GROUND PENETRATING RADAR SIGNAL PROCESSING FOR LANDMINE DETECTION

Dissertation

zur Erlangung des akademischen Grades

Doktoringenieur

(Dr.Ing.)

von **M.Sc. Fawzy Abujarad**

geb. am 13.6.1968 in Beitlahia-Gaza

genehmigt durch die Fakultät für Elektrotechnik und Informationstechnik der Otto-von-Guericke-Universität Magdeburg

Gutachter: Prof. Dr.-Ing A. S. Omar

Prof. Dr. rer. nat. Georg Rose

Dr. W. Al-Nuaimy

Promotionskolloquium am 01.03.2007

To my mother, my wife and my children.

Table of Contents

List of Abbreviations

AP   Anti-personnel

AT   Anti-tank

BSS   Blind source separation

CWT   Continuous wavelet transform

DWT   Discrete wavelet transform

EM   Expectation-Maximization algorithm

EMI   Electromagnetic induction

GLRT   Generalized likelihood ratio test

GPR   Ground penetrating radar

HOS   Higher-order-statistics

ICA   Independent component analysis

JADE   Joint approximate diagonalisation of eigenmatrices

LRT   Likelihood ratio test

MD   Metal detector

MI   Mutual information

MLE   Maximum likelihood estimator

MSR   Mean square error

NMR   Nuclear magnetic resonance

NQR   Nuclear quadrupole resonance

PCA   Principle component analysis

pdf   Probability density function

ROC   Receiver operating characteristic

SOBI   Second order blind identification

List of symbols

Symbol   Meaning

$\mathbf{A}$ $M \times N$ mixing matrix

a $M \times 1$ a column mixing vector

a_i $M \times 1$ ith column mixing vector

a_{ij} scalar   mixing coefficient of the jth source in ith observation

$\mathbf{E}$ $M \times N$ residual matrix

M scalar   number of observations x_i

N scalar   number of sources s_j

$\mathbf{S}$ $N \times K$ matrix of N sources with K samples

s $1 \times K$ a row vector consisting of a source

s_j $1 \times K$ a row vector consisting of the jth source

$s_j(k)$ scalar   value of the jth source at (time) index k.

$s(k)$ $N \times 1$ a column vector containing the values of all of the sources at time instance k

K scalar   number of samples in sources s_i, and observations x_i

θ vector   a set of model parameters

$\mathbf{\Sigma}$ $M \times M$ covariance matrix of the observed variables

$\mathbf{\Psi}$ $M \times M$ residual covariance matrix

$\mathbf{U}$ $M \times M$ unitary matrix

$\mathbf{V}$ $N \times N$ unitary matrix

$\mathbf{W}$ $N \times M$ unmixing (separating) matrix

w_j $M \times 1$ jth column unmixing vector

w_{ij} scalar   unmixing coefficient of the ith observation in the jth source

$\mathbf{X}$ $M \times K$ matrix of M observations with K samples

x_i $1 \times K$ a row vector consisting of the ith observation

$x_i(k)$ scalar   value of the ith observation at (time) index k.

$\mathbf{Y}$ $N \times K$ estimation matrix

Acknowledgements

Seldom is anything accomplished without the assistance or encouragement of others. This thesis would also not have been in its present form without the support of many people who, in one way or other, provided help to me.

First and foremost, I would like to thank my supervisor Prof. Dr.-Ing. A. S. Omar for his suggestions and constant support during this research work. I had the opportunity to learn many things from his profound knowledge and experiences and hope to continue collaboration with him in the future.

I am also grateful to Prof. Dr. rer. nat. Georg Rose, University of Magdeburg, and Dr. Waleed Al-Nuaimy, University of Liverpool, for kindly agreeing to be referees for this thesis in spite of their hectic schedules.

Deep thanks to Prof. G. Nadim, Cairo University, Cairo, Egypt for his help and support.

I always feel lucky to be with so many excellent researchers in Prof. Omars group. I sincerely thank Nick, whose enormous contributions to the measurement lab greatly facilitated the progress of my work. Again, thanks to all my friends: Atallah Balalem, Ali Ramadan Ali, Ayan Bandyopadhyay, Omar Ali, Ehab Hamad, A. Boutejdar, M. Anis, A. Teggatz, Ali Aassie, and A. Slavova for their support.

Of course, I am grateful to my mother for her patience and love. Without her this work would never have come into existence.

But, most of all, to my wife, Raessa, I owe everything for making this effort possible; without her love and support over the past 16 years, this manuscript would not exist.

Zusammenfassung

Die vorliegende Arbeit beschäftigt sich mit Detektionsalgorithmen für nichtmetallische Anti-Personen (AP) Landminen und den zugehörigen Signalverarbeitungsverfahren, wobei ein Stepped-Frequency Ground Penetrating Radar Systems (SF-GPR) verwendet wurde. Moderne Anti-Personenminen werden fast vollständig aus Plastik und Keramikmaterialien hergestellt. Deswegen ist ihre Detektion mit Hilfe von GPR Sensoren ohne geeignete Signalverarbeitung beinahe unmöglich. Viele Signalverarbeitungstechniken wurden in der Literatur präsentiert und erfolgreich für die Detektion von AP Minen verwendet. Deshalb ist es notwendig, die Algorithmen hinsichtlich ihrer Effizienz und bezüglich der benötigten Hard- und Softwareressourcen zu vergleichen.

Diese Arbeit stellt die verbreitetesten Signalverarbeitungstechniken für die SF-GPR Detektion von im Boden verborgenen Objekten vor. Diese Techniken wurden untersucht, umgesetzt und miteinander hinsichtlich der Fähigkeit verglichen, das Signal der Landminen vom Umgebungsrauschen zu unterscheiden. Der Algorithmus, der die besten Ergebnisse liefert ist die sogenannte Independent Component Analyse die dazu gehührt, dass GPR Clutter fast vollständig beseitigt konnten werden und das geforderte Zielsignal extrahiert wurde. Weiterhin konnte gezeigt werden, dass Wavelet Packet Transform Techniken in Kombination mit statistischen Verfahren höherer Ordnung sehr effektiv für die GPR Signalverarbeitung genutzt werden können.

Alle experimentellen Ergebnisse, die im Rahmen dieser Arbeit präsentiert werden, basieren auf realen Messungen am experimentellen SF-GPR Systemaufbau, welcher am Institut für Elektronik, Signalverarbeitung und Kommunikation (IESK) der Universität Magdeburg, entwickelt und aufgebaut worden ist.

Abstract

This thesis is concerned with detection logarithms of non-metallic anti-personnel (AP) land mines and the related signal processing using stepped-frequency ground penetrating radar (SF-GPR) technique. Modern land mines are essentially made out of plastic and ceramic materials. This makes their detection using GPR sensor almost impossible without proper signal processing. Many signal processing algorithms have been presented in the literature and successfully applied to GPR data for the detection of AP land mines. Therefore there is a need to compare these methods as regarding efficiency and hard- and software requirements. The thesis presents most common signal processing techniques used for SF-GPR based detection of buried objects. These techniques have been investigated, implemented and compared to each other as regarding their ability to separate the land mine and noise signals. The algorithm that performed best in these comparison is called Independent Component Analysis algorithm, which has demonstrated the ability to eliminate the GPR clutter and extract the target signal. Furthermore, combining the wavelet packet transform with the higher-order-statistics has shown to be very effective in the GPR signal processing. All experimental results presented in the thesis are based on real measured data obtained from an experimental SF-GPR system. The system has been developed and built at the Institute of Electronics, Signal Processing and Communications Engineering (IESK), University of Magdeburg, Magdeburg, Germany.

Publications Related to the Thesis

1. Fawzy Abujarad, Jöstigmeier, A. S. Omar. Clutter Removal for Landmine using Different Signal Processing Techniques. *Proc. tenth international conference on ground penetrating radar*, pp. 697-700, June, 2004.

2. Fawzy Abujarad, Galal Nadimy and Abbas Omar. Wavelet Packets for GPR Detection of Non-Metallic Anti-Personnel Land Mines Based on Higher-Order-Statistic. *Proc. 3rd international workshop on advanced ground penetrating radar*, pp.21-24, May 2005.

3. Fawzy Abujarad, Galal Nadimy and Abbas Omar. Clutter Reduction and Detection of Landmine Objects in Ground Penetrating Radar Data Using Singular Value Decomposition (SVD). *Proc. 3rd international workshop on advanced ground penetrating radar,* pp.37-41,May, 2005.

4. Fawzy Abujarad, Galal Nadimy and Abbas Omar. Combining Wavelet Packets with Higher-Order-Statistic for GPR Detection of Non-Metallic Anti-Personnel Land Mines. In *Proc. SPIE Image and Signal Processing for Remote Sensing XI*, volume 5982, Brugge, Belgium, sept 2005. p.380-390.

5. Fawzy Abujarad, A.S. Omar. Factor and Principle Component Analysis for Automatic Landmine Detection Based on Ground Penetrating Radar. *German Microwave Conference (GeMIC) 2006*, Karlsruhe, Germany, March 2006.

6. Fawzy Abujarad, A.S. Omar. GPR Data Processing Using the Component-Separation Methods PCA and ICA. *IEEE International Workshop on Imaging Systems and Techniques*, Minori, Italy, April 2006.

7. Fawzy Abujarad and Abbas Omar. GPR Detection of Landmine by FastICA. 11th International Conference on Ground Penetrating Radar, June 19-22, 2006, Columbus Ohio, USA.

8. Fawzy Abujarad and Abbas Omar. Comparison of Independent-Component-Analysis (ICA) Algorithms for GPR Detection of Non-Metallic Land Mines. I*n Proc. SPIE Image and Signal Processing for Remote Sensing XI,* volume 6365, Stockholm, Sweden, 11-14 September 2006.

Chapter 1

Introduction

Taking care about environment is a recent phenomenon that widely spread in fact after the World War II. The development of science and technology leads to fatal environmental problems especially regarding disputes and wars. One of the worst problems that face the humanity is the buried landmines and unexploded ordnance (UXO), see [1].

Millions of Landmines have already been scattered over many post-war countries [2, 3]. Complete clearance of a minefield is required to restore public confidence. Therefore, mine-detection techniques require extremely high detection rates. Although research and development of detection techniques has been going on for years, no single technique is deemed suitable for all de-mining scenarios.

The metal detector (MD) approach, that was used in World War II, is still the standard approach used to detect land mines, see [4]. Although, all metallic objects are detected and identified by a metal detector [5], the problem is heightened when MD approach is used to detect and identify the plastic or low metal content landmines. In [6], it has been shown that Ground Penetrating Radar (GPR) approach is one of the most talented technologies for detection and identification of plastic and/or low metal content buried landmines. The GPR approach uses the difference in the permittivity of both the mine and the surrounding medium to detect the target [7]. However, it is difficult for a GPR system to identify the mine if

- the target is buried close to the surface of the ground, or if it

- has very small geometrical dimensions, about 9 cm diameter, and irregular shape and

- if the object has permittivity near to that of the ground.

In all these cases the reflected signal of the target is very weak compared to and/or overlapped with the noise, making it difficult to distinguish between both without proper signal processing [6].

Thus in order to extract useful information about the target, it is necessary to apply proper signal processing to the GPR data.

Actually, the data that received from GPR system consist of:

1. Coupling induced by transmitting and receiving antennae.

2. Reflection by ground surfaces.

3. Stochastic electromagnetic. interference either from GPR system itself or from environment.

4. Reflection by subsurface targets.

5. Dispersion or reflection by the underground inhomogeneities.

Usually, the signals (1) and (2) are called "clutter", which are much more intense than the buried-target reflection due to short distance from transmitting to receiving antennae and the attenuation of the target reflection by underground medium like soil, etc. It is therefore necessary to deploy proper clutter reduction method to be able to detect and identify the target.

1.1   Landmines Humanitarian Problem

Landmines are basically explosive devices designed to explode when triggered by pressure or a tripwire. These devices are typically found on or just below the surface of the ground [8]. Landmines are one of the worst environmental problems that humanity faces long with being one of the most terrifying legacies of war. Landmines are made to be used by armed forces to disable any person or vehicle that comes close with it by an explosion or fragments released at high speeds, so they become a burden to those who have to support them.

The International Campaign to Ban Landmines (ICBL) [9] estimates that 15,000-20,000 people are killed or injured by land mines per year many of them children. The U.S. State Department estimates that a total of 45-50 million mines remain to be cleared. Worldwide, approximately 100,000 mines are cleared each year [10]. At that rate, clearing all 45-50 million mines will require 450- 500 years, assuming no new mines are laid. By some estimates, roughly 1.9 million new mines are emplaced annually, yielding an additional 19 years of mine clearance work every year.

The Landmine Monitor Report 2003 identified over 11,700 new landmine victims in 2002. Of this number at least 2,649 were children, a staggering 23 percent. More than 85 percent were innocent civilians [9]. It is estimated 300,000-400,000 people live with mine-related injuries, according to the 2005 Annual Report of Landmine Survivors Network, an organization created for and by landmines survivors [9].

It is estimated that more than 100-million landmines located in 70 countries around the world. Since 1975, there have been more than one million landmine

Table 1.1: World distribution of landmines.

