. To be considered a wavelet, a function, ψ , should have some localization in time and position [13]. Wavelets must satisfy the *admissibility condition*.

$$C_\psi = \int_{-\infty}^{+\infty} \frac{|\hat{\psi}(k)|^2}{k} dk < +\infty. \quad (2.3)$$

This implies that ψ is a function of zero average.

$$\int_{-\infty}^{+\infty} \psi(t) dt = 0. \quad (2.4)$$

The wavelet function performs signal decomposition by dilating and translating the *mother wavelet* at each scale, $a > 0$, and time, b .

$$\psi_{a,b} = a^{-\frac{1}{2}} \psi\left(\frac{t-b}{a}\right) \quad (2.5)$$

Figure 2.1 shows an example of what real wavelets looks like. The factor of $a^{-\frac{1}{2}}$ serves to normalize the wavelet so that the $L^2(\mathbb{R})$ norm is independent of the scale, a .

The wavelet transform is a convolution of the signal with the basis function ψ . Varying the scale parameter changes the time-frequency spread of the wavelet. The height and width of the wavelet are affected by changes to the parameter a , but the average remains constant. The time (position) resolution is inversely proportional to the frequency resolution.

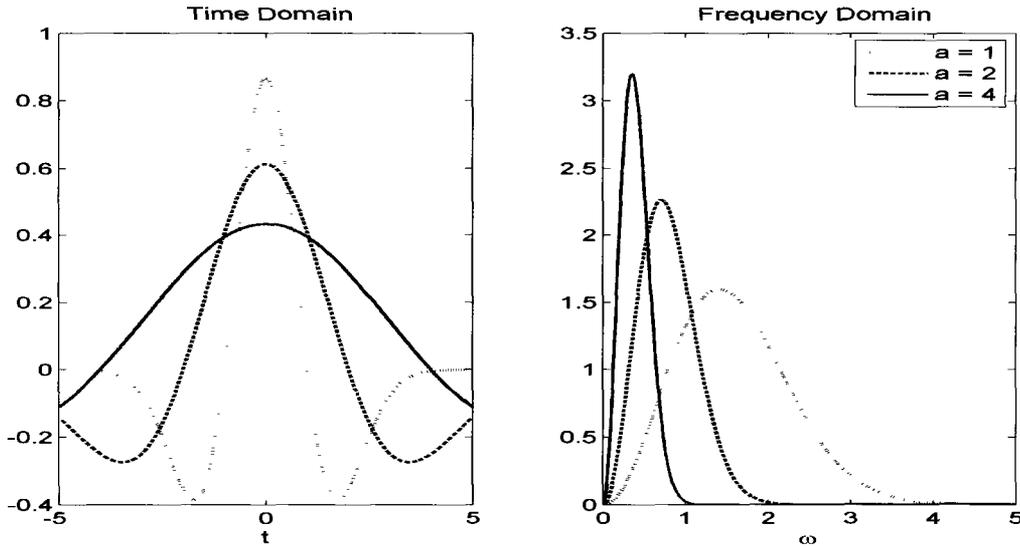


Figure 2.1: The Mexican hat wavelet basis at the position, t_0 , and scale, a .

Definition 2.2.1. *The wavelet transform:*

$$\tilde{s}(a, b) = \int_{-\infty}^{+\infty} s(t) \overline{\Psi_{a,b}(t)} dt = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} s(t) \overline{\Psi\left(\frac{t-b}{a}\right)} dt \quad (2.6)$$

$s(t)$ is the signal to be decomposed, $\Psi(t)$ is the mother wavelet.

$\tilde{s}(a, b)$ is called a wavelet coefficient. It measures the variation of the signal in a neighborhood of the time b and of size proportional to a . Large amplitude coefficients are created by sharp changes in frequency. A one dimensional time series is transformed to a two dimensional wavelet coefficient representation. Overlapping basis functions lead to a redundancy in the wavelet coefficients. So neighboring coefficients share information. This makes the CWT an excellent tool for detecting specific features of a signal.

For the transform to be complete one must be able to recover the signal completely. Perfect reconstruction of the time series make the wavelet transform well suited to applications such as data compression. Reconstruction is possible using theorem 2.2.2.

Theorem 2.2.2. [17] *If $\Psi \in L^2(\mathbb{R})$ is a real function s.t. equation 2.3 holds then any $s \in L^2(\mathbb{R})$ satisfies*

$$s(t) = \frac{1}{C_\Psi} \int_0^{+\infty} \int_{-\infty}^{+\infty} \tilde{s}(a, b) \frac{1}{\sqrt{a}} \Psi\left(\frac{t-b}{a}\right) db \frac{da}{a^2} \quad (2.7)$$

2.2.1 Implementing the Continuous Wavelet transform

FT are used to compute the wavelet coefficients. The Fourier transform of a convolution is the product of the FT and the Fourier transform of a product is the convolution of the FT [23]. If $j = 0$ represents the input signal resolution, where $a = 2^j$, then for all $a > 1$ the CWT is interpreted as a convolution of the signal with the appropriately scaled wavelet function.

$$\tilde{s} = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} s(t) \overline{\psi\left(\frac{t-b}{a}\right)} dt = s \star \frac{1}{\sqrt{a}} \psi\left(\frac{-t}{a}\right) \quad (2.8)$$

The transform can be completed by integrating the CWT given in definition 2.2.1 in physical space, but the CWT is computed most efficiently using the fast Fourier transform (FFT). The signal, and appropriately scaled wavelet, is transform into Fourier space, multiplied, and transformed back into physical space.

In practice, the CWT is computed using a finite number of convolutions. So one must choose a sufficiently small step size for a and b . The output wavelet coefficient data is one dimension larger then the input data. Inputting a signal of length N results in output of N points for each scale of the transform. The number of scales is chosen task specifically, depending on how much scale resolution is desired. Computational cost increases as the number of scales increases. Having many scales means having a great deal of redundancy too. This redundancy can be taken advantage of by choosing a small number of scales for efficiency or a large number for a highly resolved analysis.

2.3 The Discrete Wavelet Transform

The discrete wavelet transform (DWT) operates on dyadic scales, avoiding the redundancies of the continuous wavelet transform. A one-dimensional time series is transformed into a one-dimensional wavelet coefficient representation. The requirement of orthogonality in the wavelet basis distinguishes this transform from those discussed previously. Although the CWT is computed with a discrete range of values, the wavelets used are not orthogonal. Mallat introduces the discrete wavelet transform in the context of multiresolution analysis [22]. This section summarizes Mallat's work by first introducing a multiresolution approximation and then proceeding to describe the theory needed to implement the transform.

Let the scale, 2^j , be the inverse of the resolution.

Definition 2.3.1. A multiresolution approximation is a sequence $\{V_j\}_{j \in \mathbb{Z}}$ of closed subspaces satisfying:

1. $\forall (j, k) \in \mathbb{Z}, f(t) \in V_j \iff f(t - 2^j k) \in V_j$. V_j is invariant by any translation in proportion to the scale 2^j .
2. $\forall j \in \mathbb{Z}, V_{j+1} \subset V_j$.
3. $\forall j \in \mathbb{Z}, f(t) \in V_j \iff f\left(\frac{t}{2}\right) \in V_{j+1}$.
4. $\lim_{j \rightarrow +\infty} V_j = \bigcap_{-\infty}^{+\infty} V_j = \{0\}$.
5. $\lim_{j \rightarrow -\infty} V_j = \text{closure} \left(\bigcup_{-\infty}^{+\infty} V_j \right) = L^2(\mathbb{R})$.
6. $\exists \theta$ s.t. $\{\theta(t - n)\}_{n \in \mathbb{Z}}$ is a Riesz basis of V_0 .

The approximation of a function, $f(t)$, at each scale is an orthogonal projection on the space $V_j \subset V_{j-1}$. Let W_j be the orthogonal complement of V_j in V_{j-1} . Then $V_j \oplus W_j = V_{j-1}$. The orthogonal projection on the space W_j contains the *details* of the finer approximation V_{j-1} . The wavelets span the space $\{W_j\}_{j \in \mathbb{Z}}$, and *scaling functions*, ϕ , span the space $\{V_j\}_{j \in \mathbb{Z}}$. At each scale the signal approximation is decompose into a *detail* component and an *approximation* component. In effect, $V_1 = W_2 \oplus W_3 \oplus \dots \oplus W_J \oplus V_J$.

Let $A_{2^{-j}}^d f$ be the *discrete approximation* of the signal $f(t)$ at a resolution 2^{-j} . The input signal will be denoted with $j = 0$ (i.e. $A_1^d f$). Subsequent approximations are denoted with $0 > j \geq -J$. The decomposition is dyadic, at each scale the resolution decreases by half and scales are limited to a maximum $J = \log_2(N)$ where N is the number of points used to represent the input signal.

The notation used in this section mimics that of Mallat (1989) [22]. Other notations are encountered in the literature. Some authors prefer to denote the scale of the signal as 2^J . The first level of the approximation would then be at scale 2^{J-1} . Because the scales are known to be dyadic, it is not uncommon to find the scale denoted as simply j corresponding to a resolution of 2^{-j} . One should always be careful to keep in mind the notation which is being used.

Theorem 2.3.2. [22] Let $(V_{2^j})_{j \in \mathbb{Z}}$ be a multiresolution approximation of $L^2(\mathbb{R})$. \exists a function $\phi(t) \in L^2(\mathbb{R})$, called a *scaling function*, such that if $\phi_{2^j}(t) = 2^j \phi(2^j t)$ for $j \in \mathbb{Z}$ then, $\{\sqrt{2^{-j}} \phi_{2^j}(t - 2^j n)\}_{n \in \mathbb{Z}}$ is an orthonormal basis of V_{2^j} .

Since ϕ generates an orthogonal basis of each space V_j , ϕ completely characterizes the multiresolution approximation. The discrete approximation is given by an inner product which can be interpreted as a convolution product evaluated at a point $2^{-j}n$.

$$\begin{aligned}
A_{2^j}^d f &= (\langle f(u), \phi_{2^j}(u - 2^j n) \rangle) \\
&= \int_{-\infty}^{+\infty} f(u) \phi_{2^j}(u - 2^j n) du \\
&= (f(u) \star \phi_{2^j}(-u))(2^{-j} n)
\end{aligned} \tag{2.9}$$

The algorithm described above is sometimes referred to as a fast filter bank algorithm. This is because any scaling function is specified by a discrete filter called a conjugate mirror filter. A conjugate mirror filter is a discrete function whose transfer function satisfies:

$$\forall k \in \mathbb{R}, |\hat{h}(k)|^2 + |\hat{h}(k + \pi)|^2 = 2. \tag{2.10}$$

The transform is computed by passing the signal into a filter bank and downsampling at each iteration. Let $\{V_{2^j}\}_{j \in \mathbb{Z}}$ be a multiresolution approximation with scaling function $\phi(t)$. Since $\phi_{2^j}(t - 2^{-j}n) \in V_{2^j} \subset V_{2^{j+1}} \forall n \in \mathbb{Z}$, it can be expanded in an orthonormal basis of $V_{2^{j+1}}$.

$$\phi_{2^j}(t - 2^j n) = 2^{-j-1} \sum_{k=-\infty}^{+\infty} \langle \phi_{2^j}(u - 2^j n), \phi_{2^{j+1}}(u - 2^{-j-1}k) \rangle \cdot \phi_{2^{j+1}}(t - 2^{-j-1}k) \tag{2.11}$$

Change variables in the inner product integral, $u = u - 2^{-j-1}n$, and compute the inner products of $f(u)$ with both sides.

$$\langle f(u), \phi_{2^j}(u - 2^j n) \rangle = \sum_{k=-\infty}^{+\infty} \langle \phi_{2^{-1}}(u), \phi(u - (k - 2n)) \rangle \cdot \langle f(u), \phi_{2^{j+1}}(u - 2^{-j-1}k) \rangle. \tag{2.12}$$

Suppose $h(n) = \langle \phi_{2^{-1}}(u), \phi(u - n) \rangle$ for all $n \in \mathbb{Z}$. By letting H be the filter with impulse response $h(n)$ and letting \tilde{H} be the mirror filter with impulse response $\tilde{h}(n) = h(-n)$ one obtains:

$$\langle f(u), \phi_{2^j}(u - 2^j n) \rangle = \sum_{k=-\infty}^{+\infty} \tilde{h}(2n - k) \langle f(u), \phi_{2^{j+1}}(u - 2^{-j-1}k) \rangle \tag{2.13}$$

The approximation at a scale 2^j is computed by convolving the approximation at resolution 2^{j+1} with \tilde{H} and dyadically down sampling.

$$A_{2^j}^d f = \sum_{k=-\infty}^{+\infty} \tilde{h}(2n - k) A_{2^{j+1}}^d f \tag{2.14}$$

The original signal, $f(t) = A_1^d f$, is represented by $2^J = N$ points. For each scale $2^j < 1$ there are $\frac{N}{2^j}$ points in $A_{2^j}^d f$. Similarly there are $\frac{N}{2^j}$ detail coefficients, $D_{2^j} f$, at each scale 2^j . An orthogonal wavelet representation at a scale 2^J is given by the detail coefficients at all scales $2^J \geq 2^j \geq 1$ and the remaining approximation coefficient at the scale 2^J ; $A_{2^J}^d f$ and $\{D_{2^j} f\}_{0 \leq j \leq J}$. At each scale there are N wavelet coefficients. The set of $D_{2^j} f$ is computed in a similar process as above by convolving the approximation $A_{2^j}^d f$ with the appropriate filter and keeping half of the resulting points.

Theorem 2.3.3. [23] *Let ϕ be a scaling function and H the corresponding conjugate mirror filter. Let $\psi(t)$ be the function whose Fourier transform is given by*

$$\hat{\psi}(k) = \hat{G}\left(\frac{k}{2}\right)\hat{\phi}\left(\frac{k}{2}\right), \quad (2.15)$$

with $\hat{G}(k) = e^{-ik}\hat{H}^*(k + \pi)$.

$$\psi_{j,n}(t) = \frac{1}{2^j}\psi\left(\frac{t - 2^j n}{2^j}\right). \quad (2.16)$$

for any scale 2^j , $\{\psi_{j,n}\}_{j,n \in \mathbb{Z}}$ is an orthonormal basis of W_j . For all scales $\{\psi_{j,n}\}_{j,n \in \mathbb{Z}}$ is an orthonormal basis of $L^2(\mathbb{R})$

For any $n \in \mathbb{Z}$, $\psi_{2^j}(u - 2^{-j}n) \in W_{2^j} \subset V_{2^{j+1}}$. As before, one can expand this function in an orthonormal basis of $V_{2^{j+1}}$. Changing variables and computing the inner product of $f(t)$ with both sides of the resulting equation gives:

$$\langle f(u), \psi_{2^j}(u - 2^j n) \rangle = \sum_{k=-\infty}^{+\infty} \langle \psi_{2^{-j}}(u), \psi(u - (k - 2n)) \rangle \cdot \langle f(u), \psi_{2^{j+1}}(u - 2^{-j-1}k) \rangle \quad (2.17)$$

Let G be the discrete filter with impulse response $g(n)$ and \tilde{G} be the symmetric filter with impulse response $\tilde{g}(n) = g(-n)$.

$$\langle f(u), \psi_{2^j}(u - 2^j n) \rangle = \sum_{k=-\infty}^{+\infty} \tilde{g}(2n - k) \langle f(u), \psi_{2^{j+1}}(u - 2^{-j-1}k) \rangle \quad (2.18)$$

There is a nice equation for computing the detail coefficients.

$$D_{2^j} f = \sum_{k=-\infty}^{+\infty} \tilde{g}(2n - k) A_{2^{j+1}}^d f \quad (2.19)$$

This algorithm is equivalent to a two-channel, multi rate filter bank. It splits the input signal into two bands and reduces the number of points by half. H and G are called *quadrature mirror*

filters. G is a high pass filter which is the mirror filter of the low pass filter, H . The relationship between the two filters' impulse responses can be derived from the previous theorem and is given by:

$$g(n) = (-1)^{1-n}h(1-n). \quad (2.20)$$

The reconstruction uses perfect reconstruction filter banks. Obtaining the appropriate orthogonal reconstruction filter is done using conjugate mirror filters H and G . $A_{2^{j+1}}^d$ is reconstructed by putting zeros between $A_{2^j}^d f$ and $D_{2^j} f$ and convolving with H and G as follows:

$$\begin{aligned} A_{2^{j+1}}^d f &= \langle f(u), \phi_{2^{j+1}}(u - 2^{-j-1}n) \rangle \\ &= 2 \sum_{k=-\infty}^{+\infty} h(n-2k) \langle f(u), \phi_{2^j}(u - 2^{-j}k) \rangle \\ &\quad + 2 \sum_{k=-\infty}^{+\infty} g(n-2k) \langle f(u), \psi_{2^j}(u - 2^{-j}k) \rangle \\ &= 2 \sum_{k=-\infty}^{+\infty} h(n-2k) A_{2^j}^d f + 2 \sum_{k=-\infty}^{+\infty} g(n-2k) D_{2^j} f \end{aligned}$$

An orthogonal wavelet basis is used to specify the scaling functions and filter coefficients are generated from the scaling functions. Details with regard to the choice of basis are given below.

2.3.1 Implementation of DWT

One may compute the DWT using a circular convolution matrix [26]. Start with a wavelet with impulse response:

$$h = \begin{pmatrix} c_1 \\ c_2 \\ c_3 \\ c_4 \end{pmatrix} \quad (2.21)$$

Recall that a discrete convolution is given by $y_j = \sum_{k=-\infty}^{+\infty} A_k c_{j-k}$. A matrix of the following form can be used to perform the required convolution.

$$C_H = \begin{pmatrix} c_3 & c_2 & c_1 & 0 & \dots & 0 & c_4 \\ c_4 & c_3 & c_2 & c_1 & 0 & \dots & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ c_1 & 0 & \dots & 0 & c_4 & c_3 & c_2 \end{pmatrix} \quad (2.22)$$

If the signal initially contains N sample points C_H is a $N \times N$ matrix. A similar matrix, C_G is created to perform the convolution associated with the filter G . By taking the odd numbered rows of C_H and placing them into the odd rows of C_G a convolution matrix $C_{H,G}$ is formed. $C_{H,G}$, when multiplied by a vector containing the N sample points of the signal on the right, performs the circular convolutions and keeps only half of the resulting points. The result is a vector which has $A_{2^{-1}}^d f$ in the odd numbered rows and $D_{2^{-1}} f$ in the even rows. To further decompose the signal the detail coefficients must be separated from the approximation coefficients. A matrix of the form $C_{H,G}$ of size $\frac{N}{2} \times \frac{N}{2}$ is created and applied to the, $\frac{N}{2}$, $A_{2^{-1}}^d f$ coefficients.

If the signal is decomposed to the scale 2^J it is convenient to store the wavelet coefficients in a row vector with the single approximation coefficient $A_{2^{-J}}^d f$ on the left and the $D_{2^{-1}} f$ coefficients on the right.

$$\left[A_{2^{-J}}^d f \quad D_{2^{-J}} f \quad D_{2^{-(J-1)}} f \quad \dots \quad D_{2^{-1}} f \right] \quad (2.23)$$

Using the inverse of the matrix $C_{H,G}$ the signal is reconstructed from the wavelet coefficients. Since $C_{H,G}$ is orthogonal, the inverse matrix is simply the transpose matrix. For efficient computation it is convenient to use vectors to represent the convolution matrix.

The fast pyramid algorithm of Mallat requires only $O(N)$ multiplications. This is much faster than the CWT and the FFT which require $O(N^2)$, and $O(N \log_2(N))$ multiplications, respectively.

2.3.2 Higher Dimensions

The present work focuses on the one dimensional wavelet transform for a time series, none the less it is important to mention the extension of the wavelet transform to higher dimensions. The two-dimensional DWT is a direct extension of the one dimensional DWT. In this case, the signal to be examined must be a finite energy function $f(t,y) \in \mathbf{L}^2(\mathbb{R})$. Define, $(V_{2^j})_{j \in \mathbf{Z}}$, a multiresolution approximation of $\mathbf{L}^2(\mathbb{R})$. Where,

$$V_{2^j} = V_{2^j}^1 \oplus V_{2^j}^1$$

Mallat presents a concise, descriptive explanation of how the decomposition is done with orthogonal wavelets in two-dimensions [22]. The process is very similar to that of the one dimensional case. An important difference between the 1D and 2D case is that there are more detail coefficients resulting from the two dimensional decomposition. $A_{2^{j+1}}^d f$ is decomposed into $A_{2^j}^d f$, $D_{2^j}^1 f$, $D_{2^j}^2 f$, $D_{2^j}^3 f$. The 2D transform can be thought of as 1D transformations in the t and y directions. The process is as follows:

- Convolve the rows of approximation $A_{2^{j+1}}^d f$ with the one dimensional filter \tilde{H} .

- Keep half of the resulting rows.
- Convolve the columns of the down sampled matrix with the one dimensional filter \tilde{G} and separately with the filter \tilde{H} .
- Keep half of the resulting columns. The result is one of the detail components $D_{2^j}^1 f$ and the approximation component $A_{2^j}^d f$.
- Now repeat this process using the discrete filter \tilde{G} instead of \tilde{H} in the first step. The result will be the detail components $D_{2^j}^2 f$ and $D_{2^j}^3 f$.

For three dimensions the DWT can be thought of as 1d transformations in the t , y and z directions. There will be seven detail coefficients and one approximation coefficient at each scale. In this way, the wavelet transform is extended to as many dimensions as is required.

2.4 Properties of Wavelet Basis

In wavelet analysis there is a necessary trade-off between time localization and frequency localization. This can be understood by recalling the Hiesenberg uncertainty principle. The principle is often used in physics to describe the uncertainty as to the position and momentum of a free particle. In the time-frequency plane, (t, k) , one defines the *Heisenberg rectangle* with time width σ_t , and frequency width σ_k . The uncertainty principle implies ([22]):

$$\sigma_t^2 \sigma_k^2 \geq \frac{1}{4} \quad (2.24)$$

Choosing a wider window offers more frequency resolution at the expense of resolution in time. Conversely, choosing a narrower window results in better resolution in time, but less resolution in frequency. Wavelets have variable window lengths which provide wider windows for low frequencies, to increase frequency resolution, and narrower windows for high frequencies, to increase time resolution, but not all wavelets are created equal.

The continuous wavelet transform utilizes a real or complex wavelet function, while the discrete wavelet transform relies on an orthogonal basis. Wavelet basis and wavelet functions are characterized by the number of vanishing moments they possess, the size of their compact support and the smoothness of the wavelet.

Definition 2.4.1. A function $\psi(t)$ is said to have p vanishing moments if

$$\int_{-\infty}^{+\infty} t^k \psi(t) dt = 0 \quad (2.25)$$

for $0 \leq k < p$.

