Threshold Estimation Using Wavelets and Curvelets on Ground Penetrating Radar Data for Noise and Clutter Suppression

by

Dustin Harrison

B.S., South Dakota School of Mines and Technology, 2000

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

Master of Applied Science

in

THE FACULTY OF GRADUATE STUDIES

(Electrical and Computer Engineering)

The University of British Columbia

May 2005

Abstract

This thesis provides an evaluation of the Redundant Discrete Wavelet Transform with application to the removal of additive white or colored Gaussian noise on a synthetic GPR signal. Special attention is given to the parameter that controls the number of decomposition levels. Evaluation is performed using a level-dependent threshold to estimate and remove noise. Results are presented using noisy synthetic Ground Penetrating Radar pulses to compare Wiener filtering and thresholding the Redundant and Non-redundant Discrete Wavelet transform. Additional results are presented on the effects of choosing a number of decomposition levels using signal-to-noise ratio measurements, which suggest the importance of choosing this parameter. Recommendations are made and supported which determine the order of thresholding before or after the practice of trace averaging.

Using GPR images, an application of a novel 2D threshold model in the newly discovered curvelet domain is compared to average trace subtraction. Promising results are presented on both synthetic and actual landmine data, which shows thresholding as a viable method of clutter suppression.

Contents

List of Tables

List of Figures

List of Abbreviations

AGCN	Additive Gaussian Colored Noise
AGWN	Additive Gaussian White Noise
ATS	Average Trace Subtraction
DSP	Digital Signal Processing
DWT	Discrete Wavelet Transform
FWHM	Full Width Half Maximum
GCN	Gaussian Colored Noise
GPR	Ground Penetrating Radar
GWN	Gaussian White Noise
HumDem	Humanitarian Demining
JRC	Joint Research Centre
PSNR	Peak Signal-to-Noise Ratio
RDWT	Redundant Discrete Wavelet Transform
SNR	Signal-to-Noise Ratio
UWB	Ultra-wideband

Acknowledgements

- Dr. M. R. Ito provided me with funding and academic advisement which facilitated my completion.
- Dr. Felix Herrmann teaches an inspiring course on wavelets, which motivated this research and his great knowledge of the topic gave me the techniques and insight to finish.
- The University of British Columbia also saw enough potential in my determination and skills to provide additional funding for this project.
- Emmanuel Candes, Laurent Demanet, David Donoho, Lexing Ying deserve gracious thanks for making CurveLab available for research.
- Joint Research Centre for having the brilliant foresight to facilitate humanitarian landmine detection research by the creation of the JRC Landmines Signature Database [11].

Without this generous support I would not have been able to complete such an arduous task.

DUSTIN HARRISON

The University of British Columbia
May 2005

To my wife, DoraLi

for her patience and love

and my sister, Alissa

for showing me I can do it

Chapter 1

Introduction

There are millions of landmines worldwide, which are difficult to find and remove. The concept of *humanitarian demining*, which is sometimes referred to as HumDem, is to provide simple, sustainable and affordable method of detection and removal of landmines. The groups that implement these programs are often non-governmental organizations who did not place the mines and generally have little information on mine type or placement locations. The most common and highly visible aspect of demining are the activities that in poor countries devastated by war. The objective of HumDem groups in these countries are to develop programs within that country that allows present and future landmines to be found and destroyed even after the HumDem group has left.

The *sustainability* objective of many HumDem programs is critical to implementing a method of landmine clearance that can be performed by locally trained demining groups. Previous estimations state that 100 million or more landmines are currently laid in the world, however that number has been reduced to 60 million or less [61]. The number of landmines, danger of detection and the removal procedures increase the cost and time requirements to nearly impossible levels for poor

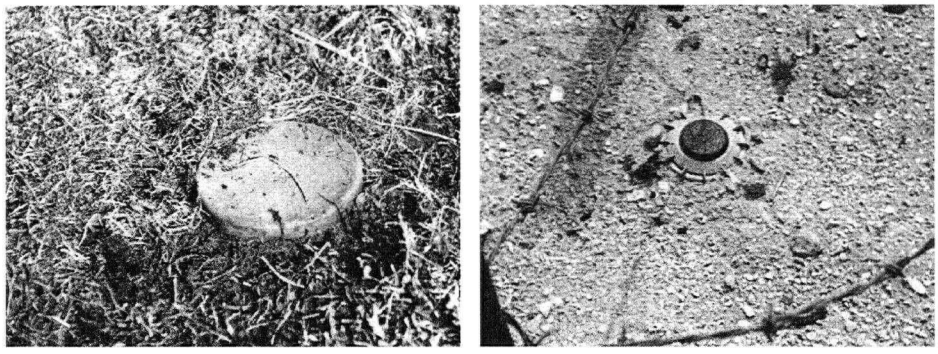
countries. The cost of detection and removal is also related to the amount of time a landmine is in the ground. Delays in landmine clearance causes vegetation coverage, which increases the challenge to detect minefields and mine locations. Due to the large amount of money and time required for a successful HumDem project, new technology is not well accepted. Therefore making simple, affordable and mature technology is necessary so that projects can continue even after external funding has ceased.

HumDem programs determine suitability of technology by the cost and status of development maturity. The most readily used and implemented technology is human protection devices, metal detectors and trained dogs. Other technologies that have been tested or evaluated with respect to HumDem are:

- Explosives detection using chemical sensors, biological sensors and nuclear quadrupole resonance
- Optical sensors such as infrared, hyper-spectral, visible and laser ranging
- Electromagnetic sensors including ultra-low frequency, microwave, SAR and ground penetrating radar
- Acoustic sensors such as water-jet echo [50], impulse and ultrasound

Bruschini and Gros discuss in detail most of the above technologies in [5]. In particular, the authors suggest that GPR is a mature technology and that it is "near ready" for use in HumDem programs.

The HumDem field is not receptive to new technologies, regardless of the promises of detection improvements [3]. For example, estimates for Cambodia suggest that mine clearance is happening at a rate of 15 square kilometers per year with the possibility of demining all land that is immediately in need for settlement,



(a) PMN Landmine (b) VS-50 Landmine

Figure 1.1: Anti-personnel landmines *in situ*, obtained from the Canadian Forces Landmine Action Database [20].

critical development or agricultural uses taking place in 5-10 years [2, 51]. Technology proponents suggest that these numbers could be drastically reduced, but the reality is that field tested technology is a difficult and slow process. Specifically in Cambodia, heat and moisture can dramatically reduce the lifespan or functionality of any electronics making new and untested technology dangerous to deminers.

Not all applications of demining take place in such demanding environments. The declining cost and size of GPR components, such as analog/digital converters, amplifiers and high speed switches contribute to the increasing interest to implementing GPR as a landmine detection device. Additionally, technology improvements in computer processors and specialized DSP chips allow additional realtime and field-based signal processing of GPR surveys.

While the processing of GPR is maturing, the limited resolution of GPR systems make the task of landmine detection difficult and ambiguous. When using GPR for HumDem tasks, the standard method of expert interpretation is generally not available. Experts are too expensive to train and support in these situations. Offline processing of GPR data is an option, which can allow computationally diffi-

cult algorithms such as imaging and feature discrimination to be done off the field. Offline processing, however, requires a thorough understanding of noise present in GPR data.

Two aspects of GPR data are examined in this thesis. The first is the ability of 1D wavelet thresholding to remove Gaussian noise from raw GPR data. The second is to suppress horizontal clutter present in GPR data using thresholding with a new transform, similar to wavelets, called curvelets. Both methods attempt to separate signal from noise by means of thresholding.

The received data from GPR is a one-dimensional trace that contains reflections of a transmitted pulse. To study this, a GPR trace is synthesized from a reasonable physical model of a GPR pulse. The type of noise and amount of noise energy added is controlled allowing numerical measurements of wavelet thresholding performance.

A separate case, in two-dimensions, applies to an image of received GPR traces. This image contains strong horizontal clutter events, which detract from image understanding. The effect of thresholding clutter events in the curvelet domain is evaluated using an image formed from synthetic GPR traces and actual GPR data from the JRC Landmine Signatures database.

1.1 Research Objectives

While 1D wavelet noise removal using thresholding is well established, the application to GPR is minimal. The type of wavelet decomposition, specific parameters and the choice of a wavelet basis are difficult to understand and is not well established. This research examines the most promising wavelet decomposition method for noise removal, which is the Redundant Discrete Wavelet Transform (RDWT).

The hypothesis is that the method of level-constant thresholding and the number of decomposition levels used by the RDWT will have a direct effect on the ability to remove additive Gaussian white and colored noise.

The second objective is to compare the ubiquitous method of subtracting an average trace from a GPR image to a proposed method that uses a threshold in the curvelet domain as a method of subtraction. This new method allows a simple noise model, like the average trace subtraction method, but without the drawbacks of global averages and subtraction. The hypothesis is that the lack of resilience of subtraction requires clutter models that are too complex, and that thresholding is more effective at suppressing horizontal clutter using a simple noise model. Additionally, both approaches to GPR processing take into account that noise removal and clutter suppression should be implemented with fast algorithms. The advantage of wavelets are that most implementations are reasonably fast. The RDWT can be implemented very easily in modern DSP chips, making it a candidate for on-line processing.

1.2 Thesis Overview

This thesis first introduces the basic mechanics of GPR technology and the unique obstacles along with a selection of signal processing techniques that have been applied to GPR data. A brief wavelet introduction precedes the two wavelet algorithms used in the methodology of this research. The first method, known as the redundant discrete wavelet transform has been demonstrated as a successful transform where noise can removed by the application of a simple threshold in the wavelet domain. The following section introduces a 2D "wavelet" transform, known as curvelets, which is a recent development in multi-resolution basis function decomposition and

similar to, but not strictly, a wavelet transform. The transform properties and construction is described after a breif overview. The remainder of Chapter 2 introduces the threshold operator that is used to perform noise removal.

The methodology is presented in Chapters 3 and 4, for A-scan noise removal and B-scan clutter suppression, respectively. Chapter 3 introduces a method for choosing bases, the parameter that controls the number of decomposition levels in the RDWT and threshold estimation. A description of the GPR trace model explains how a physical approximation of a GPR signal is made. At this point, the noise models is introduced, first additive Gaussian white noise or AGWN, and then a filtered form of AGWN, denoted as additive Gaussian colored noise (AGCN).

The clutter suppression methodology builds on the previous material, using the synthetic GPR pulse model to form a synthetic B-scan. A brief introduction to the landmine data is made, before a novel method of clutter suppression, inspired from the success of thresholding 1D signals is introduced. The remainder of Chapter 4 describes the method of evaluation and performance of the new algorithm.

The results present a comparison between a proven wavelet denoising technique and the more successful RDWT method. Additionally, demonstrations show the difference in the optimal mean-square-error linear filtering technique and wavelets to remove either AGWN or AGCN. New results show the relationship of input noise energy and decomposition levels along with a demonstration of compatibility of the method to the current GPR processing method of stacking. Simulations of clutter suppression are presented for the synthetic GPR image data and followed by an experiment on actual GPR landmine data before the thesis concludes with a discussion of results.

Chapter 2

Background

The topic of this thesis is evaluating the estimation of a threshold with respect to ground penetrating radar. The concepts that are needed to perform this evaluation are a basic physical understanding of the GPR data and how wavelets can be applied. The motivation is clear: additional insight into GPR processing should serve to enhance the field of landmine detection.

2.1 Ground Penetrating Radar

Radar operates on the principle that the range from the radar to the target can be determined by timing the round trip of a transmitted pulse reflected by a target. The assumption of radar is that the transmission medium is conducive to passing the frequency content of the pulse. The earth is not a homogeneous medium, therefore anticipating the frequency content to pass through the media of the earth is location dependent. Ground penetrating radar overcomes this by using a wide range of frequencies, which allows sub-surface targets to be "ranged".

A pulse transmission from an antenna at a single spatial location is received

over the duration of receiver gating. This resulting time-domain signal is called a trace, or A-scan. For GPR data to be meaningful, a GPR survey usually attempts to spatially sample a grid of points on the surface. The collection of traces from a GPR survey of several spatial locations forms a three-dimensional data set of two spatial coordinates and one of time called a *GPR data cube.*

The method of conducting a GPR survey varies on the type of GPR. A GPR may be bi-static, mono-static or neither. In mono-static mode, the same antenna transmits and receives the GPR pulse. The burden falls on the transceiver electronics to properly gate the transmit and receive times and protect the receiver from the large initial radiated pulse. In bi-static mode, separate antennas are used for transmit and receive, but their spatial relationship is held constant. The analog in seismic jargon for this method of operation is called a common-offset gather. Each trace represents a constant angle between the transmitter and receiver. In bi-static mode, antenna polarization relative to transmit and receive can improve target detectability [43]. Additionally, common mid-point gathers can be used to determine velocity models of the subsurface. In this survey, the angle between transmit and receive antenna is increased after each pulse transmission.

Many practical applications of GPR are being developed. High resolution GPR solutions are used to image the internal structure of roads and bridges for the appearance of cracks. Military and security applications have been suggested to image potential threats behind walls of caves and buildings [6]. Snow pack layers, internal ice structure and avalanche victim location have also been demonstrated using GPR [40, 42]. Geophysical scientists make extensive use of GPR for the study of near surface features [15, 58].

2.1.1 Principles of Operation

There are many variations of GPR systems, which are defined by the type of antenna and method of pulse generation. Impulse radar generates a pulse in the time domain and uses an antenna, which is non-dispersive. The other common GPR system creates a pulse in the frequency domain and may use either a dispersive antenna to transmit a "chirp" pulse or a non-dispersive antenna. This type of radar is called Frequency Modulated Continuous Wave (FMCW) or a popular variation in HumDem research is Stepped Frequency Continuous Wave (SFCW) [18, 53]. These radars usually fall into a broader category known as Ultra-Wideband (UWB) radar.

A UWB radar has a ratio of bandwidth to center frequency that exceeds one. This is called the *fractional bandwidth* and is defined by [62] as

$$B_F = 2\left(\frac{f_H - f_L}{f_H + f_L}\right). \tag{2.1}$$

A UWB impulse radar has several advantages, because it uses low average power and has a wide range of frequencies. This allows it to image the subsurface of a wide variety of materials. Research by Cherniakov [12] shows that as the fractional bandwith of a radar approached the theoretical limit of 2 it obtained maximal spectrum efficiency. Spectrum efficiency is defined as maximizing depth of penetration to enable detection of buried objects. It was also shown that a radar with $B_F \geq 1.1$ has a high spectrum efficiency, making UWB-GPR radars an excellent choice for depth of penetration when searching for underground events.

An impulse GPR in commercial applications commonly uses a Gaussian pulse to excite an antenna,

$$V(t) = e^{-t^2/2\sigma^2}. \tag{2.2}$$

This pulse can be created by using a transistor in avalanche mode, with results in

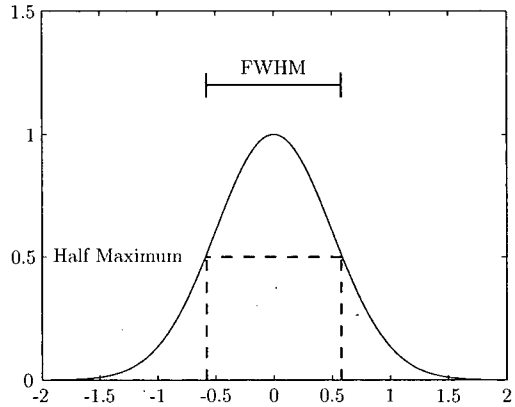


Figure 2.1: Pictorial representation of the FWHM parameter of a Gaussian bump.

an output that is an approximation of an exponential charge and discharge cycle. A typically GPR specification is the Full Width, Half Maximum (FWHM) of a pulse. The FWHM is a common measurement of the width of a gaussian pulse and is defined by

$$FWHM = 2\sigma\sqrt{2\log_e 2} \tag{2.3}$$

The width of a pulse indicates bandwidth, but usually the working bandwidth of a radar is limited by the antenna.

The response of an antenna to the input of (2.2) is to radiate a voltage wavefield which is the derivative with respect to time,

$$V'(t) = \frac{-t}{\sigma^2} e^{-t^2/2\sigma^2}. \tag{2.4}$$

An *event* is defined as the filtered form of (2.4) that occurs at some point in time of a received GPR trace. This event corresponds is the reflection of the transmitted pulse from a subsurface feature and can be related to depth if the velocity of the transmission media is known. In the case of a metallic reflector the event is the second derivative of (2.2) because a metallic reflector acts as a second transmitting antenna.

GPR antenna development is crucial to successful GPR applications. The antenna controls the interaction of the radiated waveform with the surface of the ground by impedance matching. The antenna also determines the frequency content of a radiated pulse, which dictates spectral efficiency thereby controlling depth of penetration and target visibility. HumDem researchers have developed an antenna that exhibits excellent GPR performance with respect to spectral efficiency and ground coupling [59]. Data from this laboratory GPR and antenna is used to test the algorithm presented in Chapter 4.

The characterization of an event in a GPR trace is determined by the timing and amplitude of the occurrence. Due to spherical spreading of the transmitted wavefront and attenuation of the subsurface, the amplitude of the received signal for a mono-static configuration is shown by [17] to be proportional to

$$\frac{1}{d^4 \sec^4 \theta} e^{-2\alpha d \sec \theta}. \tag{2.5}$$

The distance, d, is the line, in meters, normal with the top of the point reflector to the surface. As the GPR survey is performed, the angle θ increases, which further attenuates the signal. In practice, some commercial GPR systems attempt to compensate for this attenuation using an exponential gain applied during or after data collection [39]. While it is the prerogative of the user interpreting the data to apply a gain, it is important that noise removal be done beforehand to avoid non-linear changes to the data statistics.

The determination of timing in a homogeneous medium is based on the horizontal location, x, of the antenna relative to the target and the depth, d, of the target. For a point target, or a target size that is near the resolution limit of the radar, the round trip time is determined by twice the velocity of pulse propagation

times the line of sight distance to the target

$$t = \frac{2}{v}\sqrt{x^2 + d^2}. \tag{2.6}$$

The velocity of propagation is determined by the speed of light scaled by the relative permittivity of the ground, ϵ

$$v = \frac{c}{\epsilon}. \tag{2.7}$$

If the subsurface is a combination of media, the determination of velocity can become intractable. Additionally, the influence of water on ϵ is dramatic, capable of increasing the permittivity by an order of magnitude.

2.1.2 Data formatting and visualization

In order to process GPR data, there are three common formats for presentation of GPR data. Each format chooses a set of dimensions from the 3D cube of GPR data. The most common format to perform image processing and understanding is a B-scan. B-scans are are two dimensional images with time on the vertical axis and a spatial dimension on the horizontal. Most often the spatial dimension is the "down-track", which is the direction that the antenna traveled as each trace is collected. Specific spatial coordinates in a B-Scan represent an A-scan. An A-scan is a time-domain waveform that is measured from the voltage on the antenna after an initial pulse is transmitted into the ground. Typically an A-scan is measured in milli-Volts, but may also be presented in a normalized amplitude. The amount of time measured in an A-scan varies with the GPR system in use, and depends on the center frequency of the system. A low frequency GPR can typically penetrate further into the ground and requires a longer time to allow the pulse to reflect back to the antenna.