COUNTRY	NUMBER OF LANDMINES (Estimated)	NUMBER OF UXO (Estimated)
Afghanistan	4 million	Large
Bosnia & Herzegovina	1 million	Large
Cambodia	300,000 -1 million	2.5 million
Croatia	1 - 1.2 million	0
Ecuador	50,000 - 60,000	Small
Egypt	5 - 7.5 million	15 - 15.5 million
Ethiopia	1.5 - 2 million	Large
Vietnam	3.5 million	Large
Zimbabwe	2.5 million	Unknown
Source: United States Commitment to Humanitarian Demining [10]		

casualties most of them civilians, many of them children. Where they do not kill immediately, landmines severely maim their victims, causing trauma, lifelong pain and often social stigma. World wide there are some 250,000-landmine amputees. Survivors face terrible physical, psychological and socio-economic difficulties. The presence of mines also can cause economic decline. Most victims are males of working age, and often they are unable to return to work. It is found that "households with a mine victim were 40% more likely to report difficulty in providing food for the family." Further, the medical bills for survivors can bankrupt families. Many victims must undergo multiple surgeries. Children who lose limbs require multiple prosthetic devices over their lifetimes. Mines affect not only the victims' families but also the entire community surrounding the mined area. Even the rumor of mine presence can halt all activity in an affected area. The global distribution of landmines in the more severely affected countries is shown in Table 1.1.

There are two forms of landmines:

- Anti-personnel (AP) mines.

- Anti-tank (AT) mines.

The functions of both forms of these landmines are the same, namely to disable and kill. There are minor differences between them. Anti-tank mines are typically larger in size and contain more explosive material than anti-personnel mines. The explosive material that is found in anti-tank mines is enough to destroy a tank or truck, as well as kill people in or around the vehicle. In addition to that the anti-tank mines require more pressure to explode. Most of these mines are found on roads, bridges and large places where tanks may travel.

On the other hand an anti-personnel mine is a mine designed to be exploded by the presence, proximity or contact of a person and that will incapacitate, injure or kill one or more persons. These hidden indiscriminate weapons cannot tell the difference between the tread of a soldier or a child. They continue to kill and maim long after wars have ended. According to the International Campaign to Ban Landmines (ICBL), more than 350 different kinds of anti-personnel mines have been produced by more than 50 countries.

AP mines act to injure or kill victims by both the explosive blast and the fragmentary metal debris projected upon detonation. Anti-personnel mines fit into three basic categories:

- **Blast** - Blast mines are the most common type of mines, buried nearly at the surface of the land and are generally triggered by someone stepping on the pressure plate, requiring about 5 to 16 kg of pressure to explode. The main purpose of these mines is to destroy an object in close proximity, such as a person's foot or leg. A blast mine is designed to break the targeted object into fragments, which can cause secondary damage, such as infection and amputation.

- **Bounding** - This type of mines, usually buried with only a small part of the igniter protruding from the ground, these mines are pressure or tripwire activated. Bouncing Betty is another name referred to these type of mines. When activated, the igniter sets off a propelling charge, lifting the mine about 1 meter into the air in order to cause injury to a person's head and chest.

- **Fragmentation** - This type of mines release fragments in all directions, or can be arranged to send fragments in one direction (directional fragmentation mines). These mines can cause injury up to 200 meters away and kill at closer distances. The fragments used in these mines are either metal or glass. Fragmentation mines can be bounding or ground-based.

1.2 Detection of Buried Landmines

The landmines detection and clearance always still as a time overwhelming and unsafe task. Efficient and accurate detection of buried mines is still unsolved problem. The widely used approaches for locating mines is still Metal detection and hand prodding.

Metal detection that are the key part of the deminers tool kit employ the same principles as those first used in World War I and refined during World War II [11].

The U.S. military has used another approach that is a detector operate via a principle known as electromagnetic induction (EMI).

EMI detector can sense as little as 0.5 grams of metal and thus maintain a high detection probability. Furthermore EMI detector is lighter and easier to operate than

the World War II counterparts. However, significant limitations to this technology remain. EMI systems can still miss mines and still not accurate a hundred percent. Probabilities of detection varied remarkably by detector, location, and soil type. The best performing detector found 91 percent of the test mines in clay soil, but the same detector found only 71 percent of the mines in laterite (iron-rich) soil. The poorest-performing detector found 11 percent of mines in the clay soil and 5 percent in the laterite soil [5]. Furthermore, modern landmines can be constructed without metallic parts and would therefore not be detected with current technology. An ideal sensor would be able to detect the actual explosive material, e.g. TNT, RDX, rather than detect parts used in the construction of the landmine. It would need to be sensitive but also specific so that not to result in a large number of false alarm that would slow down the clearing process. The desire detector must be harmonizing with UN Requirement.

So, the starting point in terms of the end-user is defined by the UN Statement of Requirement. The UN's International Standards for Mine Clearance Operations defines an area as being cleared when all mines and munitions have been removed and/or destroyed. The area should be cleared of mines and UXO to a standard and depth which is agreed to be appropriate to the residual/planned use of the land, and which is achievable in terms of the resources and time available. At least 99.6 % of the agreed standard of clearance must be achieved by the contractor. The removal of all mines and UXO to a depth of 200 mm is an object for all UN-sponsored clearance programmes. Mines/UXO below this depth are therefore not necessarily cleared.

Also images of High resolution of shallow buried objects in soil remains a problem needs to be solved. Through using available technologies small AP mines are most difficult to be detected. As stated above, most humanitarian de-mining operations rely upon the use of metal detectors and hand prodding. De-mining operations occasionally employ specially trained dogs to sniff out explosives. Besides, there are at least 20 different kinds of technologies specifically aimed at detecting buried mines that are currently either under development or are potentially available. However, all of these technologies have their limitations and none of them can be used alone as a reliable mine detection tool.

Many methods and combinations thereof have been tried for the detection of land mines. These include:

Metal Detectors

The main use of the metal detector is to detect the buried landmines. It operates on the principles of electromagnetic induction. The metal detector contains one or more inductor coils that are used to induce a magnetic field. The magnetic field emitted into soil and interacts with metallic elements on the ground. The field induces an eddy current in metallic object which induce their own magnetic field. That field generates an opposite current in the coil, which induces a signal indicating the presence of

metal. Unfortunately, metal detectors are succeed only in identifying the presence of an anomaly without providing any information on whether the detected object is explosive material or not. However, landmines typically contain a small amount of metal in the firing pin while many others contain no metal at all.

Acoustic Sensors

Some approaches of landmines detection based on acoustics have been proposed [12]. Acoustic sensor sends acoustic waves into the ground. These sound waves reflected on the boundaries between materials with different acoustical properties. The reflected acoustic are used to locate and identify the body [13]. The Ultrasound detection depends on the soil density and the bulk modulus as well as the system's operating wavelength.

This type of detection has been studied. It has been shown that it is powerful in very wet and heavy ground such as clay but it is less effective in sandy soils. For wet soils where the GPR system is not effective, the acoustic sensor is a promising candidate. The problem is how to couple the waves with the ground surface without applying pressure. Due to soil inhomogeneities the accuracy of acoustic measurements is very poor. Furthermore such measurements require information about the return signals. The knowledge about these signals is suffering since acoustic energy is highly absorbed by sand and there are strong disturbances at the air-to-ground interface. However, the acoustic method can be used in combination with other techniques such as metal detectors and radar. Note that the combination with radar has been done by Scott [14].

Infrared Imaging Systems

Infrared images of sufficient temperature and spatial resolution to detect anomalies in the ground introduced by the presence of a landmine are commercially available from numerous sources [15]. All bodies emit infrared radiation that is related to the temperature of a body. In order to detect objects using a thermal imaging system a difference in the emitted infrared radiation is required.

This can be caused by either a temperature difference between the object and the background or an emittance difference of bodies at the same temperature. The thermal properties of mines are different from the surrounding medium [16]. This means that a contrast can be expected for surface laid mines when the environment and weather conditions are favorable. For buried mines the contrast occurs due to disturbances in the ground conditions. An object buried in soil change the conditions in the ground. The presence of an object alters the migration of water, which introduces a change in thermal conductivity and heat capacity, as these properties are dependent on the moisture content. The infrared imagery depend on landmine type, soil type and compaction, moisture, shadow and time of day. However, over the long term the thermal properties of the disturbed soil will return to their natural state.

This means that after mines have been buried for a long period of time, the only ones that can be detected are those that are buried within 10 cm of the surface [17]. Also, vegetation will obscure the ground surface and dissipate thermal contrast making it difficult to detect by IR.

Nuclear Magnetic Resonance (NMR)

A new technique for quick detection buried landmines has been discovered. It is called the nuclear magnetic resonance (NMR). This technique is very efficient if the target is inside the detecting coil. Unfortunately it doesn't work with buried mines. NMR can be configured to detect buried mines from outside the coil but the resulting DC and RF fields are relatively weak and non-uniform thus resulting in poor detection. It requires a superconducting coil, which in return needs a cooling system and a source of high power, thus rendering this technique unsuitable for hand-held detectors. Furthermore, NMR would not be able to detect metallic mines and would have to be used in conjunction with a metal detector.

Markus Nolte at Darmstadt University in Germany and colleagues have developed a new NMR sensor. It can be detect the nitrogen in the explosive TNT, which is particularly hard to spot [18].

Nuclear Quadrupole Resonance (NQR)

The actual explosive material such as TNT or RDX could be detected by Nuclear Quadrupole Resonance(NQR). TNT that contains the atomic nuclei of nitrogen ^{14}N have a spin property called a quadrupole moment. An oscillating nuclear magnetic moment that can be detected by sensitive receiver is generated by an externally applied RF pulse at the precession frequency [19]. The electronics required to generate precisely timed spins at the correct frequencies, to detect the very weak signals, is still being developed. At present, the time required to detect explosives varies and is slow for a hand-held mine detector - between 0.1 and one second [17].

X-Ray Backscatter

By passing the photons through the object the detection of buried landmine can be done by x-ray. The production of high resolution image of buried landmine resulted because of the wavelength of x-ray variety in compare with the size of landmine. The principle of detection by x-ray can be done by passing the photons through the object and the backscatter of x rays may still be used to provide information about buried irradiated objects.

The backscatter exploits the fact that mines and soils have slightly different mass densities and effective atomic numbers that differ by a factor of about two. To detect buried object it is necessary to use low-energy incident photons, but soil penetration of photons backscatter devices is poor. This limits detection to shallow mines whose

depth is less than 10 cm deep. Also the time required to obtain surface image is long. The x-ray technique is sensitive to source/detector standoff variations and ground-surface fluctuations. Furthermore it is difficult to achieve high spatial resolution by x-ray.

Dogs

For having extremely acute sense of smell, and after a hard process of training to detect traces of explosives, the dogs are used in mine detection. They can also be training to locate trip wires. The dogs are used for Mine detection in two ways. First, working with a a trainer, the dogs detect landmines, which are marked for investigation and removal. The second technique is having the mine detection dogs sniff air samples from suspected mine-affected areas. The dog may work in a safe environment and test various samples from large areas, thus providing quick indications of affectedness for demining crews.

Dog performance varies widely depending on the individual dog, how it was trained, and the capabilities of the handler. Further, dogs may need to be retrained periodically because they can become confused if they discover behaviors other than explosives detection that lead to a reward. An additional limitation is that when trained to detect high levels of explosives, dogs may not automatically detect much lower levels and may need to be specially trained for this purpose. Like other methods that rely on vapor detection, performance of mine detection dogs can be confounded by environmental or weather conditions that cause explosive vapors to migrate away from the mine or that result in concentrations of vapors that are too low even for dogs to detect.

Microwave Techniques

Using microwaves to detect mines requires radiation of these microwaves into the ground in order to analyze any returning signals. Return signals occur when the microwave encounters a subsurface discontinuity and is reflected back. High frequencies are capable of high resolution but are also highly attenuated in soil. Thus they are suitable for the detection of small shallow objects. Conversely, low frequencies achieve lower resolution but are less attenuated in soil. Hence they are more suitable for detecting large deep objects. Microwaves may be transmitted as continuous waves, in impulses, by stepped frequencies, pure oscillating sinusoids or as a combination of these methods. The most prominent type of active microwave is Ground Penetrating Radar GPR.
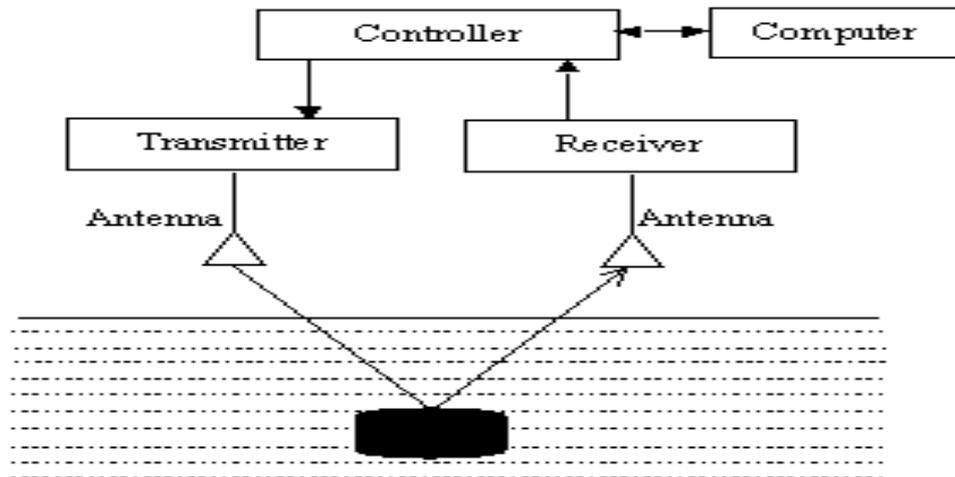
Figure 1.1: Outline of a basic GPR system.