The concept is made clear by expanding the wavelet transform given in definition 2.2.1 using a Taylor series approximation at $t = 0$ [16].

$$\tilde{s}(a, 0) = \frac{1}{\sqrt{a}} \left[\sum_{p=0}^n s^{(p)}(0) \int \frac{t^p}{p!} \psi\left(\frac{t}{a}\right) dt + O(n+1) \right] \quad (2.26)$$

Define the moments $M_p = \int t^p \psi(t) dt$. Now equation 2.26 may be written as:

$$\tilde{s}(a, 0) = \frac{1}{\sqrt{a}} \left[s(0)M_0a + s^{(1)}(0)M_1a^2 + \frac{s^{(2)}(0)}{2!}M_2a^3 + \dots + \frac{s^{(n)}(0)}{n!}M_na^{n+1} + O(a^{n+2}) \right] \quad (2.27)$$

The admissibility condition indicates that $M_0 = 0$. If moments M_1, M_2, \dots, M_n are also zero, then the wavelet has n vanishing moments and the coefficients $\tilde{s}(a, b)$ will decay as fast as a^{n+2} for a smooth signal. For a wavelet with n vanishing moments any polynomial of degree $n - 1$ can be written as a linear combination of appropriately shifted scaling function ψ . It should be noted that the values of M_n need only be close to zero for one to consider it a vanishing moment.

For many applications wavelets are designed so as to minimize the number of non-zero coefficients. Increasing the number of vanishing moments a wavelet possesses results in the removal of more moments from the signal. Wavelets with a high number of vanishing moments produce a higher number of zero, or near zero, wavelet coefficients than wavelets with fewer vanishing moments. Intuitively, one can think of the vanishing moments as a restriction on what the wavelet can “see” [19]. With one vanishing moment the wavelet does not see linear trends in the data, with two vanishing moments quadratic functions become invisible to the wavelet and so on.

Vanishing moments provide a necessary condition for the regularity of a wavelet. The number of continuous derivatives of a wavelet determines the *order of regularity*. A wavelet must have at least $n + 1$ vanishing moments to have regularity greater than or equal to n .

The relationship between vanishing moments and the support of the wavelet suggests that the choice of wavelet basis should be guided by the signal to be analyzed. For example, an isolated singularity at a point, t_i , will generate high amplitude wavelet coefficients at fine scales at that position. If the size of the compact support for the wavelet is large, then many wavelet basis functions will intersect the singularity. For ψ with compact support, K , there are K wavelet coefficients that intersect t_i . To reduce the number of high amplitude coefficients one could choose a wavelet basis with narrow compact support, or increase the number of vanishing moments. Wavelets with many vanishing moments are ideal for functions with few isolated regularities, but for signals with a greater density of singularities it is better to choose wavelets with smaller support size.

Using orthogonal wavelets Daubechies developed a series of basis functions with compact support in Fourier space. She showed that to create a orthogonal wavelet with p vanishing moments it must be of minimum length of $2p$. Daubechies wavelets are often referred to by the length of their associated filters. For example the Daubechies wavelet with 10 vanishing moments has filters of length 20 and is referred to as the D20 wavelet. The supports of ϕ and ψ , for Daubechies wavelets, are $[0, 2p - 1]$ and $[-p + 1, p]$ respectively [23]. For $p = 1$, the Haar wavelet is produced. Increasing the number of vanishing moments decreases the number of non-zero coefficients, but also increases the support size which, in turn, affects the number of non-zero coefficients.

Mallat's fast wavelet algorithm requires approximately $2KN$ operations for a signal of size N where K is the size of compact support of the wavelet. So computational efficiency is improved by reducing the size of support. This is what makes Daubechies wavelets so intriguing. They have a minimum compact support for a given number of vanishing moments.

Figure 2.2 shows several Daubechies wavelets. Notice how as the number of vanishing moments increase, the wavelet becomes more smooth.

It is not necessary, and not always possible, to derive an explicit formula to view the wavelet basis functions. Instead, simply construct a unit vector and apply the inverse wavelet transform.

Continuous wavelets and orthogonal wavelets are both derived as translations and dilations of a mother wavelet. More recently, biorthogonal wavelets were discovered and constructions have been established independent from the frequency domain.

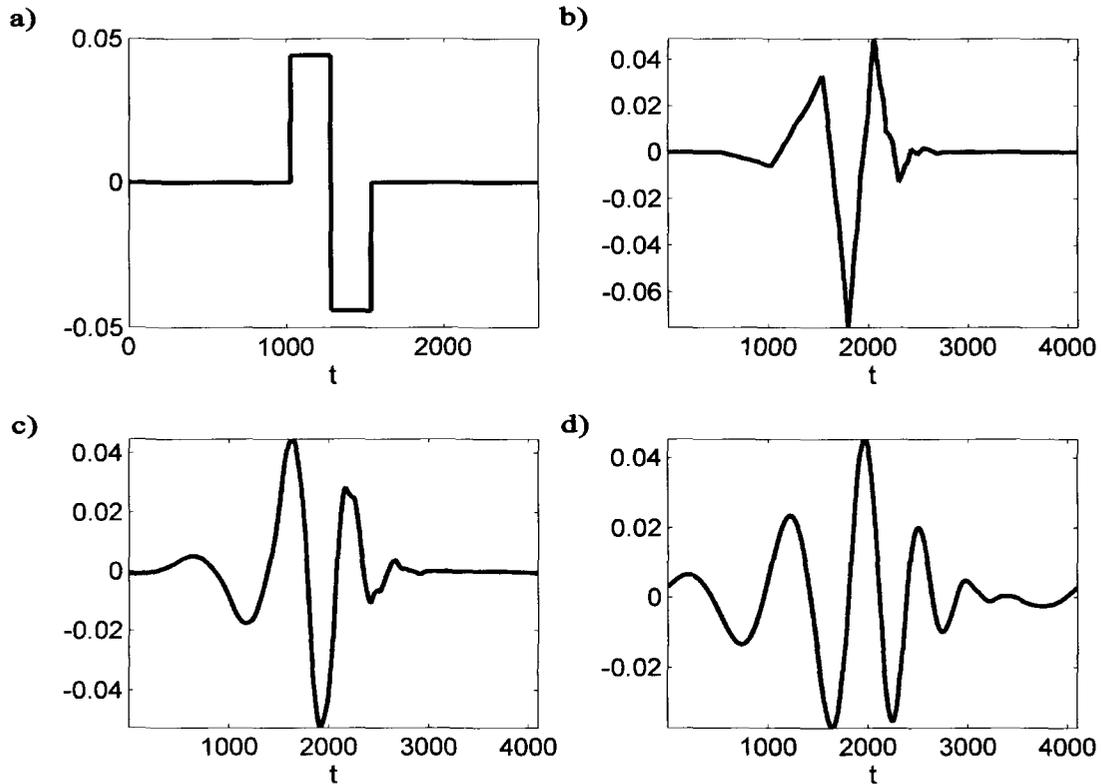


Figure 2.2: a) Daubechies 2 wavelet with 1 vanishing moment. b) Daubechies 4 wavelet with 2 vanishing moments. c) Daubechies 6 wavelet with 3 vanishing moments. d) Daubechies 20 wavelet with 10 vanishing moments.

2.5 Biorthogonal Wavelets

To give a complete overview of wavelet analysis an introduction to bi-orthogonal wavelets is necessary. The discrete wavelet transform was introduced in the context of multiresolution analysis. Bi-Orthogonal wavelets are similarly defined.

Recall that for every vector space, V , there exists a dual space, \tilde{V} , of linear functionals $f : V \rightarrow \mathbb{R}$. If the basis of V is $\{e_1, e_2, \dots, e_n\}$ then the basis of the corresponding dual space is called the dual basis and may be denoted by $\{e^1, e^2, \dots, e^n\}$. The dual basis is defined by:

$$e^i e_j = \delta_{ij} = \begin{cases} 1, & \text{if } i = j \\ 0, & \text{if } i \neq j \end{cases} \quad (2.28)$$

For the vector space V with bases $\{\phi_{j,n}\}_{n \in \mathbb{Z}}$ there exists a dual space, \tilde{V} with basis $\{\tilde{\phi}_{j,n}\}_{n \in \mathbb{Z}}$. Let $\{V_j\}_{j \in \mathbb{Z}}$ and $\{\tilde{V}_j\}_{j \in \mathbb{Z}}$ be two multiresolution approximations. The detail spaces W_j and \tilde{W}_j

have bases given by the dilated wavelets, $\{\psi_{j,n}\}_{n \in \mathbb{Z}}$, and $\{\tilde{\psi}_{j,n}\}_{n \in \mathbb{Z}}$, where $V_j \oplus W_j = V_{j-1}$ and $\tilde{V}_j \oplus \tilde{W}_j = \tilde{V}_{j-1}$. W_j is orthogonal to \tilde{V}_j , but not to V_j . \tilde{W}_j is orthogonal to V_j , but not to \tilde{V}_j . The biorthogonal wavelet transform proceeds using the method of perfect reconstruction filter banks as previously outlined for orthogonal wavelets.

Wavelets are traditionally defined as translations and dilations of a mother wavelet. These translations and dilations become algebraic operations after the Fourier transform [28]. Different choices of filter coefficients dictate wavelet properties such as vanishing moments, support size, symmetry and regularity. Orthogonal and biorthogonal wavelets were typically constructed in Fourier space until Sweldens introduced the concept of *second generation wavelets*. *First generation wavelets* are those wavelets which are translations and dilations of functions. Second generation wavelets are not necessarily translates and dilates of some function, but are constructed using *the lifting scheme* [28].

2.5.1 The Lifting Scheme

The lifting scheme is used to construct wavelets in circumstances where dilation and translation cannot be used. Solving partial differential equations on bounded domains in euclidean space is one such example. Analyzing data on curves and surfaces requires specially constructed wavelets for the manifold in question. Other applications include wavelets for irregular sampling, weighted wavelets for weighted approximations, and solving differential equations in complex geometries. For a short introduction to the lifting scheme Sweldens offers a concise review [30]. More comprehensive sources with examples are also available [29], [28].

The lifting scheme has been used effectively for solving partial differential equations in flat geometries and for solving partial differential equations on a sphere [25]. Wavelets are used to adapt the computational grid and finite differences are used to approximate derivatives. These recently developed methods offer exciting advancements in fields such as geodesy, weather prediction and climate modeling.

The process of *lifting* is represented by three stages: split, predict, update.

1. Split

- The data is split into two disjoint sets.
- If the data is split into two disjoint sets of even and odd indices, the procedure is called the *lazy wavelet* transform.
- There are no restrictions on how the data must be split, or on the size of the disjoint subsets.

2. Predict

- If the signal is highly correlated locally, then the subsets must be highly correlated as well.
- Given one of the subsets it should be possible to predict the other. For the lazy wavelet, one might choose to use the even subset to predict the odd one. (arbitrarily)
- Let $d_{j-1,l}$ be the difference between the odd sample and the predicted value. Then, the operator P is defined: $d_{j-1} = odd_{j-1} - P(even_{j-1})$.
- This step establishes the number of vanishing moments of the dual wavelet.

3. Update

- Determines the number of vanishing moments of the primal wavelet.
- The updating operator U is defined so that, for sample s_j ,

$$s_{j-1} = even_{j-1} + U(d_{j-1})$$

The following simple example due to Sweldens should clarify the method [30]. Denote the initial signal $\lambda_{0,k} = f(k)$ for $k \in \mathbb{Z}$. For data compression, one desires a transform that captures the information contained in the signal with as few points as possible. The number of points in the input signal is reduced by sub sampling the even samples. $\lambda_{-1,k} = \lambda_{0,2k}$ for $k \in \mathbb{Z}$. The idea is to avoid loss of information so it is important to keep track of how the signal can be reconstructed from the subsample. Wavelet coefficients, $\gamma_{-1,k}$, encode this information.

In this example the most obvious choice for the wavelet coefficients are the odd coefficients of the input signal $\gamma_{-1,k} = \lambda_{0,2k+1}$. More elaborate choices are possible. If it is possible to predict the odd samples from the even ones the wavelet coefficients could be defined as

$$\gamma_{-1,k} = \lambda_{0,2k+1} - \frac{1}{2}(\lambda_{-1,k} + \lambda_{-1,k+1})$$

Here the odd samples are predicted from the even samples by averaging neighboring coefficients. With this choice the wavelet coefficients provide all the necessary information to reconstruct the original signal from our sub sampling. If the signal is highly correlated these coefficients will be very small. To compress even more, one could remove coefficients below a certain threshold.

The process can be repeated to further decompose the signal. If aliasing is to be avoided the current choice of wavelet coefficients is insufficient. The average of the approximation coefficients should be the same at each level $\sum_k \lambda_{-1,k} = \frac{1}{2} \sum_k (\lambda_{0,k})$. The appropriate choice is not difficult to derive:

$$\lambda_{-1,k} = \frac{1}{4}(\lambda_{-1,k-1} + \lambda_{-1,k})$$

Second generation wavelets hold further advantages over first generation wavelets. Implementation of the wavelet transform is made more efficient with the lifting scheme as it takes advantage of similarities between high and low pass filters. Unlike the CWT, the lifting scheme has a straightforward inverse. The operations performed are simply reversed, step by step. In addition, this approach may be more easily understood by a beginner than the classical approach.

2.6 Summary

Wavelets use multiscale decomposition to transform time series. The choice of basis is dependent on the application. The continuous wavelet transform, computed with real and complex valued wavelets, is ideal for detecting specific features of a signal because of its redundancy. The discrete wavelet transform is much more efficient. It uses a fast algorithm which relies on an orthogonal basis. Biorthogonal wavelets provide more degrees of freedom than orthogonal wavelets and lead to the development of second generation wavelets. Second generation wavelets are essential for applications where dilations and translations cannot be used. These transforms are suited to a variety of applications and offer a range of tools for the analysis of non-stationary time-series.

Chapter 3

Wavelet Analysis Tools

3.1 Reproducing the Fourier Energy Spectrum

In the previous chapter, the relationship between Fourier analysis and wavelet analysis was established. Wavelets are advantageous as they possess both time and frequency localization. FT have long been utilized to study the energy spectrum associated with a signal, wavelets allow for the comparison of spectral properties at various locations in time (space). Due to the properties of wavelets, wavelet analysis depends on the wavelet being used and the signal being analyzed. To ensure consistency with previously established spectral theory, the mean power spectrum obtained from wavelet analysis is compared with the classic Fourier spectrum.

Perrier, Philopovitch and Basdevant wondered to what degree the spectral slopes computed with wavelet analysis are biased by the wavelet being used [27]. They show that the wavelet spectra is meaningful only when the analyzing wavelet has a sufficient number of vanishing moments. Using the appropriate wavelet allows one to express the wavelet spectrum as a function of the Fourier spectrum. What is presented here is largely a review of their work.

Definition 3.1.1. *The Fourier power spectrum is given by:*

$$E(k) = \frac{1}{2\pi} |\hat{s}(k)|^2 \quad (3.1)$$

The CWT can be computed as a function of the Fourier transform.

$$\tilde{s}(a, b) = \frac{\sqrt{a}}{2\pi} \int_{-\infty}^{+\infty} \hat{s}(\omega) \overline{\hat{\psi}(a\omega)} e^{i\omega b} d\omega \quad (3.2)$$

The local wavelet spectrum measures the contribution to the total energy coming from the vicinity of t and wave number k . It is given by:

$$\tilde{E}(k, t) = \frac{1}{2c_\psi k_0} \left| \tilde{s}\left(\frac{k_0}{k}, t\right) \right|^2 \text{ for } k \geq 0 \text{ and } t \in \mathbb{R}. \quad (3.3)$$

The peak wave number of the analyzing wavelet, k_0 , is the point in Fourier space where $\hat{\psi}(k)$ reaches a modulus maximum. The *mean wavelet spectrum* $\tilde{E}(k)$ is given in terms of the local spectrum.

$$\tilde{E}(k) = \int_{-\infty}^{+\infty} \tilde{E}(k, t) dt \quad (3.4)$$

From these formulas the relationship between the Fourier spectrum and the mean wavelet spectrum is readily established. That is,

$$\tilde{E}(k) = \frac{1}{c_\psi k} \int_0^{+\infty} E(\omega) \left| \hat{\psi}\left(\frac{k_0 \omega}{k}\right) \right|^2 d\omega \quad (3.5)$$

Consider a random phase signal where $\hat{u}(k) = k^{-p} e^{2\pi i \phi}$ with $\phi \in (0, 1)$. This algebraically decreasing spectrum is characterized by $E(k) \approx k^{-2p} = k^{-\alpha}$. A sufficient condition for the mean wavelet spectrum to exhibit the same behavior as the decreasing Fourier spectrum is:

$$\int_{-\infty}^{+\infty} t^n \psi(t) dt = 0, \quad 0 \leq n \leq \frac{\alpha - 1}{2} \quad (3.6)$$

This important result reveals that reproducing the Fourier energy spectrum is accomplished by choosing a wavelet basis with enough vanishing moments. To accurately describe signals with singular behavior wavelets with infinitely vanishing moments are ideal. The Paul wavelet has infinitely many vanishing moments and is expressed in Fourier space by:

$$\hat{\psi}(k) = e^{p \ln(|k|)^2}. \quad (3.7)$$

Perrier's wavelet also has infinitely many vanishing moments and is expressed in Fourier space by:

$$\hat{\pi}_n(k) = \alpha_n e^{-0.5\left(k^2 + \frac{1}{k^{2n}}\right)}, \quad n \geq 1, \quad (3.8)$$

α_n is chosen for normalization. Examples of Fourier space representations of both wavelets are found in figure 3.1. The peak wavenumber of Paul's wavelet is always 1. The peak wave number for Perrier's wavelet is $k_{n0} = n^{\frac{1}{2n+2}}$, $1 \leq k_{n0} \leq 1.15$.

The associated mean wavelet spectrum exhibits the same slope at small scales as the Fourier

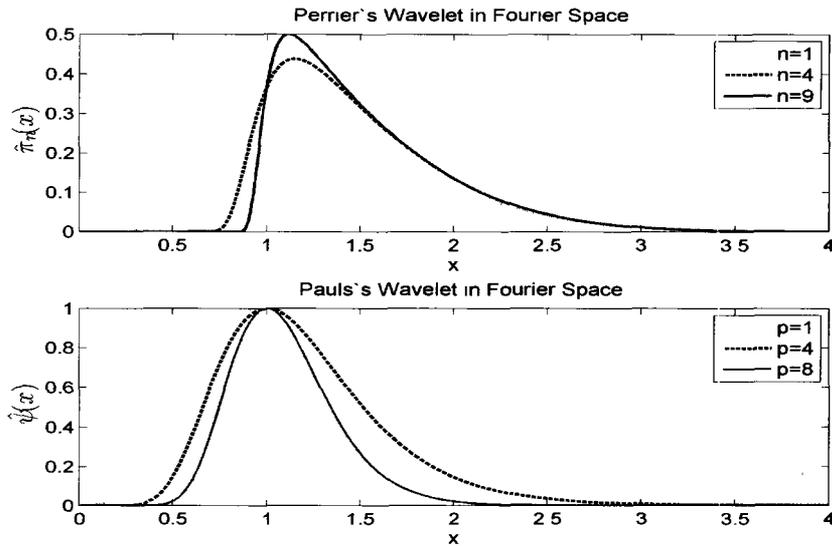


Figure 3.1: Paul and Perrier wavelets in Fourier space. Both wavelets have infinitely many vanishing moments, but Perrier's wavelet decreases faster at the origin in Fourier space

spectrum for both real wavelets. For the Paul wavelet, the power spectrum is shown in figure 3.2.

The redundancy in the frequency content of the wavelet coefficients computed with the CWT makes the CWT a better choice for reproducing the Fourier power spectrum than the DWT, but the DWT may also be used to reproduce the slope of the Fourier spectrum at fine scales. Table 3.1 verifies that the claim with regard to the relationship between reproducing the Fourier power spectrum and vanishing moments is valid [27]. That is, a wavelet with p vanishing moments will exhibit the same slope at small scales as the Fourier power spectrum if the slope of the Fourier spectrum, α , is less than $2p + 1$. In table 3.1 the slopes of the energy spectra are displayed for the Fourier spectrum and the wavelet power spectra related to the Daubechies wavelets: D2(Haar), D4, D6, and D10. Note that Daubechies wavelets were defined in section 2.4 and examples are shown in figure 2.2. If the cancellation order of the analyzing wavelet is not high enough, the wavelet power spectrum will not resemble the algebraically decreasing Fourier power spectrum. The Haar wavelet generates the correct slopes, as observed from Fourier power spectra, when $\alpha < 3$.

3.2 Signal Regularity

The capabilities of the wavelet transform have been shown to exceed the limited capacity of the Fourier transform. In this section, another useful advantage of the wavelet transform is

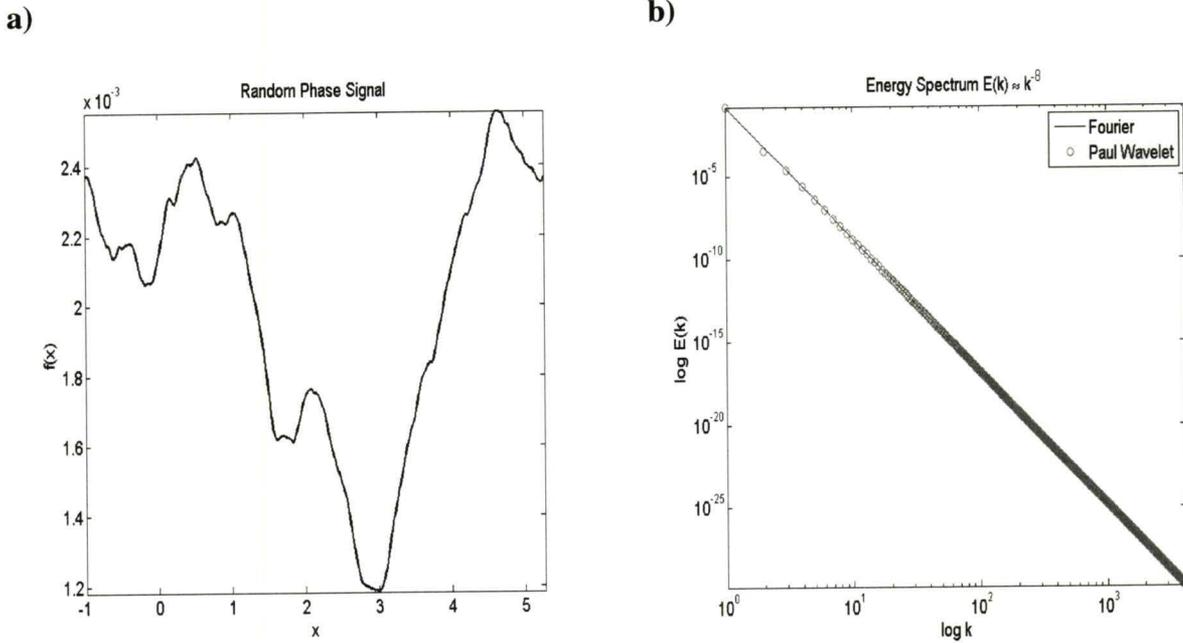


Figure 3.2: (a) A random phase signal constructed from an algebraically decreasing Fourier spectra. (b) The Paul wavelet is used to compute the power spectrum.

Fourier	D2	D4	D6	D10
-2	-2	-2	-2	-2
-4	-3	-4	-4	-4
-6	-3	-5	-6	-6
-8	-3	-5	-7	-8
-10	-3	-5	-7	-10

Table 3.1: The table displays the value of the slope, α , of the energy spectrum for the Fourier transform and various Daubechies wavelet transforms. Recall that Daubechies wavelets of length $2p$ have p vanishing moments (D2 is Daubechies wavelet of length 2). To get a reasonable correspondence $\alpha < 2p + 1$.

unmasked. The FT can be used to characterize the global regularity of a signal, but the FT gives no information with regard to the regularity of a signal at a particular point. Using the wavelet coefficients one may gather information about the local singularities present in a signal. The decay of wavelet coefficients at fine scales is used to measure the strength of local singularities in terms of Holder/Hurst exponents. Two important methods are presented: the wavelet transform Maximum Modulus method and the recently developed wavelet leaders based method.