Due to the nature of a GPR pulse containing multiple events, an A-scan is described as "wiggly" and has many smooth oscillations. This descriptive of A-scans also refers to a method of viewing B-scans called a *wiggle plot*, which is actual A-scans vertically oriented and plotted close together. This method is used most often by human operators to make interpretations of GPR data. In Figure 2.2, a pixel intensity B-scan image is presented, where each pixel color is scaled based on the amplitude in the A-scan.

Occasionally C-scans are used in image processing. A C-scan is a "depth" slice, which takes the A-scan values over all spatial locations to form a intensity image. The C-scan contains both spatial dimensions for axes, which are the down-track and cross-track. The cross-track dimension is the spatial direction that the antenna is moved after a down-track survey is completed. A C-scan is a subset of a B-scan, because it is generated from the set of all B-scans. Each B-scan is, in turn, a subset of A-scans.

2.1.3 Sources of Noise

There are many noise sources present in a GPR device. During the creation of the initial source waveform, noise may be added by interference from external sources or poor high-frequency digital design of electronics. After a pulse is transmitted, cross-coupling from the antennas may generate spurious, large amplitude events prior to any reflection events. In addition, upon reception of a GPR signal, clock jitter, analog-to-digital quantization noise and amplifier noise determine a noise floor and the operating dynamic range.

At the resolution of GPR, the majority of signal energy comes from large impedance changes in the ground. Changes in soil water content will dramatically

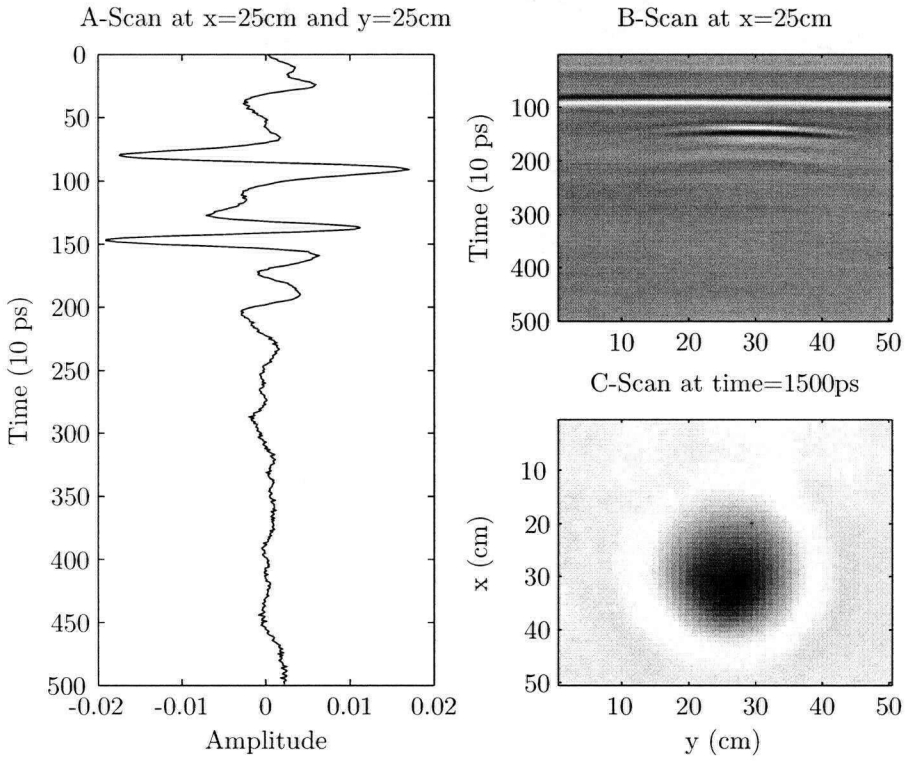


Figure 2.2: Selected Examples of different scan types. Each contains different perspectives of a GPR data cube.

attenuate the signal amplitude as it propagates into the ground. Additionally, if the subsurface contains large amounts of near wavelength objects, like glacial till, depth of penetration and target localization will also decrease. Collectively, events that are due to actual reflections of the source waveform, but reduce target signal reflections are referred to as clutter.

Positional Noise

It is important to consider that the antenna is a physical object that is never precisely positioned. In the lateral sense, the antenna position is critical if any transforms are taken on the data in the X-Y dimension. Both the wavelet and Fourier transforms expect data to be sampled on a fixed grid. In order to correct this noise, antenna position must be recorded and the data interpolated onto a fixed grid. In the application of landmine detection this becomes a problem with handheld detectors, which can be swung in arcs or irregular lines [32]. Most laboratory solutions use robots or antenna jigs to stabilize and accurately determine spatial location.

In addition to lateral movement, it is important to consider vertical and angular antenna position. Both angular and vertical positioning affect the coupling of energy into the ground. As the coupling changes, the amplitude of the received signal varies. The result is that algorithms must adapt to an unknown variance in amplitude. Additionally, vertical position uncertainty creates temporal variance of the received signal. Usually this noise is minimized by allowing the antenna to travel flush to the ground, but as ground changes occur, the hyperbolic assumption of a target response can be invalidated. Often this issue is lessened by the duration of scans, which are taken over contours that can be considered flat.

System Noise

The most well understood system noise is thermal noise. Thermal noise can be modeled by additive Gaussian noise. The noise is the cumulative effect of the antenna gain, amplifier noise figure and the transmission line. It can be quickly reduced by *trace averaging* or *stacking* of multiple traces from the same spatial location. Commercial GPR systems implement this in hardware and will provide the user with the ability to average up to 32 traces.

Other system noise, which is less understood and might not be well modeled by Gaussian noise comes from electronic design and susceptibility. For example, poor circuit board design of a transmitter can add noise to the generated pulse. Reception of external signals may also corrupt the signal. The quantization noise of the analog-to-digital converters also limits the dynamic range of the system. The combination of all sources of noise is very likely Gaussian, but not necessarily white.

Clutter noise

Clutter is defined by the user and strictly speaking is not noise. Clutter is the cumulative effects of reflections from actual boundaries that are illuminated by the GPR, but impede with analysis of the primary targets under study. The most common form of clutter in any GPR application is the ground. The occurrence of a reflection from the ground offers very little information to a GPR image and may interfere with near-surface objects. Landmines are an example of a near surface target that can be impeded by the presence of an air-ground event.

Other forms of clutter, with respect to landmines, are targets that have properties very close to the desired target and can result in a false positive determination of a primary target. These false positives can be metal fragments, stones and other

debris that produce events similar to a landmine event in either amplitude, frequency content and shape or a combination of the three. Generally, this form of clutter is filtered using discrimination algorithms. A discrimination algorithm will use some unique feature of the desired target to suppress all other possible targets that do not contain any such features.

Another form of clutter, sometimes called background noise, appears as the portion of a signal, which does not necessarily represent a physical object, but appears as a slowly varying horizontal event in a B-scan. The most common form is multiple reflection that occurs between the ground and antenna, often exacerbated by antenna ringing. Subsurface planar reflectors are sometimes grouped with background noise, because of the typically large horizontal component. The terms background noise and clutter are used interchangeably in the literature and are problem specific.

Clutter noise models are often considered additive, such that the subsurface can be represented by a train of dirac functions, each of which represent either target or clutter temporal locations. The GPR signal is then convolved with the dirac train to obtain a simulated received trace. Clutter is normally the limiting noise in GPR systems [18]. Besides target depth, clutter is the limiting factor that determines the detectability of a subsurface target.

2.2 Selected Review of GPR Signal Processing

All GPR data is processed, whether by human or machine. In a simple case of utility finding, it is usually enough for a human to interpret utility location based on the existence of a hyperbola in an unprocessed B-scan. In the case for humanitarian demining, it is usually necessary to perform additional machine-based processing

and discrimination to perform accurate and reliable landmine detection.

GPR processing can be divided into three categories, restoration, transformation and discrimination. Restoration is concerned with the removal of noise, correction of system error or correction of the sensor degradation. Transformation is used as a functional tool to perform signal processing, but in the case of migration is also an output. Discrimination is a key aspect to landmine detection. Discrimination uses a feature, hopefully unique, to separate noise and clutter from targets. For discrimination clutter often refers to anything that is not a desired target. This definition varies depending on the features used to discriminate.

A typical signal processing chain for landmine detection may include several restorative techniques, possibly a transformation and always a discrimination step. The importance of developing an excellent understanding and usefulness of restoration tools is paramount to successful landmine detection.

2.2.1 Review of A-scan Processing Methods

When processing a GPR A-scan, most methods focus on signal restoration. Restorative methods include, DC-offset removal, Gaussian noise filtering and feature enhancement. With the exception of DC-offset removal, which is sometimes performed by the equipment, most restorative techniques are chosen based on the remaining operations of the signal processing chain.

Feature enhancement is usually done for the benefit of an expert interpreter. These techniques are often times carried out at the discretion of expert, where the various outputs are visually compared. For A-scans, these methods are frequency filtering, averaging and gain adjustment. Frequency filtering consists of applying a low, high or band-pass filter to select certain features. As a GPR signal travels into

the earth it is dominated by the time-variant low pass effects, which reduces the high-frequency content of late time events. High pass filtering is performed when it is necessary to *dewow* an A-scan. A "wow" signal is the low frequency effects of transmitter and receiver coupling, which produces a slow decaying signal into the A-scan [39].

Noise filtering is not a standardized process, nor is there a correct choice for all applications. The only ubiquitous noise removal that is performed on A-scans is *stacking*. Stacking, or in GPR jargon, trace averaging, is the average of repetitive scans from the same spatial location. Commercial GPR systems often include this form of noise reduction as a primary parameter choice when performing a GPR survey. The purpose of stacking is to remove system noise.

Other methods include Wiener filtering, which assumes the noise model is well understood and attempts to perform an inverse noise filter in the frequency domain. This method is also called minimum square error filtering, because it seeks to minimize the error measure, $\mathsf{E}[f - \tilde{f}]^2$. This filter seeks to invert the degradation response, $H(\omega)$ in the frequency domain using [28]

$$\tilde{F}(\omega) = \left[\frac{1}{H(\omega)} \frac{|H(\omega)|^2}{|H(\omega)|^2 + K}\right] G(\omega). \tag{2.8}$$

The parameter K, assumes additive white Gaussian noise and can be estimated from the variance of the noise. The result is to suppress the noise and degradation in the corrupted signal, $G(\omega)$ and produce an estimate of the signal, f. Wiener filtering has been applied for different special cases, which are described in [17].

Discrimination can also be performed on A-scans using parametric target models, frequency features, or time features. Discrimination using time features is done with matched filtering and is usually not well suited to GPR, because of the time variant filtering of the subsurface. Frequency features can be discriminated by

using the unique spectral content of a target to suppress other events [37]. McClure and Carin explored the idea of compactly representing GPR waveforms using a matching pursuits algorithm in [52]. This method attempted to find compact atoms that represented fundamental structures of radiated and reflected electromagnetic waveforms with success in approximating time and frequency parameters of GPR signals in noise, which proved useful for discrimination.

2.2.2 Review of B-scan Processing Methods

The majority of B-scan processing is focused with clutter suppression and discrimination. Clutter suppression is a difficult and subjective topic that requires careful definition of what is clutter. Horizontal clutter is well studied, with current research showing moderate success at removing the air-ground event and other persistent horizontal events.

A common form of horizontal clutter reduction is to use average trace subtraction (ATS). Given a collection of traces located spatially at points in x and y. ATS uses an average trace for a particular B-scan located at position y,

$$\bar{f}_y = \frac{1}{N} \sum_{x=1}^{N} f_{xy} \tag{2.9}$$

to perform a subtraction for a given B-scan,

$$\tilde{f}_{x,y} = f_{x,y} - \bar{f}_y. \tag{2.10}$$

This is effective in enhancing the hyperbolas or any non-horizontal event in GPR data. The drawback is that all horizontal events are removed or attenuated, therefore if a target contains horizontal information ATS is not suitable. Other examples focus only on the air-ground event, such as Wu [68], where the air-ground event is

adaptively estimated using a shifted and scaled reference signal to perform subtraction.

The nature of subtraction is that ATS is not robust to phase changes or amplitude changes. If a clutter event decays to zero, ATS will continue to use the same static amplitude calculated from the average to subtract. For computational critical applications, it can be a useful method and is used in some discrimination algorithms in [69]. Gader *et. al* uses a variation that calculates the derivative in x instead of the average, which is used in the signal processing chain previous to discrimination and is highly sensitive to noise [25, 26].

The major problem of ATS is the instability when subtracting coherent noise. This can lead to artifacts that can decrease the usability of a data set. Ulrych *et. al* studied this method of clutter suppression with respect to other competing methods. In particular, the α-trimmed mean was suggested as a way to handle variable clutter, but at the additional cost of careful expert decisions to choose α. The median was also studied, which led to [66], a study on the robustness of L-moments as estimators. New research was presented on the possibility of using a James-Stein estimate for noise and clutter suppression on a B-scan. This method uses a shrinkage parameter to adjust an local observation based on the global mean [65].

After B-scan processing, GPR data is often times used to form a 3D image. This is referred to as *migration* and is the process of focusing a B-scan or GPR data cube into a coherent image. This technique is sometimes performed in landmine detection, but due to computational complexity and resolution it is not popular. It does present the most understandable output for human understanding, because the hyperbolas are collapsed and the time axis is recast to depth [71, 46, 27, 43].

2.3 Review of Wavelet Transforms

Wavelets and Wavelet transforms are an exciting topic of research, which was inspired by the collaboration of Morlet and Grossmann [30]. The principle of wavelets is that a signal can be represented by its frequency and temporal content. The relationship between time and frequency, necessarily, is a tradeoff of minimal support in one domain demanding maximal support in the other. In the world of signals, transient events present a challenge to estimation and compactly representation [48]. If a time series contains a mixture of signals with varying amount of support in both domains, there is a fundamental limit to the determination of a precise time location and frequency content for each component. The elegance of wavelets is an efficient representation that attempts to balance the inherent uncertainty between time and frequency.

Wavelets are often introduced as a continuous transform, which underlies the fact that each discrete representation of wavelets makes a unique sacrifice to obtain efficient algorithms and signal representations. The continuous wavelet transform, qualitatively, is the partitioning of a signal into time and frequency atoms by the application of translated bandpass filters. These atoms, ideally, attempt to maximize frequency and time localization.

Using the notation of Mallat [48], a wavelet has a zero average such that

$$\int_{-\infty}^{+\infty} \psi(t)dt = 0. \tag{2.11}$$

To obtain localization in time, the wavelet is translated by u and scaled by s,

$$\psi_{u,s}(t) = \frac{1}{\sqrt{s}}\psi\left(\frac{t-u}{s}\right). \tag{2.12}$$

The scale parameter is usually greater than zero. The wavelet transform can now

be defined as the convolution of the wavelet with the function under study, or

$$W_{f(u,s)} = \int_{-\infty}^{+\infty} f(t) \frac{1}{\sqrt{s}} \psi^* \left(\frac{t-u}{s} \right) dt. \tag{2.13}$$

The sampled continuous wavelet transform should not be confused with the discrete wavelet transform. Creative solutions to use the sampled continuous wavelet transform attempt to characterize the type of discontinuity. Mallat used the wavelet modulus maxima to estimate the Lipschitz exponent to classify a singularity [47]. Hsung extended this work by approximating signals without directly identifying the singularity [38]. The author uses a modified form of wavelet modulus maxima detection to estimate an inter-scale ratio and inter-scale difference of the sampled continuous wavelet transform coefficients to estimate the signal in noise.

A wavelet transform can be described as a *multi-resolution decomposition* of a function. Resolution may not be precisely correct in the continuous domain, but is certainly applicable in the discrete. When the continuous wavelet transform is discretized, the scale parameter defines the time/frequency resolution of the corresponding decomposition. This discrete multi-resolution decomposition will allow efficient signal processing to approximate, compress and estimate signals.

2.3.1 Discrete Wavelet Transform

The orthogonal discrete wavelet transform is a well studied implementation of wavelets in the discrete domain. It became popular amongst applied scientists when a fast filter bank approach was introduced. Alongside the fast computation and the ability of the DWT to compactly represent mixed signal functions, researchers were also able to derive many analytical results that suggested profound implications not accessible before.

Amongst these discoveries was the work that influenced the field of signal processing by Donoho that suggested non-linear thresholding could achieve near optimal signal estimation in white noise [23]. The substance of that claim relies on the orthogonality of the discrete wavelet transform.

The continuous wavelet transform is far from orthogonal and offers no immediate discrete solutions. To solve this problem the discrete wavelet transform makes two dramatic changes to the choice of the scale and translation parameters s and u. The first change is to choose s by choosing the number of scales dyadically. The new discrete parameter, j is

$$s = 2^j, \quad j \in \mathbb{Z} \tag{2.14}$$

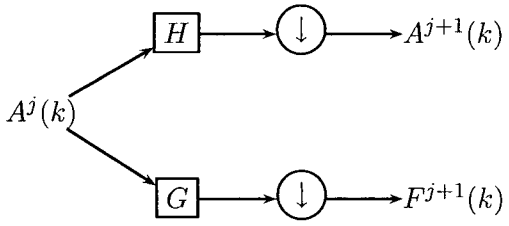
and the discrete translation parameter, k is dependent on the scale

$$u = 2^j k, \quad k \in \mathbb{Z}. \tag{2.15}$$

This choice of discretization of the continuous parameters is motivated by a fast implementation algorithm, but solidly rooted in sampling and information theory. Intuitively, the dyadic partition of the scale parameter in the DWT allows a higher localization in low frequency, where time localization is necessarily minimal. The dependence of k on the discrete scale parameter, j balances this uncertainty at higher frequencies when signals become highly localized in time.

This choice of scaling and translation forms a set of *wavelet basis functions.* The dyadic choice of j and k results in N basis functions, which is equal to the length of the signal $f[n]$. The index, m is related to each translation and dilation of the discrete wavelet basis to form $g[m]$. The discrete wavelet transform can be represented by

$$F[m] = \langle f, g_m \rangle, \tag{2.16}$$



Intialize the iteration with:

$$A^0(k) = f(n)$$
$$H = \text{Scaling Filter}$$
$$G = \text{Wavelet Filter}$$

The wavelet coefficients are F^j for $j = 1, \ldots, J$ and the scaling coefficients are $F^{J+1}(k) = A^J(n)$.