1.3 Ground Penetrating Radar

Ground Penetrating Radar (is also known as earth sounding radar, ground probing radar, subsurface radar, or georadar) (GPR) is a high-resolution electromagnetic technique used to evaluate the location and depth of buried objects and to investigate the presence and continuity of natural subsurface conditions and features, without drilling, probing, or digging, see [6] Thus GPR is used to locate the buried objects such as landmines [20], pipes, cables and reinforcement [21], the location of subsurface cavities and fractures in bedrock [22], as well as ground water and moisture [23], etc. Ground penetrating radar operates by transmitting electromagnetic wave that is radiated from a transmitting antenna down into the ground. The electromagnetic wave is reflected from various buried objects or distinct contacts between different earth materials that have contrasting dielectric properties , such as at the boundary between soil and a landmine or between soil and a large rock. The reflections are created by an abrupt change with the dielectric properties in the ground. These electrical properties are namely, relative permittivity, relative permeability and conductivity. However, not all three parameters provide useful information to the GPR. Conductivity generally affects the penetration depth of the GPR due to absorption of the radar signals in the medium. Soil with high moisture content increases the electrical conductivity, thus decreasing penetration. On the other hand, due to the lack of magnetic content in earths soil, relative permeability is hardly provides any useful information because it offers little contrast in the radiated EM pulses. Contrastingly, relative permittivity, which corresponds to the dielectric constant of the medium, provides the highest degree of contrast in the reflected wave, thus resulting in good characterization of the ground. Therefore, the contrast in permittivity usually leads

to the reflection in the EM pulse. In addition to having a sufficient electromagnetic property contrast, the boundary between the two materials needs to be sharp. The reflected wave that is back to the surface is captured by a receiving antenna, and recorded on a digital storage device. GPR is a time-dependent geophysical technique that can provide a good three-dimensional subsurface image which is useful for interpreting specific target, can include the fourth dimension of color, and can also provide accurate depth estimates for many common subsurface objects. GPR units consist principally of a control unit, which generates synchronized trigger pulses to the transmitter and receiver electronics in the antennas. These pulses control the transmitter and receiver electronics in order to generate a sampled waveform of the reflected radar pulses. Further it contains one (monostatic) or two (bistatic) antennas for the transmission and the receiving of the signals. Finally a computer is commonly used for data collection, see Figure 1.1.

Performance of GPR is depending upon the surface and subsurface conditions and its specifications include requirements for or information about reflections, depth of investigation, resolution. The ability at which GPR can detect objects is depending on the wavelength of the input signal, so the quality of the image improves as the wavelength decreases and the frequency increases. But, at high frequencies, the GPR penetration of the incident wave into the soil can be poor. At Low frequency the penetration is more but with less resolution. However, the design of GPR system must make a tradeoff between quality of the image and required penetration depth. The optimal design for maximizing image quality while ensuring sufficient penetration depth changes with environmental conditions, soil type, mine size, and mine position. Recently, alternative GPR are being explored to optimize the tradeoff between penetration depth and image quality under a wide range of conditions.

Moreover the signal-processing is considered to be the most critical part in the design of a GPR system. In fact it is the process that filters out clutter signals and selects objects to be declared as mines.

Although GPR is a mature technology, it has not received widespread implementation for mine detection. At this time, GPR was unable to meet performance targets for landmine detection established for military countermine operations Strengths. Anyway GPR has a number of advantages. Firstly: GPR can locate and characterize both metallic and non-metallic subsurface features. Therefore GPR can used to detect nonmetallic APL. Secondly: Generating an image of the mine or another buried object based on dielectric constant variations is often possible because the required radar wavelength is generally smaller than most mines at frequencies that still have reasonable penetration depth. Thirdly: GPR is an efficient technology with a long performance history of other applications. It can be combined with EMI for mine detection. Finally, GPR can be made lightweight and easy to operate, and it scans at a rate comparable to that of an EMI system.

Along with the previous advantages, GPR has some limiting factors such as its usage in inhomogeneous soil, which increase the false alarm and the depth of signal

penetration. Therefor, GPR system designer must make a tradeoff between resolution of the return signal and depth. Since high-frequency signals yield the best resolution but its penetration depth is limitted.

1.3.1 GPR Systems

GPR systems can be subdivided into four categories, depending on their operating principle, these are, see [24]

Pulsed Radar

An impulse radar transmits a narrow pulse (less than a few nano-seconds), associated with a spectrum of signals, with a large peak power at a constant pulse repetition frequency (PRF). This spectrum should be wide, in order to measure the target reflectivity profile at high resolution. The time delayed received waveform is then sampled. The radar electronics must be able to deal with the wide instantaneous bandwidth. In general, this also implies high A/D conversion rates. Improvements of the signal to noise ratio (SNR) are achieved by averaging the profiles, a process known as stacking. If we add several random noises together, some of them will cancel each other because they are usually out of phase with each other. If they are statistically independent, the standard deviation of the sum of n random signals will be proportional to $\sqrt{n}$, whereas the sum of n coherent in-phase signals will be proportional to n so that the SNR will be improved by the factor of $\sqrt{n}$ [25].

Stepped Frequency Radar

A stepped frequency system achieves the required resolution by radiating a succession of carriers stepped uniformly in frequency across a desired band. Signal to noise improvement is achieved by dwelling on each frequency and integrating them. A/D conversion rates are modest. Direct coupling of the transmit and receive signal leads to some dynamic range problems.

Pulsed Stepped Frequency Radar

A pulsed stepped frequency system transmits a spectrum of signals with a certain centre frequency and receives the return. The frequency is stepped up in the successive pulses to cover a range of frequencies to achieve high resolution. An advantage of the pulsed stepped frequency approach is the reduction of the instantaneous bandwidth and sampling rate requirements of the radar system, as well as the possibility of skipping frequencies that might be corrupted due to external interfering frequency sources [26].

Stepped Frequency Continuous Wave Radar

The stepped frequency continuous wave (SFCW) waveform is implemented by transmitting a number of single frequency tones separated in frequency by Δf. At each frequency the amplitude and phase of the received signal is sampled and recorded. The received signal is transformed into the time domain using Inverse Discrete Fourier Transform (IDFT) to obtain a synthesised pulse [27]. In the case of the stepped frequency continuous wave radar system, direct coupling of the transmit and receive signal leads to dynamic range problems. The main advantage of the SFCW ground penetrating radar over the pulsed system is that the former has a higher average radiated power for transmission.

1.3.2   Radar Data Acquisition Modes

There are four main modes of radar data acquisition. Following we define these modes.

Common offset

In Common offset mode operation, the transmitter and receiver antennae are at a fixed distance and moved over the surface simultaneously. The measured travel times to radar reflectors are displayed on the vertical axis, while the distance the antenna has travelled is displayed along the horizontal axis in a radargram display. Most GPR surveys, use a common offset survey mode. This mode of data acquisition can be used to improve the azimuth or plan resolution, where a long aperture is synthesized along the azimuth line. This type of operation, in the radar field, is called synthetic aperture radar image formation. Figure 1.2 shows a common offset bistatic mode data acquisition configuration.
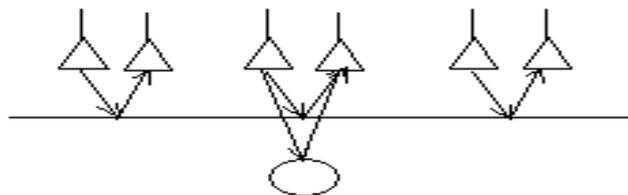
Figure 1.2: Common offset bistatic mode data acquisition configuration in which the distance between transmitter and receiver antennae is fixed and the pair moves along a horizontal line.

Common Source Point

In a common source data acquisition system, sometimes called wide-angle reflection and refraction (WARR) sounding [28], the transmitter is kept at a fixed location and the receiver is lowed away at increasing offsets. This type of data acquisition mode is most suitable in an area where the material properties are uniform and the reflectors are planar in nature. Figure 1.3 shows the antennae configuration of a common source data acquisition mode.
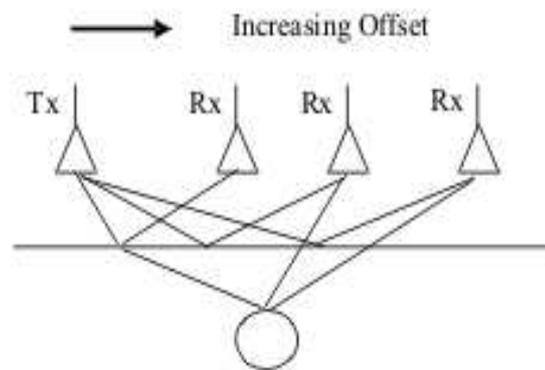
Figure 1.3: Common source antenna configuration where the transmitter is kept at a fixed location and the receiver is towed away at increasing offsets.

common Receiver Point

In common receiver point the data were acquired at the same receiver point for different source points. Figure 1.4 shows the configuration of a common source data acquisition mode.
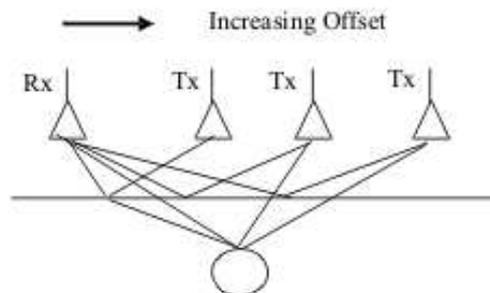
Figure 1.4: Common receiver antenna configuration where the receiver is kept at a fixed location and the transmitter is towed away at increasing offsets.

Common Midpoint

In this type of acquisition mode, the transmitter and receiver antennae are moved away at increasing offsets so that the midpoint between them stays at a fixed location. In this case, the point of reflection on each sub-surface reflector is used at each offset, and thus areal consistency at depth is not a requirement.

Figure 1.5 shows the antennae configuration of the common midpoint data acquisition mode.
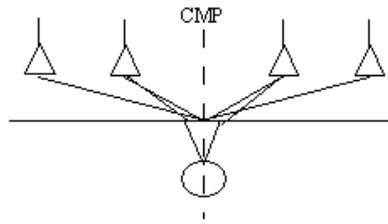
Figure 1.5: Common midpoint (CMP) measurement configuration, both the transmitter and receiver antenna are moved away at increasing offsets.

1.3.3   Visualizing GPR Data

The objective of GPR data presentation is to provide a display of the processed data that is closely approximates an image of the subsurface, with the anomalies that are associated with the objects of interest located in their proper spatial positions. Data display is central to data interpretation. In fact, producing a good display is an integral part of interpretation.

There are three types of displays of surface data, these are: a one-dimensional trace (A-scan), a two dimensional cross section (B-scan), and a three dimensional display (C-scan) [29]. Following we described briefly each of these types.

1. **A-scan**

   An A-scan (or one dimensional data presentation) is obtained by a stationary measurement, emission and collection of a signal after placing the antenna above the position of interest. The collected signal is presented as signal strength vs. time delay.

2. **B-scan**

   B-scan (or two dimensional data presentation) signal is obtained as the horizontal collection from the ensemble of A-scans.

   The horizontal axis of the two dimension image is surface position, and the vertical axis is the round-trip marvel time of the electromagnetic wave

3. **C-scan**

   C-scan (or three dimensional data presentation) signal is obtained from the
   ensemble of B-scans, measured by repeated line scans along the plane. Three
   dimensional displays are fundamentally block views of GPR traces that are
   recorded at different positions on the surface. Obtaining good three-dimensional
   images are very useful for interpreting specific targets. Targets of interest are
   generally easier to identify and isolate on three dimensional data sets than on
   conventional two dimensional profile lines.

1.4 Thesis Outline

This thesis is written as a partial fulfillment of the requirements for the degree Doctor
of Philosophy. The broad objective of this research was to study, develop and compare
signal processing techniques to improve the detectability of buried anti-personnel land
mines (APL's) using ground penetrating radar (GPR).

The thesis is presented in eight Chapters. The GPR system that has been used
to collect the measured data used in this thesis is presented in Chapter 2. The chap-
ter describes the indoor laboratory, radar device, ultra-wideband antenna and test
mines. The following chapters (3 and 4) are closely connected. Chapter 3 introduces
Blind Source Separation, and three common techniques to solve it. These algorithms
are singular value decomposition, factor analysis and principle component analysis.
Chapter 4 presents a method called Independent Component Analysis. This method
uses to solve Blind Source Separation problem. In this chapter four of the most
common Independent Component Analysis algorithms called: extended Infomax, the
FastICA, Joint Approximate Diagonalisation of Eigenmatrices, and the second order
blind identification (SOBI) have been studied and applied to GPR data. In Chapter
5 the GPR noise and the target signal have been modeled using a very simple model.
This model has been used to derive a detector based on a likelihood ratio test using
some estimated parameters.

Chapter 6 describes the fundamental properties of wavelets in signal decomposi-
tion and the application of wavelet to detect the buried objects. The combination
of the wavelet packet and the higher-order-statistics is also included in this chapter.
The results of applying all algorithms presented in this thesis to real GPR data are
presented in Chapter 7.

Finally, Chapter 8 gives the conclusions that can be drawn from this research.

Chapter 2

Data Collection

The GPR data used in this dissertation were acquired with a bistatic-stepped frequency GPR system. The system was built at Institute of Electronics, Signal Processing and Communications Engineering (IESK), University of Magdeburg, Magdeburg, Germany.

2.1 GPR System

The radar device, which is used to obtain the data in this dissertation, is a network analyzer, which generates a stepped frequency waveform. The GPR system consists of network analyzer (Rohde & Schwarz), which is controlled by a PC (type Pentium 4) and two ultra-wideband (UWB) transmitting and receiving antennas. The two-port s-parameters are measured with real and imaginary components. Via implementing a chirp z-transform, the time domain profile of the receiving data can be determined.