3.2.1 Detecting/Measuring Singularities

The *wavelet scalogram* is a visual representation of wavelet coefficients. Figure 3.3 shows an example of the wavelet scalogram for a signal with multiple discontinuities. The minimum wavelet coefficient is subtracted and the coefficients are multiplied by 256 to fully utilize the 256 bit colormap. The position and scale form the x and y axis, respectively. Sharp changes in the signal cause high amplitude wavelet coefficients to appear in a cone like region with abscissa pointing to the singularity. These high amplitude coefficients are represented by the darkest and lightest patches on the scalogram.

The wavelet spectrogram or wavelet power spectrum is given by:

$$P = |\tilde{s}(a, b)|^2 \quad (3.9)$$

The presence of light colored patches in the wavelet spectrogram or scalogram can be indicative of a local singularity [23]. Stronger singularities create larger coefficients in the energy spectrum. These strengths are measured using local holder exponents.

Definition 3.2.1. ([23])

- A function f is pointwise Lipschitz/Holder $\alpha > 0$ at v , if $\exists K > 0$ and a polynomial p_v of degree $m = \lfloor \alpha \rfloor$ so that $\forall t \in \mathbb{R}$, $|f(t) - p_v(t)| \leq K|t - v|^\alpha$.
- A function f is uniformly Lipschitz α over $[a, b]$ if it satisfies the above equation $\forall v \in [a, b]$ with a constant K which is independent of v .
- The Lipschitz/Holder regularity of f at v or over $[a, b]$ is the supremum of α so that f is Lipschitz α .

Herrmann ([18]) explains that for finite K the Lipschitz/Holder exponents measure:

$$|f(t + \Delta t) - f(t)| \leq K|\Delta t|^\alpha \quad [18].$$

- $\alpha \geq 1$, $f(t)$ is continuous and differentiable.
- $0 < \alpha < 1$, $f(t)$ continuous but not differentiable.

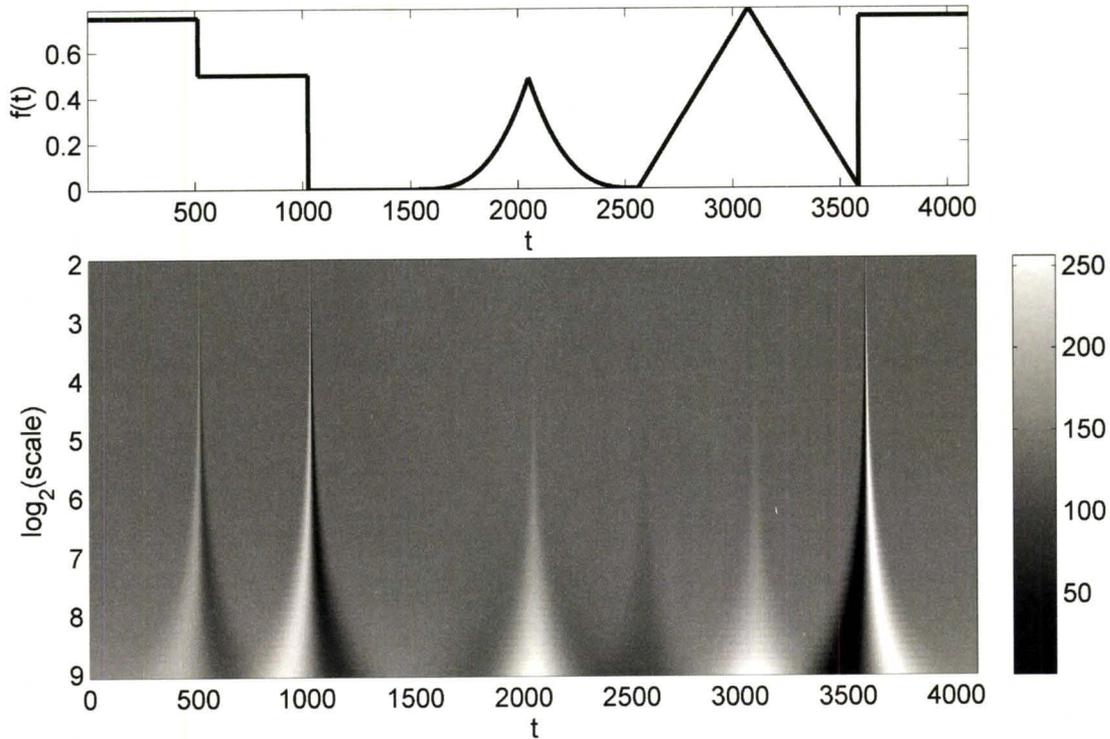


Figure 3.3: The Mexican Hat wavelet used to generate the wavelet scalogram for a signal with multiple singularities. The WC are amplified appropriately to the 256 bit color map. Lighter pigments correspond to higher values of wavelet coefficients.

- $-1 < \alpha \leq 0$, $f(t)$ discontinuous and non-differentiable.
- $\alpha \leq -1$, $f(t)$ is not locally integrable.

The uniform Lipschitz exponent is a global measurement of regularity, while the pointwise Lipschitz exponent characterizes the local regularity of a signal.

If f is uniformly Lipschitz $\alpha > m$ in a neighborhood of v then f is necessarily m times continuously differentiable in that neighborhood [23]. Assume that the wavelet ψ has n vanishing moments and is C^n with derivatives that have fast decay. The relationship between the uniform Lipschitz regularity and the decay of wavelet coefficients at fine scales is elucidated by the following theorem.

Theorem 3.2.2. ([23]) *If $f \in L^2(\mathbb{R})$ is uniformly Lipschitz $\alpha \leq n$ over $[a, b]$ then $\exists A > 0$ s.t.*

$$\forall (u, w) \in [a, b] \times \mathbb{R}^+, \quad |\tilde{s}(w, u)| \leq Aw^{\alpha + \frac{1}{2}} \quad (3.10)$$

conversely, if $\tilde{s}(w, u)$ satisfies the above equation, and if $\alpha < n$ is not an integer, then f is uniformly Lipschitz α on $[a + \varepsilon, b - \varepsilon]$ for any $\varepsilon > 0$.

For signals with several isolated singularities, the uniform Lipschitz exponent reflects the strongest singularity present in the signal. Finding the Holder exponent at a specific point can be more difficult. f may have many different singularities accumulated around the point in question. A necessary and sufficient condition for finding the Lipschitz regularity at a particular point is given by the following theorem.

Theorem 3.2.3. ([23]) *If $f \in L^2(\mathbb{R})$ is Lipschitz $\alpha \leq n$ at v then $\exists A > 0$ s.t.*

$$\forall (u, w) \in [a, b] \times \mathbb{R}^+, \quad |\tilde{s}(w, u)| \leq Aw^{\alpha + \frac{1}{2}} \left(1 + \left| \frac{u - v}{w} \right|^\alpha \right). \quad (3.11)$$

If $\alpha < n$ is not an integer then $\exists A$ and $\alpha' < \alpha$ so that

$$\forall (u, w) \in [a, b] \times \mathbb{R}^+, \quad |\tilde{s}(w, u)| \leq Aw^{\alpha + \frac{1}{2}} \left(1 + \left| \frac{u - v}{w} \right|^{\alpha'} \right). \quad (3.12)$$

Then f is Lipschitz α at v .

These conditions are more easily interpreted by assuming that the analyzing wavelet has compact support.

Definition 3.2.4. ([23]) *Suppose ψ has compact support on $[-C, C]$. The cone of influence of v is the set of point (w, u) s.t. v is included in the support of $\psi_{w,u}(t)$. Because the support of $\psi(\frac{t-u}{w})$ is $[u - Cw, u + Cw]$, the cone of influence of v is given by:*

$$|u - v| \leq Cw. \quad (3.13)$$

Recall u is the position, w is the scale.

If ψ is not compactly supported one approximates C by considering a region of effective support; i.e. where the wavelet becomes very close to zero. Two methods are now presented to show how the decay of local maxima is used to measure the local regularity.

3.2.2 Local Regularity: Maximum Modulus Method

A wavelet transform maximum modulus (WTMM) is any point, (a_0, b_0) , where $|\tilde{s}(a, b_0)|$ attains a strict local maximum.

$$\frac{\partial \tilde{s}(a, b)}{\partial a} = 0 \quad (3.14)$$

WTMM which converge to the abscissa v are used to measure the pointwise Lipschitz regularity of the signal at that point. The creation of these WTMM lines eliminates spurious maxima that may be present in the data. The existence of a singularity at a point v is linked to the presence of a convergent sequence of wavelet maxima, but the presence of a sequence of WTMM converging to a point v is not sufficient to lead to the conclusion of the existence of a singularity at that point. Noisy data, for instance, may generate high amplitude coefficients. One must measure the decay of these WTMM at fine scales to determine the pointwise Lipschitz regularity at v . If $|\tilde{s}(a, b)|$ has no modulus maxima at fine scales, then f is locally regular. In fact, it can be shown that the absence of local maxima at fine scales implies that f is uniformly Lipschitz [23].

Approximation of the Holder exponent for local, isolated singularities is accomplished by observing the decay of modulus maxima within the cone of influence of each point of interest, v . Notice that in figure 3.3 large coefficients are present in a cone-like region below abscissa which correspond to the location of a singularity in the signal.

The WTMM within the cone of influence are connected to form a WTMM line. Measuring the decay of WTMM along these lines as the scale goes to zero gives an approximation of the Holder exponent. If u is in the cone of influence of the point v then $\frac{|u-v|}{w} \leq C$. Equation 3.12 can be rewritten as:

$$|s(w, u)| \leq A' w^{\alpha + \frac{1}{2}} \quad (3.15)$$

The following equation is used to determine α by consider points in the cone of influence of v :

$$\log(|\tilde{s}(w, u)|) \leq \log(A) + \left(\alpha + \frac{1}{2}\right) \log(w) \quad (3.16)$$

3.2.3 Local Regularity: Wavelet Leaders

Lashermes Jaffard and Abry recently introduced an alternative method for computing pointwise regularity based on finite quantities called *wavelet leaders* [21]. The method improves upon the results observed from the WTMM method for chirp type singularities. The computation of wavelet leaders is more systematic and straightforward than the computation of WTMM lines.

Assume ψ_0 is a compactly supported mother wavelet and that $\{2^{j/2} \psi_0(2^{-j}t - k), j \in \mathbb{N}, k \in$

\mathbb{N} forms an orthonormal basis. It is convenient to adopt an alternate notation for the discrete wavelet transform for the signal $s(t)$:

$$d(j, k) = \int_{\mathbb{R}} s(t) 2^{-j} \psi_0(2^{-j}t - k) dt. \quad (3.17)$$

The DWT is defined on dyadic intervals $\lambda = \lambda_{(j,k)} = [k2^j, (k+1)2^j)$. Then $d_\lambda = d(j, k)$. Let 3λ denote the union of the interval λ and its two adjacent dyadic intervals:

$$3\lambda_{j,k} = \lambda_{j,k-1} \cup \lambda_{j,k} \cup \lambda_{j,k+1}. \quad (3.18)$$

For $s \in L^\infty$.

$$|d_\lambda| \leq \int |s(t)| |\psi_{j,k}(t)| dt \leq C \|\psi_0\|_{L^1} \|s\|_{L^\infty} \quad (3.19)$$

which means that the wavelet leaders given by:

$$L_s \equiv L_\lambda = \sup_{\lambda' \subset 3\lambda} |d_{\lambda'}| \quad (3.20)$$

are finite. Computing the wavelet leader at a scale j and position k involves taking the supremum over wavelet coefficients on adjacent dyadic intervals and all finer scales. To find the holder exponent one simply determines the log-linear relationship between the wavelet leaders and the scales as in 3.16. The method is very similar to the WTMM method, but replaces the modulus maxima quantities with wavelet leaders.

3.2.4 Cusp and Chirp type Singularities

Both WTMM and wavelet leaders work well for determining the strength of cusp type singularities, but only wavelet leaders provides an accurate estimate of the holder exponent for non-isolated or chirp type singularities. Cusp type singularities are of the form $X(t) = A + B|t - t_0|^\alpha$. Oscillating or chirp type singularities are of the form:

$$X(t) = |t - t_0|^\alpha \sin\left(\frac{1}{|t - t_0|^\beta}\right) \quad h, \beta > 0$$

β is referred to as the oscillating exponent.

Figures 3.4, and 3.5 shows the performance of both methods for chirp and cusp type singularities. As the signal becomes more and more irregular, it becomes more and more difficult to estimate the Holder exponent.

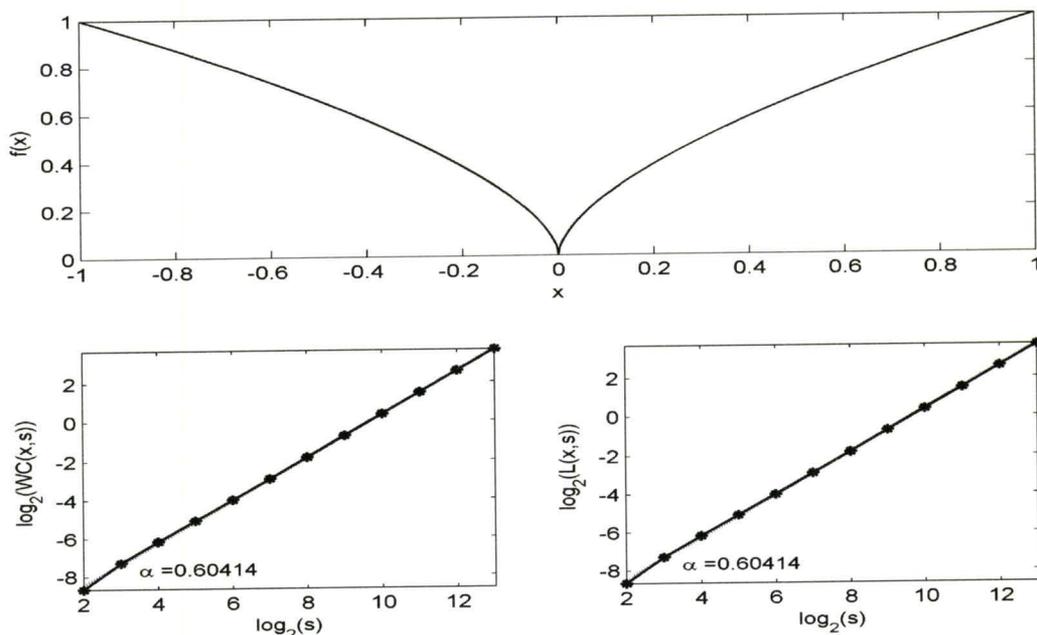


Figure 3.4: A cusp singularity with Holder exponent $\alpha = 0.6$. WTMM (left) and WL (right). Recall: $\alpha = slope - 1/2$

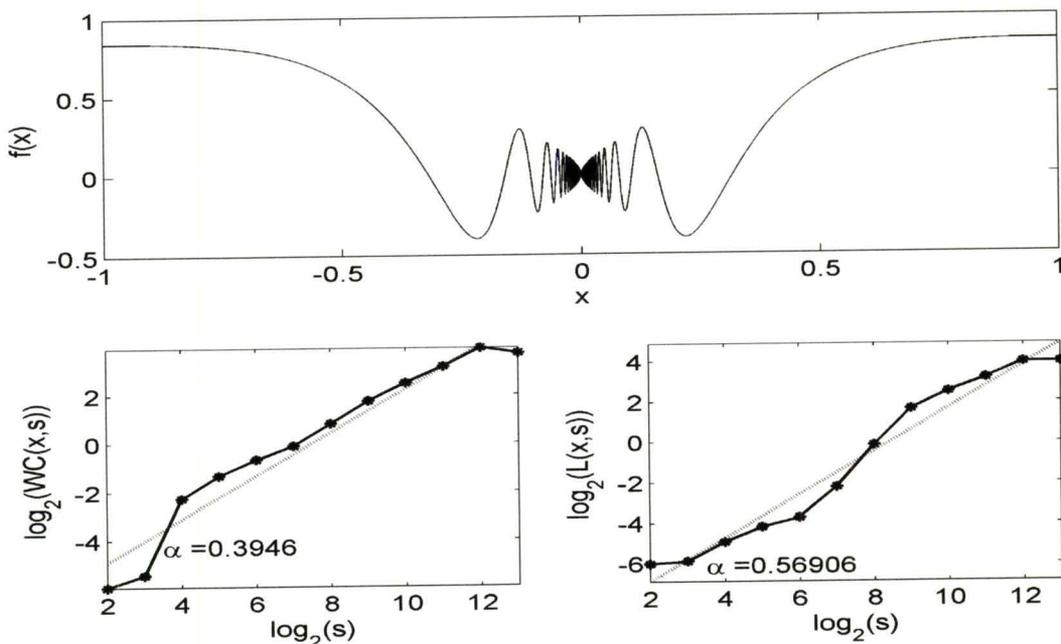


Figure 3.5: A chirp type singularity with Holder exponent $\alpha = 0.6$, $\beta = 1$. WTMM (left) and WL (right). Recall: $\alpha = slope - 1/2$

3.3 Multifractal Analysis

Abrupt changes in a signal lead to wavelet coefficients of large amplitude. The previous section showed how measuring the decay of these high amplitude coefficients allows one to characterize the local singularities present in the signal. Similarly, the decay of wavelet coefficients can be used to measure the singularity distribution for a fractal signal.

A signal is said to exhibit a *multifractal* structure when the holder exponent varies in time. In contrast, a *monofractal* signal has the same unique Holder exponent at each point in the signal. The singularity spectrum is estimated globally by taking advantage of multifractal self-similarities [23]. The self-similar structure of a multifractal signal is reflected in the wavelet transform. Before methodology for obtaining this multifractal spectrum is discussed, a brief introduction to the underlying theory of fractals is presented.

3.3.1 Fractal Sets and Functions

A proper introduction to fractals begins with an historical perspective. Mandelbrot suggests that a fractal is “[...] by definition a set for which the Hausdorff Besicovitch dimension strictly exceeds the topological dimension” [24]. The definition was thought to be too restrictive and was subsequently retracted by Mandelbrot. He proposed instead that “[...] a fractal is a shape made of parts similar to the whole in some way” [14]. A precise definition has yet to be established.

A self similar fractal function often demonstrates singular behavior [24]. *Self-affinity* is a concept meant to group together sets which are invariant under affine transformations. For all self- affine functions

$$f(t_0 + \lambda t) - f(t_0) \simeq \lambda^H (f(t_0 + t) - f(t_0)) \quad (3.21)$$

H is the Hurst exponent. When $H = 1$ the graph of the function is self-similar [3]. The smaller the value of the Hurst exponent, the more singular the function. The Holder exponent is thought of as a local Hurst exponent.

A continuous, compactly supported function f is said to be self-similar if there exists S_1, \dots, S_k so that the graph of f restricted to each S_i is an affine transformation of f . Consider the following example commonly found in the literature [2], [14]. Imagine a line of unit length. Changing the length scale by a factor of $r < 1$ generates a new line. If r is chosen appropriately $r = \frac{1}{N}$, then the original line may be covered by N segments of length r . A unit square may be covered by N^2 scaled down pieces with sides of length r and a unit box may be covered by N^3 versions of the original with side lengths r . The *similarity dimension*, d , in the preceding 3 cases is 1, 2, 3, respectively and is defined by $N_c = (\frac{1}{N})^d$. N_c is the minimum number of boxes with side length

$\frac{1}{N}$ needed to cover the set.

Notice that a line may be covered by N planes, squares or cubes of side length r . Regardless of the dimension of the box, $N = (\frac{1}{r})^{D_b}$ boxes are always required. D_b is referred to as the *box counting dimension*. For a unit length line as $r \rightarrow 0$ the length estimated by $L = Nr$ approaches the true length of the line. For higher dimensions the limit tends to zero as $r \rightarrow 0$; $A = Nr^2 \rightarrow 0$, $V = Nr^3 \rightarrow 0$, etc. If the dimension chosen for the cover differs from the dimension of the set then the limit will tend to infinity or zero depending on whether the chosen dimension is too low or too high for the set in question. There is a critical value of D_b for which the estimated measure of the set changes from zero to infinity. This value is the Hausdorff dimension.

The Hausdorff dimension, D is also referred to as the *fractional dimension* or *fractal dimension* and is approximated numerically by the capacity dimension. The capacity dimension D of the set S is given by:

$$D = - \liminf_{a \rightarrow 0} \frac{\log N(a)}{\log a} \quad (3.22)$$

Here $N(a)$ is the number of balls of radius a needed to cover S .

Wavelets offer a method for determining the fractal dimension of singularities distributed throughout the signal. This distribution must be determined from global quantities as each finite sample of the signal may contain infinitely many singularities of varying magnitude. Global measurements allow one to take advantage of self similarities present in the data to determine the fractal dimension of points with the same Lipschitz regularity.

For self affine records, a distinction is made between the local and global fractal dimension [14]. To understand why, consider using a very large box when computing the box-counting dimension. For a large enough box size, one box is enough to cover the span of the fractal record. Globally a self affine record has $D = 1$. Which means that *globally a self affine record is not a fractal*.

Singularity Spectrum

The *singularity spectrum*, $D(\alpha)$, is used to find the distribution of singularities for a signal. The spectrum gives the proportion of Lipschitz α singularities that occur at any scale a [23]. The regularity of a signal is thus characterized by the regularity of its subsets.

Definition 3.3.1. ([23]) *Suppose S_α is the set of all points $t \in \mathbb{R}$ s.t. the pointwise Lipschitz regularity of f is equal to α . The singularity spectrum $D(\alpha)$ of f is the fractal dimension of S_α . The support of $D(\alpha)$ is the set of α such that S_α is not empty.*

The singularity spectrum is computed from the real data using a multifractal formalism. The spectrum offers a description of the pointwise Lipschitz regularity of a signal.

3.3.2 WTMM Multifractal Formalism

A *multifractal formalism* (MF) is any formula that can be used to determine the spectrum of singularities of a function from an estimate of global quantities. Mallat introduced a MF based on wavelet transform modulus maxima.

As discussed previously, for a wavelet ψ with n vanishing moments if f has pointwise Lipschitz regularity $\alpha_0 < n$ at ν then the wavelet transform $s(a, b)$ has a sequence of modulus maxima that converge toward ν at fine scales. At these maxima $|\tilde{s}(a, b)| \approx a^{\alpha_0+1/2}$. Define a set $\{b_n(a)\}_{n \in \mathbb{Z}}$ to be the positions of the local maxima at a fixed scale a . Now, define a partition function Z :

$$Z(a, q) = \sum_n |\tilde{s}(a, b_n)|^q. \quad (3.23)$$

This function measures the sum at a power q of all the aforementioned local modulus maxima. The wavelet itself defines the shape of the partitions and the scale parameter dictates the size. WTMM are used to indicate how the partitions should be taken at each scale. The scaling exponent $\tau(q)$ measures the asymptotic decay of $Z(a, q)$ at fine scales a for each $q \in \mathbb{R}$.

$$\tau(q) = \liminf_{a \rightarrow 0} \frac{\log Z(q, a)}{\log a}. \quad (3.24)$$

The scaling exponent $\tau(q)$ is the Legendre transform of the singularity spectrum $D(\alpha)$ [23]. Jaffard [20] generalized the result of [4] which relates the scaling exponent, $\tau(q)$, to the singularity spectrum.