Figure 2.3: The DWT is implemented by an iterative application of high pass and low pass filters that are constructed by the wavelet and scaling function.

where g_m is an orthogonal set of basis functions,

$$\langle g_m, g_n \rangle = C_{mn}\delta_{mn}. \tag{2.17}$$

Often, wavelet bases are chosen such that $C_{mn} = 1$, which makes a basis orthonormal. However, the DWT is better understood as an implementation of cascaded digital filters, which is shown in Figure 2.3.

The limitation of the DWT is the lack of shift invariance. For an intuitive understanding, consider the second coarsest scale, which only contains two coefficients that represent the signal frequency and temporal content at this particular scale. These two coefficients are sufficient for perfect reconstruction of the signal using the DWT, but if approximations or estimations in noise are made the lack of shift invariance becomes evident. Signal estimations can be performed by thresh-

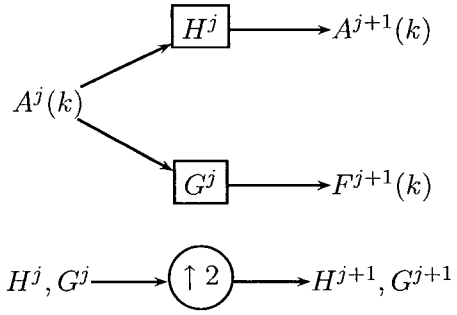
olding which chooses a particular coefficient based on the coefficients amplitude. Shift invariance implies that a shifted version of the signal under study will produce a shift in the wavelet coefficients, while not significantly changing the amplitude of the coefficients. Without shift invariance, the success of thresholding the DWT coefficients to remove noise will depend on the signal's location in time.

2.3.2 Redundant Discrete Wavelet Transform

The Redundant Discrete Wavelet Transform (RDWT) is a fast, convenient method of obtaining shift invariance. The term RDWT is ambiguous, because the nature of obtaining coefficient redundancy is not unique. A literature review suggests that the undecimated version of the discrete dyadic wavelet transform is the RDWT. This is equivalent to other names, such as the Shift-Invariant Discrete Wavelet Transform (SIDTW) [45], Undecimated Wavelet Transform (UDWT), Stationary Wavelet Transform (SWT) [54, 56] and the original method *Algorithme à Trous* by Mallat. These methods are not equivalent to the CycleSpin [13] or Shifted Orthogonal Wavelet Transforms, which obtain redundancy by means of shifting the input signal and re-calculating the wavelet transform.

A diagram of the RDWT in Figure 2.4 shows the similarity to the DWT. Each level, j, is a bandpass filter of the original signal. The bandwidth of each filter is chosen dyadically, which is identical to the DWT. In the orthogonal DWT, the signal is decimated at each stage of decomposition, because the exact location at lower frequencies is ambiguous. However, without decimation the RDWT coefficients fully represent the uncertainty of low frequency events with the same sampling resolution of the signal, thus producing linearly dependent coefficients.

The choice to use a redundant representation is a double-edged sword. Proven



Intialize the iteration with:

$$A^0(k) = f(n)$$
$$H^0 = \text{Scaling Filter}$$
$$G^0 = \text{Wavelet Filter}$$

The wavelet coefficients are F^j for $j = 1, \ldots, J$ and the scaling coefficients are $F^{J+1}(k) = A^J(n)$.

Figure 2.4: Rather then decimating the signal at the output of the filters, the filters are up-sampled which preserves the frequency resolution at all wavelet scales.

results for the DWT do not transfer to the RDWT, though experiments suggest that the results are similar. The nature of a redundant transform is that the approximation of the signal is no longer unique. Therefore, the complexity of choosing the smallest set of coefficients for a given quality of reconstruction becomes NP-complete.

Current research by Donoho [22] and Gribonval [29] suggest that if certain conditions for sparsity are met, a set of coefficients can be chosen by the minimization of a linear program. Herrmann [34] has demonstrated these results by using linear programming to maintain high quality reconstruction using a sparsity constraint when choosing a set of redundant coefficients. Tropp [64] has instead used a greedy algorithm to choose a set of coefficients. Another method used by Pizurica [57] uses the redundancy of the RDWT to estimate a local spatial indicator. This indicator allows coefficients to be chosen across scales, exploiting the similarities that are present at different scales.

2.3.3 Curvelet Transform

Curvelets provide an excellent representation of GPR events in B-scans for signal processing. The nature of curvelets is that they efficiently represent a piecewise smooth curve which contain singularities normal to the direction of smoothness. This property is the precise nature of GPR data, which suggests that curvelet coefficients can provide a localized sparse representation where non-linear operations, such as thresholding, can be performed. The same arguments that apply in the 1D wavelet case work in this 2D representation, namely that optimal linear denoising methods, in the sense of minimum risk, can be further improved by non-linear estimators using curvelets [7].

The seminal curvelet papers [7] and [8] describe the nature of curvelets and their properties. An approach to an intuitive understanding is to compare the shape of the curvelet atoms. The separable 2D wavelet bases are nearly isotropic and well contained within a the support of the scale, j. However, curvelets are designed such that $width = length^2$. This makes a curvelet anisotropic and requires an additional orientation parameter, θ. Curvelets also require a scale parameter, j, and position vector, k, to represent the location and frequency content.

The building blocks for sparse representation of curves can be made using these three parameters: scale, spatial location and orientation. To insure sparsity, the curvelets are constructed using a parabolic law scaling law, which provides a natural extension to smooth curves. Therefore, unlike 2D wavelets, which tend to represent a curve as a collection of points in the wavelet domain, curvelets minimize the number of coefficients needed to represent the curve by adapting to the scale of the curve.

The most important result presented in [7] is the rate of energy decay, in an L-2 sense, of a incomplete reconstruction of a signal, f, using the m largest amplitude coefficients of a a curvelet decomposition is

$$\|f - f_m^C\|_2^2 \leq Cm^{-2}(\log n)^3. \tag{2.18}$$

The strength of this statement is that adaptive methods, which are not necessarily tractable can approach m^{-2} as $m \to \infty$, whereas wavelets approach m^{-1}. Despite the log term, a curvelet representation approaches the adaptive limit. The benefit is that curvelets are not adaptive and do not require the iterative computations necessary to adapt to a signal. With the results in [9], a new technique of constructing curvelets has introduced a tight frame, which is used in this work, and the same approximation rates hold.

Tight frames will preserve the energy in both domains, known as Parseval's relation for the Fourier domain and echoed here in the curvelet domain as

$$\sum_m |\langle f, \gamma_m \rangle|^2 = \|f\|^2_{\mathbf{L}_2(\mathbb{R}^2)}, \quad \forall f \in \mathbf{L}_2\left(\mathbb{R}^2\right), \tag{2.19}$$

which makes reconstruction straight forward by using the adjoint. As expected, the reconstruction is the same for curvelets as for wavelets, which is the sum of coefficients in the curvelet domain projected onto the basis functions,

$$f = \sum_m \langle f, \gamma_m \rangle \gamma_m. \tag{2.20}$$

The partitioning of curvelets has some similarity to wavelets in terms of scale and translation, but not in orientation. When wavelets are extended to two dimensions, the frequency partitioning does not represent the fullness of the extra dimension. In a 1D signal, a singularity can be described by a point in time. Two dimensional wavelets operate under the same premise, that singularities occur at localized points in an image. However, edges are singularities that are oriented on a line. The angular partitioning of curvelets allows these singularities to be represented by a location and orientation. With a wavelet representation, this edge would be represented by a collection of points rather than a smooth line.

Curvelets overcome this inefficient representation through a frequency partition that includes not only dyadic frequency partitioning, but also a doubling of angular partitions at every other scale. Angular localization with respect to frequency is subject to the same uncertainty as temporal localization. Therefore, the angular parameter, θ, provides finer resolution at higher frequencies. The result is a partitioning, which allows a piecewise-smooth edge containing a singularity to be represented by a sparse set of coefficients.

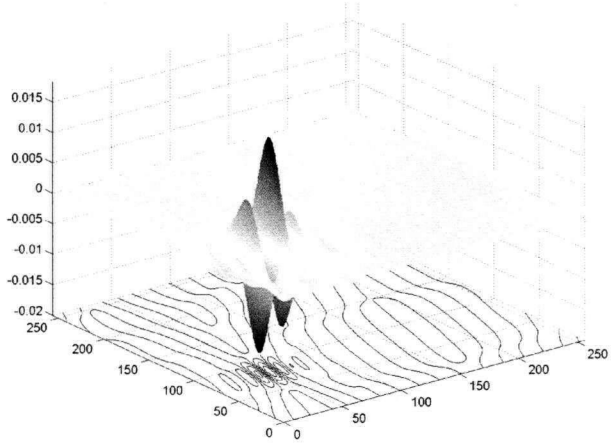


Figure 2.5: An image of a single curvelet is shown. The curvelet is directional and localized with a definite angular component.

These new atoms can be described as "needle-like" at fine scales [9], whose rectangular dimensions follow the parabolic scaling law of $width = length^2$. When these functions overlap an edge that follows a smooth curve, they produce large coefficients. This produces a sparse representation for edges, which can be adapted to the hyperbolic events of GPR data.

2.4 Noise Removal using Threshold Estimations

The foundation for wavelets is laid and the extensions to various discrete implementations are described. The next step explores the application of wavelets and curvelets to the estimation of a signal in the presence of noise. Statisticians refer to this problem as non-parametric regression, or in engineer speak: denoising.

The goal of noise removal will be to estimate a signal, f from a noisy observation d. The model is additive so $d = f + n$, where n is the noise present in the data observation. If the data is transformed into the wavelet domain using a set of

orthogonal basis functions g_m, by

$$D[m] = \langle d, g_m \rangle, \tag{2.21}$$

then the underlying model is also transformed such that

$$D[m] = \langle f, g_m \rangle + \langle n, g_m \rangle. \tag{2.22}$$

If it was possible to know the value of $N(m)$ it would be a simple matter of subtraction. Instead, a decision operator is introduced, $\Omega[m]$, which will shrink each wavelet coefficient of the data. The application is

$$\tilde{f} = \sum_m \Omega[m] D[m] g_m, \tag{2.23}$$

which produces an estimate, f that can be measured with reference to f by the the risk:

$$r_\Omega(f) = \mathsf{E}[\|f - \tilde{f}\|^2]. \tag{2.24}$$

The minimum risk can be obtained if the nature of the noise is known in the wavelet domain using *oracle estimation* [48] with a decision operator defined as

$$\Omega[m] = \frac{|D[m]|^2}{|D[m]|^2 + \sigma^2}. \tag{2.25}$$

This holds only if the noise is Gaussian and white with a zero mean. The remarkable result shown by Donoho [23] was that the decision operator defined in (2.26) produces a risk which is nearly minimal with respect to oracle estimation. This nonlinear *threshold* operator would instead make a decision to either keep a coefficient that is above a threshold, T, or kill it, by multiplying with a zero. More formally, hard thresholding is,

$$\Omega[m] = \begin{cases} 1 & \text{if } |D[m]| > T, \\ 0 & \text{if } |D[m]| \leq T \end{cases}. \tag{2.26}$$

This produces a lower risk in the estimate $\tilde{f}$, but higher amounts of spurious oscillations, where f contains transients. An alternative to hard thresholding is soft thresholding. Soft thresholding applies the same principle of keep or kill, but removes the sharp discontinuities of hard thresholding by shrinking coefficients above T. Soft thresholding is defined as

$$\Omega[m] = \begin{cases} 0 & \text{if } |D[m]| < T, \\ D[m] - T & \text{if } D[m] \geq T \\ D[m] + T & \text{if } D[m] \leq -T \end{cases} . \tag{2.27}$$

Donoho showed in [23] that T could be chosen by simply knowing the standard deviation of the noise and the number of points,

$$T = \sigma\sqrt{2\log_e N}. \tag{2.28}$$

This result is coupled with the ability to estimate the standard deviation from the finest scale coefficients of $D[m]$. If the wavelet coefficients are arranged, such that $0 < m \leq N/2$ then an estimate of the standard deviation is

$$\tilde{\sigma} = \text{median}|D[m]|/0.6745, \quad \text{for } 0 < m \leq N/2[48]. \tag{2.29}$$

These results led to a surge of research in the application of threshold estimators. The results were promising, but a nagging problem remained: shift invariance. The above results hold for additive Gaussian white noise thresholded from the DWT. Due to the lack of shift invariance, the resulting estimations were excellent in the sense of risk, but lacking in smoothness.

In [45, 44], Lang *et. al* were inspired by R. Coifman to threshold a shift invariant wavelet transform to remove additive Gaussian white noise. Using the RDWT, they did a performance comparison on several of the 1D and 2D signals. Comparing

these results to Donoho's work [21] with soft thresholding they were able to show that as the energy of the additive noise decreased, thresholding the RDWT began to outperform soft thresholding of the DWT. These results were demonstrated by fixing a number of the parameters available in the algorithm. Donoho and Coifman also attempted their own versions of a shift-invariant wavelet threshold algorithm. In [13], the authors introduced the SpinCycle algorithm which used the results from [1] to produce a set of redundant coefficients to threshold. These results also outperformed classic DWT thresholding methods.

Thresholding in the wavelet framework is summarized by theoretical and experimental results present in current research. The questions that have not been answered in present research are the significance of choosing the number of decomposition levels and if RDWT thresholding can be applied to colored Gaussian noise using a threshold estimator. The next two chapters address an aspect of these questions by exploring methods of evaluating and applying a threshold estimator to GPR data.

Chapter 3

Removing Noise from a GPR A-scan

The objective of the following methodology is to evaluate and understand specific practical questions that have not been addressed in the application of wavelet noise removal to GPR signals. Current research suggests that wavelet denoising can be effectively performed by redundant transforms, but it is still unclear how to choose a threshold estimator, scale parameter or basis. It is beyond the scope of this thesis to explore exhaustively each choice, but instead focus is placed on an the scale parameter after a wavelet basis and threshold estimator is chosen. The hypothesis is a level-constant threshold estimator applied to the RDWT transform of a GPR signal is a suitable method for removing white and colored Gaussian noise if the scale parameter is properly chosen.

There are four ingredients, which are necessary to test the suitability of the thresholded RDWT for GPR signals. Each of the following sections discusses the specific choices based on the following questions:

- How can a synthetic GPR pulse be generated?
- How can a threshold be estimated and applied to the RDWT?
- How can a wavelet basis be chosen?
- How will the scale parameter affect Gaussian noise removal of the thresholded RDWT?

In order to evaluate the suitability of RDWT thresholding a direct comparison to the existing theoretical results is made. The performance is juxtaposed to the choice of thresholding the RDWT and Wiener filtering on GPR signals. The metrics provide comparisons can be made to the existing literature. Specifically, the theoretical results of wavelet literature focus on the use of a risk metric, introduced in (2.24). Proven results in wavelet denoising show the decrease in a normalized risk metric,

$$\frac{\|f - \tilde{f}\|^2}{N\sigma^2} = \frac{r(f)}{N\sigma^2}. \tag{3.1}$$

This metric assumes that zero-mean Gaussian noise, has been added to the signal f. The difference of the signal and its estimate $\tilde{f}$ are normalized by the length, N of the signal and the variance σ^2 of the Gaussian noise.

Evaluation of the maximum scale parameter, J is presented in the familiar engineering metric of signal-to-noise ratio,

$$SNR = 10\log_{10}\left(\frac{\|f\|^2}{\|f - \tilde{f}\|^2}\right). \tag{3.2}$$

When comparing the signal, f to the estimate, $\tilde{f}$, emphasis is placed on the improvement of the output signal-to-noise ration, SNR_{out} relative to the signal-to-noise ratio of the noisy input data,

$$SNR_{imp} = SNR_{out} - SNR_{in}. \tag{3.3}$$

This emphasizes the changes in performance when experimental parameters are chosen.

The experiments are conducted in MATLAB using Rice University's implementation of the RDWT from the Rice Wavelet Toolbox [31] and the wavelets are generated with Wavelab [24]. A description of the ingredients to the RDWT thresholding process with a discussion of the rationale and relevance follows. The parameters of RDWT thresholding are explained before the GPR and noise model is described.

3.1 Determining a Basis for A-Scan Approximation

In Mallat's *Wavelet Tour of Signal Processing* he labels a chapter: ESTIMATIONS ARE APPROXIMATIONS. Among the properties of the wavelet transform, one elegant result is that signals are sparsely represented for a given wavelet basis function. This property makes *estimation* of a signal in noise simple, because the correlation of the signal to the wavelet basis provides high amplitude coefficients, which are easily chosen via a threshold. The foundation of successful estimation relies on a small number of coefficients being an effective *approximation* of the signal.

A basis function should be chosen such that the decomposition presents a sparse set of coefficients and the signal is well approximated with a small number of coefficients. Data adaptive methods [14, 49] choose a "best" basis from a dictionary of bases using a sparsity measure to make a decision. Rather than form a best-basis, this work instead uses the log-energy distribution measure,

$$\sum_{m=1}^{N} |F[m]|^2 \log |F[m]|^2. \tag{3.4}$$

to choose a particular mother wavelet.

When picking a basis function, the support of the bases affect sparsity, but usually at a trade-off in the smoothness of reconstruction. This trade-off is related to the to wavelet construction, which demands that the wavelet support is determined by the number of vanishing moments. The wavelet has p vanishing moments if

$$\int_{-\infty}^{+\infty} t^p \psi(t) dt = 0. \tag{3.5}$$

An intuitive understanding of vanishing moments is that a wavelet acts a differentiator. If a singularity is differentiable q times, then a wavelet will be able to "detect" it when the vanishing moments, p is greater. If a wavelet is chosen for the number of vanishing moments, then the minimum support is $2p - 1$.

For GPR signals, the underlying signal is highly differentiable and it is more important to produce large amplitude coefficients, which make signal estimation possible. Using (3.4) as a guide to determine the best choice for a basis, minimal support should increase sparsity. Table 3.1 verifies that the Daubechies wavelet with 2 vanishing moments produces the sparsest representation with a log-entropy metric.

Basis	Entropy
Daubechies4	-302.3
Symmlet4	-294.6
Coiflet2	-290.1
Battle1	-289.5
Symmlet6	-273.5
Coiflet3	-272.6
Daubechies6	-270.4
Coiflet1	-268.5
Symmlet5	-266.2
Battle3	-225.8

Table 3.1: Summary of entropy measurements on selected basis functions.

Table 3.1 is generated by decomposing a 2048 point synthetic GPR trace

with one reflection event. Each entry represents a particular basis for decomposition calculated with the Wavelab function `MakeONFilter`.

The sparsity of a decomposition does not guarantee the ability to approximate a signal. However, the alternative to using a sparsity metric is to find the smallest set of coefficients for a given approximation rate. Searching for this optimal set of coefficients which satisfies a certain quality of reconstruction is an NP-complete problem. Therefore, the sparsity property becomes an efficient and reasonable assumption for choosing a wavelet basis.