An indoor laboratory test facility consisting of a sandbox made of wood has been designed at our institute (IESK). The dimension of the sandbox is 1.1 meter wide, 1.1 meter long, and 1.1 meter deep. The internal sides of sandbox are covered with absorption material and it is filled with sand of 0.5 m depth with a relatively flat surface. The measurement grid covers the area bounded by x = 27 → 76 cm and y= 39 → 89 cm with distance between the measurements of 1 cm in both x and y directions , the setup is demonstrated in Figure 2.1.

The transmitting and receiving antennas are mounted on the 2D scanning system, and placed above the ground surface at height 35 cm. The GPR data can be represented using three different forms, A, B, C-Scans. The A-Scan (radar trace) signal is obtained by a stationary measurement after placing the antennas above a specific position. The collected signal is presented in the form of a group of signal strength versus time delay. Figures 2.7, 2.9 and 2.11 show an examples of A-scans in the presence and absence of a landmine. As the antennas are moving along a survey time,

(a) Network analyzer.

(b) Sandbox.
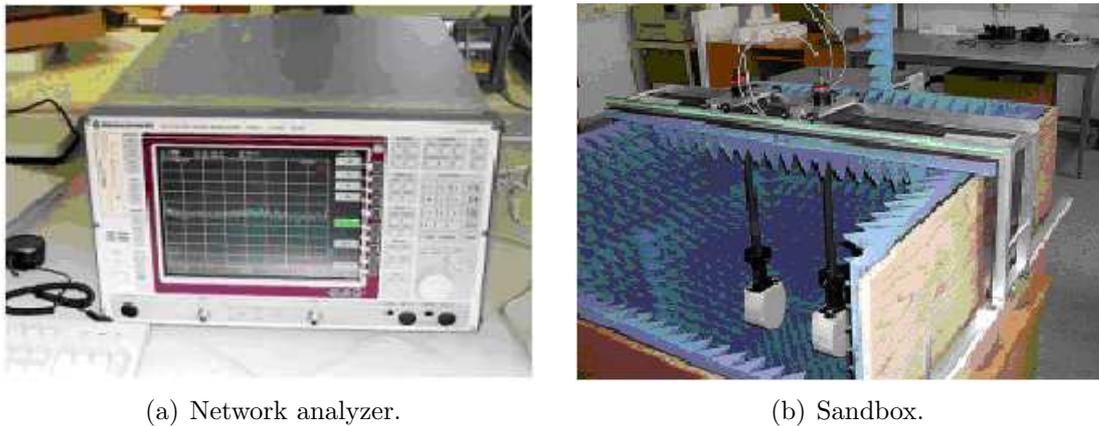
Figure 2.1: Network analyzer and test box for acquisition of the data.
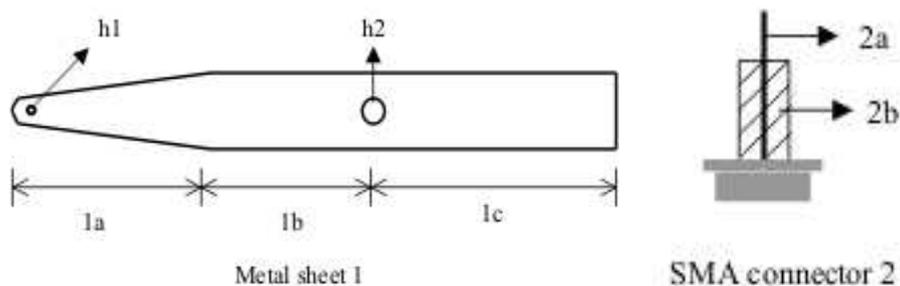
Figure 2.2: The metal sheet and the SMA connector used to construct the antenna.

a series of traces or A-Scans are collected at discrete points along the line. These A-Scans a long a line on the surface is called B-Scan (radargram for radar data or profile of the subsurface). An example of B-Scan shows in Figure 2.8.

2.2   Description of the Antenna System

A new wideband antenna were used in the GPR measurements in this thesis have been designed and fabricated in our institute. The outstanding properties of the proposed wideband antenna are achieved by combining an electric monopole, a magnetic dipole and a TEM horn in one single structure. The entire antenna is simply made of one thin metallic sheet and a standard SMA connector is used to feed the structure. Its first part consists of a loop formed by a metallic sheet and the second one is similar to a TEM horn. The horn is formed by extending the loop further and flaring it. The antenna radiates a linearly polarized field with a medium gain over a very wide band. The operating frequency band can be changed by linearly scaling the size of the structure.

A schematic diagram of the individual parts of the antenna and its cross section
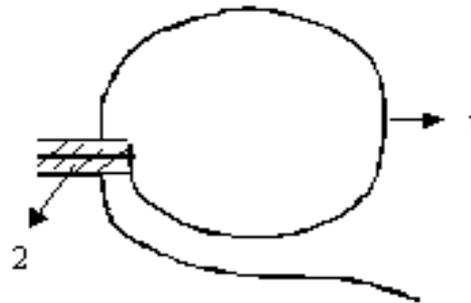
Figure 2.3: Cross-section of the antenna at the symmetry plane.

Figure 2.4: Fabricated prototype of the proposed antenna with SMA feed..
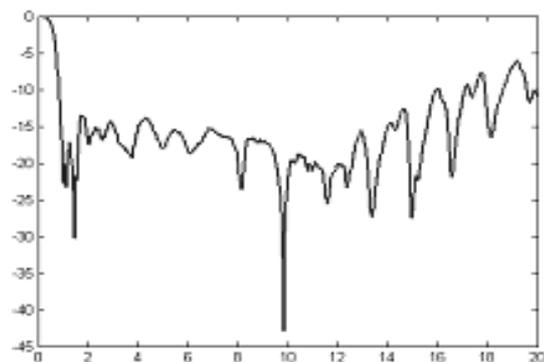
Figure 2.5: Return loss (S_{11}) at the antenna input. x-axis is the frequency rang in GHz and y-axis is the return loss (S_{11}) at the antenna input in dB .

are shown in Figures 2.2 and 2.3. The shape of the metal sheet 1 and the schematic drawing of the used SMA connector 2 are presented in Figure 2.2. The tapered portion 1a along with the straight portion 1b are bent to form a loop. The straight portion 1c is flared outwards to form the shape of a horn antenna. The small hole h1 is used to connect the metal pin 2a of the SMA connector 2 with the metal sheet 1. The Teflon isolation 2b fits exactly into the hole h2 where the flange 2c of the SMA connector is soldered to the metal sheet 1. In the case of a 50 Ohm feeding line the hole h2 has a diameter of 4.1 mm. The cross section of the complete structure is shown in Figure 2.3. The input impedance of the antenna is the same as that of the feeding line, which is just 50 Ohm in this example. Any other input impedance can be obtained using different feeding lines. The dimensions of the parts 1a, 1b and 1c are 90 mm, 110 mm and 130 mm, respectively, while the width of the metal sheet, denoted by w, is 60 mm. The antenna can be viewed as a combination of an electric monopole formed by section 1a, a magnetic dipole formed by section 1a and 1b and a TEM horn formed by section 1a and 1c. The upper part of the TEM horn is simultaneously the electric monopole. In order to provide stability to the whole antenna, it can be placed inside a case or holder or packed with foam or thermocol (materials with relative permittivity close to 1.0). A prototype of such an antenna is presented in Figure 2.4. The thickness of the metal sheet used in this example is 0.5 mm, but metal sheets with other thickness can also be used.

The measured return loss of the antenna in the frequency band from 1 GHz up to 18 GHz is shown in Figure 2.5. The measurements clearly demonstrate that the return loss is less than -10 dB in the entire frequency range.

Since the proposed type of antenna is wideband, lightweight, small in size and easy to fabricate, it has a high potential for the application in ground penetrating radars. Furthermore it can be used for all other applications that require such an extremely large bandwidth, e.g., in applications, which are based on the transmission of short pulses in the time domain or in the communication systems, which requires a single antenna for multiple frequency bands.

2.3   Description of the Test Mines

A set of test objects, representing dummy anti-personnel landmines (APLs), and false targets have been placed at various depths (1-10 cm) in the test sandbox. The APLs are made of various materials and have different shapes and sizes. Non-metal APLs, mines with a casing made out of plastic, metal or another material are included. The test mines have signatures close to those of real mines. An example of some test mines is given in Figure 2.6. A complete overview of possible test targets is given in Table 2.1.

The GPR data used in this dissertation were obtained from a series of experiments with different frequency ranges.

(a) PMN.

(b) PMN2.

(c) PPM2.
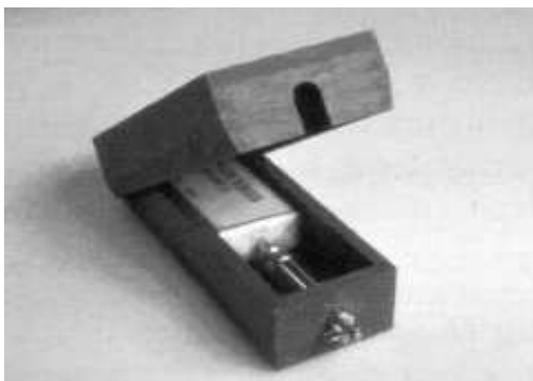
(d) PFM1.

(e) PMD6.

(f) M14.

Figure 2.6: Anti-personnel land mines (a) PMN (b) PMN2 (c) PPM2 (d) PFM1 (e) PMD6 (f) M14

Table 2.1: Mines for testing, available at IESK:

	Type	Case material	Diameter	Height	Depth
1	PMN	Plastic	11.2	5.6	1,3,10 cm
2	PMN2	Plastic	11.5	5.4	1,3,10 cm
3	PPM 2	Plastic	12.5	6.5	1,3,10 cm
4	M14	plastic	5.6	4	1,3,10 cm
5	Cylindrical metal 1	metal	10	5	1,3,10 cm
6	Cylindrical metal 2	metal	5	15	1,3,10 cm
7	Cylindrical metal 3	metal	5	3	1,3,10 cm
8	Cylindrical wood 1	wood	10	5	1,3,10 cm
9	Cylindrical wood 2	wood	5	15	1,3,10 cm
10	Cylindrical wood 3	wood	5	3	1,3,10 cm
11	Cylindrical Plastic 1	plastic	10	5	1,3,10 cm
12	Cylindrical Plastic 2	plastic	5	15	1,3,10 cm
13	Cylindrical Plastic 3	plastic	5	3	1,3,10 cm
14	Metallic mine	metal	10	13	1,3,10 cm
15	PFM-1	plastic	12(L)* 2(W)*6.1(H)		1,3,10 cm
16	MNO 50	plastic	22(L)*6.2(W)*14.6(H)		1,3,10 cm
17	PMD-6	wood	19.6(L) * 8.7(W)*5(H)		1,3,10 cm
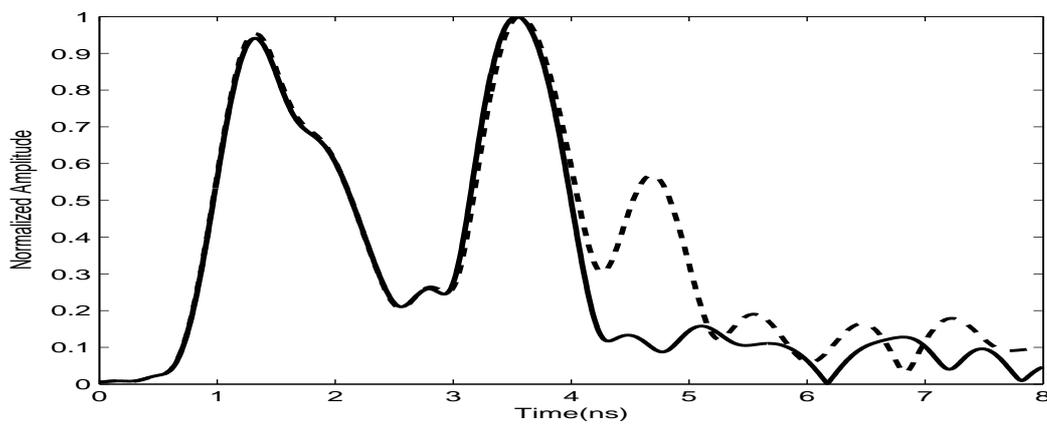
Figure 2.7: A-scans in the presence (dashed) and absence (solid) of a mine for data with 2 GHz bandwidth.
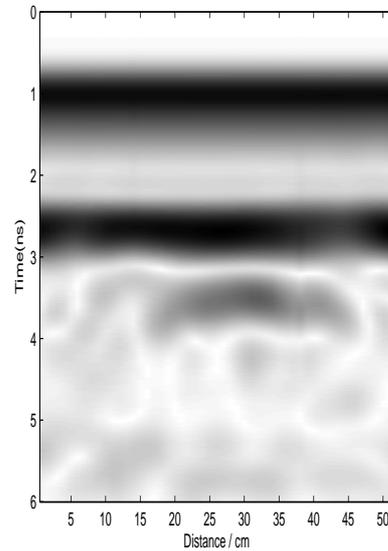
Figure 2.8: B-Scan of PMN landmine at depth 5 cm for data with 2 GHz bandwidth.
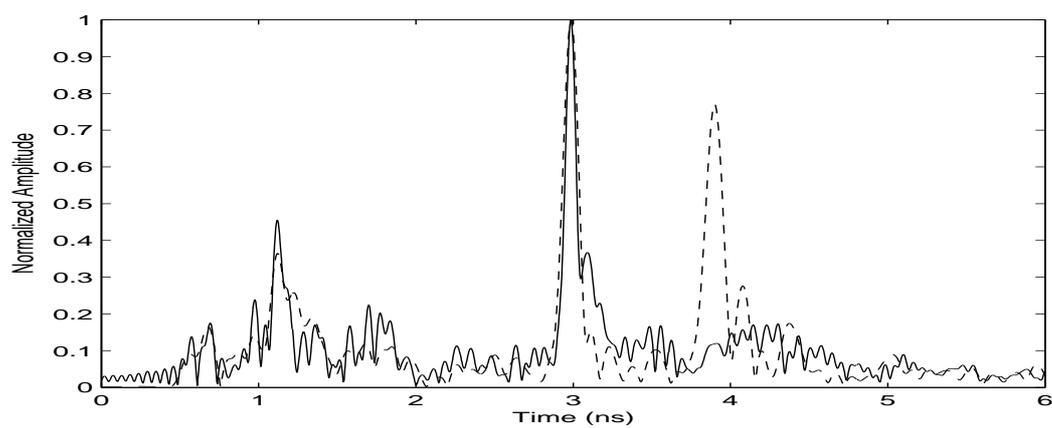
Figure 2.9: A-Scans in the presence (dashed) and absence (solid) of a mine for data with 19 GHz bandwidth.
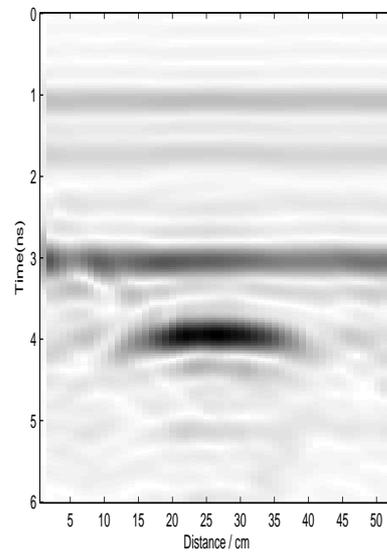
Figure 2.10: B-scan of PMN landmine at depth 10 cm for data with 19 GHz band-width.
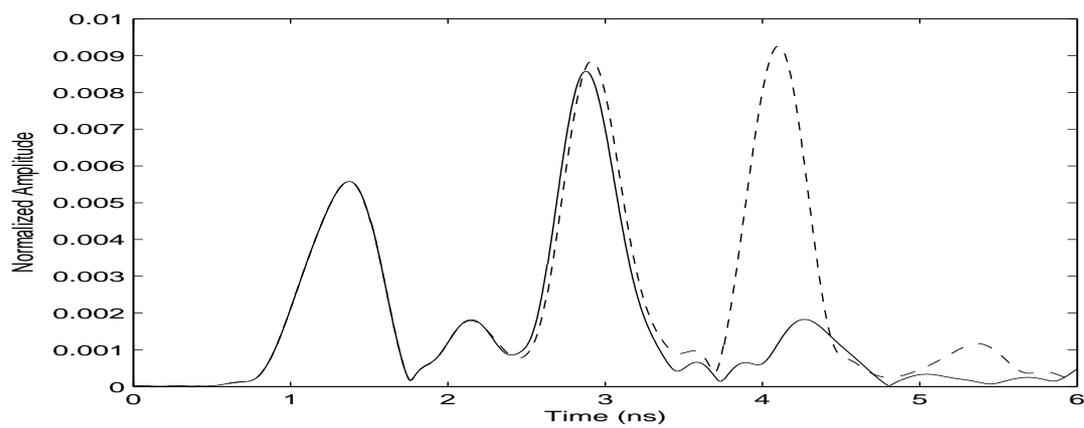
Figure 2.11: A-Scans in the presence (dashed) and absence (solid) of a mine for data with 5 GHz bandwidth.
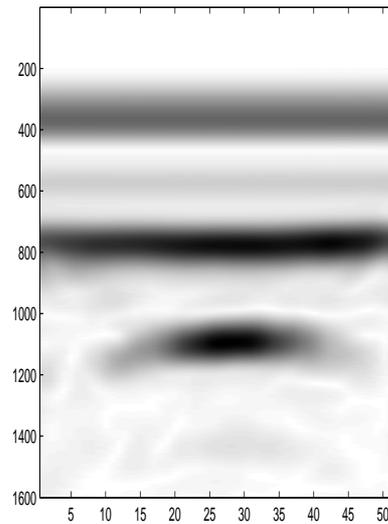
Figure 2.12: B-Scan of PPM landmine at depth 10 cm for data with 5GHz bandwidth.