Theorem 3.3.2. *Suppose the support of $D(\alpha)$ is $\Lambda = [\alpha_{\min}, \alpha_{\max}]$. Let ψ be a wavelet with $n > \alpha_{\max}$ vanishing moments. If f is self-similar then:*

$$\tau(q) = \min_{\alpha \in \Lambda} \left(q \left(\alpha + \frac{1}{2} \right) - D(\alpha) \right). \quad (3.25)$$

Computing the derivative of equation 3.25 reveals $q(\alpha) = \frac{dD}{d\alpha}$. From this computation and using the fact that $\tau(q)$ is at a minimum we derive that $D(\alpha)$ is a convex function, and $\tau(q)$ is an increasing and convex function. For the Legendre transform to be invertible, $D(\alpha)$ must be convex. Details of the proof are given in [20]. Note that the negative of the scaling function is used in the computation giving a *concave spectrum*.

$$D(\alpha) = \min_{q \in \mathbb{R}} q \left(\alpha + \frac{1}{2} \right) - \tau(q) \quad (3.26)$$

$D(\alpha)$ is the fractal dimension of the set with Holder exponent α . If the set of points where the signal is Lipschitz/Holder α is an empty one, by convention $D(\alpha) = -\infty$. The approximation to the singularity spectrum is not necessarily exact one. Some multifractals possess multifractal

spectra which are not concave. The computed $D(\alpha)$ is an upper bound for the actual spectrum [23].

A closer look at equation 3.25 shows that the maximum or peak of the singularity spectrum occurs at $q = 0$.

$$-\tau(0) = \max_{\alpha} f(\alpha) \quad (3.27)$$

The right side of $D(\alpha)$ is computed from negative q , and the left from positive values of q .

3.3.3 Examples

Fractional Brownian Motion

Brownian motion describes the movement, or random walk, of particles in a fluid. *Fractional Brownian Motion* (FBm) is a monofractal process which can be thought of as a natural extension of Brownian motion. In Brownian motion, the increments and position of a Brownian particle is given by:

$$X(t) - X(t_0) \approx \xi |t - t_0|^H \quad (t \geq t_0) \quad (3.28)$$

$H = 1/2$ gives ordinary Brownian motion. Here ξ is a normalized independent Gaussian random process; $X(t)$ is a random function of time. Choosing $H \neq 1/2$, $0 < H < 1$, gives FBm [14].

It can be shown that any realization of the process is almost everywhere singular with point-wise Lipschitz exponent $\alpha = H$ [23]. B_H is often used instead of $X(t)$ to denote the particle position. Hurst found that $H = 1/2$ characterizes a independent random process. $H > 1/2$ gives persistence; an increasing trend in the past implies an increasing trend in the future [14]. $H < 1/2$ gives antipersistence; an increasing trend in the past implies a decreasing trend in the future [14].

The monofractal structure of FBm makes it an excellent test case as its theoretical multifractal spectrum $D(\alpha)$ reduces to a single point, $D(\alpha) = 1$. Figure 3.6 shows the scaling exponent and singularity spectrum computed for FBm with $H = 0.7$.

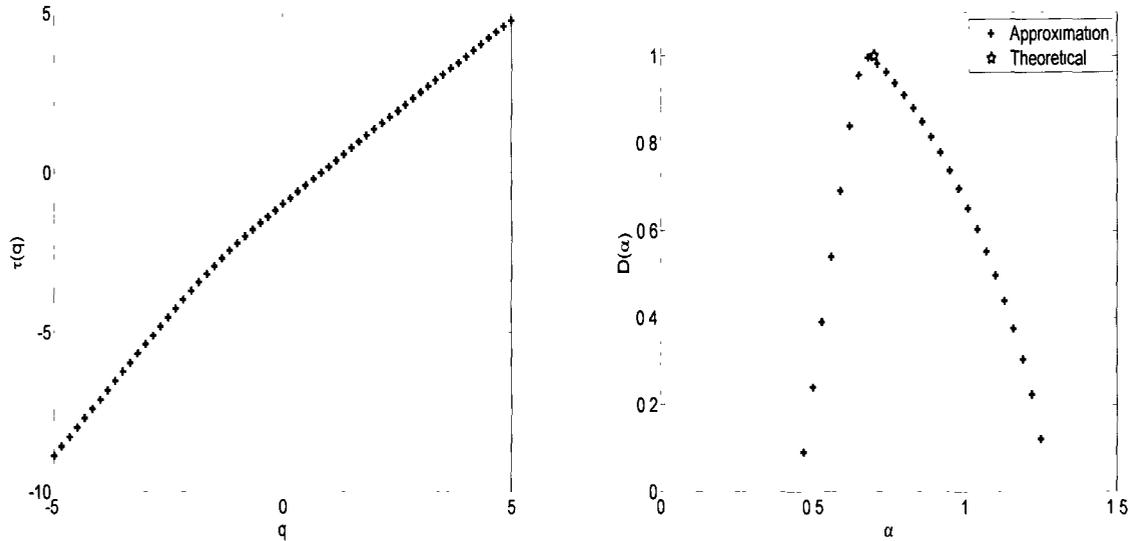


Figure 3.6: The singularity spectrum (right) for $B_{0.7}$ is generated using the Mexican hat wavelet and 2^{14} points. The scaling exponent (left) is computed with $q \in [-5, 5]$.

The Devil’s staircase

The Devil’s Staircase is also referred to as the Cantor function as it is constructed from the Cantor set. The Cantor set is obtained as the intersection of sets c_i defined recursively by setting $c_0 = [0, 1]$, c_{i+1} is obtained from c_i by removing the middle third of each interval in c_i . The set $C = \bigcap_{i \geq 1} c_i$ is closed bounded and non-empty. The Cantor function is constant on those intervals removed in the construction of the Cantor set, $[0, 1] \setminus c_i$. Mallat gives the following definition [23]:

$$f_{n+1}(x) = \begin{cases} p_1 f_n(3t) & t \in [0, \frac{1}{3}] \\ p_1 & t \in [\frac{1}{3}, \frac{2}{3}] \\ p_1 + p_2 f_n(3t - 2) & t \in [\frac{2}{3}, 1] \end{cases} \quad (3.29)$$

f_0 is any function such that $f_0(0) = 0$ and $f_0(1) = 1$. This function is sometimes referred to as the Devil’s staircase for different values of the weights p_1 and p_2 .

The example shown in figure 3.7 uses the parameters $p_1 = 0.4$, $p_2 = 0.6$ and is computed with $N = 2^{20}$ points.

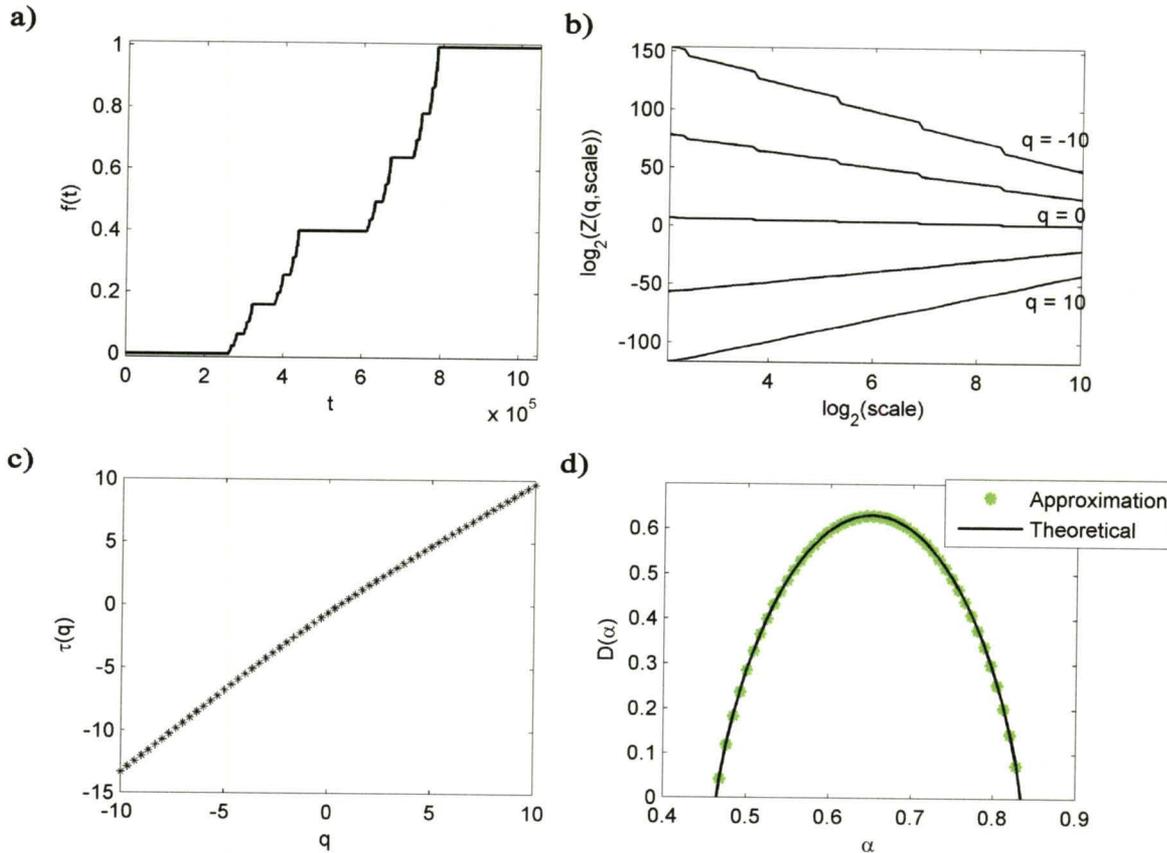


Figure 3.7: a) The Cantor function with $p_1 = 0.4$, $p_2 = 0.6$. 2^{20} points were used. b) Partition function. c) Scaling exponent. d) Singularity spectrum.

Distinguishing Monofractals from Multifractals

Classifying a signal as either monofractal or multifractal is an important aspect of multifractal analysis. The concave multifractal spectrum, which is generated by the multifractal formalisms discussed, can make it difficult to identify a signal as a monofractal. This is due to the fact that these wavelet based multifractal formalism often generate spurious data points in the resulting singularity spectrum. The singularity spectrum shown in figure 3.6 for FBm is computed for $q \in [-5, 5]$. The theoretical spectrum is a single point, while the approximation consists of a range of values. This may lead to the false conclusion that a signal is a multifractal, when it is a monofractal. Fortunately, the peak of the resulting convex multifractal spectrum ‘points to’ the correct value of the Hurst exponent for these monofractal signals. In order for this to be useful one must have a method for identifying, *a priori*, the monofractal nature of a signal.

The singularity spectrum of a monofractal reduces to a single point, $D(\alpha_0) = 1$. Notice from equation 3.25 there is an important relationship between $\tau(q)$ and the Hurst exponent, $h =$

$\alpha_0 + 1/2$, for monofractal signals.

$$\tau(q) \approx qh - 1. \quad (3.30)$$

Linear behavior of $\tau(q)$ indicates the presence of a monofractal, while non-linear behavior indicates multifractality.

The fractal nature of a signal may thus be determined by first evaluating $\tau(q)$ for linearity and second computing the multifractal spectrum. If $\tau(q)$ is indeed linear, then the peak of $f(\alpha)$ corresponds to the parameter α_0 which characterized the signal. If $\tau(q)$ is non-linear, $f(\alpha) = D(\alpha)$ is an approximation of the singularity spectrum.

As a first step to examining the linearity of the scaling functions, the test functions described above are log transformed and plotted with a line of best fit, figure 3.8. For FBM a strong linear relationship between τ and q is observed. The estimated scaling exponent for $B_{0.7}(t)$ deviates from an exact linear relationship with q , but the deviation is relatively small. For the Cantor function, the scaling exponent exhibits a much more prominent deviation from linearity. Examining these functions derivatives should accentuate the non-linear behavior of the functions and may also give some indication of the shape of the scaling exponent for multifractals.

A linear function must have a constant derivative. Scaling exponents which are nearly linear in q should have derivatives very near a constant value. Numerical derivatives are easily computed with a finite difference scheme. $\tau'(q)$ is computed for FBM and the devil's staircase and shown in figure 3.8. These examples show that the behavior of $\tau(q)$ for the multifractal Cantor function is noticeably less linear than the behavior of the scaling exponent in the FBM example. What is needed here is a quantitative distinction. Some measurable degree of linearity should be required for monofractal signals.

Linear behavior is expected for the scaling exponents of monofractal signals, but what type of behavior is expected for multifractals? The answer to this question is unknown suggesting that more research is needed. We noticed that the behavior of $\tau'(q)$ is closely modeled by the function:

$$f(x) = 1.149 - \frac{\tanh\left(\frac{q}{4.5045}\right)}{5.607}. \quad (3.31)$$

Error Convergence

The reason for examining these test functions is to establish a framework for numerical approximations of singularity spectra for observed time series. Many of these time series possess only a small number of points, so, it is imperative that one has some idea of the error in the approximation. This information is not found in the literature. If the error in the approximation for a small number of points is unknown, then our computations for a particular data set may

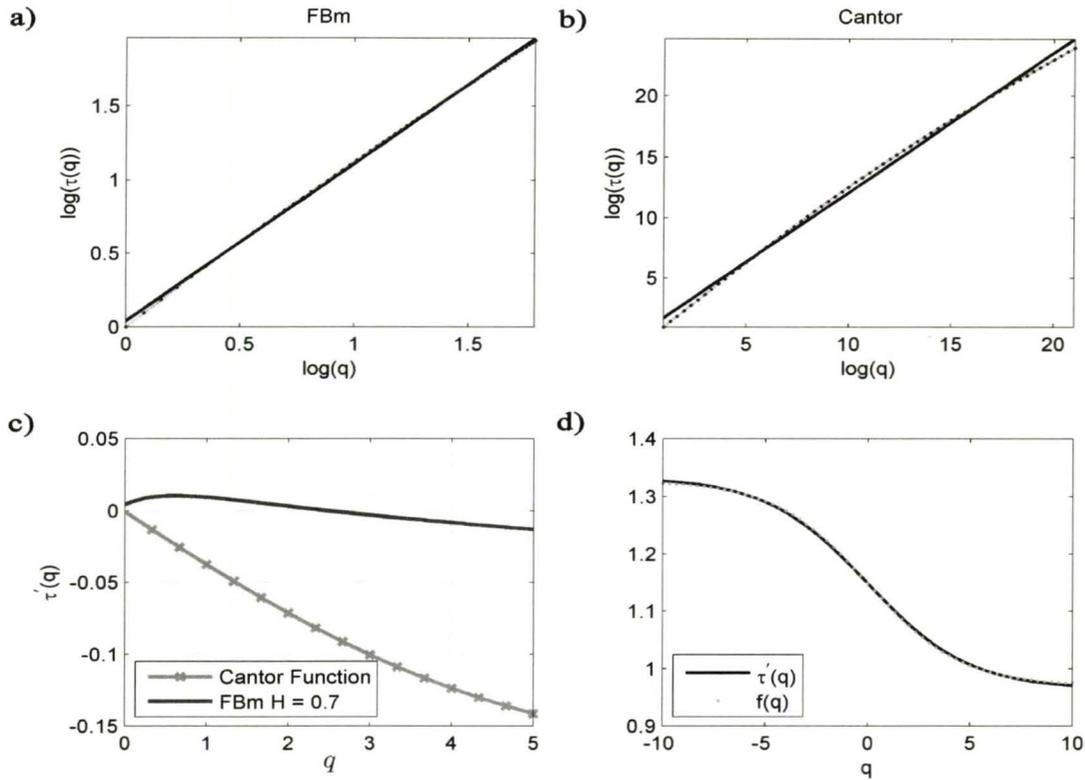


Figure 3.8: a) Log transformed scaling exponent for FBM, $H = 0.7$ (2^{20} points). b) Log transformed scaling exponent for the Cantor function with $p_1 = 0.4$, $p_2 = 0.6$ (2^{20} points). c) Derivative of the scaling exponent for both functions. The mean is subtracted for easy comparison. d) The behavior of $\tau'(q)$ for the Cantor function as compared with equation 3.31.

be erroneous. We examine the error in the approximated multifractal spectrum for the devils staircase, as a function of the number of points, N .

The WTMM implemented with the CWT gives a good approximation of the multifractal spectrum for both monofractals and multifractals. It is possible to use the discrete wavelet transform, however the DWT based method converges more slowly than the CWT. Figure 3.9 shows error convergence for both transforms. The error is computed as the norm of the difference between theoretical spectrum and the approximated spectrum for the Cantor function. The figures show that the error does decrease as resolution increases, but the rate of convergence is quite slow.

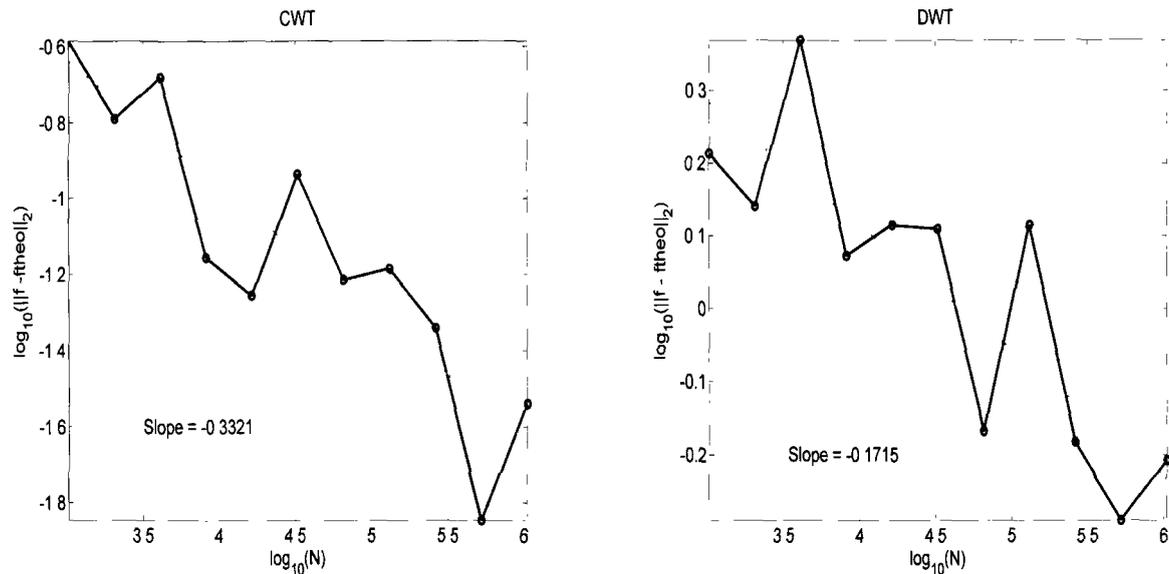


Figure 3.9: The convergence of the approximated spectrum to the theoretical spectrum as the number of points used increases.

3.3.4 Wavelet Denoising

This section presents an overview of wavelet de-noising using orthogonal wavelets. The term *wavelet de-noising* refers to the process of using the wavelet transform to eliminate noise present in a given data set. David Donoho provides a method for removing additive Gaussian noise using thresholding in the wavelet domain [11]. Two distinct approaches are examined: *soft* thresholding and *hard* thresholding. Hard thresholding involves computing the threshold, t_n , and rejecting all wavelet coefficients below that threshold by setting them to zero. Soft thresholding is slightly more complicated as it involves first rejecting those coefficients below the threshold and subsequently shrinking the remaining coefficients using the formula:

$$n_t(w) = \text{sign}(w)(|w| - t)_+ \quad (3.32)$$

The procedure for both methods are described by Donoho [11] as follows:

1. Apply the discrete wavelet transform to the noisy signal to obtain the wavelet coefficients.
2. Set the threshold.

$$t_n = \frac{\sigma}{\sqrt{n}} \sqrt{2 \log(n)} \quad (3.33)$$

n is the length of the input signal. σ is the noise level.

3. Apply the inverse wavelet transform to recover the signal.

If the noise level, σ , is unknown it can be estimated by computing the median of the absolute value of the wavelet coefficients (MAD) at the finest scale:

$$\hat{\sigma} = \frac{MAD}{0.6745} \quad (3.34)$$

For soft thresholding, one does not “shrink” the coarsest scale coefficients as they contain information describing the overall structure of the data. These coarse scale coefficients are needed to reconstruct essential features of the signal, whereas the finer scale coefficients contain information about the details of the signal and may be altered. This method of soft thresholding is referred to as *VisuShrink*. The name refers to the good visual quality obtained via wavelet shrinkage when de-noising images.

Both of the thresholding techniques described here are known to exhibit visual artifacts. Artifacts are deviations that occur as a result of interactions between the input function and the wavelet basis [7]. A signal that exhibits visual artifacts after de-noising, may not exhibit those artifacts if shifted in time or frequency. Donoho and Coifman observe that size of the artifact is closely related to the spatial location of the singularity in the signal. This led Coifman and Donoho to develop a translation invariant transform. The idea being, that the signal could be appropriately shifted to avoid artifacts and subsequently shifted back to reconstruct the signal.

For signals with multiple singularities, there is no way to apply a single shift that will eliminate artifacts for all singularities. Instead, the approach is to apply a range of shifts and average over the results. A extensive account will not be given here. Coifman and Donoho present a detailed introduction to the method with computational examples in their 1995 paper[7]. Matlab programs implementing translation invariant de-noising are available as part of Wavelab850 [1].

3.4 Testing the Capabilities of Wavelet Thresholding

To test the effectiveness of wavelet denoising we apply the method to several test functions. The diagnostics proceed by adding Gaussian white noise with a known amplitude to the signal, denoising with wavelet thresholding, and comparing the de-noised signal with the original signal. The noise vector is generated by first creating a vector of signal length containing random numbers drawn from a normal distribution of zero mean and variance 1; $N(0, 1)$. This vector is multiplied by the amplitude of the noise, σ . Note that σ is also referred to as the the standard deviation of the noise. In what follows, the Cantor function with $p_1 = 0.4$, $p_2 = 0.6$ and the sine wave, $f(x) = \sin(2x)$, are the primary test functions used.

Qualitative Comparisons

Gaussian white noise is added to a sine wave and then de-noised using both the Haar (D2) and the Daubechies 20 wavelet. As the variance of the GWN increases the D20 wavelet is observed to outperform the Haar wavelet. This could be due to the smoothness of the wavelet basis. Figure 3.10 shows the results of wavelet denoising using hard thresholding.

The VisuShrink method typically leads to noise-free reconstructions, but may shrink important features of the data. Figure 3.11 demonstrates that “shrinkage” increases as the standard deviation, σ , of the additive noise increases.