3.2 Determining the Number of Levels of Decomposition

In addition to determining a basis function, it is important to determine a useful number of decomposition levels. The maximum number of decomposition levels, J, determines how many different wavelet scales are calculated. It is possible to decompose a signal using the Redundant Discrete Wavelet Transform to a maximum of $J = \log_2 N$ levels. However, it is not clear if a signal *should* be decomposed to the maximum scale. As the parameter J increases, the support of the wavelet increases to a maximum of N when $J = \log_2 N$. Because the efficiency of a median operator increases with N and the advantage of thresholding decreases with sparsity, it is expected that the optimal value of J is less then $\log_2 N$.

There appears to be no research that suggests an appropriate value of J. The choice of scale parameter depends on the signal, noise and wavelet basis function. The wavelet basis controls the balance between frequency and temporal localization. The amount of energy that the noise and signal under study have at particular scale,

j determines if threshold estimation and application is appropriate.

The evaluation of the scale parameter compares the signal-to-noise improvement calculated from (3.3). The number of levels, additive white noise and colored noise, and the energy of the noise is varied to illustrate the scale parameter effect. The measurements are performed on a synthetic GPR trace.

Evaluation of scale parameter experiment:

1. Generate noisy data from the addition of a model and Gaussian Noise.
2. Decompose to J levels using the RDWT.
3. Estimate a threshold and apply.
4. Reconstruct the signal and measure SNR_{imp}.
5. Repeat steps 1-4 for a new realization of noise with larger SNR_{in}.
6. Repeat steps 1-5 to obtain an average SNR_{imp}.

The reference model is a synthetic GPR pulse that models an approximation of a physically realistic version of equation (2.4). Two different experiments are conducted, one to examine the effects of Gaussian white noise and the second for Gaussian colored noise. The coloration of the noise is described in Section 3.4.2, which assumes the difficult case where the noise has nearly the same characteristics as the signal.

The input noise is measured with (3.2) and varies from -9dB up to 30dB in 3dB increments. The range of SNR_{in} is representative of the signal attenuation as it propagates into the ground, which makes late time reflection events decay into the noise floor of the receiver. For each value of SNR_{in}, the number of levels is

varied from the minimum of one to the maximum of $\log_2 N$. Each experiment is conducted twenty times and the average SNR_{imp} is calculated with the mean.

The result form a surface where an optimal number of levels can be chosen for the particular signal and noise type. However, because choosing the number of levels is equivalent to bandpass filtering, a more generalized understanding can be implied based on the frequency content of the signal and the noise under study.

3.3 Estimation of a Threshold from Noisy Data

The choice of a threshold estimator should only suppress a coefficient if it is below a predicted value of noise. For the DWT a global threshold estimator can approximated for all coefficients. This global threshold is valid, because the noise remains white in the wavelet domain with a constant standard deviation. When thresholding the RDWT, the noise does not remain white and a threshold should be estimated for each coefficient individually.

Donoho proposes in [23] that the optimal threshold for the DWT is,

$$T = \tilde{\sigma}\sqrt{2\log_e N}, \tag{3.6}$$

which, for certain classes of signals, achieves nearly minimax risk with respect to the optimal non-linear decision operator. This result makes the removal of white noise in the DWT domain a simple matter of estimating the standard deviation of the noise. Using the equation (2.29), an estimate of $\tilde{\sigma}$ is reasonable if the signal is sparse relative to the noise. For the DWT, the finest scale provides $N/2$ coefficients to estimate the standard deviation. The support of the wavelet at the fine scale is very narrow, which makes the signal sparse relative to the noise. For the RDWT, each scale provides N coefficients to make an estimate, but the sparsity of the signal

coefficients decreases with j. Therefore to generate a threshold estimator requires either, advance knowledge of the noise behavior or multiple realizations of a signal with identical noise distributions.

Using a hybrid method, which neither sets a threshold for each coefficient, nor uses a global threshold for all coefficients will be called a *level-constant threshold.* This threshold is constant for a given level and is calculated from an estimate of the standard deviation of the coefficients at level j,

$$T_j = \tilde{\sigma}_j \sqrt{\log_e N}. \tag{3.7}$$

This is applied to the wavelet coefficients using a hard thresholding technique given in equation (2.26) to the matching level of coefficients. The assumption is that the decomposition is reasonably sparse and an estimate of $\tilde{\sigma}$ will remove the majority of noise without being too aggressive. This assumption is tested by the scale parameter experiments outlined above and a calculation of risk with respect to the competing methods. The expectation is that the scale parameter has a direct effect on the ability to remove noise. As the J increases, the decreasing sparsity introduces more bias in the coefficients making the median estimator too aggressive.

The alternative, which is to estimate a global threshold from only the finest coefficients is sufficient and successful for white noise. The results of Lang [44], Mallat [48] and Donoho [21] justify this method of thresholding. However, in the case of colored noise, the average standard deviation will vary drastically for each level, j. Therefore, a global threshold does not apply. Thus, a level-constant threshold should prove to be a reasonable signal estimator in the presence of colored noise.

The verification of the level-constant threshold is tested against the performance of classic wavelet denoising using the DWT and Wiener filtering. A calculation of the normalized risk for the three methods is presented as N increases. This

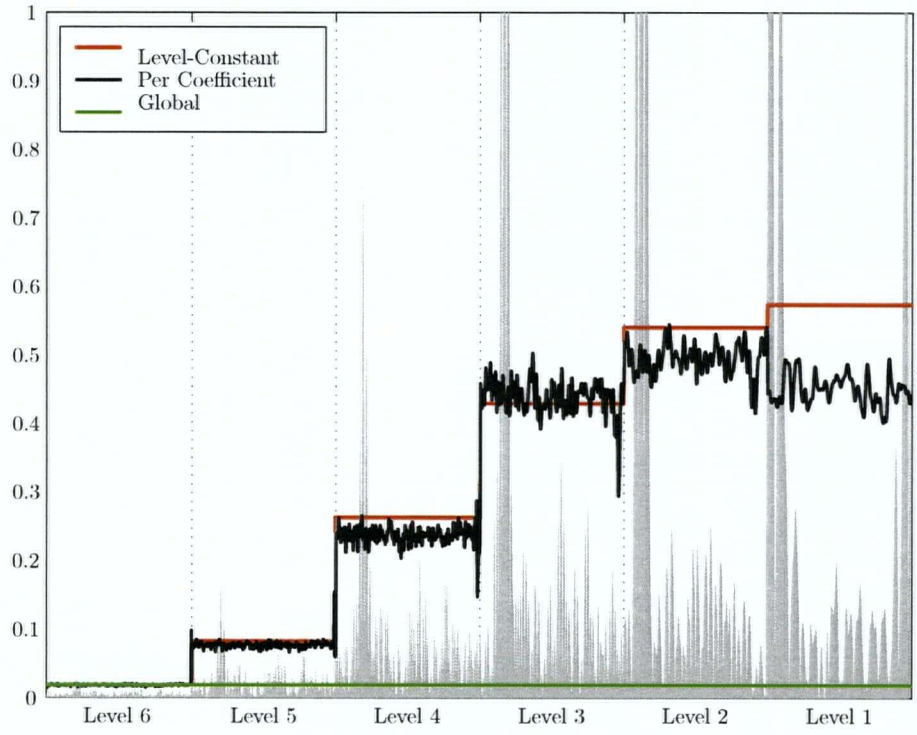


Figure 3.1: Using the DWT, a global threshold is estimated, but is ineffective for removing colored noise, which is clearly above the green threshold. A level-constant threshold approximates a threshold that is calculated for each coefficient, which adapts to the noise coloration.

allows a direct comparison of the proven DWT thresholding results in [48] to the RDWT level-constant threshold method.

3.4 GPR A-Scan Model

A GPR A-scan can be measured as a voltage, V_{rx} from a receive antenna. The result is a convolution of several impulse responses from the environment and the system with a source function, V_s,

$$V_{rx} = V_s(t) * h_{rx}(t) * h_{tx}(t) * h_c(t) * h_g(t) * h_t(t) + n(t). \quad [18] \tag{3.8}$$

The impulse response from cross-coupling, $h_c(t)$ is described in [18] and is one source of horizontal clutter in a B-scan.

The impulse response of the ground, $h_g(t)$ is quite complicated as it depends on the material, moisture content and granularity. Due to the variation with time, accurate ground models should be considered uniquely for specific application. A simplification of the ground model can be constructed using a constant attenuation for a time-invariant frequency response. This methodology assumes the ground is constant for all frequencies. Refer to the work of Irving [41] for a frequency dependent inversion of simple ground cases.

The target impulse response, $h_t(t)$ is also complex, but simple reflectors such as the ground or point reflectors can be roughly approximated as specular. The input signal, V_s is modeled as the derivative of a Gaussian bump like in (2.4). This is a simplification, because a real pulse tends to be asymmetric and long-tailed, but depends upon the transmitter design and operating characteristics.

The impulse responses of the receiving and transmitting antenna are the same for a GPR in mono-static operation, which makes $h_{rx}(t) = h_{tx}(t)$. The development

of a synthetic antenna model, $h_n(t)$ is described in Section 3.4.1. Using the antenna model, the source pulse is filtered by $h_n(t)$ to form a synthetic transmitted pulse. The reflection events are specular, which corresponds to a dirac function in time making the final GPR pulse model,

$$V_{rx} = V_s(t) * h_n(t) * h_n(t) * \delta(t - t_i). \tag{3.9}$$

The target locations are located at the times in the array t_i, which is the round trip time from the antenna. The pulse must be filtered twice by the antenna, once during transmission and once upon reception, thus the two antenna responses in (3.9).

The physical parameter, FWHM, used to construct V_s is 100ns, which is approximately the FWHM of the experimental GPR demonstrated by the Belgium Royal Military Academy HumDem research group [59]. This compliments the antenna model, which is a simulation of the same experimental GPR.

The simplifications and assumptions made to produce the GPR impulse model are based on method of noise removal. Wavelets can be thought of as a singularity detector. Within limits, the singularity can undergo linear filtering without dramatically affecting the ability of a wavelet to efficiently represent it. Therefore, the simplification of the ground and target to specular reflectors is reasonable. Additionally, ignoring the cross-coupling term is justified, because it should have similar frequency and temporal support as reflection events making the results on the air-ground event transferable.

3.4.1 Antenna Model

Bart Scheers, in his doctoral thesis [60], developed a normalized antenna impulse response

$$h_{N,tx} = \sqrt{\frac{Z_c}{Z_a}} \frac{\tau_{tx}}{\sqrt{f_g}} h_{tx} \text{ and } h_{N,rx} = \sqrt{\frac{Z_c}{Z_a}} \frac{\tau_{rx}}{\sqrt{f_g}} h_{rx} \tag{3.10}$$

which describes the frequency dependent effects of the ground, f_g, impedance of the system and antenna Z_c, Z_a and the transmission coefficient τ_{tx}, τ_{rx}.

The goal of the synthetic antenna model is to simulate a reasonable GPR pulse in the time domain. For the purpose of determining the suitability of the RDWT for noise removal it is not critical to understand each parameter of the model. Instead, the experiment results from an antenna gain plot from Fig. 4-11 in Scheers' ([60]) thesis are used to calculate h_N directly. Using points that estimated by sight from the gain response of Antenna 2 of the referenced figure, a piecewise model of $G_t(\omega)$ is generated. The normalized frequency response, $H_N(\omega)$ is related to the gain by

$$G_t(\omega) = \frac{4\pi}{\lambda^2} |H_N(\omega)|^2. \tag{3.11}$$

To find a the coefficients for a digital filter to approximate h_N, the piecewise model is fed into a Yule-Walker regression estimator to generate the coefficients b_i, a_i of a digital filter,

$$H_N(z) = \frac{\sum_{i=0}^{l} b_i z^{-i}}{1 + \sum_{i=1}^{k} a_i z^{-i}}. \tag{3.12}$$

Using (3.12) it is straight forward to find the normalized impulse response and consequently filter coefficients of the theoretical pulse. In Figure 3.2 the Yule-Walker fit is compared to the piece-wise estimation.

The final assumption is that $h_{N,tx}(t) = h_{N,rx}(t)$ and notation is simplified to $h_n(t)$. Using this model a MATLAB script is created to filter the synthetic input

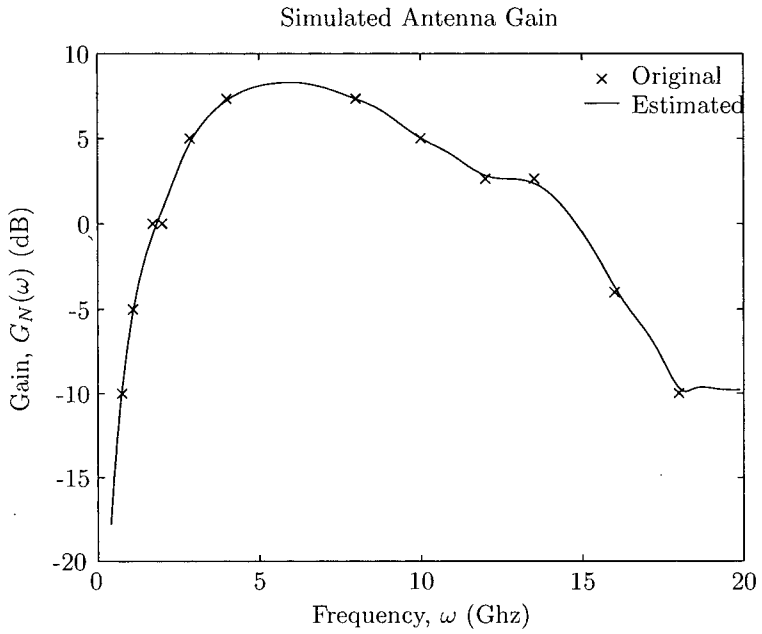


Figure 3.2: Using the Yule-Walker method to estimate an example antenna response.

pulse by the simulated antenna response. Figure 3.3 shows the result of the physical approximation of the GPR pulse.

3.4.2 Noise Model

The noise model, $n(t)$ is straight forward to model in MATLAB. To produce Gaussian white noise, the MATLAB function `randn` generates an arbitrary matrix values which is normally distributed, zero mean and has a standard deviation of one. This is referred to as the additive Gaussian white noise, or AGWN model. The AGWN model is used to verify the performance of the RDWT level-constant threshold method to the classic method using discrete wavelet thresholding.

The colored noise model is created using the antenna model from the previous section and GWN. The additive Gaussian colored noise, or AGCN is filtered white

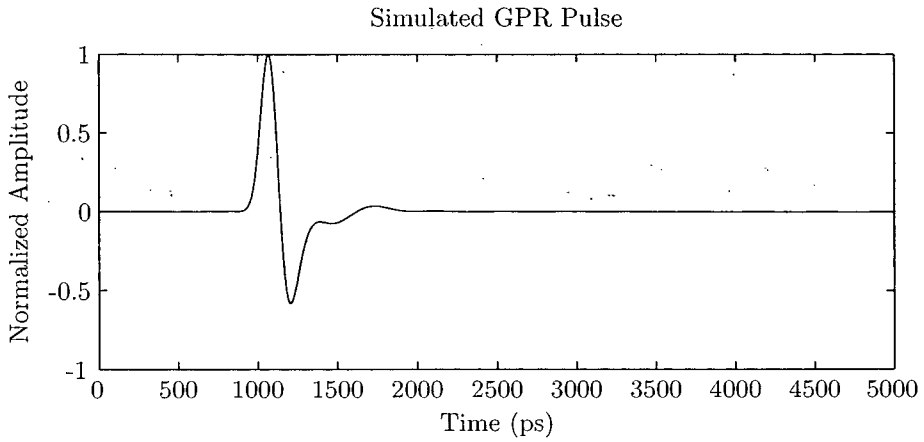


Figure 3.3: This pulse shows an typical pulse that might be received with only one reflector at 1000ns.

noise and is represented as an operator on the additive noise model,

$$d(t) = f(t) + h_n(t) * n(t). \tag{3.13}$$

The antenna model ensures that the GCN is "in-band", meaning that the noise has a similar frequency content as the signal, f. This presents a challenging case for the verification of the level-constant thresholding of the noisy data, d.

The AGCN model is not unique and therefore does not represent an exhaustive study of AGCN noise removal. Conceptually, an AGWN model has a flat spectrum. In the context of thresholding, a flat spectrum means that if the spectral content of the signal is less then the noise, it is thresholded to zero. For the AGCN, the spectral content is similar to the signal. This makes the estimation and suppression much more difficult. These two cases should garner a general intuitive understanding, which will allow an extension to other cases.

3.5 Summary

This chapter introduced the methodology for testing the RDWT in the presence of Gaussian colored noise. The next chapter will introduce the methodology for suppressing clutter in B-scans before the final results are presented. The main points from this chapter are:

- A simulated GPR pulse is the signal under study.
- A level-constant threshold is defined.
- A wavelet basis is chosen based on sparsity.
- A risk and signal-to-noise metric is used to evaluate results.
- A noise model for both colored and white Gaussian noise is shown.

The experimental results use these building blocks to evaluate the effect of choosing the maximum number of decomposition levels on the ability of the RDWT to estimate a GPR signal in white or colored noise. In order to establish a baseline performance, experiments are designed to compare the DWT and RDWT in the presence of white noise, before a performance evaluation on colored noise.

Chapter 4

Clutter Suppression in GPR B-Scans

Clutter suppression is a common step in the signal processing chain towards the goal of landmine detection. The purpose is to minimize damage of the target signal event, while suppressing anything that is not related to the target. There are several reasonably successful methods of clutter reduction present in current literature. For a general overview the book written by D. Daniels, *Surface-Penetrating Radar* [18] and a more comprehensive review in the recent book edited by the same author, *Ground Penetrating Radar* [19] are recommended. A brief overview of clutter suppression methods will be presented here to develop a motivation for the proposed method.

Adaptive methods, such as the one introduced by Brooks *et. al* [4] uses a linear time-invariant model to perform a non-parametric regression of the system response to separate clutter. Another adaptive method of clutter suppression attempts to suppress multiple reflection events, where the author uses an L-1 weighted

estimator instead of the traditional least squares filter to derive a clutter model [33]. Van Kempen *et. al* examines the data for a set of features to form a model and suppresses clutter with a modified Wiener filter [67].

A non-adaptive method is presented in [63] where a principal component decomposition is used with linear estimation. The assumption is that the target is contained within the largest principle components and a linear approximation will achieve clutter suppression. An alternative to regression models, is discrimination, where clutter is suppressed by choosing a target, which forces all other data to be accepted as clutter. Zhao proposes a successful method of discrimination using a discrete hidden Markov model, which finds both metal and non-metal landmines in GPR data [70].

This work is derived from the results of Herrmann in the seismic realm [36]. Herrmann uses the parsimonious curvelet representation of seismic data to iteratively suppress multiple target reflections. Herrmann's work tackles the difficult problem of selectively removing events that are nearly identical to the target, but smaller in amplitude and later in time [34]. Herrmann's method has also been extended to removing temporal clutter using multiple data sets from the same location, but several years apart [16] and reducing the dimensionality of the seismic imaging problem [35].