1. Frequency range from 2 GHz to 4 GHz and the number of the samples is 1024 per A-Scan (2 GHz bandwidth). Figure 2.7 shows an example of A-Scans in the presence and absence of a landmine for this experimental. An example of B-Scan shows in Figure 2.8.

2. Frequency range of 1 GHz to 20 GHz and the number of the samples is 1601 per A-Scan (19 GHz bandwidth). Figure 2.9 shows an example of A-Scans in the presence and absence of a landmine for this experimental. An example of B-Scan shows in Figure 2.10.

3. Frequency range from 1 GHz to 6 GHz and the number of the samples is 1601 per A-Scan (5 GHz bandwidth). Figure 2.11 shows an example of A-Scans in the presence and absence of a landmine for this experimental. An example of B-Scan shows in Figure 2.12.

Chapter 3

Blind Separation of Linearly Mixed Sources

Blind Source Separation (BSS), or Blind Signal Separation, is concerned with the separation of a set of signals called source signals from their mixtures signals, without acquaintance of any information (or with very little information) about mixing background and sources. They are called 'blind' because they goal to estimate source signals from mixtures with unknown mixing coefficients. This problem can be solved by assuming that, the source signals are non-redundant. For example, the signals may be decorrelated or mutually statistically independent. Other assumptions of non-redundancy can be used. However, Blind Source separation is the separation of a set of signals into a set of other signals, which the regularity between the signals is minimized (decorrelation is minimized) or the regularity between the signals is maximized (statistical independence is maximized).
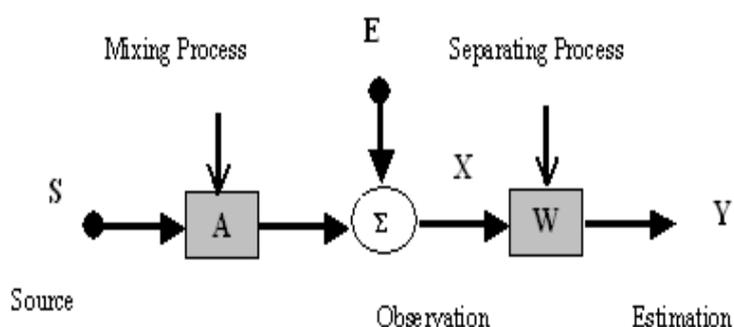
Figure 3.1: Block diagram illustrating the blind source separation (BSS).

These days the BSS became a well known and an important signal processing problem and applications [30], such as signal analysis, processing of speech, biomedical image analysis [30, 31] and seismic signals [32]. The most BSS problem is studied under the linear data model, where the observation data are linear mixtures, with unknown coefficients, of the source signals. To solve the BSS problems, mixing and unimixing processes of the unknown input sources $\mathbf{s}_j(k)(j = 1; 2; ...; n)$ may have different mathematical or physical models, depending on specific applications. The simplest case, m mixed signals $\mathbf{x}_i(k)(i = 1; 2; ...; m)$ or observed variables (generally correlated) are linear combinations of n (typically $m \geq n$) unknown source signals $\mathbf{s}_j(k)$. This can be written as

$$\mathbf{x}_i(k) = \sum_{j=1}^{n} a_{ij}\mathbf{s}_j(k) + \mathbf{e}_i(k) \tag{3.1}$$

$i = 1, ..., m$, or in the matrix notation

$$\mathbf{X} = \mathbf{AS} + \mathbf{E} \tag{3.2}$$

Where $\mathbf{x}(k) = [x_1(k), x_2(k), ..., x_m(k)]^T$ is a vector of sensor signals (observed variables), $\mathbf{s}(k) = [s_1(k), s_2(k), ..., s_n(k)]^T$ is a vector of sources, $\mathbf{e}(k) = [e_1(k), e_2(k), ..., e_m(k)]^T$ is a vector of additive noise, and $\mathbf{A}$ is an unknown full rank $m \times n$ mixing matrix [31].

BSS consists of identifying mixing matrix $\mathbf{A}$ and/or retrieving the source signals without resorting to any *a priori* information about $\mathbf{A}$; it uses only the information carried by the received signals themselves, hence, the term *blind*. Therefore, The objective of BSS is to find an inverse (unmixing) system, sometimes termed a reconstruction system. In order to estimate the all primary source signals $\mathbf{s}(k)$ or only some of them with specific properties. This estimation is performed on the basis of only the output signals $\mathbf{y}(k) = [y_1(k), ..., y_n(k)]^T$ and the sensor signals. Then the BSS problem is to determine a constant unmixing (weight, separating) matrix $\mathbf{W}$,(where $\mathbf{W}$ is $n \times m$) so that the solution is sought in the form

$$\hat{\mathbf{s}}(k) = \mathbf{y}(k) = \mathbf{Wx}(k) \tag{3.3}$$

has some suitable properties as shown in BSS block diagram Figure 3.1.

Different methods have been developed to find such a linear representation, including singular value decomposition [33], factor analysis [34], principal components analysis [35], independent component analysis [36], dependent component analysis [37], non-negative matrix factorization [38], projection pursuit [39],etc.

3.1 Singular Value Decomposition (SVD)

The main applications of signal processing in many fields (underwater acoustics, geophysics, etc.), are usually separated the initial data set into complementary subspaces

called signal and noise subspaces in order to enhance the signal-to-noise ratio. The singular value decomposition (SVD) is a useful tool to achieve this separation. It provides two orthogonal matrices that convey information on normalized wavelets and propagation vectors [40].

It is a matrix decomposition theorem of linear algebra that produces a two-dimensional separable unitary transformation of a matrix [41], see Figure 3.2 . Also SVD is a known algorithm for "diagonalizing" rectangular matrices [42]. The diagonalizing process is such that it promises that the diagonalizing matrices will be orthonormal. The SVD has had extensive application in signal [43, 44] and image processing [45] and can be used for a variety of purposes: noise reduction, information retrieval, compression, and patterns detection [46]. In this contribution, the SVD
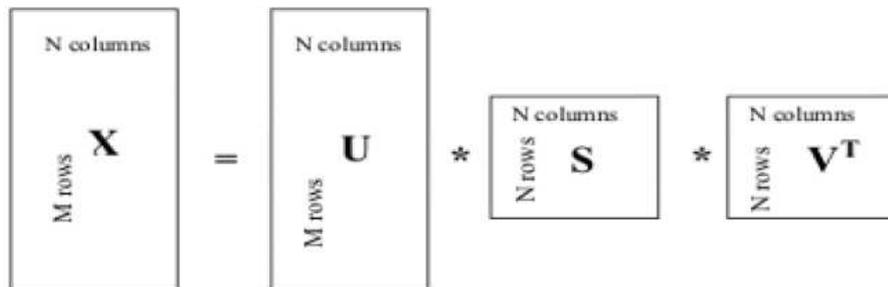
Figure 3.2: Graphical depiction of SVD of a matrix **X**

technique will be used for anti-personal landmine detection by using GPR system. For this purpose two algorithms have been developed: In the first algorithm, the SVD is used directly after mean subtraction method. A threshold is then applied to separate the target (mine) signal from the clutter. The success of this method relies on the choice of an optimum threshold value. A new formula has been developed to obtain an optimum threshold value. In addition the SNR has been analyzed and compared before and after the application of the SVD method with and without applying the threshold. In the second algorithm, the target signal has been estimated using two different methods. In the first method, the background is estimated from the data after SVD and the target is then calculated by subtracting this background from the data. In the second method, the target is calculated directly from the data after SVD. The SNR is analyzed in both cases.

3.1.1 Mathematical Definition of the SVD

Let us consider a B-scan image to be represented by a rectangular matrix $\mathbf{X}_{ij}$, whose dimension is $M \times N$, $(i = 1, 2..., M; j = 1, 2, ..., N)$. Here i denotes the time index and j denotes the antenna position index. We will assume that $M \geq N$. Then, the

rank of the matrix $r \leq N$. SVD of $\mathbf{X}$ (see Figure 3.2) is then given by:

$$\mathbf{X} = \mathbf{U}\mathbf{S}\mathbf{V}^T \tag{3.4}$$

Where $\mathbf{U}$ and $\mathbf{V}$ are $(M \times M)$ and $(N \times N)$ unitary matrices, respectively, and $\mathbf{S} = \text{diag}(\sigma_1, \sigma_2, ..., \sigma_r)$ with $\sigma_1 \geq \sigma_2 \geq ... \geq \sigma_r \geq 0$. The columns of $\mathbf{U}$ and $\mathbf{V}$ (namely $\{\mathbf{u}_M\}$ and $\{\mathbf{v}_N\}$) are called the left and right singular vectors, respectively. They are in fact the eigenvectors of $\mathbf{X}\mathbf{X}^T$ and $\mathbf{X}^T\mathbf{X}$, respectively. In addition, the two sets $\{\mathbf{u}\}$ and $\{\mathbf{v}\}$ fully span $\Re^M$ and $\Re^N$, respectively. Because $\mathbf{X}\mathbf{X}^T$ and $\mathbf{X}^T\mathbf{X}$ are both square and symmetric, their eigenvalues are real and positive, and their eigenvectors can be made orthonormal. Consequently

$$\mathbf{U}\mathbf{U}^T = \mathbf{U}^T\mathbf{U} = \mathbf{I}_M \tag{3.5a}$$

and

$$\mathbf{V}\mathbf{V}^T = \mathbf{V}^T\mathbf{V} = \mathbf{I}_N \tag{3.5b}$$

where $\mathbf{I}_M$ and $\mathbf{I}_N$ are the identity matrices in $\Re^M$ and $\Re^N$ respectively.