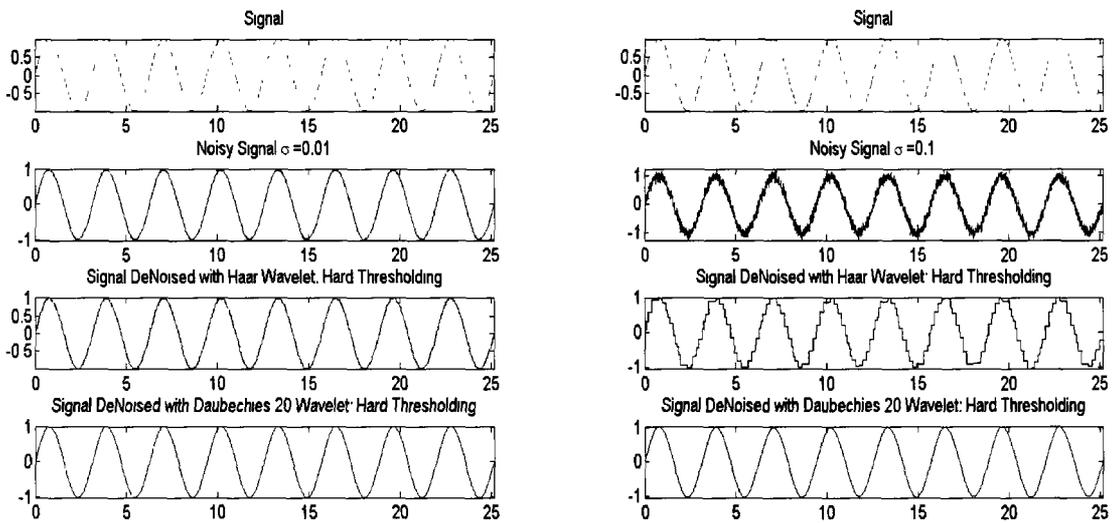


Figure 3.10: Wavelet de-noising via Hard thresholding.

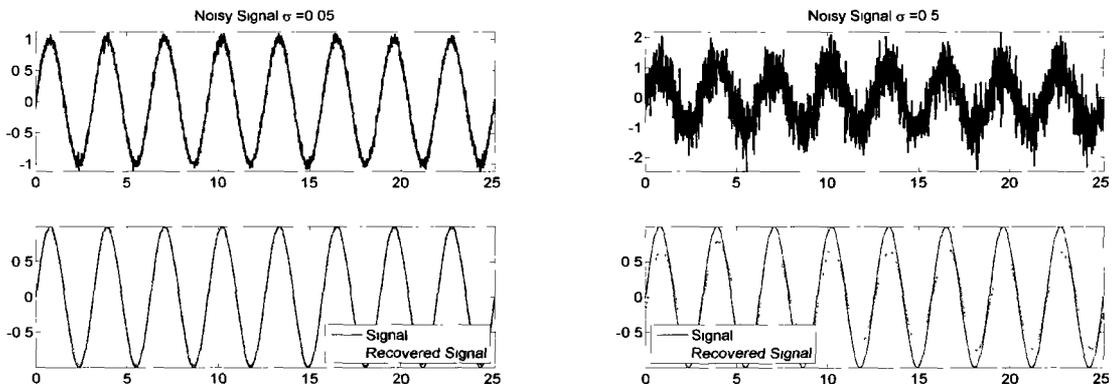


Figure 3.11: Wavelet de-noising via Visushrink. The recovered sine wave “shrinks” more for higher noise levels, σ .

Figure 3.10 shows that the D20 wavelet is a better choice for de-noising the sine function than the Haar wavelet. The analyzing wavelet used should be chosen according to the regularity of the signal. As the noise level increases, the D20 wavelet does a better job over all of recovering the signal. For high amplitude noise, the recovered signals in figure 3.12 exhibit significant deviations from the input signal. The Haar wavelet is a better choice for the Cantor function given its singular behavior. Comparing the figures shows that for smaller noise levels ($\sigma \approx 0.01$) both hard and soft thresholding work well. For larger values of σ , neither method recovers the finer details of the Cantor function. In general, we find that Hard thresholding does a better job of recovering the signal by preserving signal amplitude.

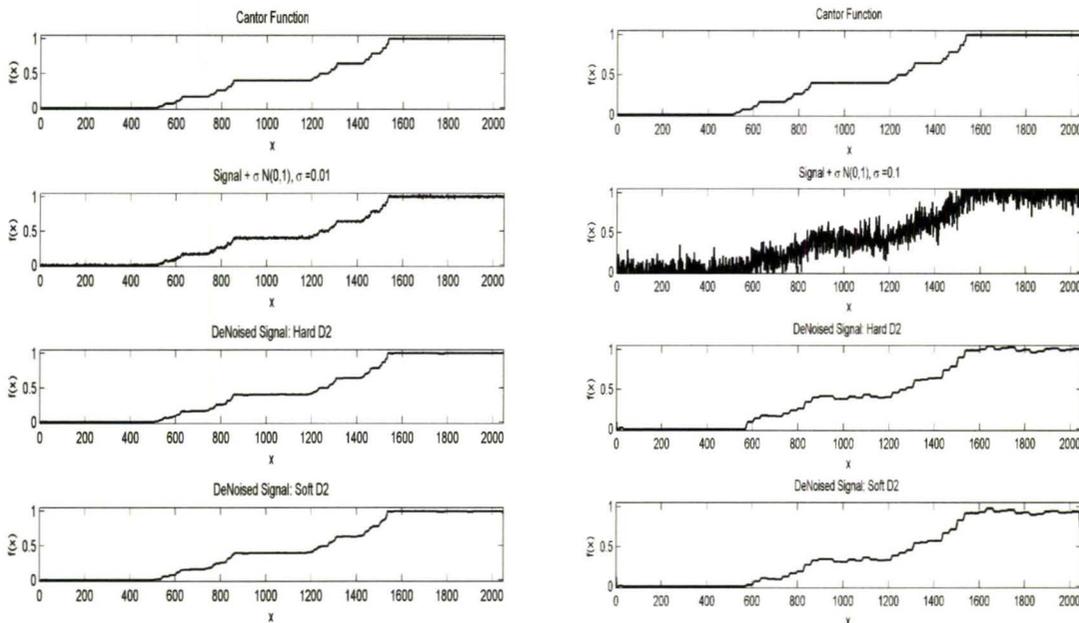


Figure 3.12: The results of signal denoising using the Haar wavelet are shown. The length of the input signal is $N = 2^{11}$.

Another visual comparison is made possible through the wavelet scalogram. In figure 3.13 the scalogram for the Cantor function is compared with scalograms produced by adding and removing Gaussian noise. The de-noised function's scalograms are hard to distinguish from the original signal's spectrum. Subtle differences are observed, particularly for the signal de-noised with soft thresholding. As the noise level is increased the recovered spectra exhibit more dramatic deviations from the original spectrum; figure 3.14. Spurious maxima lines appear in the de-noised signals scalograms. Some noise remains, but the de-noised scalograms certainly reveal more about the function than the noisy spectrogram. As expected, as σ increases it becomes harder to recover the signal from the noisy test functions.

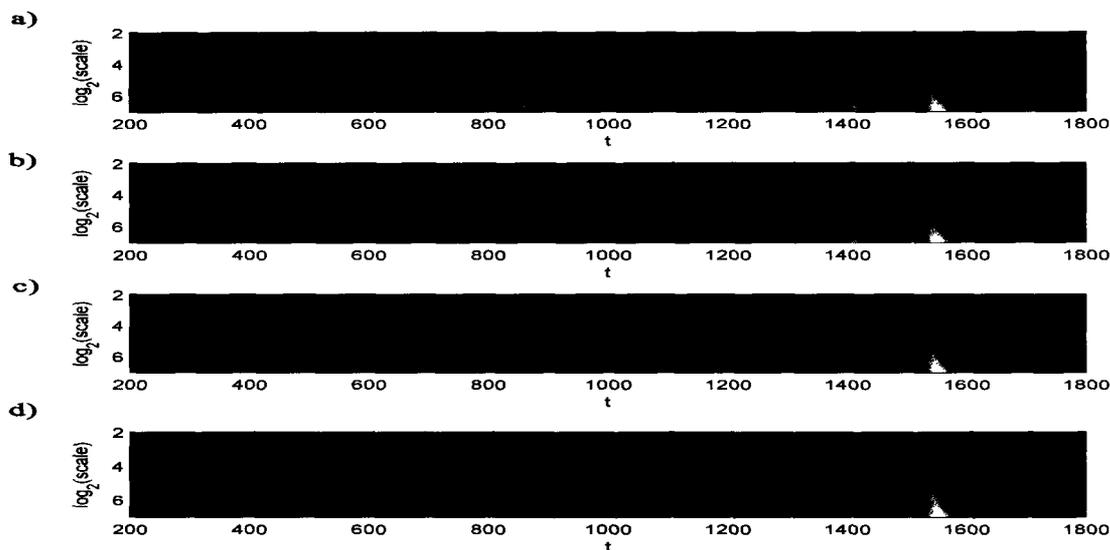


Figure 3.13: The affect of wavelet de-noising on the wavelet scalogram. a) Cantor function plus GWN with $\sigma = 0.01$. b) The Cantor function. c) The Cantor function recovered with hard thresholding. d) Cantor function recovered with soft thresholding. The scalograms are plotted within the cone of influence to eliminate edge effects.

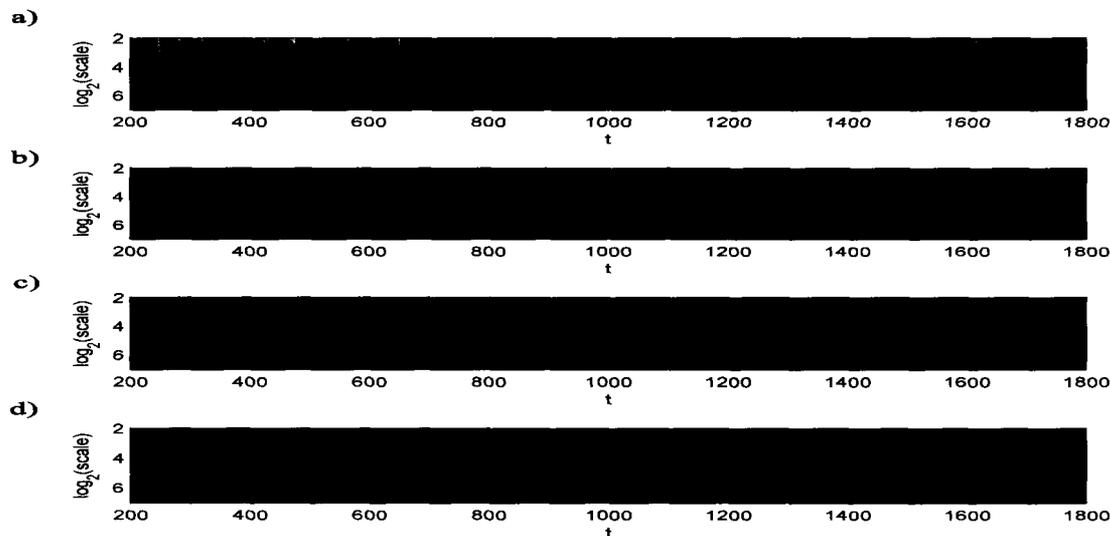


Figure 3.14: The affect of wavelet de-noising on the wavelet scalogram. a) Cantor function plus GWN with $\sigma = 0.1$. b) The Cantor function. c) The Cantor function recovered with hard thresholding. d) Cantor function recovered with soft thresholding. The scalograms are plotted within the cone of influence to eliminate edge effects.

Pointwise Error

Visual comparisons offer a qualitative analysis of the effectiveness of wavelet denoising. Quantitative analysis provides a more concrete understanding of the performance of translation invariant thresholding. Figure 3.15 shows the L^2 norm plotted against the amplitude of the noise for the sine function. For comparison, 2^{11} points were used in each case. The figures show that the D20 wavelet with 10 vanishing moments de-noises more effectively than the Haar wavelet (D2) for the sine wave; having lower values of error. Hard thresholding outperforms soft thresholding in this example.

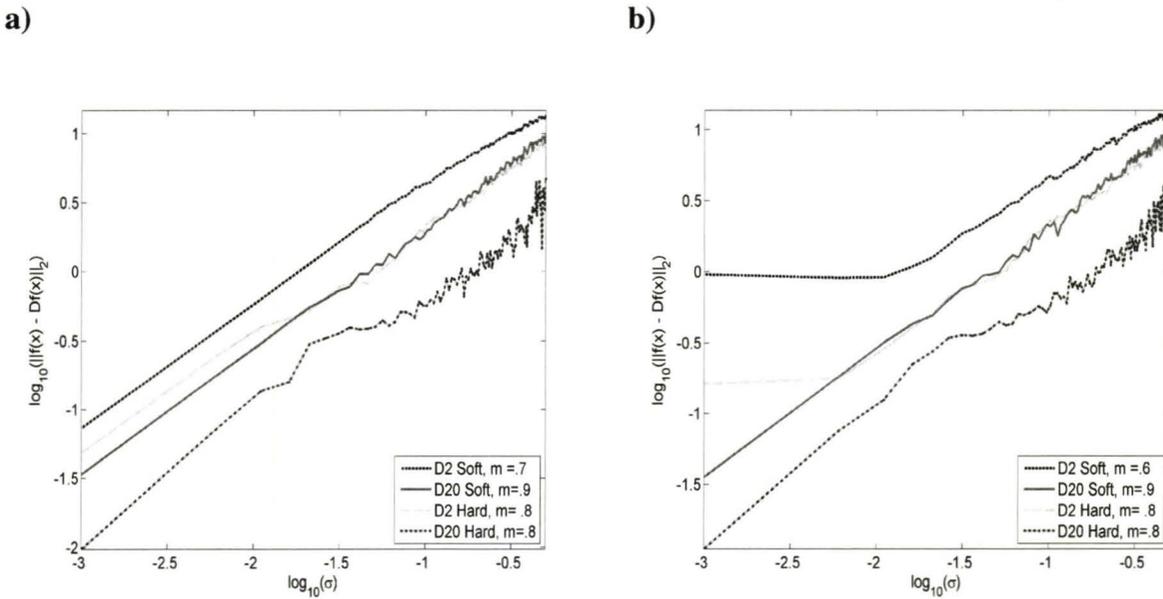


Figure 3.15: Pointwise error analysis for translation invariant de-noising of the Sine function. The slopes of the associated best fit lines, m are found in the legend. a) The known amplitude of the noise, σ , is used. b) σ is estimated.

Figure 3.16 displays the pointwise error for denoising of the Cantor function. Again, 2^{11} points are used. In the case of the Cantor function, using a Haar wavelet to de-noise produces lower Errors. Again, hard thresholding outperforms soft thresholding. These results are not unexpected since the essential difference between hard and soft thresholding is that in soft thresholding the amplitude of the wavelet coefficients are “shrunk”.

Figures 3.16 and 3.15 also show the error associated with using the estimation procedure for the noise level previously discussed (3.34). For the sine wave, the Haar wavelet exhibits much larger errors when σ is unknown. The error is lower if the variance of the noise is known a priori. Our analysis shows that when the variance must be estimated the method still provides a reasonably accurate recovery of the signal, provided the appropriate wavelet is chosen.

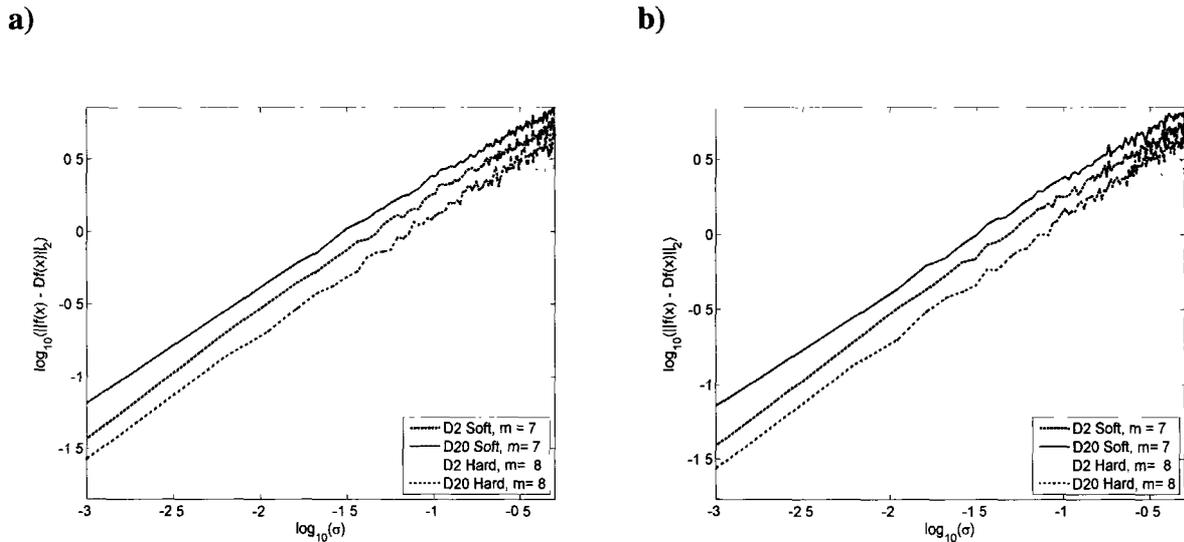


Figure 3.16: Pointwise error analysis for translation invariant de-noising of the Cantor function. The slopes of the associated best fit lines, m are found in the legend. a) The known amplitude of the noise, σ , is used. b) σ is estimated.

Power Spectrum

In section 3.1 the ability of the wavelet transform to reproduce the Fourier power spectrum was tested. In this section, the de-noising capabilities are tested by examining the associated power spectra. To begin, the wavelet power spectra is computed for a random phase signal with $E(k) \approx k^{-\alpha}$. Power spectra are computed and compared for the random phase signal, for the signal plus Gaussian white noise, and for the signals recovered via soft and hard thresholding. We also compare the power spectra de-noised using the estimated noise level.

Two distinct random phase signals are considered one with $\alpha = 8$, and the other with $\alpha = 5/3$. The function corresponding to $\alpha = 8$ is much smoother than the function corresponding to $\alpha = 5/3$. This choice of test functions provides insight into the effectiveness of de-noising for signals with contrasting regularities.

Since our goal is to reproduce the power spectrum, one may wonder whether translation invariant denoising is appropriate. The averaging procedure used may alter the resulting power spectrum giving an average power spectrum instead. The numerical results agree with this hypothesis. Much larger values of error are observed for translation invariant denoising.

The error computed in this analysis is the error in the slope of the computed power spectrum. Other choices for error analysis are possible such as pointwise error estimation, however it is the slope which is most critical in many applications. It is interesting to remark that in some instances the power spectrum may be well approximated, but visual comparisons reveal that the de-noised signal looks significantly different from the original.

Throughout our investigation, we have seen that the ability to recover a signal from noisy data depends largely on the signal to noise ratio (SNR). The SNR is a ratio of the average power in the signal to the average power of the noise.

$$SNR = \frac{P_{signal}}{P_{noise}} = \left(\frac{A_{signal}}{A_{noise}} \right)^2 \quad (3.35)$$

To compute A , the root mean square amplitude, simply find the $L^2(\mathbb{R})$ norm for the signal and divide by the square root of the length of the signal. As a convention, the SNR is computed using a logarithmic decibel scale, $SNR(db) = 20 \log_{10} \left(\frac{A_{signal}}{A_{noise}} \right)$. For the same function, the higher the variance of the noise, σ , the lower the signal to noise ratio.

Figure 3.17 shows that, for the random phase signal with $\alpha = 8$, the error in slope is greater when the SNR is lower. If the average power of the signal is greater than the average power of the noise, then the power spectrum is recovered from the signal with minimal error. For smaller SNR the power of the noise dominates the signal making the underlying signal's power spectrum more difficult to recover. Our observations indicate that a signal can be recovered from noisy data most effectively when $SNR > 1$.

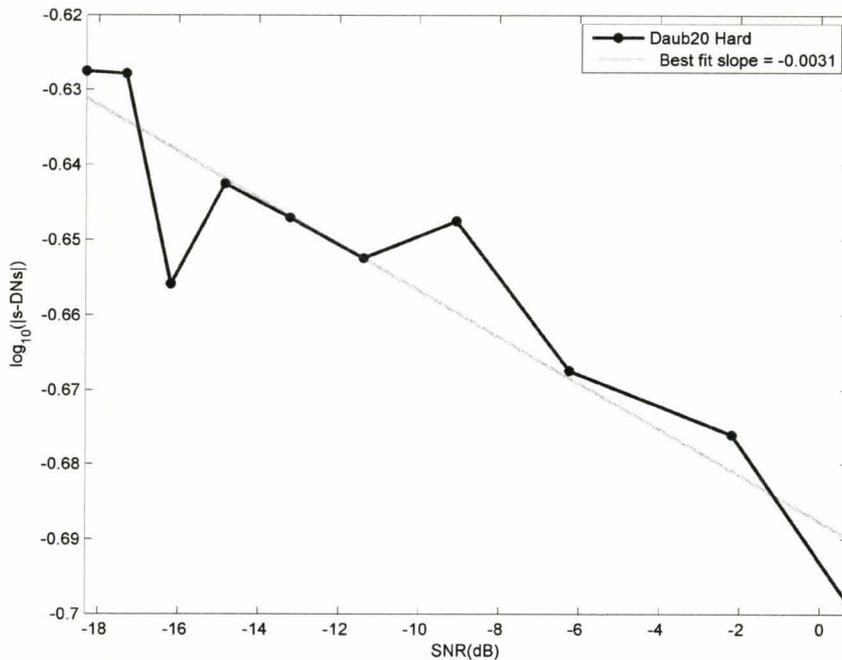


Figure 3.17: The error in slope as a function of the SNR for a random phase signal with $E(k) \approx k^{-8}$.

Figure 3.18 shows what these power spectra look like. Both graphs display the results for

$\sigma = 0.1$. For $\alpha = 8$, $SNR \approx 10^{-6}$. In the case of $\alpha = 5/3$, the ratio is higher, $SNR \approx 7.7 \times 10^{-5}$. Notice the qualitative differences between the initial power spectrum and the spectrum recovered from noise. For $\alpha = 5/3$, the Haar wavelet does a much better job of signal recovery than the D20 wavelet. The signal is discontinuous, so it makes sense to choose a wavelet with comparable regularity.

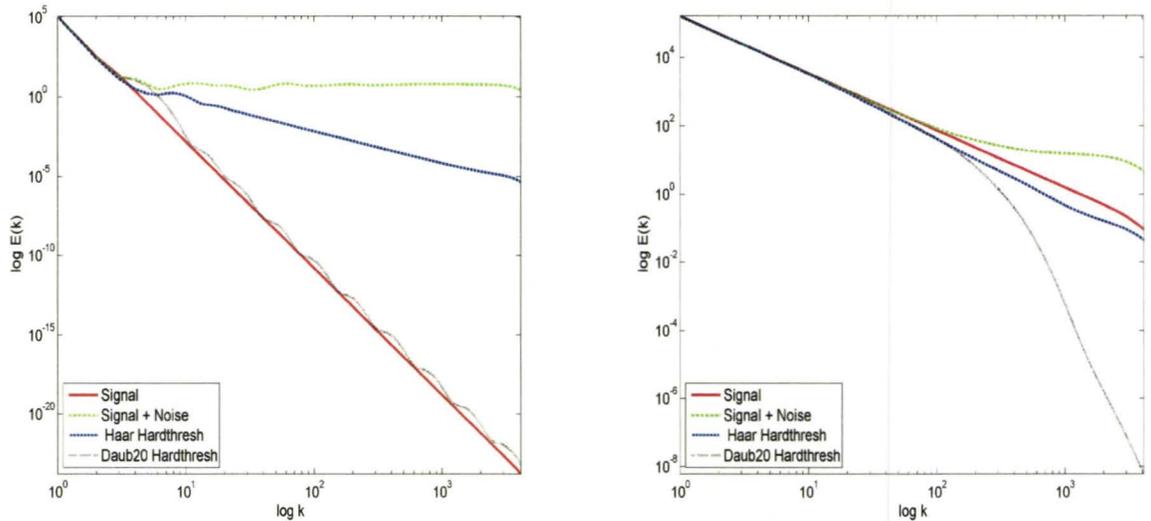


Figure 3.18: Both figures show an instance of de-noising the power spectrum for a random phase signal with algebraically decreasing spectra ($\sigma = 0.1$). Left: $E(k) \approx k^{-8}$. Right: $E(k) \approx k^{-5/3}$.