The method proposed in this work also supplements the results presented by Nuzzo and Quarta [55], which compares τ-p and wavelet transforms for effective horizontal clutter suppression. Nuzzo *et. al* used the 2D DWT to selectively threshold only the horizontal components. The two main features presented in their method are:

- Identify the wavelet scales that clutter appears in.

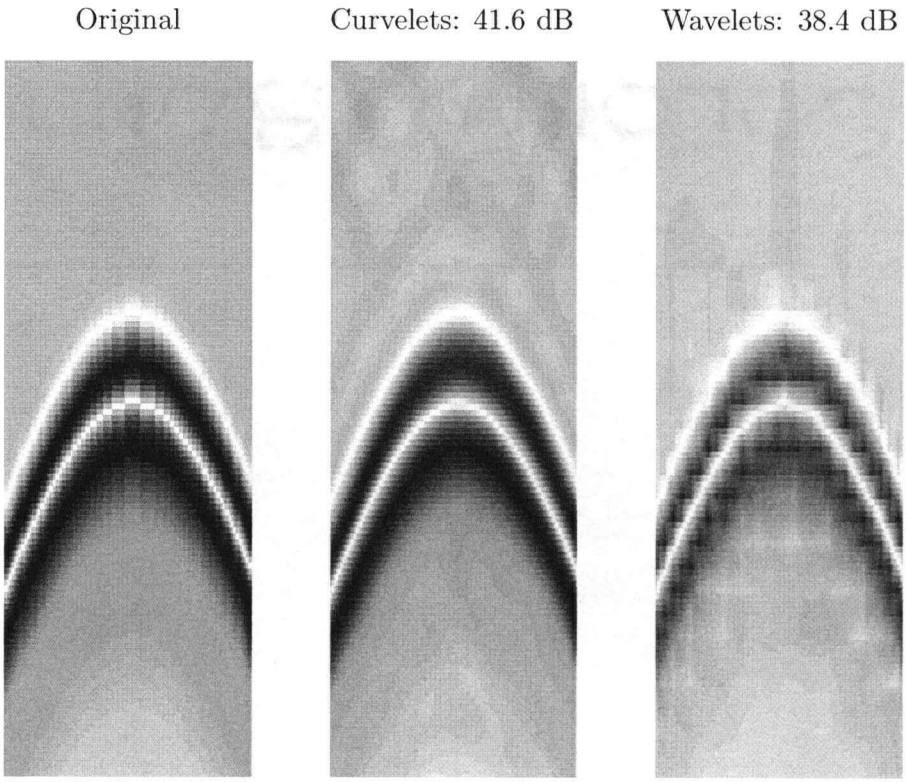


Figure 4.1: This example demonstrates reconstruction of a simulated hyperbola with only 0.5% of the curvelet and wavelet coefficients. Notice how curvelets maintain the smoothness of curve and still reproduce the discontinuity across the curve.

- Suppress horizontal wavelet coefficients at identified scales.

The algorithm proposed makes several important changes to this general approach. First, the curvelet transform is used, instead of the 2D wavelet transform. Figure 4.1 demonstrates the motivation of using curvelets to approximate GPR reflection events in B-scans. Second, a noise model is generated using features from the data set, which intrinsically define the wavelet scales and horizontal clutter features. Additionally, the noise model forms the threshold operator, which puts the algorithm into the well researched and understood domain of wavelet thresholding.

Like clutter, the noise model should be determined for each application.

Therefore, a synthetic clutter and target model is formed from actual landmine data. To verify the noise model, the synthetic B-scan is generated using the physical approximation of A-scans developed in Chapter 3.

The description of clutter suppression methodology begins with the synthetic B-scan model. Afterwards, the clutter and target models are described and the proposed algorithm is introduced. The chapter concludes with a discussion of performance metrics for evaluation of the proposed method.

4.1 B-scan Model using Synthetic GPR Pulse

The synthetic B-scan is carefully modeled such that reasonable clutter and target approximations are made. This allows a careful study of the efficiency and limitations of the proposed method of clutter suppression.

In order to create a B-scan, targets are placed into a time vector and convolved with the impulse response of a GPR pulse. The target trace is a vector of dirac functions scaled by the predicted target amplitude. For example, the ground might be a dirac function with unit amplitude occurring at 1000ps. Due to the FWHM of 100ps, the sampling rate is 10ps, and an index of 100 would contains the target.

4.1.1 Clutter Model

The clutter model used for simulation is an approximation of clutter types that exist in a real GPR scan. Figure 4.2 shows a real GPR survey of soil with no targets. The synthetic clutter model reflects the three highlighted events of 4.2.

The clutter is scaled to unit amplitude for all events. Each event is similar to the physical events present in Figure 4.2. The specific time locations of clutter

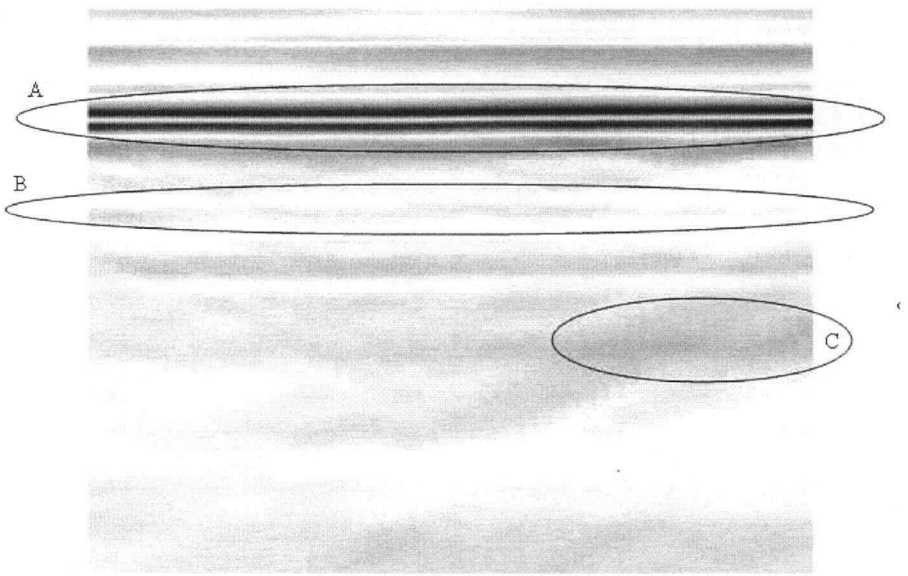


Figure 4.2: This GPR B-scan contains examples of clutter without targets. Annotation A shows the air-ground event, the clutter in oval B is a highly variable amplitude event and C shows a clutter event that does not exist for the duration of the scan.

are not critical and arbitrarily defined as:

- Synthetic Event, E_A: Ground, at 1000ps with constant amplitude
- Synthetic Event, E_B: Clutter at 1250ps with varying amplitude
- Synthetic Event, E_C: Clutter at 1500ps which does not persist

The amplitude of each event is defined by a function, before it is convolved with the synthetic pulse,

$$E_A = 1 \tag{4.1}$$

$$E_B = cos(2\pi 0.02x), \quad 1 \geq x \leq 50 \tag{4.2}$$

$$E_C = \begin{cases} 0, & 1 \geq x \leq 24 \\ e^{(x-25)}/e^4, & 25 \geq x \leq 29 \\ 1, & 30 \geq x \leq 50 \end{cases} \tag{4.3}$$

4.1.2 Target Model

The target model is generated by calculating a spatial hyperbola attenuated by an exponential attenuation function dependent on the distance from the antenna to the target center. The spatial hyperbola is calculated by creating an array of time shifts,

$$t(x) = \frac{2}{v}\sqrt{Z^2 + (x - x_t)^2}, \quad \text{for } x = 0, 0.01, \ldots, 0.25 \text{ cm}. \tag{4.4}$$

The velocity is defined in (2.7), the target center, x_t, is 25 cm and the target depth, Z, is the difference between target depth and ground, which is 5 cm. The target time is converted to a MATLAB vector index by scaling (4.4) with the time sampling rate, 10ps and rounding to the nearest integer,

$$\texttt{idx} = \text{round}\left(\frac{t(x)}{10 \times 10^{12}}\right). \tag{4.5}$$

(a) Point Target (b) Large Target

Figure 4.3: Example of synthetic targets used in the experiments.

In MATLAB, the image is indexed by, `idx` and the attenuation is calculated by (2.5) and assigned to the image array,

$$\texttt{0(idx)} = \frac{1}{G}\frac{e^{-2\alpha d}}{d^4}. \tag{4.6}$$

The attenuation coefficient, α was chosen to represent sandy soil, which has a low attenuation of around 10dB/m. The distance from the antenna to the target is found by $d = \sqrt{Z^2 + (x - x_t)^2}$ and the target gain function,

$$G = \frac{e^{-2\alpha Z}}{Z^4}, \tag{4.7}$$

normalizes the peak of the the hyperbola to unit amplitude. The normalization puts the clutter, target and ground on similar amplitude scales, which is often the case.

This represents a reasonable model for a point target. An additional target is used, which has larger dimensions. This will result in a horizontal component that facilitates testing of the average trace subtraction method. It is calculated in the same manner, except the center of the hyperbola is extended horizontally, as presented in Figure 4.3.

4.2 Proposed Method of Clutter Suppression

A new method is proposed for suppressing horizontal clutter. The foundation is that the existing prevalent method, average trace subtraction (ATS) is often times sufficient for highlighting hyperbolas in GPR data for reasons of computational complexity. Therefore, new methods must be reasonably fast or especially successful in suppressing clutter. The new wrapped curvelet transform, which is released in CurveLab [10] is sufficiently fast to consider as an application clutter suppression. Additionally, because the curvelet domain is efficient at representing GPR traces, thresholding should prove to be competing method to ATS.

The act of thresholding "mutes" an event, which meets the criteria of the threshold. Unlike subtraction, thresholding does not add energy, therefore eliminating the possibility of constructive interference. If an appropriate threshold can be calculated that represents the clutter contained in the B-scan then a simple scaled threshold can be applied in the curvelet domain. This will suppress events that are clutter like and ignore events that do not match the estimated model. The clutter model becomes the key to successful implementation of the proposed method.

The clutter model for ATS is the average trace, which has two major assumptions. The first is that the resulting average amplitude is representative of all traces. This is usually not possible for a GPR system. Amplitude varies as a result of positional noise and subsurface material changes. The second assumption is that the phase remains the same. This is usually true, however, not all horizontal clutter events are present over the window of a B-scan. If a clutter event only exists in a portion of a B-scan the resulting subtraction will actually *place* an event where none exists. This would be a similar case if the clutter event changed phase 180 degrees resulting in constructive interference.

The proposed method overcomes the pit-falls of ATS by adding amplitude and phase resilience, but retaining a simple noise model. If a quasi-accurate clutter model is generated, the phase and amplitude resilience of the method comes from the parsimonious representation of clutter. This enables clutter to be muted with a high probability that the target is preserved. The amount of muting comes from a shrinking parameter, λ, which must be empirically discovered, or controlled by an expert user to determine the tradeoff between clutter removal and target signal preservation.

The clutter model must be quasi-accurate, which is the same assumption that ATS uses. In the average trace case, the clutter is assumed to be stationary in phase and amplitude over the spatial dimension of the B-scan window. An alternative clutter model is dubbed the *edge model*. The edge model uses the average of only two traces located at the edges of the B-scan. If there are M traces in a B-scan, which are column vectors, then the 1-D edge noise model is

$$\mathbf{b}_{edge} = \frac{1}{2}\left(\mathbf{b}_1 + \mathbf{b}_M\right). \tag{4.8}$$

To form the threshold operator, the noise model is first transformed into an image. The noise model image has dimensions identical to the B-scan under study, N for time and M for space. For each trace in the 2-D noise model image, Ω is the calculated edge model trace multiplied by a length-M vector of ones,

$$\Omega = \mathbf{b}_{edge}\mathbf{1}^T. \tag{4.9}$$

There are several reasons that make a noise model calculated from the edge traces a reasonable attempt to model horizontal clutter. The first is that the clutter is highly likely to be present in the edge of a B-scan, while the target is not. The second is that a large target will have a significant horizontal energy that would

be detected by an noise model calculated from the average trace. However, if the clutter is not present at the edges, the edge trace model does not work.

The proposed method of clutter suppression is summarized:

1. Take the curvelet transform of the data B-scan.
2. Calculate the edge noise model and form a B-scan noise model.
3. Take the curvelet transform of the noise model B-scan.
4. Use the noise model in the curvelet domain as a threshold.
5. Scale the threshold and apply to the data curvelet coefficients.
6. Inverse transform the thresholded data.

In notation, the set of curvelet basis vectors, g_m produces a set of coefficients from the data, d and noise model, Ω,

$$\hat{d}_m = \langle d, g_m \rangle \tag{4.10}$$

$$\hat{\Omega}_m = \langle \Omega, g_m \rangle. \tag{4.11}$$

The threshold operator formed from the noise model and the scaling parameter, λ,

$$\Gamma_\Omega(x) = \begin{cases} 0 & \text{if } |x| < \lambda |\Omega_m|, \\ 1 & \text{if } |x| \geq \lambda |\Omega_m|. \end{cases} \tag{4.12}$$

An estimate of the the clutter free data, $\tilde{d}$ is made from thresholding with Γ, and reconstructing,

$$\tilde{d} = \sum_m \Gamma(d_m) g_m. \tag{4.13}$$

This method has the advantage of being quite easy to visualize the result. The noise model is "muted" from the data by zeroing only the coefficients that are

estimated to belong to noise. The phase of the data is not important, because the absolute value of the coefficients is considered when thresholding. This is allows *incoherent* estimation, because the noise model only provides a spatial location and approximate amplitude. Coherent estimation, such as subtraction, requires careful alignment of phase and amplitude to prevent constructive interference. The amplitude resilience of thresholding comes from control of the λ parameter. A large value increases the tolerance for amplitude variation at the cost of damaging closely spaced events.

4.3 Method of Performance Evaluation

To verify that the proposed method is successful, several means of verification are presented. Using the synthetic data, numerical results are presented by calculating the PSNR,

$$20\log_{10}\left(\frac{\Delta}{\|d-\tilde{d}\|^2}\right). \tag{4.14}$$

The traditional image processing definition of PSNR uses $\Delta = 255$, which is the difference between the maximum and minimum values of a 255 intensity level image. To calculate Δ in this application, the difference of the maximum and minimum amplitudes are taken from the input data,

$$\Delta = \max|d| - \min|d|. \tag{4.15}$$

Numerical metrics are useful for comparing the amount of *energy* removed from the noisy data, but visual quality is not well represented. Therefore, the numerical results are presented to verify that clutter suppression is successful in removing the clutter energy, but visual experiments are conducted to demonstrate the difference in artifacts.

The visual experiments allow a direct comparison of the proposed method and ATS. Image artifacts, like target energy damage, are best viewed by comparison of the input and output images of each algorithm. A residual image, $d - \tilde{d}$ is also calculated, which emphasizes the artifacts either created, or not removed by each method.

Chapter 5

Experimental Results and Discussion

The results presented in this chapter are from two separate methods. The first set of results show the suitability of a level-constant thresholded RDWT for removing white or colored Gaussian noise. The second set of results are generated by the application of clutter suppression on real and synthetic data. A discussion follows each experimental section.

The experimental results for A-scan noise address these issues:

- Performance when additive noise is Gaussian and white.
- Performance when additive noise is Gaussian and colored.
- Evaluation of scale parameter and input noise energy.
- Compatibility of proposed method to trace stacking.

In addition to the 1D results, experimental results for B-scan clutter suppression will compare two methods, the proposed curvelet thresholding method and

average traces subtraction. The experiments present:

- Visual performance for synthetic data.
- Residual images for synthetic data.
- Peak SNR results for synthetic data.
- Visual performance for real data.

5.1 Experimental Results of Noise Removal in A-scans

The main results of this section are presented in Table 5.1. This experiment uses the best choice for a scale parameter determined *a priori*. The input noise is additive with a signal-to-noise ratio of 6dB. The signal is a simulated GPR pulse at 1000ps containing 2048 points.

	White Noise	Colored Noise
RDWT	17.8 dB	13.9 dB
DWT	13.5 dB	0.1 dB

Table 5.1: Summary of the average SNR_{imp} of the level-constant thresholded RDWT and global thresholded DWT with respect to white noise and colored noise.

The results show that the global thresholded DWT can not effectively remove colored noise, while the RDWT makes significant improvements. This establishes the suitability of level-constant RDWT thresholding as a method of noise removal. The next set of experiments evaluate the specific choice of parameters and the influence on this method.

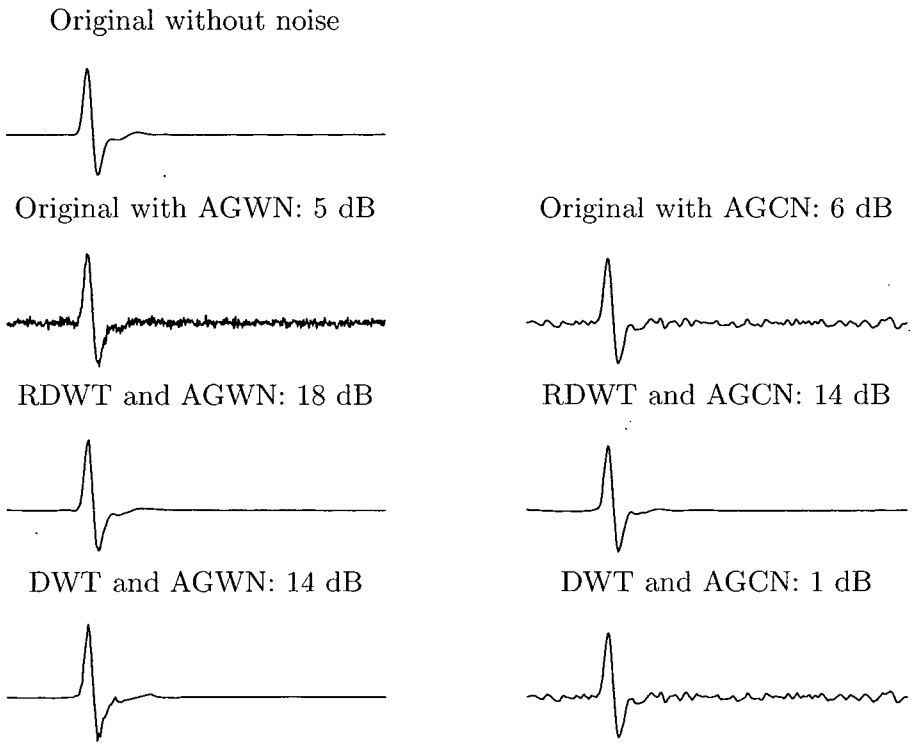


Figure 5.1: The RDWT visually removes the noise and preserves the signal using a level-constant threshold. The DWT with a global threshold can effectively remove the white noise, but not colored noise.