The matrix $\mathbf{S}$ is an $N \times N$ diagonal matrix $\mathbf{S} = \text{diag}(\sigma_1, \sigma_2, ..., \sigma_r)$. The diagonal elements of $\mathbf{S}$ are called the singular values. Furthermore, $\sigma_k > 0$ for $1 \leq k \leq r$, and $\sigma_k = 0$ for $r + 1 \leq k \leq N$. The matrix $\mathbf{S}$ can then be partitioned as,

$$S = \begin{pmatrix} \mathbf{D} \\ \mathbf{0} \end{pmatrix} \tag{3.6}$$

where $\mathbf{D}$ is an $N \times N$ diagonal matrix, and $\mathbf{0}$ is an $(M - N) \times N$ zero matrix. The main equations, in valuing $\mathbf{X}, \mathbf{U}, \mathbf{V}$ and $\mathbf{S}$ are:

$$\mathbf{X}\mathbf{X}^T u_i = \sigma_i^2 u_i \tag{3.7a}$$

and

$$\mathbf{X}^T\mathbf{X} v_i = \sigma_i^2 v_i \tag{3.7b}$$

The range of the index i requires some more explanation. The largest possible rank of $\mathbf{X}$ is N and the remainder $\mathbf{U}$ must be a null space. For the case of full-rank matrix $\mathbf{X}$, i.e. $r = N$, all N diagonal elements of $\mathbf{D}$ are nonzero. Consequently, the columns U_i; $N + 1 \leq i \leq M$; span the nullspace comprising all $\Re^M$ vectors that cannot be expressed as a linear combination of the columns of $\mathbf{X}$. The above limit on the range is then acceptable, despite the fact that there are more columns than this limit in $\mathbf{U}$. For $(r = N < M)$, the SVD is given schematically by

$$\mathbf{X} = \sigma_1 \begin{pmatrix} \vdots \\ u_1^T \\ \vdots \end{pmatrix} \begin{pmatrix} ... & v_1^T & ... \end{pmatrix} + \sigma_2 \begin{pmatrix} \vdots \\ u_2^T \\ \vdots \end{pmatrix} \begin{pmatrix} ... & v_2^T & ... \end{pmatrix} + ...+$$

$$\sigma_N \begin{pmatrix} \vdots \\ u_N^T \\ \vdots \end{pmatrix} \begin{pmatrix} ... & v_N^T & ... \end{pmatrix} \tag{3.8a}$$

or

$$\mathbf{X} = \sum_{i=1}^{N} \sigma_i u_i v_i^T \tag{3.8b}$$

or

$$\mathbf{X} = \mathbf{M}_1 + \mathbf{M}_2 + ... + \mathbf{M}_N \tag{3.8c}$$

where $\mathbf{M}_i$ are matrices of the same dimensions as $\mathbf{X}$, that are called its modes or *ith* eigenimage of $\mathbf{X}$. This representation of $\mathbf{X}$, as a series of matrices representing the various modes, is one of the most important aspects of SVD [47]. For example, the data matrix can be decomposed into two subspace signal and noise as

$$\mathbf{X} = \mathbf{X}_{signal} + \mathbf{X}_{noise} = \sum_{i=1}^{k} \sigma_i u_i v_i^T + \sum_{i=k+1}^{N} \sigma_i u_i v_i^T \tag{3.9}$$

This mean that the SVD can be used to separate the data to signal and noise subspace.

3.1.2 Application of the SVD to Clutter Reduction

Clutter reduction can be achieved using the SVD technique which is known as the subspace method. It is based on dividing the data into signal and noise subspaces, and let us consider a B-scan data (when the target is present) represented by the matrix $\mathbf{A}_{ij}$ (target and clutter) with dimensions $M \times N (M \geq N)$ and $(i = 1, 2, ..., M, j = 1, 2, ..., N)$, where the indices i and j are the time and the antenna position respectively. Let b_{ij} be B-scan data without the target, i.e., the reference B-Scan data without any buried object obtained from the laboratory measurements. Two algorithms are discussed here for reducing the clutter:

First algorithm

Considering the received data being consisting of four signal components, namely, the cross-talks between the transmitting and receiving antennas, the air-ground interface reflection, the background resulting from scatterers within the soil and the target scattered signal. In order to reduce the effect of the first two types of signals, the mean subtraction method has been applied. The mean subtraction has been done by calculating the mean vector of the B-scan, and then subtract it from each individual A-scan. After applying the mean subtraction method, we have the following:

$$X_{ij} = A_{ij} - \frac{1}{J} \sum_{j=1}^{J} A_{ij} \tag{3.10}$$

The singular value decomposition of the data $\mathbf{X}$ is given by

$$\mathbf{X} = \mathbf{USV}^T \tag{3.11a}$$

or

$$\mathbf{X} = \sum_{i=1}^{N} \sigma_i u_i v_i^T \tag{3.11b}$$

or

$$\mathbf{X} = \mathbf{M}_1 + \mathbf{M}_2 + ... + \mathbf{M}_N \tag{3.11c}$$

where the size of $\mathbf{M}_i$ is of the same size of the original data. The data $\mathbf{X}$ can be considered now as a sum of two signals

$$\mathbf{X} = \mathbf{R} + \mathbf{n} \tag{3.12}$$

where $\mathbf{R}$ contains the target signal and $\mathbf{n}$ is the undesired (clutter) signal. Our aim is to split $\mathbf{X}$ into its modes according to Equation (3.11c) such that the target signal and the clutter signal are contained in the first k and remaining $N - k$ modes respectively. Our task is to select k, which will optimize the partitioning as a function of SNR. Calculation of the SNR is done according to the empirical formula explained in [12] (description of SNR formula will be in Section 7.2.1). Table 3.1 gives the SNR for different k. For $k = 1$,

$$\mathbf{X} = \mathbf{M}_1 \quad \text{or} \quad \mathbf{X} = \sigma_1 u_1 v_1 \tag{3.13}$$

To improve the SNR, a properly chosen threshold can be applied to all samples in the $\mathbf{M}_1$ matrix, where all the samples that are less than the threshold value will be deleted. For this purpose two hypotheses are assumed

$$H_0 : \quad \mathbf{X} = \mathbf{n}$$

$$H_1 : \quad \mathbf{X} = \mathbf{R} + \mathbf{n} \tag{3.14}$$

where $\mathbf{R}$ and $\mathbf{n}$ represent the scattered field from a target, and the undesired signals, respectively. H_0 and H_1 are two hypotheses, which are tested against some threshold T. Our goal is to find a value for the threshold T such that the target signal is preserved after applying the threshold and the clutter is removed. The chosen threshold value depends on σ_1 (the first element of the $\mathbf{S}$ matrix diagonal) [48]. For this purpose we suggest the following new formula:

$$T = \sigma_1 \times \Theta_1 \times \Theta_2 \tag{3.15}$$

Where Θ_1 and Θ_2 are the standard deviation of u_1 and v_1 respectively. The SNR before and after applying the threshold are listed in Table 3.1. Any choice of the value other than T will decrease the value of SNR. Figure 3.3 shows the result after applying the threshold to the data in Figure 2.8. The imperative question is that, would T change according to data value and/or to the data length ? To answer this question, another test data have been analyzed, which have been acquired by IESK laboratory. The results of applying the last algorithm to the another data have given consistent results.
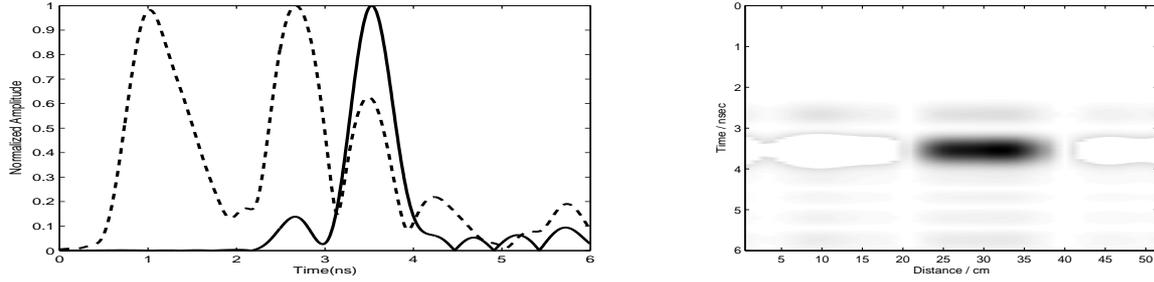
(a) A-scan of the received data (dashed) and the data after applying the threshold Equation (3.15)(solid).

(b) The B-Scans after applying the threshold Equation(3.15) to the data in Figure 2.8.

Figure 3.3: The data after applying SVD technique (First Algorithm).
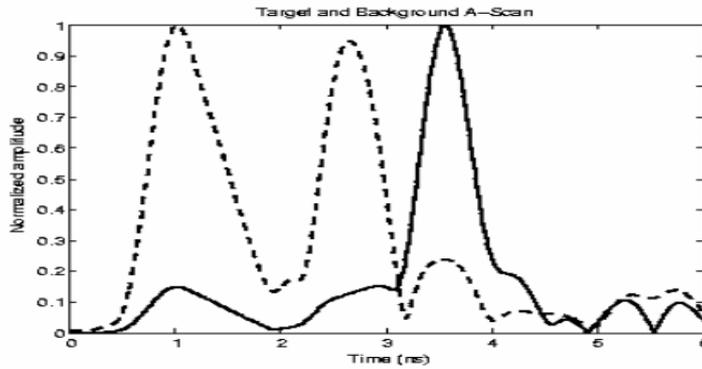
Figure 3.4: u1 component (dashed) and u2 component (solid) for data Figure 2.7.

Second algorithm

The estimation of the target signal $\mathbf{M}_t$ from the received data $\mathbf{X}$ in this algorithm can be done using two different methods. In the first method, the background $\mathbf{b}$ has been estimated from the received data $\mathbf{X}$, and then subtracted from $\mathbf{X}$. In the second method, the target signal has been estimated directly from $\mathbf{X}$. In the first method, the received data X can be splitted into three parts:

$$\mathbf{X} = \mathbf{R} + \mathbf{b} + \mathbf{n} \tag{3.16}$$

In other words, the received data matrix $\mathbf{X}$ can be splitted into three matrices representing target, clutter and noise signals as in Equation (3.4). Let us try to find out, whether there is a relation between these two representations. For this purpose we will assume that some of M_i refer to the background, some to the noise and some to the target, i.e.

$$\mathbf{X} = \mathbf{M}_t, \mathbf{M}_b, \mathbf{M}_n \tag{3.17}$$

where $\mathbf{M}_t, \mathbf{M}_b, \mathbf{M}_n$ are the target, background and noise signals, respectively. To estimate $\mathbf{M}_b$ from the original data, we follow the following steps: In the first step,
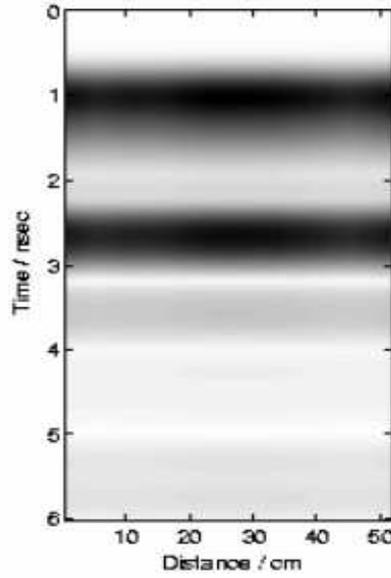
Figure 3.5: The clutter estimation by using Equation(3.18) for data in Figure 2.8.

some of the A-scans from the received data matrix $\mathbf{X}$ with some columns of the matrix $\mathbf{U}$ in Equation (3.4). As an example, let us compared the first column of $\mathbf{U}$ (i.e. u_1 shown in Figure 3.4 with an A-Scan without target Figure 2.7. In this case we can conclude that the first column of $\mathbf{U}$ refers to clutter only. In the second step, we will expand the comparison to B-scan, if we compare the reference clutter image $\mathbf{b}$ with the image $M_1 = \sigma_1 \times u_1 \times v_1^T$, we conclude that $\mathbf{M}_1$ refers to the clutter image. The background $\mathbf{b}$ can be estimated as

$$\hat{\mathbf{b}} = \mathbf{M}_1 = \sigma_1 \times u_1 \times v_1^T \tag{3.18}$$

where $\hat{\mathbf{b}}$ is the background estimation Figure 3.5. The mean square error (MSE) between the clutter estimation and the reference B-Scan (clutter) is given by

$$MSE = \frac{1}{M \times N} \sum_{i=1}^{M} \sum_{j=1}^{N} (\hat{b}_{ij} - b_{ij})^2 \tag{3.19}$$

$$\hat{R} = X - \hat{b} \tag{3.20}$$

The target signal can be obtained by subtracting the background form the original data $\mathbf{X}$ as shown in Figure 3.6.

$$\hat{\mathbf{R}} = \mathbf{X} - \hat{\mathbf{b}} \tag{3.21}$$

In the second method, we can estimate the target signal in the same way as in the above technique for the background estimation. If u_2 (Figure 3.4) is compared with A-scan in the presence of the target (Figure 2.7), we conclude that the second column
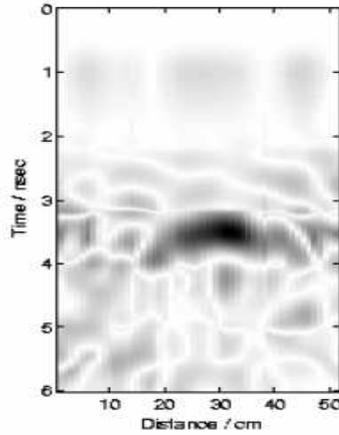
Figure 3.6: The target estimation by subtract the clutter estimation from the data by using Equation (3.21) for data in Figure 2.8.