For $\alpha = 8$, the Daubechies 20 wavelet does a better job of signal recovery than the D2 (Haar) wavelet. In section 3.1, a criterion was established for the wavelet to accurately reproduce the Fourier power spectrum. The Paul wavelet was used to compute the de-noised power spectra. This wavelet certainly has enough vanishing moments to fit the criterion, however it is evident from our observations that processing the signal with the Haar wavelet has affected the computation of the power spectrum. To recover a signals power spectrum via wavelet thresholding one should use an appropriate number of vanishing moments, see equation 3.6.

The same analysis was done using an estimated noise level, but is not shown here. Using a known variance leads more precise signal recovery, but a reasonable approximation is obtained using the estimation procedure (3.34) provided that the SNR is close to, or greater than, one.

Chapter 4

Applications in Epidemiology

In this chapter, the wavelet tools discussed earlier are applied to epidemiological times series. To familiarize the reader with the subject area, the first section introduces the basic concepts of epidemiological studies. Epidemiological time series are then studied from a wavelet perspective with the goal of answering the question: how can wavelet analysis further our understanding of disease dynamics and epidemic modeling?

4.1 Modeling in Epidemiology

A concise introduction to the central ideas and concepts of infectious disease modeling is given by Earn [12]. Mathematical modeling has become an important tool for epidemiological researchers. An effective model helps to develop strategies for the eradication and control of infectious diseases. The first step in developing an accurate model is to test it against previously observed data. If the models can successfully predict patterns of epidemics which occurred in the past, they should be useful for predicting the effects of changes in control strategies such as vaccination [12]. There are two standard approaches to modeling in epidemiology one statistical and one mechanistic.

Mechanistic modeling is often preferred to statistical modeling. The main reason for this preference is that mechanistic models take advantage of our understanding of the processes that generate the time series. Statistical time series models treat the time series itself as the only information. Compartmental models are examples of mechanistic models. In these model, individuals are distributed into mutually exclusive compartments based on infection status, age, social group etc. The SIR model with demographics assigns compartments to the number of infected (I), susceptible(S) and recovered (R) individuals. Susceptibles have no immunity to the infection and may catch the infection from an infected individual. Having lived through the infection, the individual develops immunity and enters the recovered class. The model makes several, or all of the

following assumptions. Individuals are assumed to give birth and die at a constant per capita rate. Infected individuals are assumed to recover at a constant rate and retain lifelong immunity. The incidence of infection is assumed to be proportional to the product of the number of infected and susceptible individuals. These assumptions lead to the following set of differential equations.

$$\begin{aligned}\dot{S} &= \nu N - \beta IS - \mu S \\ \dot{I} &= \beta IS - \gamma I - \mu I \\ \dot{R} &= \gamma I - \mu R\end{aligned}$$

Here $N = S + I + R$ is the population size. β is the mean rate that an infected individual transmits the infection to a susceptible. μ is the constant per capita rate that individuals in each compartment die and ν is the constant per capita birth rate. $1/\gamma$ is the mean duration of infection. It is often assumed that $\nu = \mu$ so that N is constant. These simplifications ignore some real world demography and epidemiology, but the model quantitatively captures many features of real epidemics.

Consider a situation where everyone is susceptible, $S(0) = N$. Now, introduce one infected individual. The infected person is expected to infect other at a rate of βN , according to the model. The number of individuals that this one person is expected to infect is given by $R_0 = \frac{\beta N}{\gamma}$ [12]. R_0 is called the *basic reproduction number*. When $R_0 > 1$ each infected case more than replaces itself, incidence grows, causing an epidemic.

The *SEIR* is an extension of the SIR model. The *exposed* compartment, E , is added to represent the number of individuals who are infected, but not yet infectious.

$$\begin{aligned}\dot{S} &= \nu(1 - p) - \beta IS - \mu S \\ \dot{E} &= \beta SI - \sigma E - \mu E \\ \dot{I} &= \sigma E - \gamma I - \mu I \\ \dot{R} &= \nu p + \gamma I - \mu R\end{aligned}$$

σ is the rate at which exposed individuals become infectious and p is the proportion of the population that has been vaccinated. The vaccination term is added to account for the fact that many children are vaccinated before ever being exposed to the disease. The proportion of vaccinated individuals immediately enters the immune compartment while the portion that is not vaccinated enters the susceptible compartment.

4.1.1 Noise and Stochasticity

Noise can be thought of as a disturbance which distorts the essential features of a signal and may be random or systematic. The presence of noise in epidemiological data may mask important features taking place on long timescales. For this reason the data is often smoothed [12].

Data generated by compartmental models produces observable effects from the use of a discrete population size. To investigate the effects of demographics the models are recast as stochastic processes. The *Gillespie algorithm* is used to stochastize the relevant processes for compartmental models[15]. The addition of a small amount of demographic noise is enough to prevent the models from damping out [12].

In recent years some modelers have suggested that, in addition to demographic stochasticity, environmental stochasticity should be incorporated [5]. Environmental stochasticity refers to events or effects which have an impact on an entire population. Fluctuations in climate, for instance, can lead to observable effects on disease incidence.

Gillespies Algorithm

This description of the algorithm summarizes Earn [12]. Let a_0 be the overall event rate $a_0 = \sum_i a_i$. Each a_i is the rate at which a specific process occurs, birthrate, for example. The average time between events is $1/a_0$. In these models, the transition from state to state is governed by a Poisson process. The probability that the next event occurs in the interval $[t, t + \delta t]$ is $a_0 e^{-a_0 t}$. For $u \in [0, 1]$ let $u = e^{-a_0 t}$ so that the time until the next event can be found by sampling u from a uniform distribution in $[0, 1]$ and setting $t = \frac{1}{a_0} \ln(\frac{1}{u})$. By sampling a point from a uniform distribution on $[0, a_0]$ the nature of the event is determined. The event is set to type i if it lies in the i^{th} interval:

$$[0, a_1), [a_1, a_1 + a_2), \dots, [a_1 + \dots + a_{i-1}, a_1 + \dots + a_i), \dots$$

4.1.2 Seasonal Forcing

Transmission rates are dependent on human interaction. Greater human interaction leads to higher transmission rates amongst individuals. These rates tend to vary from season to season. Consider the example of measles. Cases of measles are most often reported in children around age five. Transmission rates are, in theory, much higher when school is in session. It is appropriate to introduce a transmission rate that varies in time. Seasonal forcing is crudely approximated

by:

$$\tilde{\beta}(t) = \beta_0(1 + \beta \cos 2\pi t) \quad (4.1)$$

Where β_0 is the mean transmission, t is the time in years, and $0 \leq \beta \leq 1$ is the amplitude of seasonal forcing. A compartmental model is said to possess *seasonal forcing* if $\tilde{\beta}$ is a periodic function with an annual period. Without seasonal forcing the deterministic SEIR model exhibits damped oscillations onto a state of equilibrium. Periodically forcing the model prevents this damping and may lead to complex dynamics such as chaos for high amplitudes.

The idea that transmission rates for measles are higher when school is in session and lower when school's out lead to the development of a non-sinusoidal forcing term. The forcing is generated by incorporating a correspondence to the specific school calendar.

$$\tilde{\beta}(t) = \begin{cases} \beta_H & \text{on school days,} \\ \beta_L & \text{on non-school days} \end{cases} = \begin{cases} [1 + 2(1 - p_s)\beta] \langle \tilde{\beta} \rangle & \text{on school days,} \\ [1 - 2p_s\beta] \langle \tilde{\beta} \rangle & \text{on non-school days} \end{cases} \quad (4.2)$$

The parameters are defined as follows: p_s is the proportion of school days in a year,

$$p_s = \frac{\text{number of school days}}{365} \quad (4.3)$$

The amplitude of seasonal forcing is given by $\alpha = \frac{1}{2} \left(\frac{\beta_H - \beta_L}{\langle \tilde{\beta} \rangle} \right)$ and $\langle \tilde{\beta} \rangle = p_s \beta_H + (1 - p_s) \beta_L$. This model is found to be quite successful at reproducing many of the features observed in the data.

4.2 Power Spectra

Since the 1980s epidemic modelers have used spectral analysis on incidence time series to understand epidemic recurrence [5]. Traditionally, the approach is to detrend the data and compute the windowed Fourier transform of the autocorrelation function. The dominant peaks in the power spectrum indicate the period between successive outbreaks. Power spectra provide information about the nature of the data that is not available upon inspection. Bauch points out that power spectra of time series of childhood disease incidence reveals that spectral peaks can be both seasonal and non-seasonal [5], but the limitations of the windowed Fourier transform make this method insufficient for complete characterization of the time series. Fourier power spectra contain no phase information, and may contain spurious peaks if the window selected is too narrow.

The localization properties of the wavelet transform are used to compute power spectra in

a range of interesting applications. Most notably Torrence and Compo used wavelet spectral analysis to analyze ecological time series [31]. They offer a guide for wavelet spectral analysis complete with a wavelet toolbox available on line. Torrence and Compo not only provide a summary of the procedures for computing wavelet power spectra, they also develop significance tests to highlight important features of the data.

In chapter two, the wavelet power spectrum was defined: $P = |\tilde{s}(a, b)|^2$. Torrence and Compo suggest that in the interest of reproducible research a common normalization should be used in the wavelet transform [31]. The continuous wavelet transform is computed using the convolution theorem as the inverse Fourier transform of a product.

$$W_n(s) = \sum_{k=0}^{N-1} \hat{x}_k \bar{\hat{\psi}}(s\omega_k) e^{i\omega_k n \delta t} \quad (4.4)$$

Normalizing the wavelet to have unit energy allows for comparisons across scales, as well as, with other data sets. Explicitly, the normalization is:

$$\hat{\psi}(s\omega_k) = \left(\frac{2\pi s}{\delta t} \right) \hat{\psi}_0(s\omega_k). \quad (4.5)$$

The wavelet coefficients are further normalized by subtracting the mean from the time series and dividing by the square of the variance of the time series, ρ^2 .

The Fourier transform is known to introduce edge effects into the power spectrum because the data is treated as though it is cyclic. To avoid this problem zeros can be added to both ends of the time series before transforming and subsequently removed. The cone of influence is often plotted to demarcate the regions where edge effects are significant.

Typical wavelet spectrograms plot the time versus the scale with power coefficients color coded to indicate amplitude. For many applications, including epidemiology, what is needed is not a view of the relationship between scale and time, but instead a view of the relationship between period and time. It is important to ensure that the results of wavelet analysis are directly comparable with those of Fourier analysis. To derive the relationship between the Fourier period and the wavelet scale compute the wavelet transform of a cosine wave with known frequency and find the scale at which the wavelet power spectrum reaches a maximum. For the Morlet wavelet the value of the Fourier period is $\lambda = 1.03$; scale and period are nearly equivalent [31].

Significance testing can be used to highlight significant features of the data. The data is assumed to have a mean background spectrum. Peaks in the wavelet power spectrum which are significantly above the expected value of the background spectrum are assumed to be true features of the data with a certain degree of confidence. Torrence and Compo suggest using the

discrete Fourier power spectrum to find the expected spectra for an assumed mean background spectrum (i.e. white or red noise).

$$P_k = \frac{1 - \theta^2}{1 + \theta^2 - 2\theta \cos(2\pi k/N)}. \quad (4.6)$$

$k = 0 \dots N/2$ is the frequency index and θ is the assumed lag-1 correlation. If the lag-1 coefficient is zero, equation 4.6 can be used to model a white noise Fourier spectrum. A white noise process is expected to have $|W_n(s)|^2 = \sigma^2$ at all positions, n and scales s ; σ is the standard deviation of the noise. Because the computed spectra is first normalized by subtracting the mean and dividing by the variance, white noise has an expectation value of 1 at all frequencies. So, wavelet coefficients with values significantly larger than 1 are assumed to be true features of time series.

For the epidemiological examples a background spectrum of white noise is assumed, the 95 % confidence level is indicated by dark contours. The analysis which follows will test the validity of this assumption.

4.3 Wavelet Analysis of Epidemiological Time Series

4.3.1 The Data

The incidence data used in this study can be found at the International Infectious Disease Data Archive (IIDDA), an on line resource for infectious disease data [9]. Where data points were missing, cubic splines were used to interpolate.

The stochastic SEIR model is examined for comparison with measles incidence data. The data was collected monthly in Ontario from 1904-1989. The stochastic model used was developed by Wagner [32]. Both types of seasonal forcing discussed previously are tested; sinusoidal and term-time. The following parameters are implemented (time in years):

$$\begin{aligned} \sigma &= 48.6 \\ \gamma &= 56.15 \\ R_0 &= 17.5 \end{aligned}$$

The number of live births and population sizes in Ontario from 1921 to present are found at the Canadian Human Mortality Database [10]. These statistics were used to compute the birth rates. The relevant data for years before 1921 is not widely available. In both stochastic models birth rates are continuous functions of time computed using the derived annual birth rates. Model data is taken in daily samples from 1921 - 1989. Figures 4.1 and 4.2 show several realizations of the models.

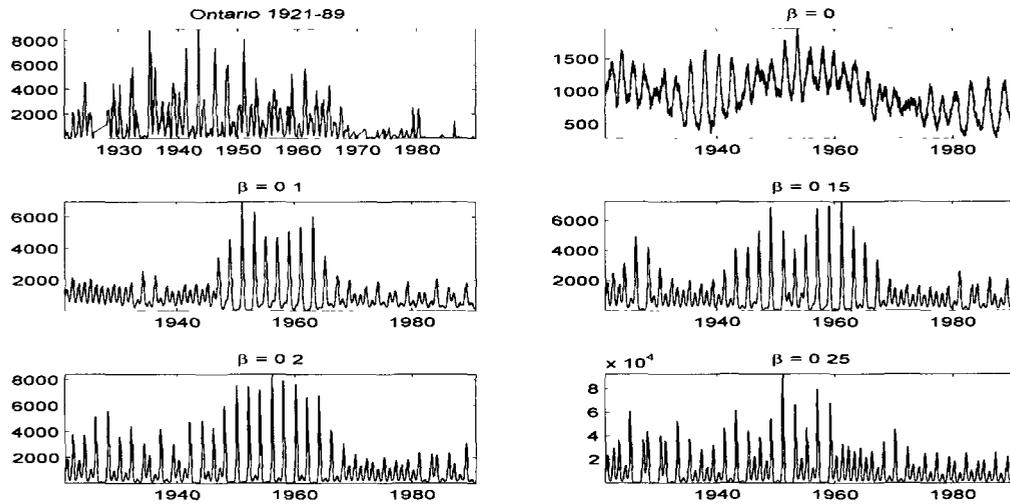


Figure 4.1: Comparison of the sinusoidally forced stochastic model incidence with Ontario incidence.

The actual data exhibits larger peaks before 1940 than the stochastic models. This could be due to high immigration rates around that time. Immigration of school age children introduces susceptibles and possibly infecteds into the system. The introduction of the measles vaccination around 1970 had significant effects on the incidence time series. The effects of immigration and vaccination are not included in the stochastic models so we do not expect to see the same pattern of recurrence from the stochastic models'.

It is difficult to draw conclusions about which type of seasonal forcing is to be preferred based on qualitative comparisons of the plotted time series'. In figure 4.1 we see that for $\beta = 0$ the time series appearance differs a great deal from the appearance of the Ontario data. Clearly, the addition of seasonal forcing of either type improves the model's ability to reproduce the Ontario data. We shall soon see how wavelets can help us to determine the appropriate level of seasonal forcing.

4.3.2 Denoising the Data

Wavelet thresholding is applied to the measles incidence data and the stochastic models. The purpose is to both observe the affect that de-noising has on the singularity and power spectra, as well as, to determine the nature of the extracted "noise". Figure 4.3 shows the singularity spectrum approximated from the Ontario measles data. Most of the singularity spectrum lies in the non-smooth range, ($\alpha < 0$) and the continuous subsets of the data are anti-persistent ($0 \leq \alpha < .5$). Recall antipersistence is the idea that an increasing trend in the past implies a decreasing

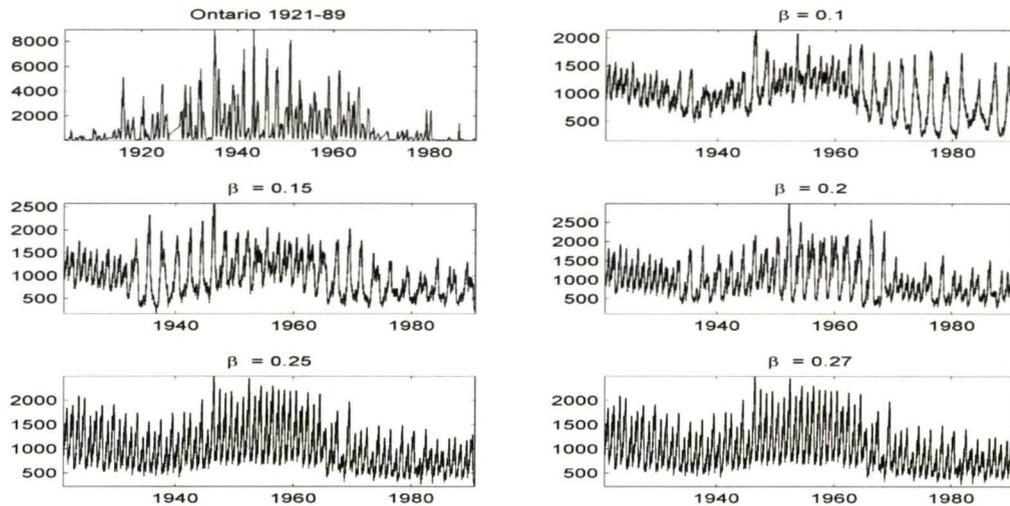


Figure 4.2: Comparison of the term-time model incidence with Ontario incidence.

trend in the future and an increasing trend is made probable by a decreasing trend in the past [14].

Soft and hard thresholding are applied using the Daubechies 20 wavelet and the Daubechies 2 (Haar) wavelet. As evidenced by the singularity spectrum of the ON incidence time series, the data is not continuous or differentiable. The appropriate wavelet basis is, therefore, one with narrow compact support such as the Haar basis. For the stochastic models, which we will see are much more monofractal than the Ontario data, a wavelet with many vanishing moments may be more appropriate; such as the Daubechies 20 wavelet.

The spectra of the de-noised time series' are also shown in figure 4.3. Virtually no change is observed in the overall singularity spectrum upon de-noising. The data can be de-noised using wavelet thresholding without altering the regularity of the signal.

Another tool for evaluating the affect of de-noising on the time series is to compute the wavelet power spectrum for the de-noised ON data. The data is de-noised using an estimated value for the noise level, σ . Hard thresholding causes no visible changes in the power spectrum at all, while soft thresholding leads to shrinkage of the contours in some areas. De-noising the stochastic models gives similar results; no significant changes are observed, but power spectra does appear lighter in color than the original in some areas.

A simple linear decomposition of our signal isolates the extracted portion of the data. Figure 4.4 displays the results. For the same estimated noise level, soft thresholding removes more noise than hard thresholding by shrinking the wavelet coefficients.

To characterize the nature of the removed portion of the data, the probability density histograms are plotted in figure 4.5. The PDF for the noise removed using soft thresholding fits

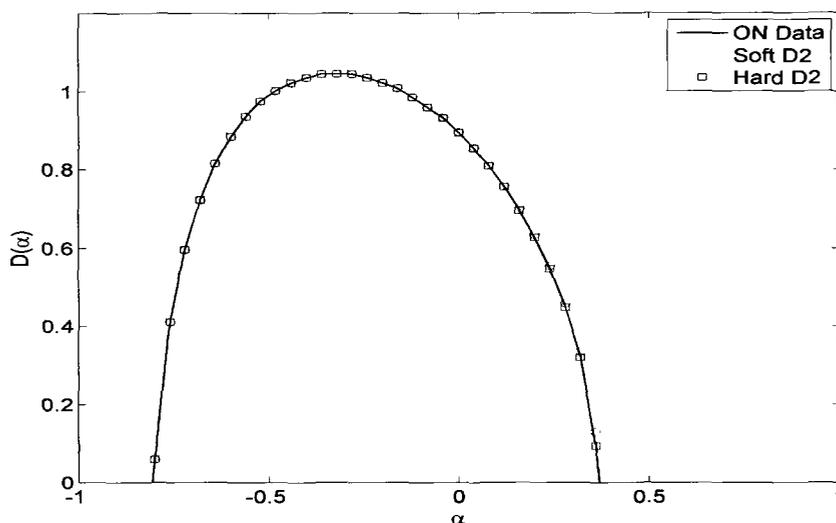


Figure 4.3: The singularity spectrum for Ontario measles incidence 1921-1989 plotted with spectra for the de-noised data. The Haar wavelet is used to de-noise the time series and the Mexican hat wavelet is used to plot the multifractal spectra.

well with a Gaussian distribution. The estimated energy spectrum of the noise is nearly flat. Therefore, the noise may be classified as Gaussian white noise. For the noise removed with hard thresholding, the PDFs cannot be fit to Gaussian distributions. This does not necessarily mean that the removed data is not Gaussian. The Ontario incidence data consists of only 1032 points of monthly data. It could be that the noise removed via hard thresholding is simply too small to properly fit a distribution to.

4.3.3 Measles Power Spectra

Figure 4.6 shows the wavelet power spectrum for measles incidence of infection data collected monthly in Ontario Canada from 1904-1989. Spectral analysis was first performed on this time series by Bauch and is reproduced by altering the codes provided by Torrence and Compo [5], [31]. A seasonal peak of one year is clearly seen in the power spectrum. These seasonal peaks are understood to be due to an increase in transmission when school is in session and children are interacting more. The presence of significant coefficients at non-integer periods indicates non-seasonal recurrence.

In the year 1970, after the vaccine was introduced, the behavior of the power spectrum noticeably changes. The period 2-5 recurrence seen in the power spectrum before 1970 diminishes reflecting longer periods between epidemics.