5.1.1 Performance of Gaussian White Noise Removal

This experiment uses the normalized risk measure from equation (3.1) to evaluate the performance of the RDWT to the DWT and Wiener filtering. This normalized risk, $\|f - \tilde{f}\|^2/N\sigma^2$, is used in the wavelet literature to show that thresholding exceeds linear methods of noise removal as the signal length, N, increases.

The experiment sets a constant value for σ and generates the input Gaussian white noise for a simulated GPR signal with increasing values of N. This makes SNR_{in} of the noise lower as N increases, because the energy of white noise is determined by $\|N\sigma^2\|$. The normalized risk is similar to a signal-to-noise measurement, but measures remaining noise energy on a scale of 0 to 1. A normalized risk measurement of one means that no noise energy is removed and a zero means all noise is removed, while preserving the integrity of the signal.

For each value of N, the experiment is conducted 20 times and the average risk is calculated according to (3.1). The DWT thresholding method is performed by estimating a global threshold from the noise variance of the finest scale coefficients. Wiener filtering is performed by the function included with Wavelab [24]. The RDWT threshold is determined by the level-constant method using an estimate of each level with $J = 8$.

The results show a decay for DWT thresholding which exceeds Wiener filtering, confirming the results presented in wavelet literature [48]. The RDWT removes more noise initially, but decays slightly slower then the DWT. This shows that the RDWT is more effective then the DWT at small values of N when using a level-constant threshold.

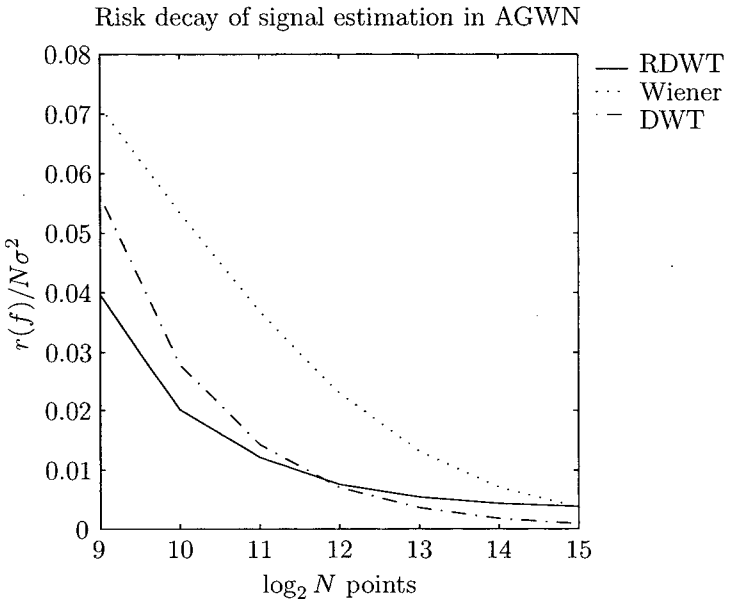


Figure 5.2: Level-constant thresholding of the noisy RDWT coefficients is effective even when the signal length is short. Level-constant thresholds are less effective when the signal length is large.

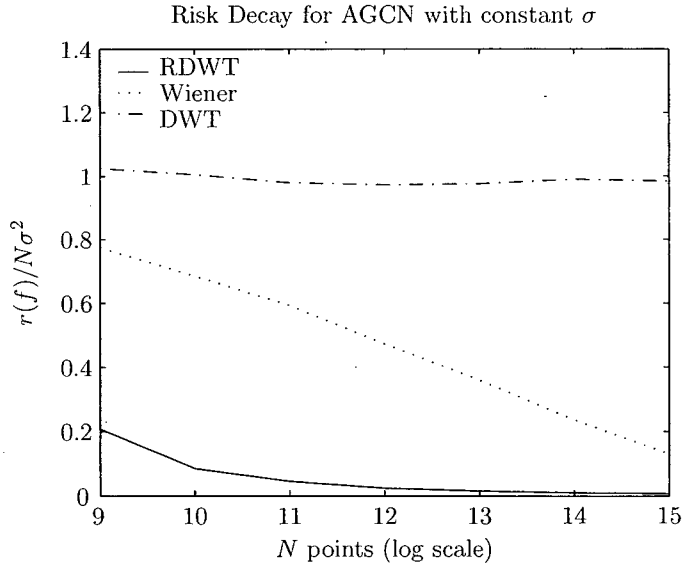


Figure 5.3: Global thresholding of the DWT removes no measurable noise, thus producing no risk decay. Wiener filtering is marginally effective for low noise, but level-constant thresholding is successful at removing colored noise.

5.1.2 Performance of Gaussian Colored Noise Removal

This experimental result demonstrates the noise removal ability of each competing method in the presence of "in-band" colored Gaussian noise. The experiment uses the normalized risk metric so that a comparison can be made to the previous results of AGWN noise removal.

The experiment is conducted in the same manner as the previous section using a synthetic GPR pulse with an increasing amount of points. The value of σ remained constant, while N increased. This choice reduces the amount of noise energy as the number of points increased. The noise is colored by the antenna model, h_n, and generated randomly for each trial. The results represent the average normalized risk for 20 trials.

There are three outstanding results in Figure 5.2. First, the DWT method of global thresholding appears to by unable to reduce the amount of noise, regardless of the number of points available. The RDWT method, however, still produces a decay in risk as N increases in the presences of GCN. Finally, Wiener filtering produces marginal results, while the risk for RDWT thresholding remains effective, though five times larger with respect to AGWN performance.

5.1.3 Dependencies of Number of Decomposition Levels

This experiment evaluates the dependence of the scale parameter and SNR_{in} using the SNR_{imp} metric. For Figure 5.4, the 2048 point simulated GPR signal is used to test the performance of the RDWT level-constant thresholding method in the presence of white noise.

The results show that as SNR_{in} decreases to zero, decomposing to a higher number of levels yields slightly better results. In the case of minimal noise, like 39dB,

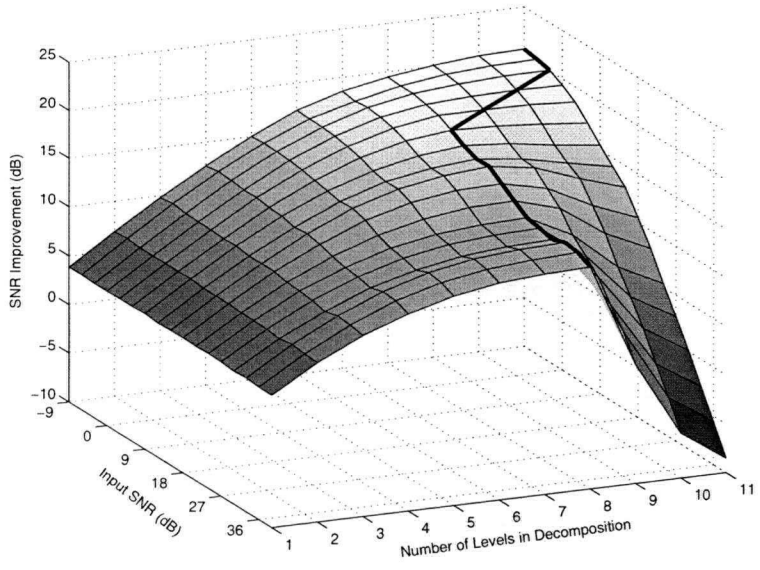


Figure 5.4: The thick black line shows the maximum SNR improvement for a given number of levels of decomposition and additive GWN. A notable feature of the figure is the sharp drop of SNR improvement when the noise energy is small.

the coefficients are not sparse at high levels and show a negative improvement at high values of J for AGWN.

The same test is performed on AGCN using the simulated GPR signal. The results are quite different for the larger values of SNR_{in}. The optimal choice of J remains the same, but the metric SNR_{imp} falls off as noise energy increases. The choice of in-band colored noise influences this drop, because the large noise events have similar characteristics to the GPR signal and thresholding becomes ambiguous. However, it is important to note that if clutter is modeled as in-band noise, then it would be suppressed by approximately 10dB.

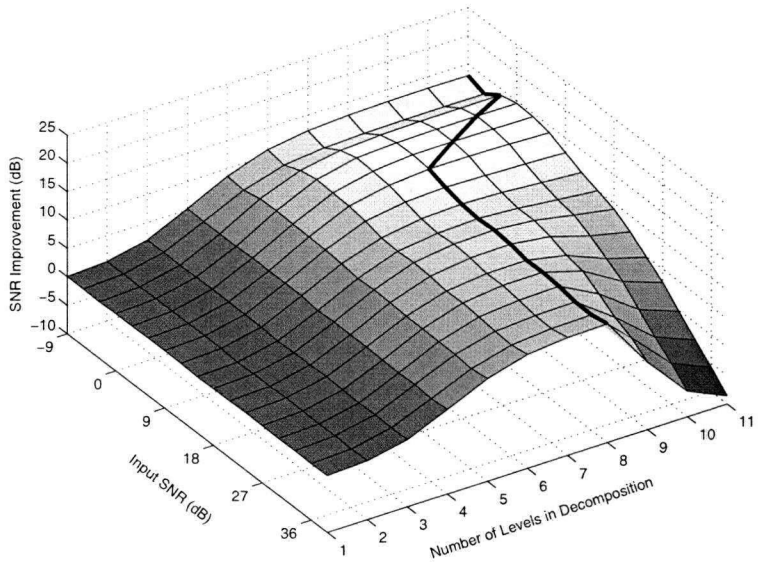


Figure 5.5: As the input SNR of the colored noise reaches 0dB, the ability to distinguish noise from signal diminishes, because noise events appear nearly identical to signal events, however the optimal choice of J remains the same.

5.1.4 Thresholding Multiple A-scans before Stacking

The first experiment is designed to test the efficacy of operator order. Thresholding is a non-linear operator, which means the order of operations will likely produce different results. The ubiquitous nature of stacking suggests that the order of thresholding and stacking should be examined. Figure 5.6 demonstrates the difference between the three different cases:

- Perform no wavelet thresholding, just stack.
- Perform wavelet thresholding before stacking.
- Perform wavelet thresholding after stacking.

The noise is AGWN applied to a 2048 synthetic GPR signal. Wavelet denoising is performed using a level-constant RDWT threshold. The input noise is 9dB and

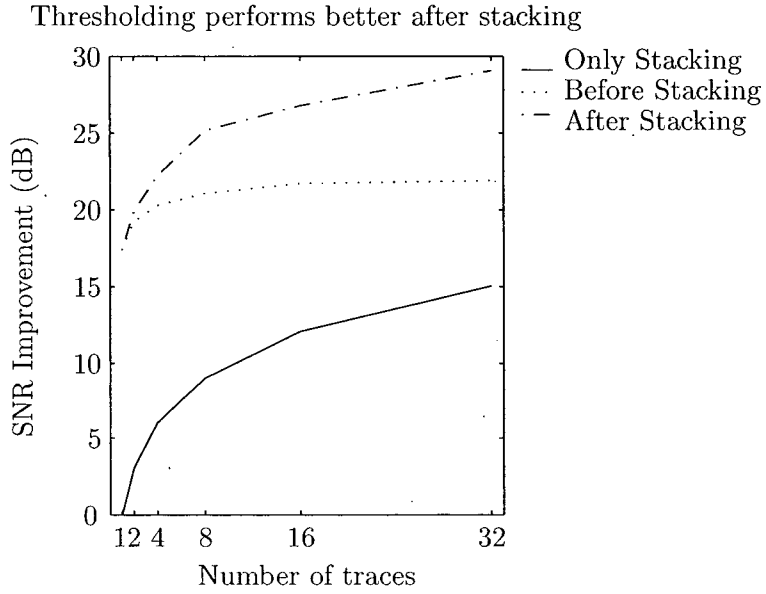
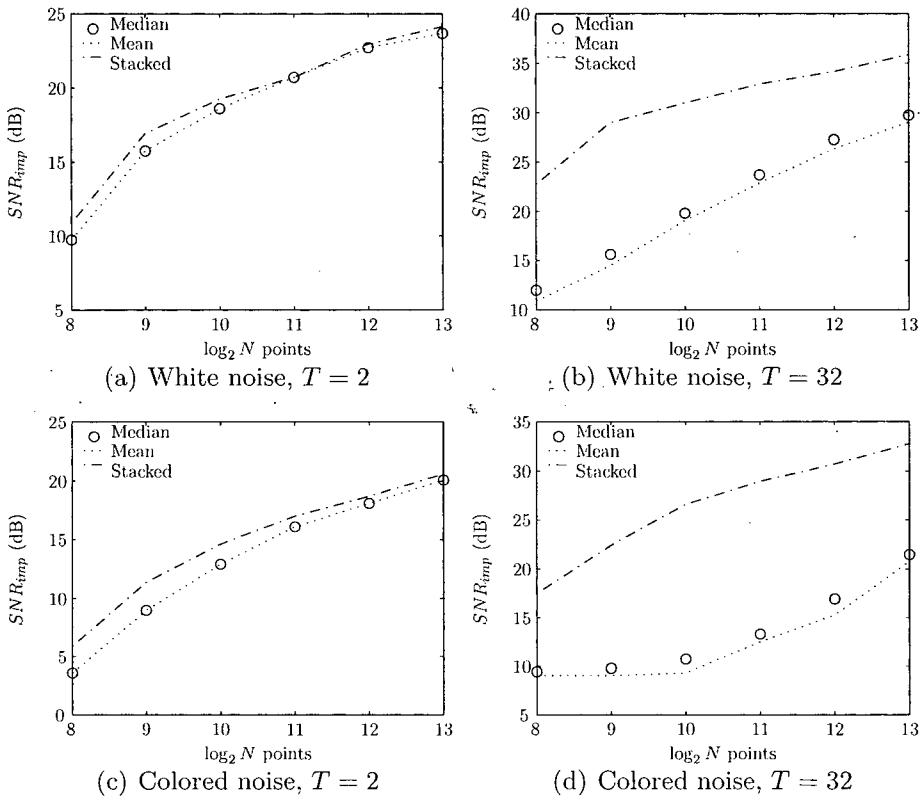


Figure 5.6: Stacking scales the input noise σ by $1/\sqrt{T}$, however, this scaling is diminished by stacking before thresholding. Regardless of the order, wavelet noise removal still provides a significant improvement.

$J = 8$, which corresponds to the optimal scale parameter from Figure 5.4. The SNR_{imp} metric is chosen to determine the amount of noise energy removed.

The results indicate that level-constant thresholding should be performed after stacking. This is an important and fortunate feature, because most GPR data is only available post-stack. The experiment does not test the assumption that each trace is aligned in time and phase, which is a basic assumption of stacking. However, if only one trace is available the results show that wavelet thresholding can still provide noise removal, whereas stacking cannot be performed.

The second experiment confirms that the number of points, N, of a trace and the noise coloring has little effect on the results presented in Figure 5.6. This set of experiments tested four cases: AGWN with 2 and 32 input traces and AGCN with 2 and 32 traces. Each case was tested as the number of points in the input



(a) White noise, $T = 2$ (b) White noise, $T = 32$

(c) Colored noise, $T = 2$ (d) Colored noise, $T = 32$

Figure 5.7: Stacking before thresholding is the better. If stacking is performed in the wavelet domain, the median is an improvement over the mean when $T = 32$.

signal varied from 256 to 32,768. In addition to testing the order of stacking and thresholding, a modification to the stacking operator was made. The median operator is used to perform a different estimate of a mean trace after thresholding in the wavelet domain. The median operator is more suitable to determining the mean if a probability distribution function is skewed. The nature of thresholding suggests that the efficiency of median stacking should be tested.

The results in Figure 5.7 show that as the number of points decreases the difference between pre-stacking and post-stacking shrinks. The median stacking operator offers no advantage for a small number of traces, but 3dB of improvement

in large stacks. Minor differences aside, noise coloring has no effect on the order of operations. Even in the presence of colored noise it is better to pre-stack GPR data.

5.2 Discussion of Noise Removal Results

The results show that the level-constant thresholded RDWT is useful for denoising a synthetic GPR signal in Gaussian noise. Several aspects are tested and the scale parameter, J, is shown to have a dramatic effect on the ability to perform noise removal. Noise coloring is shown to reduce the performance of level-constant thresholding, although performance still exceeds other methods tested. The order of operations is shown to be best performed by finding the stacked trace before wavelet denoising.

The performance difference of the level-constant thresholded RDWT between white and colored noise is due to the ability to estimate the noise in the redundant wavelet domain. Using the redundant transform, a different estimate of $\tilde{\sigma}$ is obtained at each level. In Figure 5.3, we can see that this allows the threshold to adapt to the coloring of the noise. With the understanding that the RDWT is the application of bandpass filters, the level-constant threshold is an estimate of the standard deviation for a particular frequency band of the additive colored noise, which allows it to adapt to the coloring of the noise.

This choice of level-constant thresholding has limitations, because if the estimation is wrong, the corresponding threshold will not be effective. The results from testing the scale parameter imply that when the noise is especially low, the threshold operator removes more signal energy then noise. The extreme case is in Figure 5.4, where the result of denoising makes the corresponding signal worse then the input.

When comparing RDWT denoising to Wiener filtering and DWT thresholding in Figure 5.2, the normalized risk of the RDWT shows better performance initially then either method. As N increases, the difference shrinks and DWT denoising overtakes the RDWT method. This is attributed to the choice of level-constant thresholding of the RDWT. The median, which is used to estimate the threshold for each level, is no longer efficient for low noise and large N. The level constant threshold attempts to estimate the standard deviation of the noise at the low frequency levels where the sparsity property no longer holds. This means the threshold predicts a large value for $\tilde{\sigma}_j$, which removes more signal then noise at that level. For white noise, it would be better to use a global threshold estimated from the finest scale coefficients of the RDWT for very low noise applications.

When applying thresholding techniques to GPR data, results show that thresholding should be performed after trace averaging. The reason is that an estimate of $\tilde{\sigma}$ for pre-stacked traces will be larger then post-stack. Because the threshold is dependent on $\tilde{\sigma}$, it will be larger and remove large amounts of signal energy. When the traces are averaged the value of σ is reduced by a factor of $1/\sqrt{T}$. The corresponding estimate of $\tilde{\sigma}$ is smaller and the threshold is smaller. This only holds when the distribution is Gaussian, because the mean is an excellent point estimator.

When N is increased the difference between pre- and post-stacking is reduced, especially when $T = 32$. This is because the estimation of the threshold is improved and quality of noise removal increases. However, the difference can not overcome the tremendous gain of pre-stacking, which makes thresholding more effective. However, this result does not take into account that a threshold could be estimated for each coefficient in the redundant domain. A level-constant threshold

assumes all coefficients at a particular level have a similar standard deviation. When only one trace is available this assumption is reasonable, but as the number of traces increases better estimations of $\tilde{\sigma}_m$ are possible.