of $\mathbf{U}$ (u_2) refers to the target signal. The target image estimated (Figure 3.7), can then be obtained from

$$\hat{\mathbf{R}} = \mathbf{M}_2 = \sigma_2 \times u_2 \times v_2^T \tag{3.22}$$

In the above procedure we assumed that the estimated background is $\mathbf{M}_1$, and the estimated target is $\mathbf{M}_2$. It is found that when $\mathbf{b}$ is chosen as $\mathbf{M}_1 + \mathbf{M}_i (i = 3 \rightarrow N)$, the MSE is increased. In Table 3.2 some different modes of $\mathbf{X}$ and the corresponding MSE are given. We conclude that the minimum value of MSE is:

$$\hat{\mathbf{b}} = M_1 \tag{3.23}$$

For images that contain two targets, the background also has been estimated from $\mathbf{M}_1 = \sigma_1 \times u_1 \times v_1^T$ as well and the first eigenvalue is found to be approximately unchanged. The first and the second target images are $\mathbf{M}_2 = \sigma_2 \times u_2 \times v_2^T$, and $\mathbf{M}_3 = \sigma_3 \times u_3 \times v_3^T$, respectively. Study of the data corresponding eigenvalues demonstrates the following: The first three normalized eigenvalues are 0.55, 0.18, and 0.07, where a normalized eigenvalue represents the percentage of the total variance associated to its corresponding structure (The structures are the elements of the eigenvectors of the data variance-covariance matrix). Using Equation (3.18), we will only be concerned with the first eigenvalue, which is 55% of the total variance. This means that the estimated clutter image ($\mathbf{M}_1$) has 55% of the total variance present in the original data and the target image ($\mathbf{M}_2$) has 18% of the total variance present in the original data. Figure 3.8 shows the speed with which the eigenvalues decay. Yet, the first two images (clutter and target) account for over 73% of the variance.
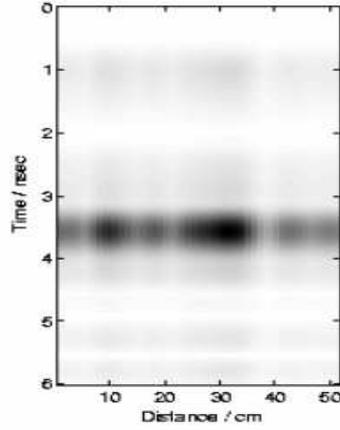
Figure 3.7: The target estimation by using Equation (3.22).
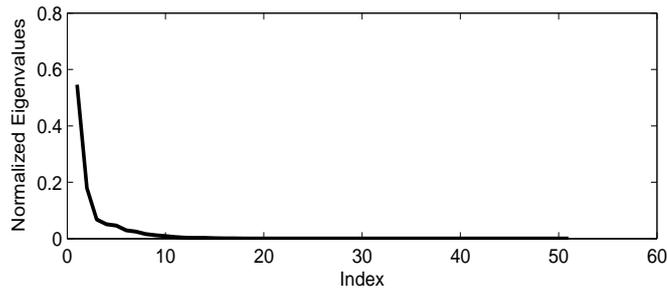
Figure 3.8: Normalized eigenvalues vs. eigenvectors of SVD.

k	1	2	3	5	10	15
SNR	8.3	6.9	6.3	5.7	5.5	5.5
SNR	11.3	10.8	10.4	9.37	9.3	9.3

Table 3.1: SNR in dB for SVD according to Equation (3.18). First and second row are SNR before and after thresholding.

X'Modes	MSE
M_1	8.3×10^{-3}
$M_1 + M_3$	8.3×10^{-3}
$M_1 + M_{3:4}$	8.3×10^{-3}
$M_1 + M_{3:5}$	8.3×10^{-3}
$M_1 + M_{3:6}$	8.3×10^{-3}
$M_1 + M_{3:8}$	8.3×10^{-3}

Table 3.2: MSE for different X' modes.

3.1.3 Experimental Results

In this contribution, the study of SVD technique for clutter reduction and non-metallic landmine detection of SF-GPR signals has been presented. Good results for automatic selection of SVD components have been achieved. Two techniques of SVD algorithms have been deployed. Ranking of SVD components shows that the background information and the target information contained in the second components respectively. On the other hand, when mean subtraction method has been used; the first component contains the target information. The threshold has been employed to improve the SNR, where a new threshold formula has been developed.

3.2 Factor Analysis

Factor analysis is a general term for a family of statistical techniques, which uses the correlations between observed variables to estimate common factors. Also, factor analysis is concerned with the dimensional (number of variables) reduction of a set of observed data in terms of a small number of latent factors.

The different methods of factor analysis first extract a set a factors from the data set. These factors are almost orthogonal and are ordered according to the proportion of the variance of the original data that these factors explain. In general, only a (small) subset of factors is kept for further consideration and the remaining factors are considered as either irrelevant or nonexistent (i.e., they are assumed to reflect measurement error or noise) [49]. The observed variables are modeled as linear combinations of the factors, plus "error" terms.

Let us consider the GPR data being represented by the rectangular matrix $\mathbf{X}$, whose dimension is $M \times K$, $(i = 1, 2, .., M; k = 1, 2, ..., K)$. In general, factor analysis assumes that every $\mathbf{x}_i$ is a linear combination of each $\mathbf{s}_j$ as follows:

$$\mathbf{x}_i = \sum_{j}^{N} a_{ij} \mathbf{s}_j + \mathbf{e}_i \tag{3.24}$$

$j = 1, 2, ..., N$ or in the matrix notation

$$\mathbf{X} = \mathbf{A}\mathbf{S} + \mathbf{E} \tag{3.25}$$

where $\mathbf{X}$ is the matrix holding the M A-scans in each row with k time samples, $\mathbf{S}$ is the matrix of factor scores (latent variables) with dimensions is $N \times k$, A is the matrix of factor loading with dimensions $M \times N$, and $\mathbf{E}$ is a matrix of residuals or error terms. Here N is a scalar denoting the number of factors to be used and is generally much smaller than M. An illustration of the factor analysis model is provided in Figure 3.9. Equation (3.24) is the fundamental model equation for many forms of factor analysis. In this model, it is assumed that the underlying distribution of $\mathbf{s}$ is $\mathcal{N}(0; \mathbf{I})$ and $\mathbf{e}$ follows $\mathcal{N}(0; \Psi)$, where $\Psi = \mathrm{diag}(\psi_1, \psi_2, ..., \psi_M)$. Since it is assumed that the $e_i's$

are uncorrelated and the $x_i's$ are conditionally uncorrelated given **S**. Another way of approaching the factor model is from the consideration of the covariance matrix. The factor model in terms of variances and covariances reads [49]
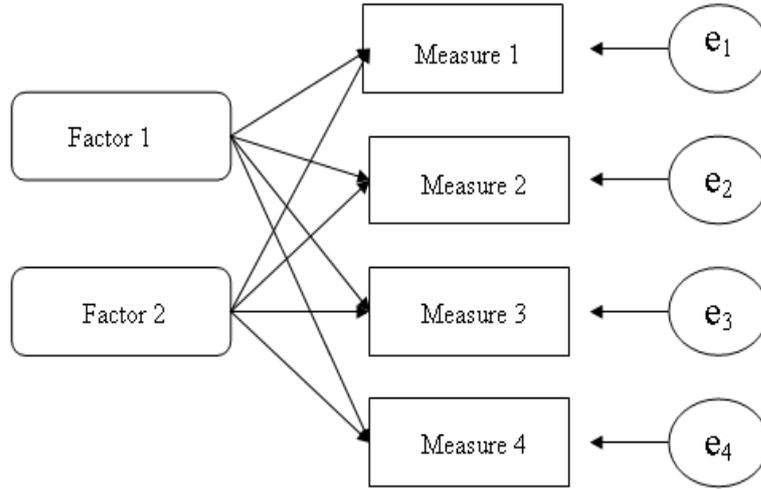
Figure 3.9: The model for factor analysis

$$\mathbf{\Sigma} = \mathbf{A}\mathbf{\Phi}\mathbf{A}^T + \mathbf{\Psi},$$

where $\mathbf{\Sigma}$ is the $M \times M$ population covariance matrix of the observed variables, $\mathbf{\Phi}$ is the $N \times N$ covariance matrix of the factors and $\mathbf{\Psi}$ is the $M \times M$ residual covariance matrix. Let us now assume that the factors are uncorrelated, that is, $\Phi = \mathbf{I}$, and the M-dimensional **e** is distributed according to $\mathcal{N}(0; \mathbf{\Psi})$, where $\mathbf{\Psi}$ is diagonal. The diagonality of $\mathbf{\Psi}$ is in fact one of the key assumptions of factor analysis. It means that the observed variables are conditionally independent (given the factors). According to this model, the distribution of **x** has zero mean and covariance $\mathbf{\Sigma}$.

The goal of the factor analysis is to find **A** and $\mathbf{\Psi}$ that best model the covariance structure of **x** . Several algorithms have been proposed in the literature for computing the matrices **A** and $\mathbf{\Psi}$. For example the squared multiple correlation coefficient (SMC) model assumed that [49]

$$\hat{\mathbf{\Psi}} = \left(diag C_x^{-1}\right)^{-1}$$

where $C_x = \mathbf{X}^T\mathbf{X}/K$. More efficient methods of estimating **A** and $\mathbf{\Psi}$ are the generalized least squares (GLS), which minimizes

$$tr\left(C_x^{-1}\mathbf{\Sigma} - \mathbf{I}\right)^2,$$

where tr is the trace operator and the maximum likelihood method (ML). In this work, the factor loadings matrix **A** and $\mathbf{\Psi}$ have been computed by using the Expectation-Maximization (EM) algorithm. An EM algorithm for a maximum likelihood estimation of factor analysis has been proposed in [50] and reviewed in [51] and [52].

3.2.1 The Maximum Likelihood (ML)

The Maximum likelihood is the procedure of finding the most probable parameter setting in a model given some data, which makes the known likelihood distribution a maximum. The maximum likelihood estimate for a parameter θ is denoted $\hat{\theta}$.

$$\hat{\theta} = arg \max_{\theta} P(\mathbf{X}|\theta). \tag{3.26}$$

$\theta = \{\theta_1, ..., \theta_k\}$ denotes unknown parameter to be estimate, and $P(\mathbf{X}|\theta)$ joint probability density function (PDF) or likelihood function. If the observed data is independent and identically distributed (iid) then the probability $P(\mathbf{X}|\theta) = \Pi_{i=1}^{N} P(\mathbf{x_i}|\theta)$.

The likelihood function can be used to obtain an estimator of the parameters θ. A natural idea is, for a given set of observations, to maximize the probability of those observations, i.e. maximize the likelihood function w.r.t. θ. For this it is more convenient to work with the logarithm of the likelihood which is defined as [53]

$$\ell(\theta) \equiv \ln P(\mathbf{X}|\theta) = \sum_{i=1}^{N} \ln P(\mathbf{x}_i|\theta). \tag{3.27}$$

In addition to $\mathbf{X}$, we also consider a complete data set $\mathbf{Z} = (\mathbf{X}, \mathbf{S})$ by including hiden variables $\mathbf{S}$, therefore, the joint probability density

$$P(\mathbf{z}|\theta) = P(\mathbf{x}, \mathbf{s}|\theta) = P(\mathbf{s}|\mathbf{x}, \theta)P(\mathbf{x}|\theta) \tag{3.28}$$

Associated with the joint density is the joint likelihood function

$$\ell(\theta|\mathbf{Z}) = \ell(\theta|\mathbf{X}, \mathbf{S}) = \prod_{i=1}^{N} P(\mathbf{x}_i, \mathbf{s}_i|\theta), \tag{3.29}$$

which is often called the complete likelihood. For the sake of computational simplicity it is often more convenient to consider the logarithm of the complete likelihood

$$l(\theta|\mathbf{Z}) = \log \ell(\theta|\mathbf{X}, \mathbf{S}) = \sum_{i=1}^{n} P(\mathbf{x}_i, \mathbf{s}_i|\theta) \tag{3.30}$$

EM algorithm is a two-step procedure repeated iteratively.

1. The Expectation step (E-step) the algorithm first finds the expected value of the complete log-likelihood function $\log \ell(\theta|\mathbf{X}, \mathbf{S})$ with respect to the unknown data $\mathbf{S}$.