The idea that localized power spectra will lead to a better characterization of recurrence is

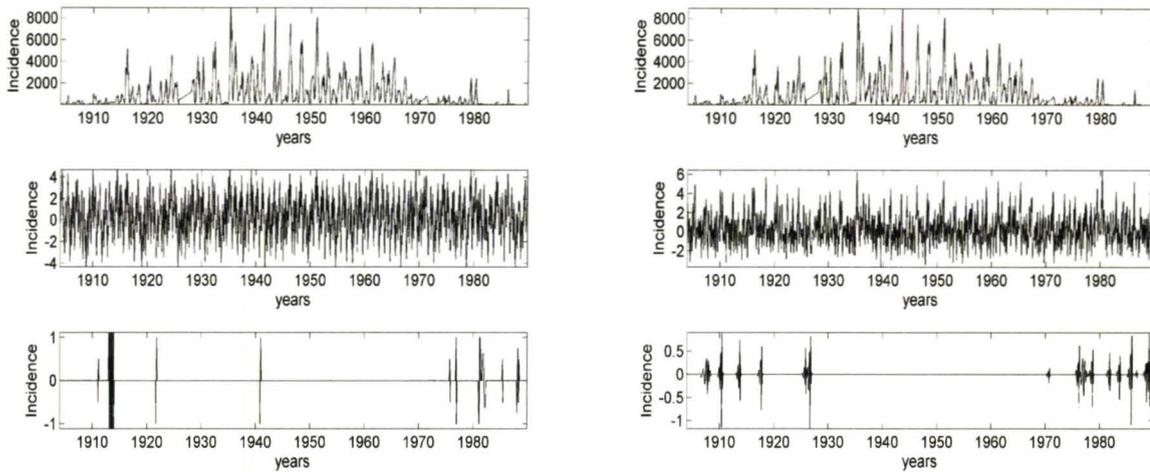


Figure 4.4: De-noising of measles incidence data 1904-1989. On the left, the Haar wavelet is used to decompose the signal. The top graph shows the actual incidence data. Below, the noised removed via Soft thresholding is shown. On the bottom the noised removed via Hard thresholding is shown. The graphs on the right use the Daubechies 20 wavelet, the ones on the left use the Haar wavelet.

the motivation for using the power spectra to fit models to the data. Several realizations of the stochastic model's power spectra are shown in figure 4.7. If the same parameters as above for the lag1 coefficient and confidence interval are used then the contour lines are not useful, they indicate that most of the spectrum is significant. On the other hand, if a much higher lag1 coefficient is chosen the contours will reveal the highest amplitude coefficients present; the significant features of the data. In these examples, $lag1 = 0.9$, and a 95% confidence level is used.

The recurrence of incidence observed in the power spectra for the stochastic models was different for different amplitudes of seasonal forcing. It seems that a minimum amplitude of seasonal forcing is required to reproduce the strong period one recurrence seen in the power spectrum for the Ontario data. For the term-time model the minimum amplitude is around $\beta = 0.15$, but much lower for the sinusoidally forced model, $\beta \approx 0.06$. Both seasonal and non-seasonal peaks appear in the power spectra for the stochastic models. Since the model we chose to use is relatively simple, we expected that the recurrence of incidence generated by our models would not fit exceptionally well with the actual data.

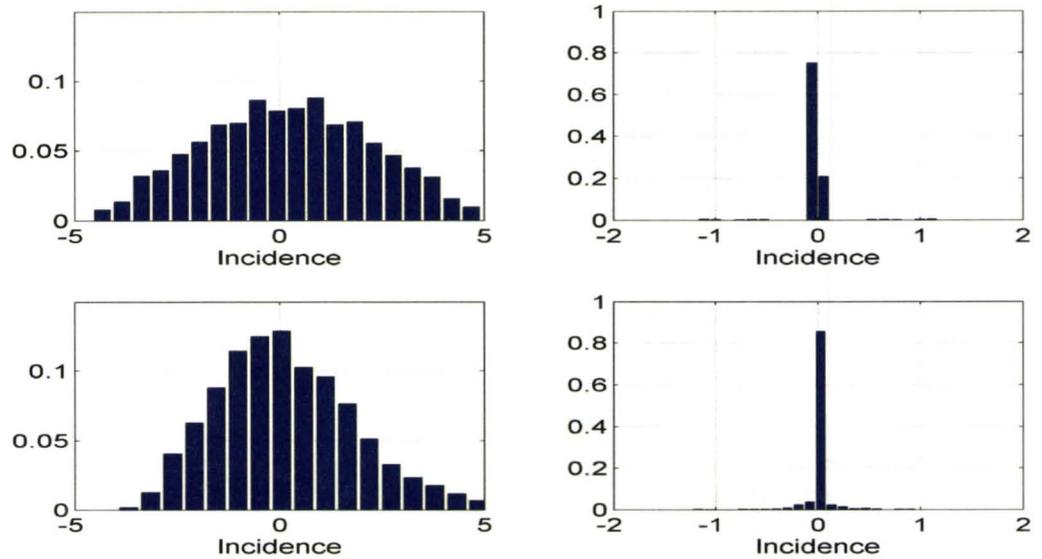


Figure 4.5: Probability density histograms of noise extracted from Ontario incidence data. The top two figures are associated with the Haar wavelet for soft (left) and hard(right) thresholding. The bottom two are associated with the Daubechies 20 wavelet.

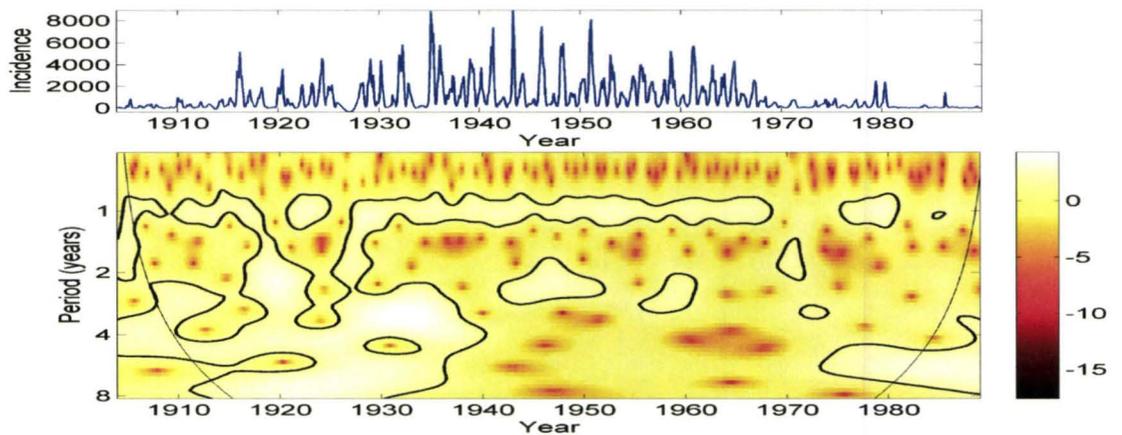


Figure 4.6: The local wavelet power spectrum for Ontario measles incidence 1904-1989. The data is log transform, the mean is subtracted and the data is normalized by the variance, $1/\rho^2$, $\text{lag}1 = 0$, 95% confidence interval shown in thick contours. The Morlet wavelet is used.

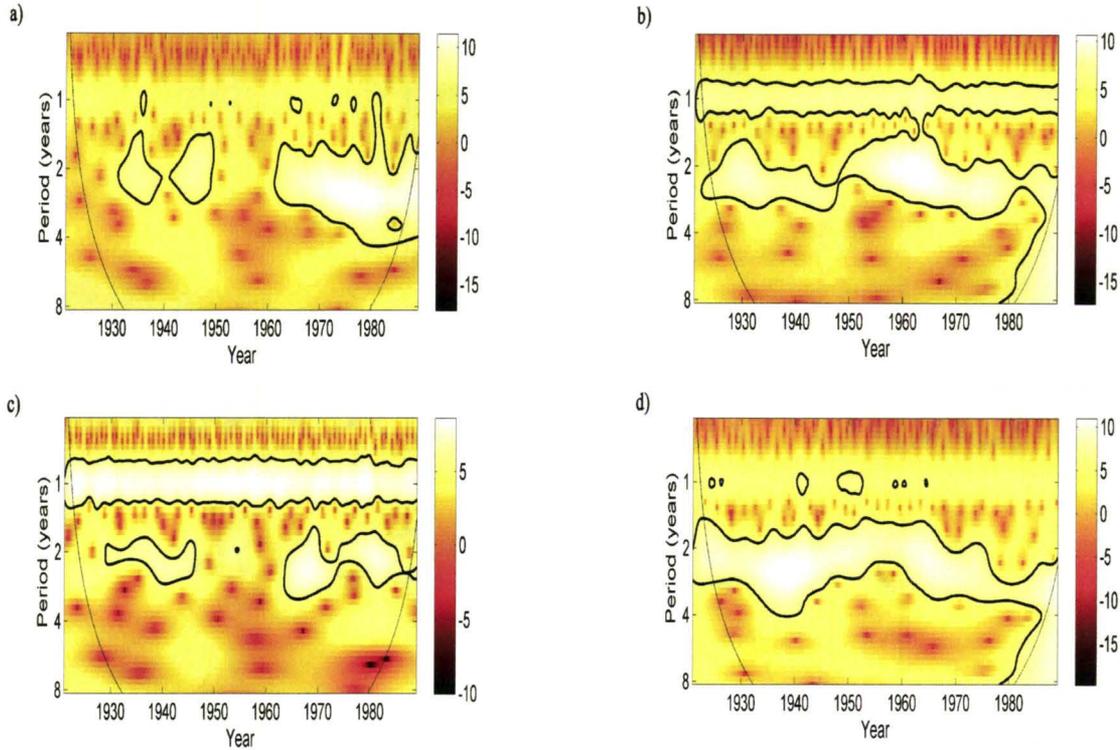


Figure 4.7: Local wavelet power spectra for the stochastic SEIR models. For the term-time forced model (left) $\beta = 0.1$ (a), 0.25 (c). For the sinusoidally forced model, (right) $\beta = 0.05$ (b), 0.08 (d) are shown. The data is log transform, the mean is subtracted and the data is normalized by the variance, $1/\rho^2$, $\text{lag}1 = 0.9$, 95 % confidence interval shown in thick contours. The analyzing wavelet is the Morlet wavelet.

4.3.4 Singularity Spectra

Fitting Models to Data

The wavelet transform maximum modulus multifractal formalism is used to find the singularity spectrum for the Ontario incidence of infection time series. After comparing several realizations of the SEIR model for different amplitudes of seasonal forcing, β , the best matches to the Ontario spectrum were selected. Spectra for the seasonally forced models are plotted with $\beta_{\text{time-term}} = 0.25$, and $\beta = 0.2$, figure 4.8. The line $x = 1/2$ is plotted to divide the regions of anti-persistence and persistence. The line $x = 0$ marks the transition from smooth ($\alpha \geq 0$) to non-smooth signals ($\alpha < 0$).

Two measures are used to quantify the differences between the singularity spectra: the relative locations of spectrum maxima and the width of the spectrum. The location of the peak of the singularity spectrum indicates the kind of regularity which dominates the signal. In other

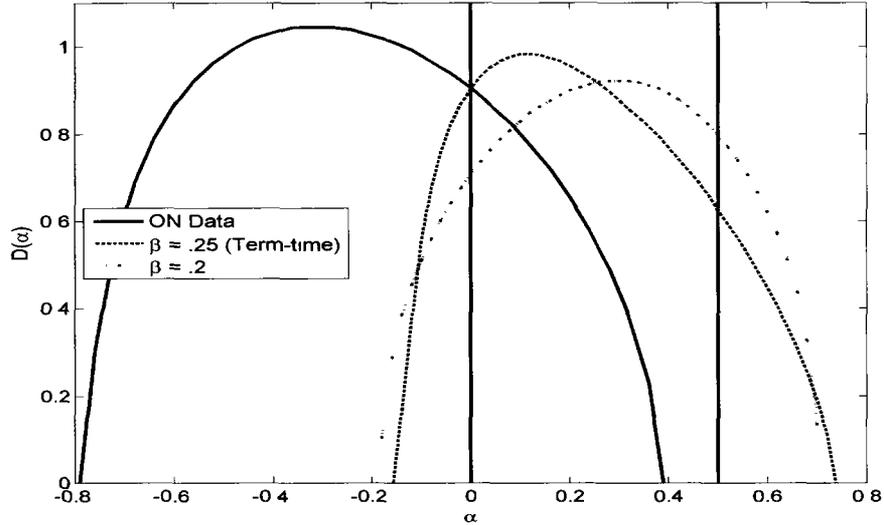


Figure 4.8: Singularity spectra of measles data compared with stochastic data. The line $x = 0$ is plotted to demarcate the smooth ($\alpha \geq 0$) and non-smooth ($\alpha < 0$) regions. The line $x = 1/2$ is plotted to divide regions of persistent and anti-persistent behaviors. The Mexican Hat wavelet was used to compute all singularity spectra.

Table 4.1: Differences in the singularity spectra for figure 4.8

Data	α_{max}	width
ON	-0.32	1.12
$\beta = 0.25$ (term-time)	0.12	0.88
$\beta = 0.2$ (sine)	0.32	0.86

words, it tells us which Holder exponent characterizes the largest subset of the time series. The width of the spectrum provides a measure of the multifractal behavior of the time series. The wider spectra is, the more the Holder exponent varies in time. The values of these parameters, for figure 4.8, are given in table 4.1. Widths are taken at the base of the spectrum with points nearest to $D(\alpha) = 0$.

In general, the stochastic model spectra are narrower and centered further to the right than the Ontario singularity spectrum. The Ontario spectrum is dominated by a Holder exponent in the discontinuous non-differentiable range, while the models' spectra are dominated by Holder exponents in the continuous and non-differentiable range. Compared to the sinusoidal model, the term-time model tends to have a peak spectrum value closer to the actual data's. The skewed shape of the Ontario spectrum is mimicked by the term-time model's spectrum. Though neither stochastic model spectrum matches particularly well with the Ontario data, the term-time model gives a better fit.

Recall that monofractals are identified by the linearity of the associated scaling exponents. Figure 4.9 displays the scaling exponent associated with the Ontario measles data and “best” stochastic model spectra. For the Ontario data, the behavior of $\tau(q)$ is distinctly non-linear. The behavior of τ for the stochastic models is more linear than that of the Ontario data, but is still not linear enough to be considered multifractal. This leads us to the conclusion that the actual data is more multifractal than the models, but it is encouraging to find that the models and the Ontario data are both multifractal in structure.

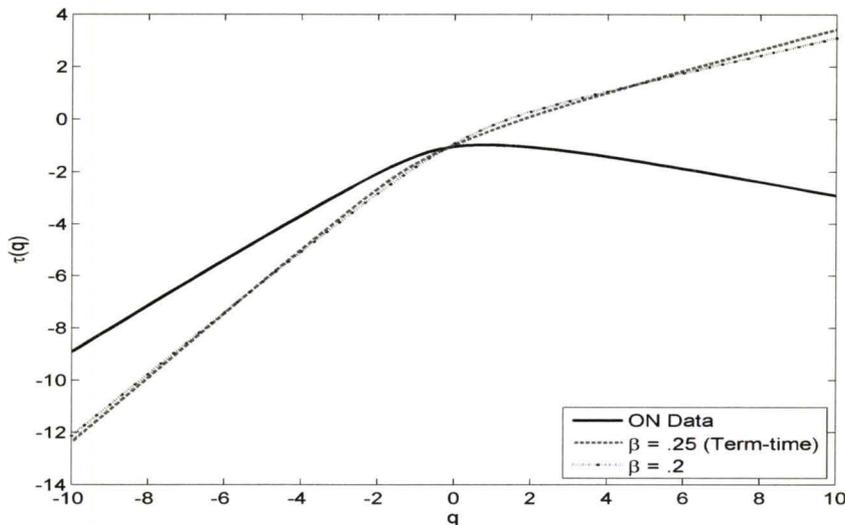


Figure 4.9: Scaling exponents for Ontario measles incidence of infection data and stochastic model incidence of infection data. Non-linear behavior of the scaling functions indicates that all of the data exhibits a multifractal structure.

To complete the analysis of epidemiological models, the singularity spectrum for the deterministic SEIR model is computed for comparison with the stochastic versions. Several amplitudes of seasonal forcing are considered. The scaling exponent is observed to be subtly non-linear in q for all realizations. Thus, the model is weakly multifractal. For $\beta = 0.15$, most of the spectrum lies to the right of $\alpha = 1$, with a peak at $\alpha = 1.5$. Increasing the amplitude of seasonal forcing shifts the peak of the spectrum to the left, taking on values $1 < \alpha_{max} < 1.5$. From this analysis it is obvious that the deterministic SEIR model is much too smooth to model measles dynamics accurately.

A Closer Look at Seasonal Forcing

To examine the effects of seasonal forcing in the term-time model more closely, the multifractal spectrum is computed for increasing values of the amplitude of seasonal forcing, β . In

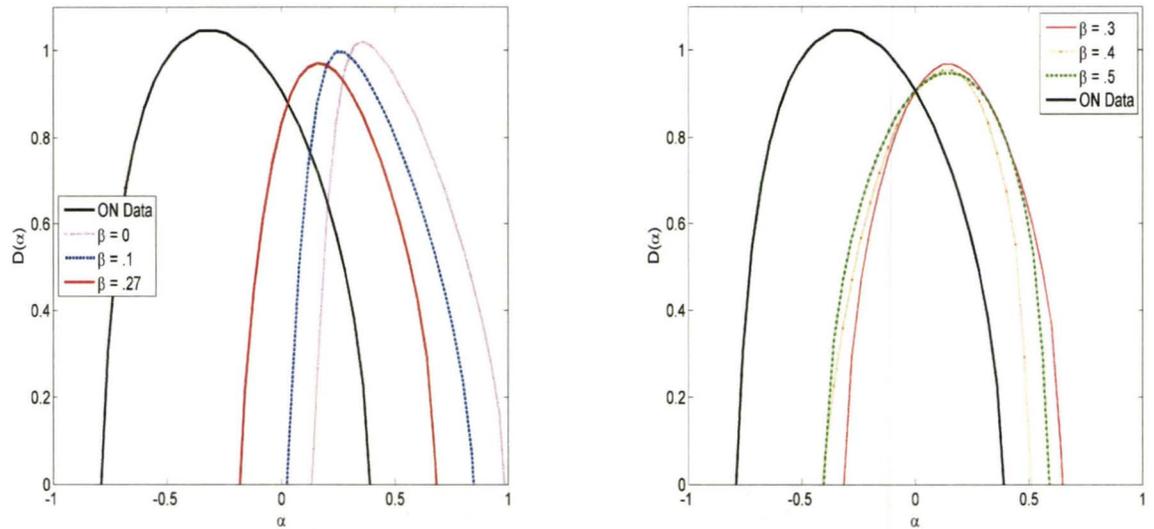


Figure 4.10: Singularity spectra of the time-term stochastic model for different amplitudes of seasonal forcing compared with ON data. The Mexican Hat wavelet is used to plot the Multi-fractal spectra.

this case, spectrum for the Ontario data is skewed to the left, figure 4.10. A similar shape is observed for spectra with $\beta \leq 0.3$. Larger values of β still shift the stochastic model spectrum, but the location of the peak of the spectrum changes very little. The base of the spectrum moves to the left creating a more symmetrical shape.

It is evident from the singularity distributions that the addition of a seasonal forcing term does improve the model's ability to reproduce the structure of the actual data, within limits. Observe that increasing β leads to a more accurate singularity spectrum up to a certain amplitude of seasonal forcing, $\beta \approx 0.3$. Further increasing β does not seem to improve the fit of the model spectra with the Ontario data spectra. In fact, the movement of the base of the spectrum to the left, for high values of β , yields spectra which have more symmetrical shapes than those observed from the actual data.

For high amplitudes of seasonal forcing, the models are said to exhibit chaotic behavior. The term chaos is here meant to refer to seemingly random, complex, aperiodic behavior with a sensitivity to initial conditions. The changes in the behavior of the singularity spectrum of the term-time model as seasonal forcing is increased might be explained as a transition to a chaotic regime. To verify this assertion, close comparisons between the singularity spectra and the associated bifurcation diagrams should be made. Time constraints prevent such an analysis here, but it is a topic for future research.

So far, all observations indicate that although seasonal forcing does improve the fit of the

model with the data, modifying this parameter is not enough to properly fit the singularity spectra of the models to the data. The addition of other model parameters may improve the fit of the model.

To test the effect of the addition of Gaussian white noise, different amplitudes of noise were added to the signal. The singularity spectra are then compared. Adding a large amplitude of noise to the “best” choices in figure 4.8, $\sigma = 1000$, does move the peak of the spectrum closer to the peak of the ON data spectrum. Figure 4.11 shows the results of adding GWN of amplitudes, $\sigma = 1000, 2500$. Recall that, for a specified signal, the SNR is lower for higher noise levels.

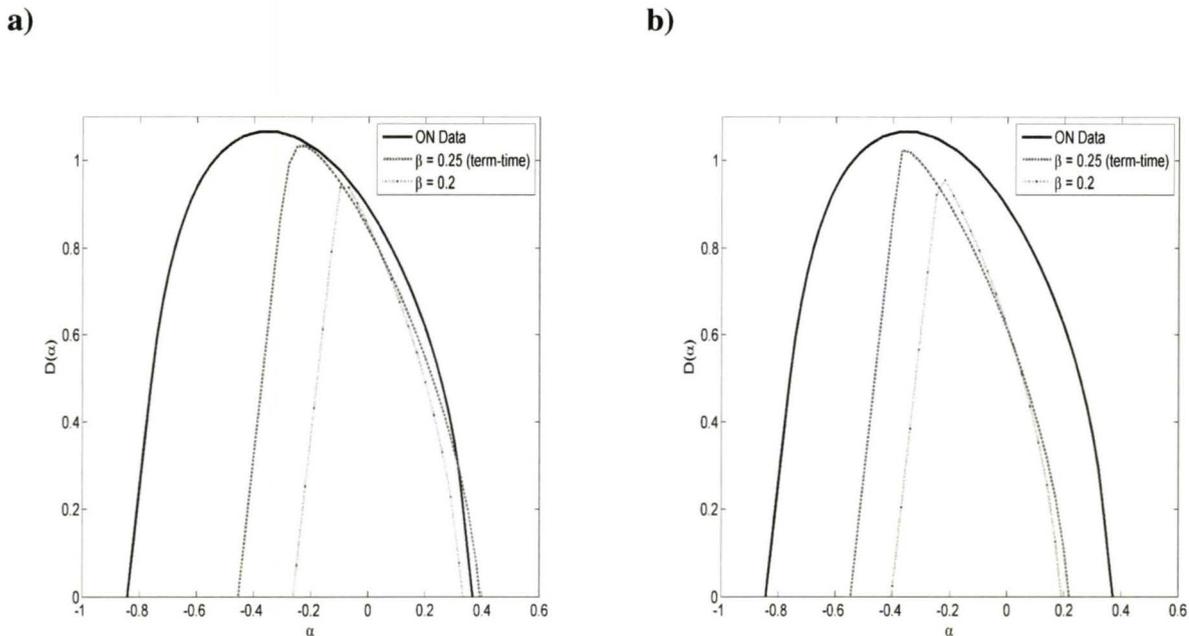


Figure 4.11: The effect of adding Gaussian white noise to the models on the singularity spectra for the stochastic SEIR models. a) $\sigma = 1000$, b) $\sigma = 2500$. Figure 4.8 shows the singularity spectra for the models with no noise added.