The benefits of level-constant thresholding seem to outweigh the negatives. Signals with small number of points are better estimated by the RDWT method then either DWT or Wiener filtering. Often for real GPR systems the number of points is under 2048, which increases the advantage of level-constant thresholding in the RDWT domain. This implies the level-constant RDWT threshold would be a candidate for the removal of Gaussian noise, white or colored of GPR A-scans.

5.3 Simulation of Clutter Suppression in B-scans

The following experimental results compare the baseline method of average trace subtraction (ATS) to a the proposed method of thresholding in the curvelet domain. The methods are compared on synthetic data to establish an understanding of the proposed method and limitations. Real data is then tested to see the visual effects of the proposed method.

The results of the synthetic tests are summarized by Table 5.2. This experiment compares the effectiveness of ATS to the proposed method using both a point target and a large target. While a PSNR metric is not a visual indicator of quality, it shows the amount of noise energy removed. The results is that the proposed method removes more clutter energy then the ATS does.

The remaining experiments focus on visual quality. The output of each algorithm and the residuals for the synthetic data are compared. Following the synthetic test, a test on real landmine data is presented.

	Original	Average Trace Subtraction	Proposed
Point Target	17.6	21.8	40.2
Large Target	17.6	21.3	40.1

Table 5.2: The peak-SNR of ATS shows an improvement, but much less then the proposed method for both targets.

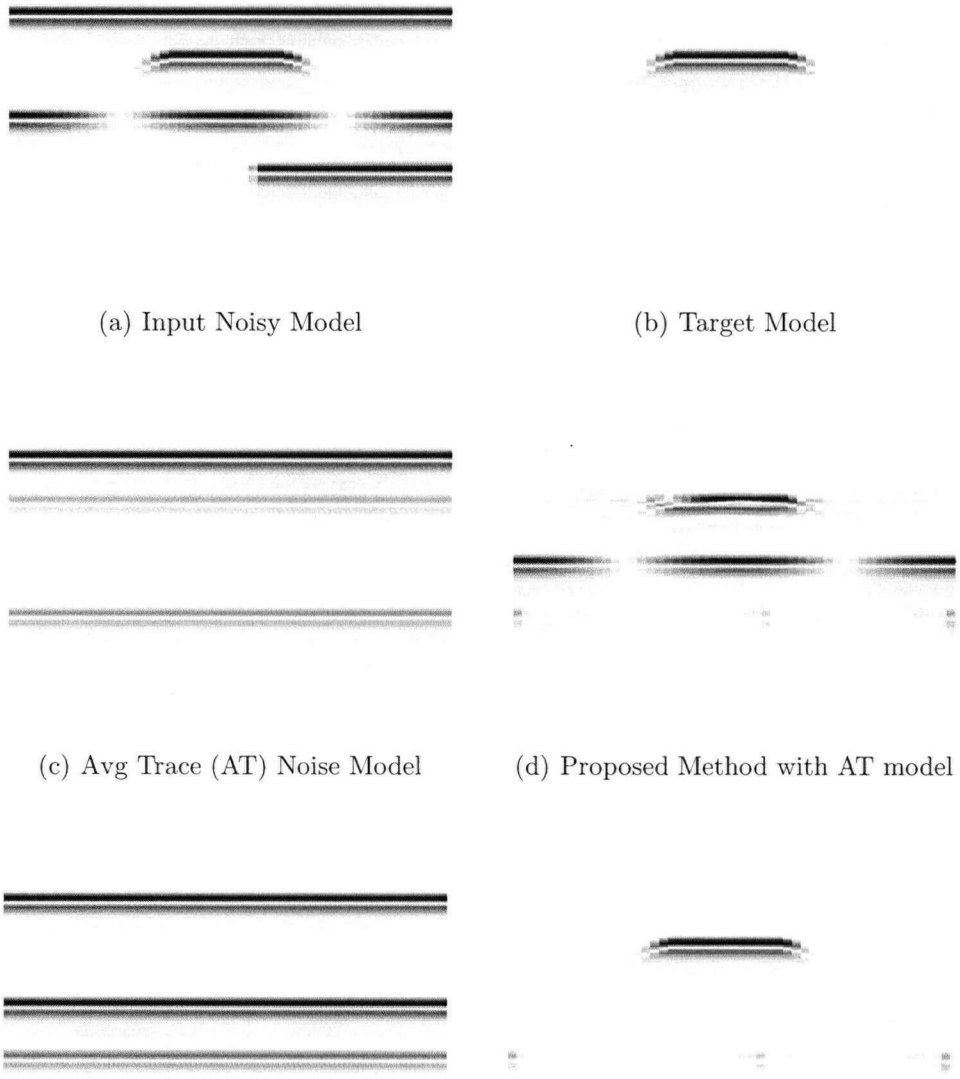
5.3.1 Synthetic B-scan Experiments

The visual synthetic experiments are conducted to establish an understanding of the method. The first experiment explores the effect of choosing a noise model. The average trace subtraction method assumes a noise model that is an image containing the average trace at every spatial point. This experiment compares visual output of ATS and then performs the proposed method also using an average trace (AT) model. As a comparison, the edge noise model is also presented and the results are shown for the proposed method.

The experiment uses a large target, which penalizes the use of an average noise model. Figure 5.8(b) shows that the target energy has been damaged, and additional artifacts have been introduced. The average noise model is not a good choice for the proposed method either. Target energy is also damaged by the thresholding, but no additional artifacts are introduced.

The next experiment in Figure 5.9 uses a point target and the same clutter model to compare ATS and the the residuals. The residuals emphasize that ATS cannot remove clutter which varies in amplitude or phase. However, thresholding is not subjected to that limitation and effectively removes all clutter events. The residual of the proposed method contains only artifacts at the sharp edges of the clutter. The ground, having constant amplitude, is easily removed by both methods.

The final synthetic experiment compares the large target to both methods.



(a) Input Noisy Model

(b) Target Model

(c) Avg Trace (AT) Noise Model

(d) Proposed Method with AT model

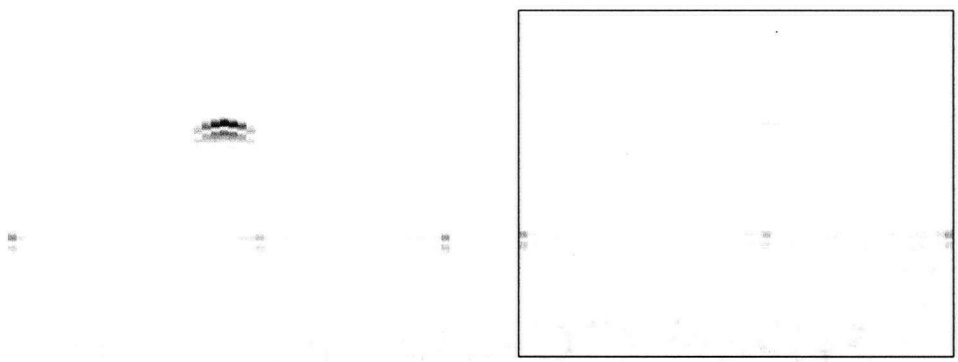
(e) Edge Noise Model

(f) Proposed Method with edge model

Figure 5.8: In 5.8(a) the original data with clutter is presented, but the result should be just the target as in 5.8(b). The proposed method is calculated with an average noise model, 5.8(d) and edge noise model, 5.8(f). To understand the output of the proposed method, the noise model used for the threshold estimation is shown in 5.8(c) and 5.8(e).

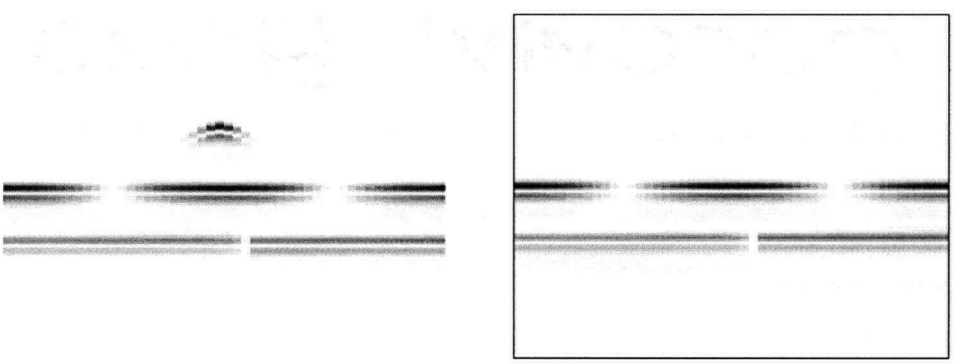
(a) Target Model

(b) Input Noisy Model



(c) Proposed Method Result

(d) Proposed Method Residual



(e) Average Trace Subtraction Result

(f) Average Trace Subtraction Residual

Figure 5.9: In 5.9(d) the proposed method leaves virtually no clutter energy and the target is still present. ATS leaves tremendous artifacts, but does manage to remove the ground clutter.

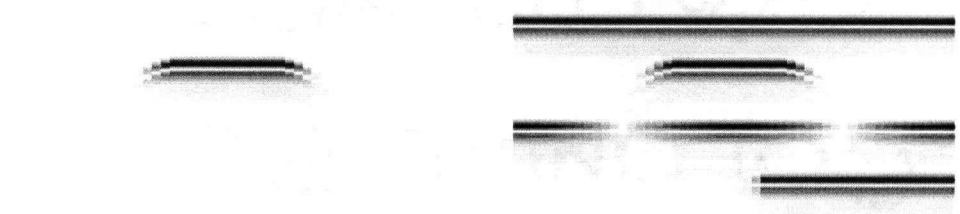
This provides results for common subsurface targets, which are larger than point targets. With these common targets, using ATS shows that significant artifacts are introduced, seen in Figure 5.10(e). The proposed method in Figure 5.10(c) demonstrates similar results for the large target as the point target. The method provides near complete removal of all clutter and very little change to the target energy.

5.3.2 Landmine B-scan Experiments

This experiment uses landmine data from the JRC landmine signatures database. The B-scan under study represents a physical location at 25cm from a survey that done on 50cm X 50cm area with 1cm resolution. The landmine is a PMN-type anti-personnel device buried at 5cms in loamy soil. The data can be obtained from [11] and the system is described in [59].

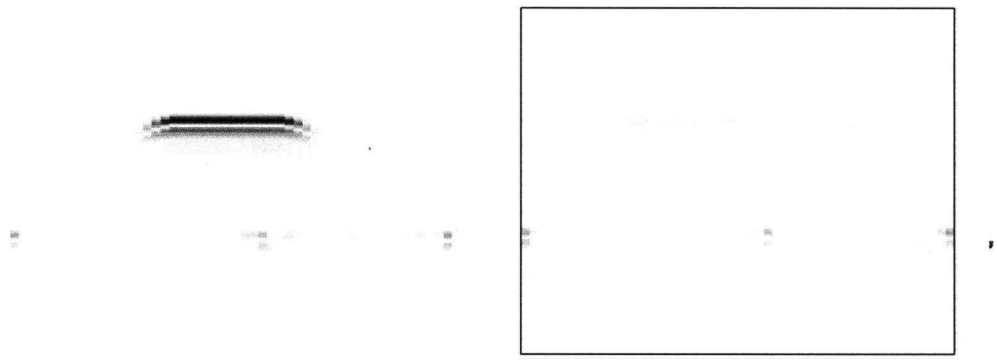
The results of clutter suppression are shown in Figure 5.11, which use the edge noise model shown in 5.11(b) to create a threshold operator. Figure 5.11(b) is the edge noise model calculated from the data domain. Any event that is strongly horizontal and exists near the edge will be shown in the noise model. The result from the proposed method will be to remove any events that are approximated by the calculated noise model. The result is Figure 5.11(c), which indeed shows superior clutter suppression to the average trace subtraction method.

As demonstrated before, targets that are large suffer when ATS is used to suppress clutter. There are artifacts introduced around the target as a result of the average noise model. This does not occur with the proposed method. The large target remains intact and becomes the prominent feature in the result. Also note the ground clutter is not completely removed by either method.



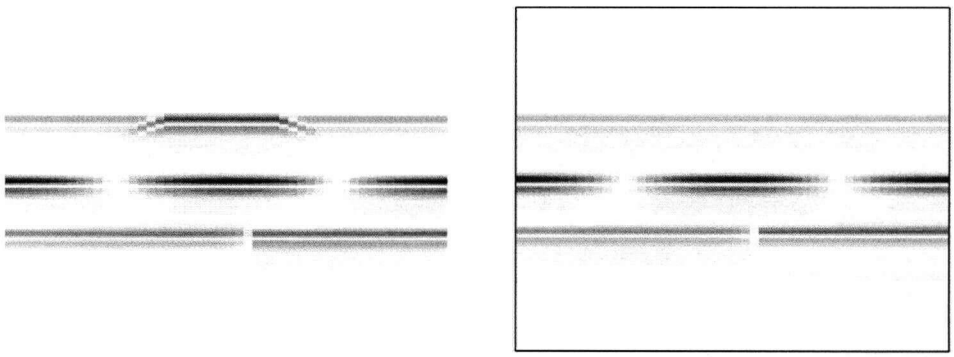
(a) Target Model

(b) Input Noisy Model



(c) Proposed Method Result

(d) Proposed Method Residual



(e) Average Trace Subtraction Result

(f) Average Trace Subtraction Residual

Figure 5.10: The large target does not effect the ability of the proposed method to remove clutter. Average trace subtraction removes a portion of target energy, adds artifacts near the target and does not remove the amplitude varying clutter.

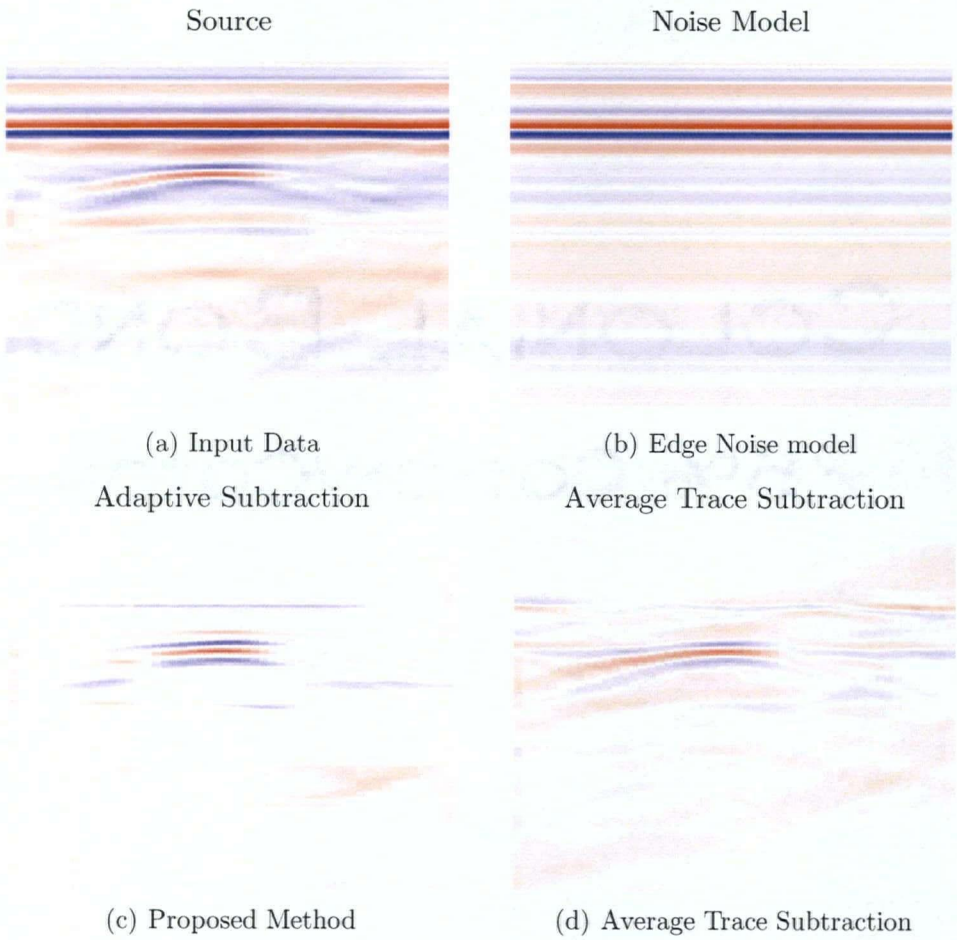


Figure 5.11: The clutter is suppressed in the proposed method, whereas significant artifacts are introduced in the ATS method. The noise model in 5.11(b) shows what likely events will be suppressed in the result of the proposed method.

5.4 Discussion of Clutter Suppression Results

Developing an understanding of clutter is a difficult problem, therefore simple solutions are sought. The computational complexity and simplicity of ATS is low, which has led to the widespread use of ATS as an initial signal processing step. Also, ATS is effective at removing the air-ground event and other strong horizontal events. In a usual GPR B-scan, the horizontal clutter is the dominant visual feature and can often impede interpretation of the data. The largest horizontal event is the ground, but multiple reflections of the ground, antenna ringing and horizontal reflectors are also significant.

The advantage of using the proposed method is that events only need to be roughly approximated. This is similar to the ATS assumption that clutter can be represented by an average of all traces. The drawback of simple noise models is that they are likely to fail if the broad assumptions are invalidated. In particular, the edge noise model used in the proposed method suffers if the clutter does not exist in the edge trace. The edge model also can not handle temporal changes in the clutter. In Figure 5.11, an example where slight variation in the temporal location of the ground begins to challenge the assumptions of the edge noise model. However, the proposed method is visually more successful at removing horizontal events and minimizing artifacts then the ATS method.

The point target, in Figure 5.9 is well fit to ATS clutter suppression because the horizontal component is minimal. The advantage of the noise edge model and the proposed method are evident in Figure 5.8(f) and Figure 5.10(c). The proposed method does not require the use of an average trace noise model, therefore no target damage occurs in Figure 5.10(c). When the average trace model is used on the same example, in Figure 5.8(f), no additional artifacts are introduced. The ATS method

struggles with this, because the subtraction operator is blind to any changes in data. The proposed method uses a threshold operator, which provides a quasi-adaptive method. This prevents the proposed method from adding artifacts even when the estimation of the noise model is wrong.

The drawback of using the proposed method is that the scaling parameter λ must be expertly controlled to balance the weighting of the noise model. The results presented here used a weight of 2.8, which corresponds to thresholding any coefficients that exceed 3 times the noise model coefficients. It was found in these experiments that the scaling parameter only required rough adjustments and was easy to understand and control. In the case of the landmine data, a much less aggressive setting was used. Sufficient clutter suppression was obtained by only scaling the noise model 1.8 times.

Chapter 6

Conclusions

This thesis made two primary contributions, which furthered an understanding of thresholding in the wavelet and curvelet domain. The first contribution is an evaluation of level-thresholding the RDWT coefficients for noise removal in a GPR A-scan. The second evaluates the use of thresholding curvelets for clutter suppression in a GPR B-scan.