We observe that the more noise is added, the less multifractal the models become; the spectra become narrower and more pointed at the top. Additive GWN does improve the fit of the stochastic models to the data by shifting the peaks of the singularity spectra to the left, however the addition of GWN is not enough to cause the stochastic model spectra to mimic the wide, strongly multifractal shape of the Ontario data’s spectrum.

Unique Multifractal Signatures

The qualitative differences observed between the measles data spectrum and the stochastic models inspired a comparative analysis of the singularity spectra for different types of diseases.

Do the singularity spectra for different data sets for the same disease vary a great deal, or is there a characteristic *multifractal signature* associated with a particular disease?

As a preliminary analysis, the singularity spectra for measles, chicken pox, rubella, and whooping cough are plotted in figure 4.12. There are certainly distinctive differences between each of the spectra, but to identify a unique multifractal signature several data sets for the same disease should be compared.

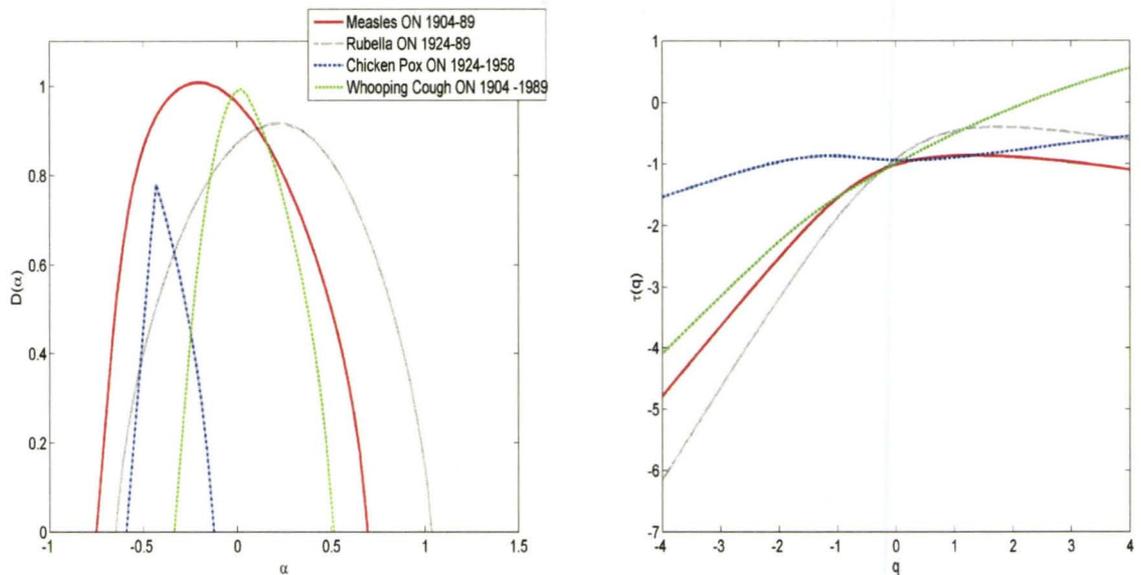


Figure 4.12: Left: Singularity spectra of incidence of infection for several diseases. Right: corresponding scaling exponents.

Five separate chicken pox data sets and three measles data sets are used to compute singularity spectra, figure 4.13. The distribution of singularities is identical for the chicken pox incidence series from New York City and British Columbia. The other three from Manitoba, Saskatchewan and Ontario are more noticeably different, but all of them share a similar shape. So too, the measles data sets share a similar shape. For measles, the spectra is smooth on the top with a wide base, whereas chicken pox spectra have pointed tops with a narrow base.

The singularity spectra for the chicken pox data have more monofractal behavior than the measles spectra; the spectra are much narrower. Figure 4.14 shows non-linear behavior of the scaling exponents for both measles and chicken pox, but the scaling exponents for chicken pox are more linear than those for measles.

The singularity spectrum provides a distinctive signature that can be used to characterize different diseases. The multifractal spectra are qualitatively similar for the same disease. An accurate numerical model should be able to reproduce this signature. Thus, we have demonstrated yet another way in which wavelets can be used to fit models to data in epidemiology. We wonder

what differences in the disease dynamics may cause the distinctive singularity spectra observed for different diseases and if they can be quantified.

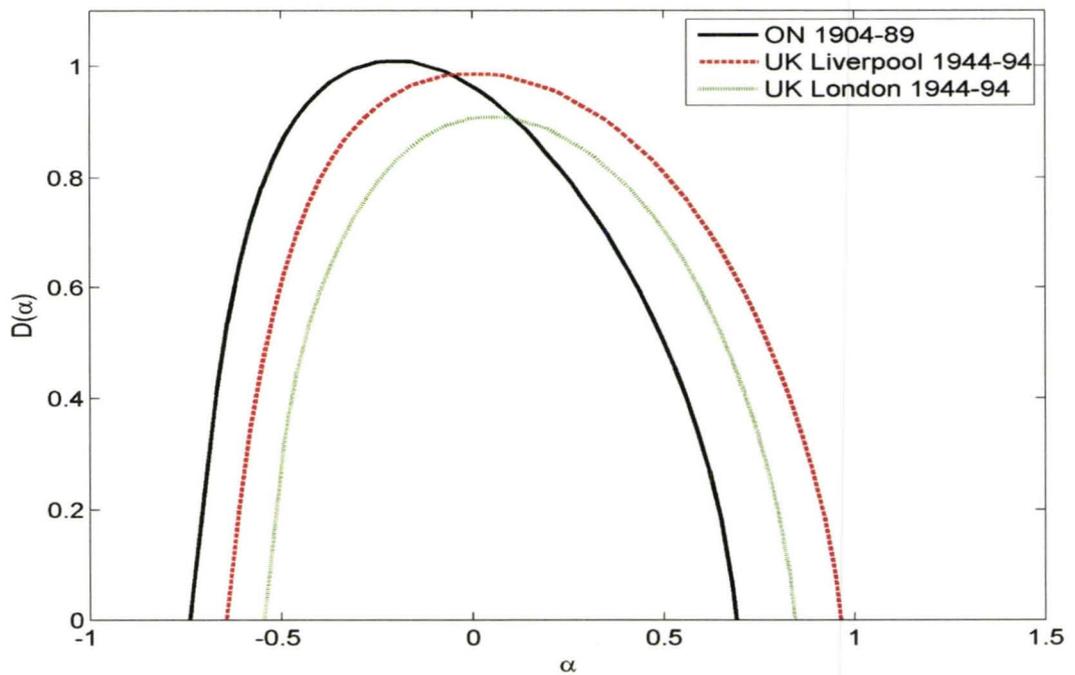
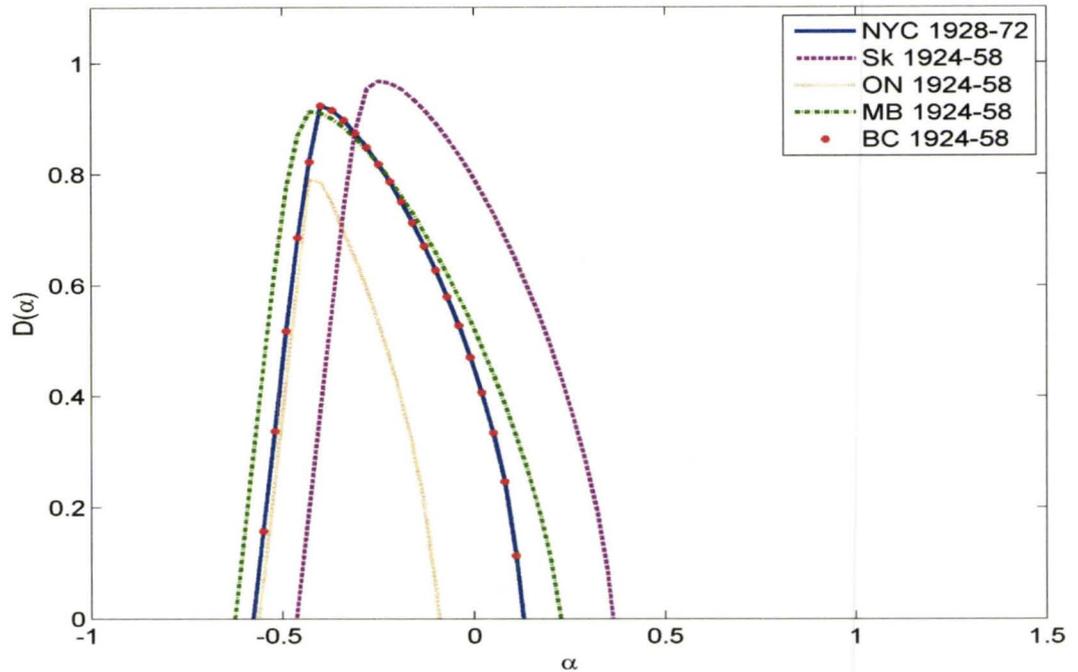


Figure 4.13: Top: Singularity spectra for incidence of chicken pox. Bottom: Singularity spectra for incidence of measles. The similarities between the various Chicken pox data sets are contrasted with their differences from the measles data sets.

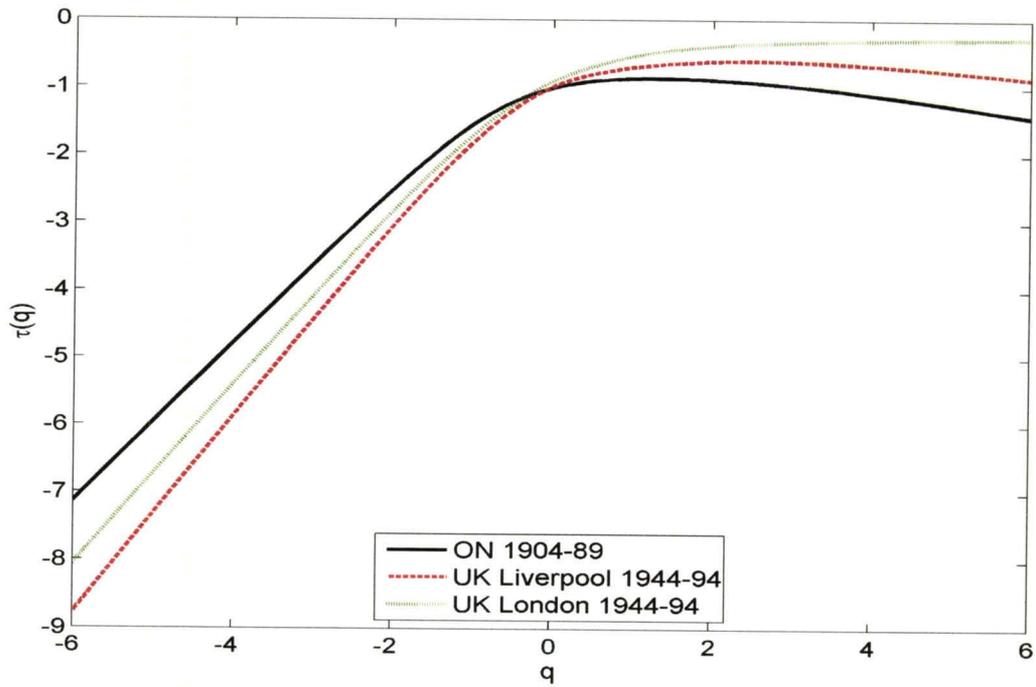
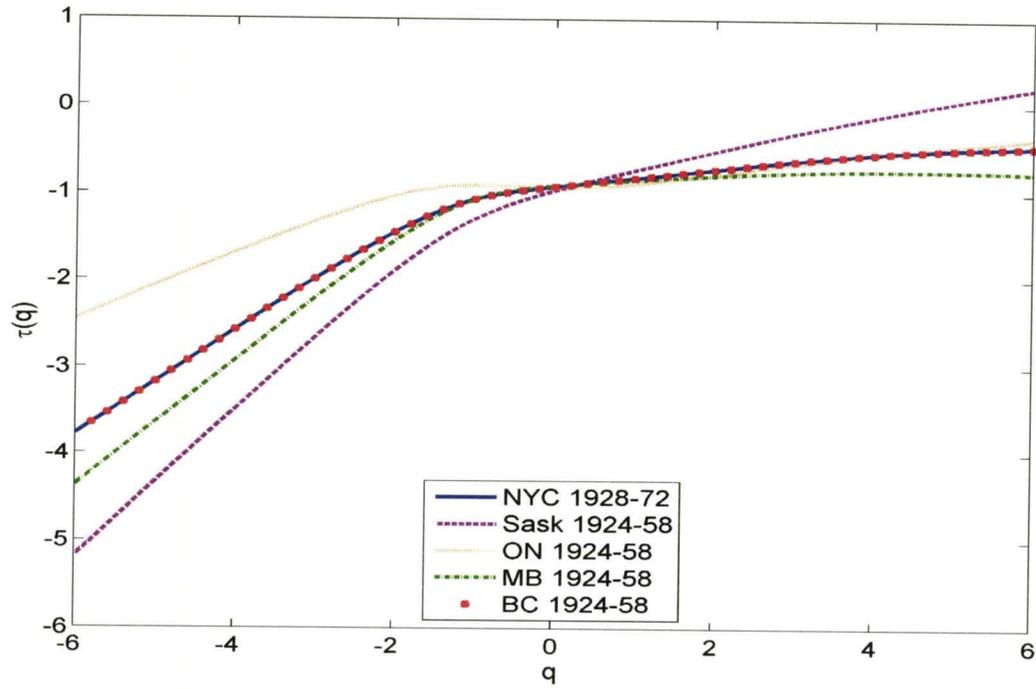


Figure 4.14: Left: Scaling exponent for chicken pox incidence. Right: Scaling exponent for measles incidence.

Chapter 5

Conclusions

The time has come to reflect on the goals initially set forth. Wavelet de-noising via thresholding was tested using power spectra and singularity spectra. The efficacy of stochastic models for predicting epidemics was tested using both wavelet power spectra and singularity spectra. We saw that wavelets provide qualitative comparisons of epidemiological time series. Similar behavior of the singularity spectra for several data sets corresponding to the same disease was observed. This is contrasted with the differences between the spectra for different diseases. These results have led to a new description of incidence of infection data using the language of multifractals.

Wavelet thresholding was applied, successfully, to both the stochastic and Ontario data. Soft thresholding removes a fine scale structure which can be characterized as Gaussian white noise. The application of wavelet thresholding, for the purpose of de-noising, does not significantly alter the singularity spectrum or the wavelet power spectrum. Based on the results given here, hard thresholding and soft thresholding are appropriate methods of signal recovery for epidemiology. Hard thresholding is a better choice for spectral analysis for its ability to recover the power spectrum.

Our analysis supports the hypothesis that wavelets offer new methods for qualitatively comparing and fitting models to data. By comparing power spectra one ensures that the patterns of recurrence in the model are consistent with observed recurrence. The singularity spectra provide a way of describing the autocorrelation of the data and the overall regularity.

It is possible to make some conclusions with regard to the stochastic models. The stochastic model using term-time forcing was found to improve upon the sinusoidally forced model in that the fundamental structure is more similar to actual data. Specifically, the distribution of singularities in the Ontario data is most similar to the term-time forced model's spectra. Neither stochastic model fits well with the actual data indicating the seasonal forcing parameter alone, is not enough to reproduce the singularity structure observed in the actual data. Even with a high degree of seasonal forcing, spectral peaks for the stochastic models' singularity spectra lie

much too far to the right. The behavior of the model data is continuous and not differentiable while the actual data is not continuous or differentiable. Still, the performance of these models is a clear improvement on the deterministic SEIR model which gives unrealistic, continuous and differentiable incidence data.

Wavelet analysis, although relatively new, offers a variety of tools and techniques for understanding, comparing, describing and manipulating time series. The application of wavelet spectral analysis, singularity analysis and wavelet de-noising to infectious disease incidence data was successful. Perhaps the most interesting result is the distinctive multifractal signature associated with each disease. What are the biological mechanisms that will help to explain these signatures? In the future, we would like to see this work extended in the following ways:

- **More sophisticated model parameters** may shift the spectrum to the left making the signal less smooth. Environmental stochasticity, the incorporation of immigration statistics, and a detailed account of age structure are all examples of suggested improvements to the stochastic SEIR model with demographics [5], [12]. Implementing some or all of the techniques mentioned may yield a closer singularity spectrum to that of the actual data.
- In this study, a single realization of the SEIR stochastic models was used for each β . One could extend the current approach by **averaging over several realizations** for comparison with the actual incidence data. Ensemble averaging was not performed here because, although each independent realization of the stochastic SEIR model does differ, the differences in the resulting singularity spectrum are quite small. By averaging over several realizations one can rest assured that the resulting data set is not an extreme case; a case where the data deviates significantly from the average realization.
- **Averaging the singularity spectra** for particular diseases is another interesting idea for future research. It was shown that different diseases have unique singularity spectra, or that they share unique multifractal signatures. These signatures may be averaged to establish a basis for comparison with the stochastic models.
- **Understanding the biological mechanisms** which cause the observed similarities in singularity structure for each disease. What are the differences between the diseases which may cause differences in their singularity spectra?
- **Studying the bifurcation structure** of the stochastic models as a function of the amplitude of seasonal forcing. This would help to clarify differences in singularity spectra observed for various amplitudes of seasonal forcing. Can the differences be explained in terms of chaotic dynamics?

The open questions described above all relate to the results of our analysis of epidemiological time series. Throughout the course of our investigation into wavelet analysis several issues were encountered which may be a source for future research:

- Hurst's work in relating the Hurst exponent to the correlation of FBM was used here to describe the time series. Literature in this area is somewhat scarce. An interesting question arises which will not be addressed here: what can be said about the correlation of time series with values of $\alpha \notin (0, 1)$ and is it valid to generalize the results for FBM to any given time series?
- It is known that the scaling exponent, $\tau(q)$ is linear for monofractals and non-linear for multifractals. For the Cantor function $\tau'(q)$ was found to be very similar to the hyperbolic tangent function. Do other multifractals have similar shapes? What more can be said, in general, about the overall shape of $\tau(q)$ and $\tau'(q)$?
- Comprehensive error analysis for the WTMM multifractal formalism is not readily available. In the previous chapter, error convergence for the Cantor function was examined. It would be interesting to see convergence results for a wide range of test functions.

This research should be looked at as the first step toward using wavelet analysis to understand epidemiological time series. Future work, testing more sophisticated models and improving interpretations of singularity spectra may become an area of active research. This new way of describing incidence data has potential. It is a small step toward predicting future epidemics and understanding the dynamics of infectious diseases.

Bibliography

- [1] Wavelabe850. www-stat.stanford.edu/~wavelab, 2007.
- [2] P. S. Addison. *Fractals and Chaos: An Illustrated Course*. Taylor and Francis, London, 1997.
- [3] A. Arneodo. Wavelet analysis of fractals: From the mathematical concepts to experimental reality. In *Wavelets: Theory and Applications*. Oxford University Press, New York, 1996.
- [4] Bacry, E., J. Muzy and A. Arneodo. Singularity spectrum of fractal signals from wavelet analysis: exact results. *Journ. of Statistical Physics*, 70, 1993.
- [5] C. T. Bauch. The role of mathematical models in explaining recurrent outbreaks of infectious childhood diseases. *Mathematical Epidemiology*, pages 297 – 319, 2008.
- [6] J. Kappey B.T. Grenfell, O. N. Bjornstad. Travelling waves and spacial hierarchies in measles epidemics. *Nature*, 414:716–723, 2001.
- [7] R. R. Coifman and D. L. Donoho. Translation-invariant de-noising. *Lecture notes in statistics-New York- Springer Verlag*, 1(103):125, 1995.
- [8] I. Daubechies. Where do wavelets come from: A personal point-of-view. *Proceedings of IEEE*, 84(4):510–513, April 1996.
- [9] Department of Mathematics David Earn and McMaster University Statistics. International infectious disease data archive, February 01 2008.
- [10] Université de Montréal (Canada) Department of Demography. Canadian human mortality database, February 01 2008.
- [11] David L. Donoho. De-noising by soft-thresholding. *IEEE Transactions on Information Theory*, 41(3):613–627, 1995.
- [12] David Earn. Mathematical epidemiology of infectious diseases. *IAS/Park City Mathematics Series*, XV, 2005.

- [13] Marie Farge. Wavelet transforms and their application to turbulence. *Annual Review of Fluid Mechanics*, 24:395–457, 1992.
- [14] J. Feder. *Fractals*. Plenum, New York, 1988.
- [15] Daniel T. Gillespie. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *J. Chem. Physics*, 22:403–434, 1976.
- [16] Amara Graps. An introduction to wavelets. *IEEE Computational Science & Engineering*, 2(2):50–61, Summer 1995.
- [17] A. Grossman and J. Morlet. Decomposition of Hardy functions into square integrable wavelets of constant shape. *SIAM J. Math. Anal.*, 15(4):723–736, 1984.
- [18] F. Herrmann. Scale analysis by the continuous wavelet transform, 2002. [Online; accessed 10-Dec-2007].
- [19] B. B. Hubbard. *The World According to Wavelets: The Story of a Mathematical Technique in the Making*. A.K. Peters, 1998.
- [20] S. Jaffard. Multifractal formalism for functions. part II: Self-similar functions. *SIAM Journal on Mathematical Analysis*, 28(4):971–998, July 1997.
- [21] B. Lashermes, S. Jaffard, and P. Abry. Wavelet leader based multifractal analysis. *Acoustics, Speech, and Signal Processing, 2005. Proceedings. (ICASSP '05). IEEE International Conference on*, 4:iv/161–iv/164 Vol. 4, 18-23 March 2005.
- [22] S. Mallat. A theory of multiresolution signal decomposition: the wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11:674–693, 1989.
- [23] S. Mallat. *A Wavelet Tour of Signal Processing*. Academic Press, 1998.
- [24] B. B. Mandelbrot. *The Fractal Geometry of Nature*. Freeman, 1968.
- [25] Mani Mehra and Nicholas K.-R. Kevlahan. An adaptive wavelet collocation method for the solution of partial differential equations on the sphere. To appear in *J. Comput. Phys.*, 2008.
- [26] Numerical Recipes Software. *Numerical Recipes Fortran*. Cambridge University Press, pub-CUP:adr, January 1986.
- [27] Valérie Perrier, Thierry Philipovitch, and Claude Basdevant. Wavelet spectra compared to Fourier spectra. *Journal of Mathematical Physics*, 36(3):1506–1519, 1995.

- [28] W. Sweldens. The lifting scheme: A new philosophy in biorthogonal wavelet constructions. In Andrew F. Laine, Michael A. Unser, and Mladen V. Wickerhauser, editors, *Wavelet applications in signal and image processing III*, volume 2569 of *Proceedings of SPIE*, pages 68–79, 1995.
- [29] W. Sweldens. The lifting scheme: A construction of second generation wavelets. *SIAM J. Math. Anal.*, 29(2):511–546, 1997.
- [30] Wim Sweldens. Wavelets and the lifting scheme: A 5 minute tour, December 05 1995.
- [31] C. Torrence and G. P. Compo. A practical guide to wavelet analysis. *Bulletin of the American Meteorological Society*, 4(3):415–447, 1998.
- [32] Bradley G. Wagner and David J. D. Earn. Circulating vaccine derived polio viruses and their impact on global polio eradication. preprint submitted to *Bulletin of Math Biology*, June 2007.
- [33] G. Wallace. JPEG still picture compression standard. *Communications of the ACM*, 34(4), 1991.