There are four secondary contributions that stemmed from the results of the evaluation of removing noise using level-constant thresholding in the RDWT domain:

- RDWT level-constant thresholding is suitable for use on the synthetic GPR signal presented here, and likely suitable for most GPR signals.
- The scale parameter, which controls the number of decomposition levels is critical to the performance of a level-constant threshold method.
- The amount of noise energy present in the noisy signal affects the performance and choice of the scale parameter.

- The order of operations, if stacking is performed, is to stack first and threshold afterwards.

This work has built on the success of previous work in wavelet thresholding. Previous achievements presented results on global thresholding the RDWT to remove white noise. These results have extended the use of RDWT to perform non-parametric signal regression using very little *a priori* information. The threshold is determined by the data, but the scale parameter must be estimated by other means.

This work has also presented a reasonable physical model for a GPR A-scan. Using this model and other practical aspects of GPR technology the results can interpreted and applied to real data. Specifically, the knowledge of stacking before thresholding is useful, because the RDWT denoising can be applied directly to current GPR data.

In the realm of clutter suppression, there are three secondary contributions that stemmed from evaluating the proposed method of thresholding an edge noise model in the curvelet domain:

- Clutter suppression can be performed by thresholding in the curvelet domain.
- An approximation of clutter using a simple model, such as the edge noise model is sufficient when using thresholding.
- Amplitude and phase resilience was demonstrated using thresholding.

This work uses the well established theory of thresholding in the wavelet domain to re-position the problem of clutter suppression as threshold estimation problem. Because of the phase and amplitude resilience of thresholding effective clutter suppression can be achieved by only estimating an approximate model of

clutter. This enables users to arbitrarily design a clutter model that suits a specific need.

6.1 Future Work

This work can be extended in three different ways. The first is to perform a physical evaluation of level-constant thresholding to accurate backscatter models. Second, complex clutter models could be created to enhance the effectiveness of clutter thresholding. Finally, combining both level-constant thresholding and clutter suppression in the curvelet domain to perform noise removal and clutter suppression simultaneously.

The most important extension is the combination of noise removal and clutter suppression. The suitability of methods has been evaluated, therefore the combination seems quite natural and exciting. Because the curvelet domain functions in generally the same way as the redundant discrete wavelet domain, this extension should be quite natural. The scale parameters will have to be carefully chosen such that signal preservation and noise removal is balanced.

The evaluation of physical data with respect to an accurate radar backscatter model is important to determine if the critical features of the signal are being preserved when noise removal takes place. In addition, a physical analysis of random clutter would allow the possibility of extending the idea of colored Gaussian noise to clutter models.

The development of complex clutter models could allow temporal clutter variance to be accounted for. The possibility of an interactive noise model could allow an operator to "select" features for removal. A clutter feature could be first identified by a user and then the angle and area of influence is then chosen to create

a noise model and suppress the clutter feature using thresholding.

Bibliography

[1] G. Beylkin. On the Representation of Operators in Bases of Compactly Supported Wavelets. *Siam Journal on Numerical Analysis*, 29(6):1716–1740, 1992.

[2] Ilaria Bottigliero. *120 Million Landmines Deployed Worldwide: Fact or Fiction?* Leo Cooper, Pen & Sword Books, Barnsley, South Yorkshire, England, 2000.

[3] J. W. Brooks. Applications of GPR Technology to Humanitarian Demining Operations in Cambodia: Some Lessons Learned. In *Internation Symposium on Technology and the Mine Problem*, Monterey, CA, 1998.

[4] John W. Brooks, Luc van Kempen, and Hichem Sahli. Primary study in adaptive clutter reduction and buried minelike target enhancement from GPR data. In *Detection and Remdiation Technologies for Mines and Minelike Targets V*, volume 4038 of *Proceedings SPIE*, pages 1183–1192, Orlando, FA USA, 2000. SPIE.

[5] C. Bruschini and B. Gros. A Survey of Current Technology Research for the Detection of Landmines. *Sustainable Humanitarian Demining: Trends, Techniques and Technologies*, pages 172–187, 1998.

[6] G.J. Burton and G.P. Ohlke. Exploitation of Millimeter Waves for Through-Wall Surveillance during Military Operations in Urban Terrain. Technical report, Royal Military College of Canada, May 24, 2000 2000.

[7] E. J. Candes and D. L. Donoho, editors. *Curvelets - A suprisingly Effective Nonadaptive Representation for Objects with Edges.* Curves and Surfaces. Vanderbilt University Press, Nashville, TN, 1999.

[8] E. J. Candes and D. L. Donoho. Curvelets and curvilinear integrals. *Journal of Approximation Theory*, 113(1):59–90, 2001.

[9] E. J. Candes and D. L. Donoho. New tight frames of curvelets and optimal representations of objects with piecewise C-2 singularities. *Communications on Pure and Applied Mathematics*, 57(2):219–266, 2004.

[10] Emmanuel Candes, Laurent Demanet, David Donoho, and Lexing Ying. CurveLab 1.0, 2005. Available from: `http://www.curvelet.org/`.

[11] Joint Research Centre. Anti-personnel Landmine Signatures Database. Available from: `http://apl-database.jrc.it/`.

[12] M. Cherniakov and L. Donskoi. Frequency band selection of radars for buried object detection. *Ieee Transactions on Geoscience and Remote Sensing*, 37(2):838–845, 1999.

[13] R. R. Coifman and D. L. Donoho. Translation Invariant De-Noising. In *Wavelets and Statistics*, volume Lecture Notes in Statistics, pages 125–150. Springer-Verlag, New York, 1995.

[14] R. R. Coifman and M. V. Wickerhauser. Entropy-Based Algorithms for Best Basis Selection. *Ieee Transactions on Information Theory*, 38(2):713–718, 1992.

[15] Lawrence B. Conyers and Dean Goodman. *Ground-penetrating radar : an introduction for archaeologists.* AltaMira Press, Walnut Creek, CA, 1997.

[16] Jamin Cristall, Moritz Beyreuther, and Felix Herrmann. Curvelet processing and imaging: 4D adaptive subtraction. In *CSEG National Convention.* 2004.

[17] D. J. Daniels, D. J. Gunton, and H. F. Scott. Introduction to Subsurface Radar. *Iee Proceedings-F Radar and Signal Processing*, 135(4):278–320, 1988.

[18] D. J. Daniels and Institution of Electrical Engineers. *Surface-penetrating radar.* IEE radar, sonar, navigation and avionics series ; 6. Institution of Electrical Engineers, London, 1996.

[19] D. J. Daniels and Institution of Electrical Engineers. *Ground penetrating radar.* IEE radar, sonar, navigation, and avionics series ; 15. Institution of Electrical Engineers, London, 2nd edition, 2004.

[20] Canadian Department of National Defense. Canadian Forces Landmine Database, Dec 18, 2003. Available from: `http://ndmic-cidnm.forces.gc.ca/index.asp?lang=e`.

[21] D. L. Donoho. De-Noising by Soft-Thresholding. *Ieee Transactions on Information Theory*, 41(3):613–627, 1995.

[22] D. L. Donoho and M. Elad. Optimally sparse representation in general (nonorthogonal) dictionaries via l(1) minimization. *Proceedings of the National Academy of Sciences of the United States of America*, 100(5):2197–2202, 2003.

[23] D. L. Donoho and I. M. Johnstone. Ideal Spatial Adaptation by Wavelet Shrinkage. *Biometrika*, 81(3):425–455, 1994.

[24] David Donoho, Mark Reynold Duncan, Xiaoming Huo, and Ofer Levi. Wavelab 802. Available from: `http://www-stat.stanford.edu/ wavelab/`.

[25] M.; Yunxin Zhao Gader, P.D.; Mystkowski. Landmine detection with ground penetrating radar using hidden Markov models. *Geoscience and Remote Sensing, IEEE Transactions on*, 39(6):1231–1244, 2001.

[26] P. D. Gader, B. N. Nelson, H. Frigui, G. Vaillette, and J. M. Keller. Fuzzy logic detection of landmines with ground penetrating radar. *Signal Processing*, 80(6):1069–1084, 2000.

[27] J. Gazdag. Wave-Equation Migration with Phase-Shift Method. *Geophysics*, 43(7):1342–1351, 1978.

[28] Rafael C. Gonzalez and Richard E. Woods. *Digital image processing.* Prentice Hall, Upper Saddle River, N.J., 2nd edition, 2002.

[29] R. Gribonval and M. Nielsen. On the strong uniqueness of highly sparse representations from redundant dictionaries. *Independent Component Analysis and Blind Signal Separation*, 3195:201–208, 2004.

[30] A. Grossmann and J. Morlet. Decomposition of Hardy Functions into Square Integrable Wavelets of Constant Shape. *Siam Journal on Mathematical Analysis*, 15(4):723–736, 1984.

[31] Digital Signal Processing Group. Rice Wavelet Toolbox v2.4, 2001. Available from: `http://www-dsp.rice.edu/software/rwt.shtml`.

[32] Frdric Guerne, Bertrand Gros, Marc Schreiber, and Jean-Daniel Nicoud. GPR Mine Sensors for Data Acquisition in the Field. In *International Workshop on Sustainable Humanitarian Demining*, Zagreb, Croatia, 1997.

[33] A. Guitton and D. J. Verschuur. Adaptive subtraction of multiples using the L-1-norm. *Geophysical Prospecting*, 52(1):27–38, 2004.

[34] Felix J. Herrmann. Curvelet imaging and processing: an overview. In *CSEG National Convention.* 2004.

[35] F.J. Herrmann and P.P. Moghaddam. Sparseness- and continuity-constrained seismic imaging with Curvelet frames. *Submitted for publication*, 2005.

[36] F.J. Herrmann and D. J. Verschuur. 3-D Curvelet Domain multiple elimination with sparseness constraints. *Submitted for publication*, 2005.

[37] K.C. Ho, P.D. Gader, and J.N. Wilson. Improving landmine detection using frequency domain features from ground penetrating radar. In *Geoscience and Remote Sensing Symposium, 2004. IGARSS '04. Proceedings. 2004 IEEE International*, volume 3, pages 1617–1620 vol.3, 2004.

[38] Tai-Chiu Hsung, D.P.-K. Lun, and Wan-Chi Siu. Denoising by singularity detection. *Signal Processing, IEEE Transactions on [see also Acoustics, Speech, and Signal Processing, IEEE Transactions on]*, 47(11):3139–3144, 1999.

[39] Sensors & Software Inc. EKKO-for-DVL pulseEKKO 100 Manual, 2001.

[40] A. Instanes, I. Lonne, and K. Sandaker. Location of avalanche victims with ground-penetrating radar. *Cold Regions Science and Technology*, 38(1):55–61, 2004.

[41] J. D. Irving and R. J. Knight. Removal of wavelet dispersion from ground-penetrating radar data. *Geophysics*, 68(3):960–970, 2003.

[42] C. Jaedicke. Snow mass quantification and avalanche victim search by ground penetrating radar. *Surveys in Geophysics*, 24(5-6):431–445, 2003.

[43] J. van der Kruk, C. P. A. Wapenaar, J. T. Fokkema, and P. M. van den Berg. Three-dimensional imaging of multicomponent ground-penetrating radar data. *Geophysics*, 68(4):1241–1254, 2003.

[44] M. Lang, H. Guo, J. Odegard, C. Burrus, and R. Wells. Nonlinear Processing of a Shift Invariant DWT for Noise Reduction. In *SPIE Symp. OE/Aerospace Sensing and Dual Use Photonics, Algorithm for Synthetic Aperture Radar Image*, Orlando, FL, 1995. SPIE.

[45] M. Lang, H. Guo, J. E. Odegard, C. S. Burrus, and R. O. Wells. Noise reduction using an undecimated discrete wavelet transform. *Ieee Signal Processing Letters*, 3(1):10–12, 1996.

[46] K. H. Lee and G. Q. Xie. A New Approach to Imaging with Low-Frequency Electromagnetic-Fields. *Geophysics*, 58(6):780–796, 1993.

[47] S. Mallat and S. Zhong. Characterization of signals from multiscale edges. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 14(7):710–732, 1992.

[48] S. G. Mallat. *A wavelet tour of signal processing.* Academic Press, San Diego, 2nd edition, 1999.

[49] S. G. Mallat and Z. F. Zhang. Matching Pursuits with Time-Frequency Dictionaries. *Ieee Transactions on Signal Processing*, 41(12):3397–3415, 1993.

[50] Adam Marcus. Inventions to Lift the Scourge of Landmines, May 6, 1998 1998. Available from: `http://www.csmonitor.com/durable/1998/05/06/f-p14s1.htm`.

[51] Rohan Maxwell. Cambodia: A Country Profile. *Journal of Mine Action*, 5(1), 2001.

[52] M. R. McClure and L. Carin. Matching pursuits with a wave-based dictionary. *Ieee Transactions on Signal Processing*, 45(12):2912–2927, 1997.

[53] I. L. Morrow and P. van Genderen. Effective imaging of buried dielectric objects. *Ieee Transactions on Geoscience and Remote Sensing*, 40(4):943–949, 2002.

[54] G. P. Nason and B. W. Silverman. The Stationary Wavelet Transform and some Statistical Applications. In A Antoniadis and G. Oppenheim, editors, *Wavelets and Statistics*, volume Lecture Notes in Statistics, pages 281–299. Springer-Verlag, New York, 1995.

[55] L. Nuzzo and T. Quarta. Improvement in GPR coherent noise attenuation using tau-p and wavelet transforms. *Geophysics*, 69(3):789–802, 2004.

[56] J. C. Pesquet, H. Krim, and H. Carfantan. Time-invariant orthonormal wavelet representations. *Ieee Transactions on Signal Processing*, 44(8):1964–1970, 1996.

[57] Aleksandra Pizurica. *Image Denoising Using Wavelets and Spatial Context Modeling.* Ph.d, Univeristy of Gent, 2002.

[58] Jane Mary Anastasia Rea. *Ground penetrating radar applications in hydrology.* University of British Columbia, Vancouver, 1997.

[59] B. Scheers, M. Acheroy, and A. Vander Vorst. Time-domain simulation and characterisation of TEM horns using a normalised impulse response. *Iee Proceedings-Microwaves Antennas and Propagation*, 147(6):463–468, 2000.

[60] Bart Scheers. *Ultra-Wideband Ground Penetrating Radar, With Application to the Detection of Anti-Personnel Landmines.* Ph.d, Royal Military Academy, 2001.

[61] US Department of State. Hidden Killers: The Global Landmine Crisis. Deparment of State Publication 10575, Bureau of Political-Military Affairs, Office of Humanitarian Demining Progams, September 1998 1998.

[62] James D. Taylor. *Ultra-wideband radar technology.* CRC Press, Boca Raton, FL, 2001.

[63] S. Tjora, E. Eide, and L. Lundheim. Evaluation of methods for ground bounce removal in GPR utility mapping. In *Ground Penetrating Radar, 2004. GPR 2004. Proceedings of the Tenth International Conference on*, volume 1, pages 379–382, 2004.

[64] J. A. Tropp. Greed is good: Algorithmic results for sparse approximation. *Ieee Transactions on Information Theory*, 50(10):2231–2242, 2004.

[65] T. J. Ulrych, M. D. Sacchi, and J. M. Graul. Signal and noise separation: Art and science. *Geophysics*, 64(5):1648–1656, 1999.

[66] T. J. Ulrych, D. R. Velis, A. D. Woodbury, and R. D. Sacchi. L-moments and C-moments. *Stochastic Environmental Research and Risk Assessment*, 14(1):50–68, 2000.

[67] L. Van Kempen, H. Sahli, E. Nyssen, and J. Cornelis. Signal processing and pattern recognition methods for radar AP mine detection and identification. In *Detection of Abandoned Land Mines, 1998. Second International Conference on the (IEE Conf. Publ. No. 458)*, pages 81–85, 1998.

[68] R. Wu, A. Clement, J. Li, E.G. Larsson, M. Bradley, J. Habersat, and G. Maksymonko. Adaptive ground bounce removal. *Electronics Letters*, 37(20):1250–1252, 2001.

[69] Xu Xiaoyin, E.L. Miller, C.M. Rappaport, and G.D. Sower. Statistical method to detect subsurface objects using array ground-penetrating radar data. *Geoscience and Remote Sensing, IEEE Transactions on*, 40(4):963–976, 2002.

[70] Y. X. Zhao, P. Gader, P. Chen, and Y. Zhang. Training DHMMs of mine and clutter to minimize landmine detection errors. *Ieee Transactions on Geoscience and Remote Sensing*, 41(5):1016–1024, 2003.

[71] M. S. Zhdanov, P. Traynin, and J. R. Booker. Underground imaging by frequency-domain electromagnetic migration. *Geophysics*, 61(3):666-682, 1996.

Appendix A

Image Plotting Function

Due to the subjectivity of plotting intensity images, a detailed description of the plotting function which was used to generate B-scan plots is described. The important features are the color scaling, image scaling and image translation. The function used to re-scale and translate is called `rel_im`, shortened from, *relative image*. The purpose was to allow results to be plotted at the same intensity scale as the input.

MATLAB includes a image plotting function called `image`, which interprets an entry in a matrix as an index to a *colormap*. The colormap used in this thesis for plotting B-scans is white at the middle index of the colormap matrix and linearly scales in shades of gray to black at the beginning and end of the array using the code:

```
function [cmap] = seisgray;
  if nargin <1,
    ncolor = 32;
  end
  tmp1 = [[0:(ncolor-1)]/(ncolor-1) ; ...
          [0:(ncolor-1)]/(ncolor-1) ; ...
          [0:(ncolor-1)]/(ncolor-1)]';
```

```
tmp2 = [[(ncolor-1):-1:0]/(ncolor-1) ; ...
        [(ncolor-1):-1:0]/31 ; ...
        [(ncolor-1):-1:0]/31]';
cmap = [tmp1 ; tmp2];
```

This choice of color mapping makes no distinction between negative and positive events if the image array is translated such that the zero indexes the middle index of the colormap. This is the purpose of the `rel_im` function, which scales and translates an input image array relative to a second input:

```
% Image something relative to another image
% Usage:
% rel_im(source, image)
function im = rel_im(source, im)
scale = .5/max(max(abs(source)));
im = im.*scale;
image(round(im.*64+32))
```

The purpose is to compare images at a common scale relative to a particular source image. This thesis uses the input image as the base scale and all other images are scaled accordingly. If the input to the function is an array, it takes the maximum value and normalizes the second argument to this value. Otherwise, a simple scalar can be used as the first argument to choose an arbitrary scale.

Note there are only 16 levels of gray to choose from, between printing and the dynamic range of the color scale, some small amplitude information will be lost. The choice to plot the images this way was carefully made in order to convey useful information that is not biased. It is a difficult topic to present results fairly, but every effort was made to remove this bias.