2. The Maximization step (M-step) finds the optimal parameter values that maximize expectation which computed in the first step E-step. This procedure can be shown to increase the log-likelihood in Equation (3.30).
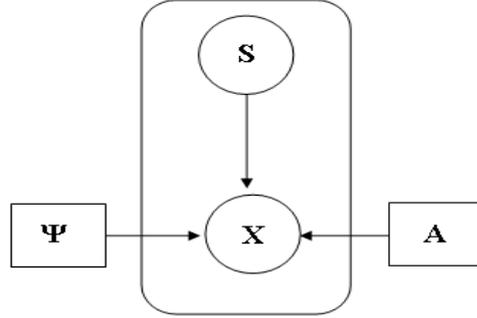
Figure 3.10: Graphical model for maximum likelihood factor analysis

Now, the log of the likelihood function can be written as [51]

$$\ell(\theta) = \log \left[\prod_{i=1}^{K} (2\pi)^{-M/2} |\mathbf{\Psi}|^{-1/2} \exp\left\{ -\frac{1}{2}(\mathbf{x}_i - \mathbf{A}\mathbf{s}_i)^T \mathbf{\Psi}^{-1} (\mathbf{x}_i - \mathbf{A}\mathbf{s}_i) \right\} \right]$$

The expectation of log likelihood function is

$$Q = -\frac{K}{2}\log|\mathbf{\Psi}| - \sum_i \left(\frac{1}{2}\mathbf{x}_i^T \mathbf{\Psi}^{-1}\mathbf{x}_i - \mathbf{x}_i^T \mathbf{\Psi}^{-1}\mathbf{A} \langle \mathbf{s}_i \rangle + \frac{1}{2}tr\left[\mathbf{A}^T \mathbf{\Psi}^{-1}\mathbf{A} \langle \mathbf{s}\mathbf{s}_i^T \rangle\right] \right) + c,$$

(3.31)

where c is constant, tr is the rance operator, $\langle \mathbf{s} \rangle = \beta\mathbf{x}$, $\beta = \mathbf{A}^T(\mathbf{\Psi} + \mathbf{A}\mathbf{A}^T)^{-1}$ and $\langle \mathbf{s}\mathbf{s}^T \rangle = \mathbf{I} - \beta\mathbf{A} + \beta\mathbf{x}\mathbf{x}^T\beta^T$.

Now, taking partial derivatives with respect to the parameters we get

$$\frac{\partial Q}{\partial \mathbf{A}} = -\sum_i^K \mathbf{\Psi}^{-1}\mathbf{x}_i \langle \mathbf{s}_i \rangle^T - \mathbf{\Psi}^{-1}\mathbf{A} \langle \mathbf{s}\mathbf{s}_i^T \rangle$$

(3.32)

$$\frac{\partial Q}{\partial \mathbf{\Psi}^{-1}} = \frac{K}{2}\mathbf{\Psi} - \sum_i^K \left(\frac{1}{2}\mathbf{x}_i\mathbf{x}_i^T - \mathbf{A} \langle \mathbf{s}_i \rangle \mathbf{x}_i^T + \frac{1}{2}\mathbf{A} \langle \mathbf{s}\mathbf{s}_i^T \rangle \mathbf{A}^T \right)$$

(3.33)

Setting the above equations to zero. The EM algorithm maximizes ℓ by iterating through E-step and M-step: In E-step, $\langle \mathbf{s} \rangle$ and $\langle \mathbf{s}\mathbf{s}^T \rangle$ are computed using current $\mathbf{A}$ and $\mathbf{\Psi}$. $\mathbf{A}$ and $\mathbf{\Psi}$ are then updated using the newly computed $\langle \mathbf{s} \rangle$ and $\langle \mathbf{s}\mathbf{s}^T \rangle$ in M-step (for the details see [53] and [52])

To ensure that $\mathbf{\Psi}$ is diagonal we set off diagonal terms to zero by the diag[.] operator as seen in Equation (3.35) of the M-step.

we obtain

$$\mathbf{A} = \mathbf{X} \langle \mathbf{S} \rangle \langle \mathbf{S}\mathbf{S}^T \rangle^{-1}$$

(3.34)

and

$$\mathbf{\Psi} = \frac{1}{K}\text{diag}\left[\mathbf{X}\mathbf{X}^T - \mathbf{A} \langle \mathbf{S} \rangle \mathbf{X}^T\right],$$

(3.35)

where $\langle \mathbf{S} \rangle = \mathbf{\Xi}\mathbf{A}^T\mathbf{\Psi}^{-1}\mathbf{X}$ and $\langle \mathbf{S}\mathbf{S}^T \rangle = K\mathbf{\Xi} + \langle \mathbf{S} \rangle \langle \mathbf{S} \rangle^T$, and $\mathbf{\Xi} = \left(\mathbf{I} + \mathbf{A}^T\mathbf{\Psi}^{-1}\mathbf{A}\right)^{-1}$
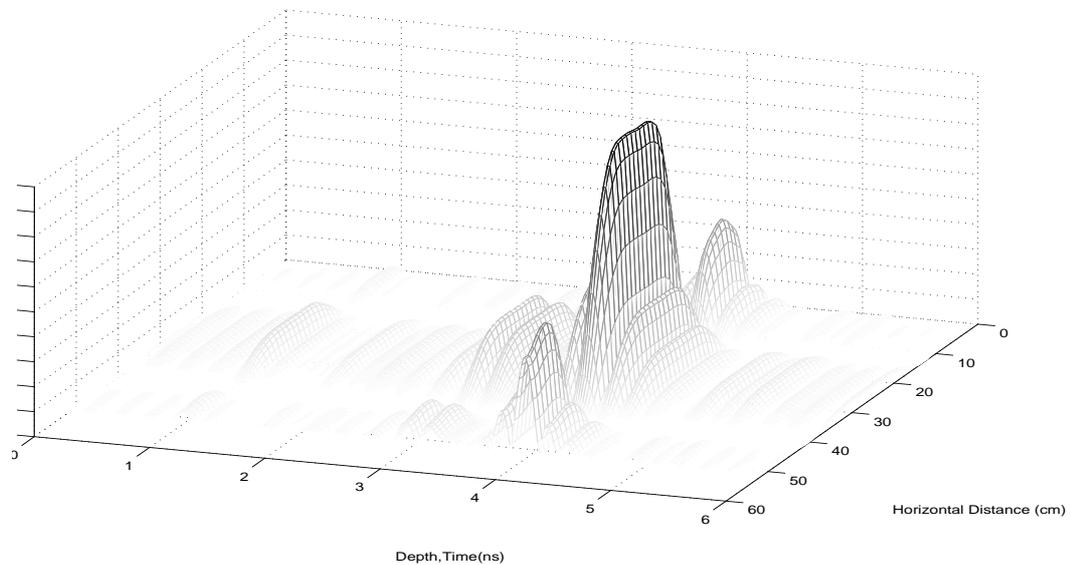
Figure 3.11: C-scan of one non-metallic Anti-personnel landmines (PMN) at location 3.9 ns after applying the factor analysis to the data in Figure 2.10.

3.2.2 Experimental Results

In this contribution, the study of factor analysis technique for clutter reduction and non-metallic landmine detection of SF-GPR signals has been presented. Figure 3.11 shows the C-scan of PMN landmine after applying the factor analysis to the data in Figure 2.10. The results have indicated that the factor analysis technique can be used to separate the air-ground interface from the target signal. The factor analysis is tested on relatively large anti-personnel landmine dummies (PMN). In all experiments, the target signal has been extracted and that of the air-ground interface has approximately been eliminated.

3.3 Principal Component Analysis

3.3.1 Introduction

The GPR data, consists of a large number of variables and large data sizes. Therefore, it is very likely that the variables are highly correlated with each other. The accuracy and reliability of detection methods will suffer if you include highly correlated variables or variables that are unrelated to the outcome of interest. Unnecessary variables can increase the data-collection and data-processing costs of deploying a model on
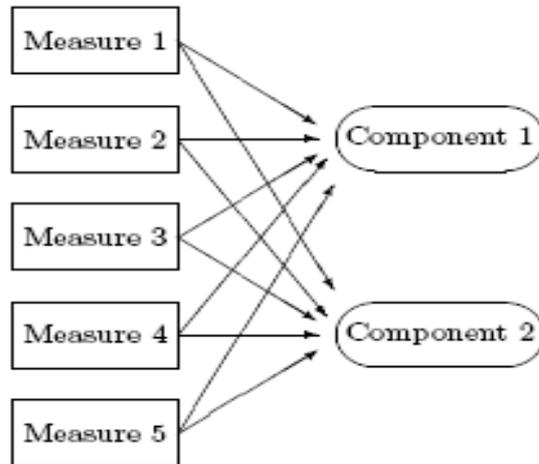
Figure 3.12: Graphical model for principal component analysis

a large database. One of the key steps in GPR data is finding ways to reduce dimensionality without sacrificing accuracy. The principal component analysis (PCA) is one of the most popular techniques for signal processing, data compressing, data visualization, image analysis and pattern recognition [54]. PCA has proven to be an exceedingly popular technique for dimensionality (number of variables) reduction and is discussed at length in most texts on multivariate analysis.

Also by applying PCA technique for dimensionality reduction, the noise from the data can be removed. Principal component techniques have been applied to GPR data analysis for detection of landmines by [55, 56].

As factor analysis, PCA is second-order statistical methods, only covariances between the observed variables are used in the estimation. Both methods have similar goals; namely to express data sets in terms of linear combinations of a small set of factors that are assumed to be uncorrelated and gaussianly distributed. However, there exists a crucial difference between the two methods. PCA is in fact a data reduction rather than a factor-analysis method. It became popular decades ago when computers were slow and expensive to use. It was a quicker, cheaper alternative to real factor analysis. It is computed without regard to any underlying structure caused by latent variables; components are calculated using all of the variance of the manifest variables. The first principal component accounts for as much of the variability in the data as possible, and each succeeding component accounts for as much of the remaining variability as possible. In most cases, the representation is sought as a linear transform of the observed variables.

3.3.2 Noise Reduction

Let us consider the GPR data being represented by an M-dimensional vector $\mathbf{x} = (x_1(k), x_2(k), ..., x_M(k))^T$ ($i = 1, 2, .., M; k = 1, 2, ..., K$), where i denotes the antenna position index, k denotes the time index. N principal components ($N < M$) of $\mathbf{x}$ are N (univariate) random variables $y_1(k), y_2(k), ..., y_N(k)$ which are defined by the following formula.

$$\mathbf{y} = \mathbf{A}^T\mathbf{x} \tag{3.36}$$

Where $\mathbf{x} = [x_1, x_2, ..., x_M]^T$ is the zero-mean input vector, $\mathbf{y} = [y_1, y_2, ..., y_N]^T$ is the output vector called the vector of principal components (**PCs**), and $\mathbf{A}$ is an $M \times N$ matrix that transforms $\mathbf{x}$ into $\mathbf{y}$.

The purpose of PCA is to derive a relatively small number of decorrelated linear combination (principal component) of a set of random zero-mean variables while retaining as much of the information from the original variables as possible [31]. The basic idea in PCA is to find the rows of the $y_1^T, y_2^T, ..., y_N^T$. PCA assumes that $\mathbf{A}$ is an orthonormal matrix ($A_i^T.A_j = \delta_{ij}$) such that the covariance matrix of Y (C_y) is diagonalized. In practice, the computation of A can be simply accomplished using the covariance matrix. Let $\mathbf{x}$ is the observed data after normalization to unity norm and subtracting the grand mean. The covariance matrix (C_x) reads then

$$C_x = \frac{1}{K} \sum_{k=1}^{K} \mathbf{x}(k)\mathbf{x}^T(k) \tag{3.37}$$

The eigenvector and eigenvalue matrices of (C_x) Φ and Λ, respectively, are computed as:

$$C_x\Phi = \Phi\Lambda, \tag{3.38}$$

where $\Lambda = diag(\lambda_1, \lambda_2, ..., \lambda_N)$, and $\lambda_1, \lambda_2, ..., \lambda_N$ are the eigenvales. If one assumes that eigenvalues are sorted in a decreasing order, $\lambda_1 \geq \lambda_2 \geq, ..., \geq \lambda_N$, then the N leading eigenvectors matrix A is given by

$$A = [\Phi_1, \Phi_2, ..., \Phi_N] \tag{3.39}$$

In the application of landmine detection using GPR signals, the PCA can be used to detect landmines and reduce the noise. This is done by selecting some components that mainly carry mine information, say A_p which the rest represent the clutter. The reconstructed clutter-free signal space can be extracted from the original GPR entire signal space according to [57]

$$\mathbf{s} = \mathbf{A}_p^T\mathbf{x} \tag{3.40}$$

We select the matrix A to be a matrix whose rows A_p are the eigenvectors of C_x (the principal components of $\mathbf{x}$). Computing PCA of a data set $\mathbf{x}$ entails (1) subtracting the mean of each measurement type and (2) computing the eigenvectors of C_x [58].

3.3.3   Experimental Results

From Equation (3.40), let us select the PC components which mainly contain information about the mine object i.e. $\mathbf{s} = \mathbf{A}_p^T\mathbf{x}$, where the size of $\mathbf{s}$ as the same dimension of $\mathbf{x}$. Here the selection can be done by inspecting the structure of the eigenimage or by the time course of $s_i(k)$. The reduction of the clutter can be done by reconstructing $\mathbf{x} = \mathbf{A}\mathbf{s}$ as following: Let us assume that the matrix $\mathbf{A}$ is deterministic and full rank. Furthermore, $\mathbf{A}$ is orthonormal, multiplying both sides of $\mathbf{x} = \mathbf{A}\mathbf{s}$ from the left by A^T immediately yields the coefficients of the expansion [59]:

$$s_i(k) = A_i^T\mathbf{x}(k) \Rightarrow \mathbf{s}(k) = \mathbf{A}^T\mathbf{x}(k) \tag{3.41}$$

The s_i are called the principal components. Now, assume the eigenvalues have been arranged in decreasing order, so that λ_1 is the largest eigenvalue, and let $\mathbf{A} = [A_t A_n]$, where the columns of $\mathbf{A}$ have been separated into the t (refer to the target) and n (refer to the noise) eigenvectors, respectively, and $N = t + n$.

Now, the target eigenimage is [60]

$$\mathbf{s}_t(k) = A_t^T\mathbf{x}(k) \tag{3.42}$$

The matrix $\mathbf{A}_t$ is not full rank, and thus the input can only be approximately reconstructed. Figure 3.13 shows the raw data in the first row. The eigenimages and associated PCs are shown in the second and third rows. The first eigenimage refer to the clutter and the second eigenimage refer to the target. Therefore, only two components have been contributed. From figures and in all measurements, the reflection from the surface has been removed or has been at least reduced.
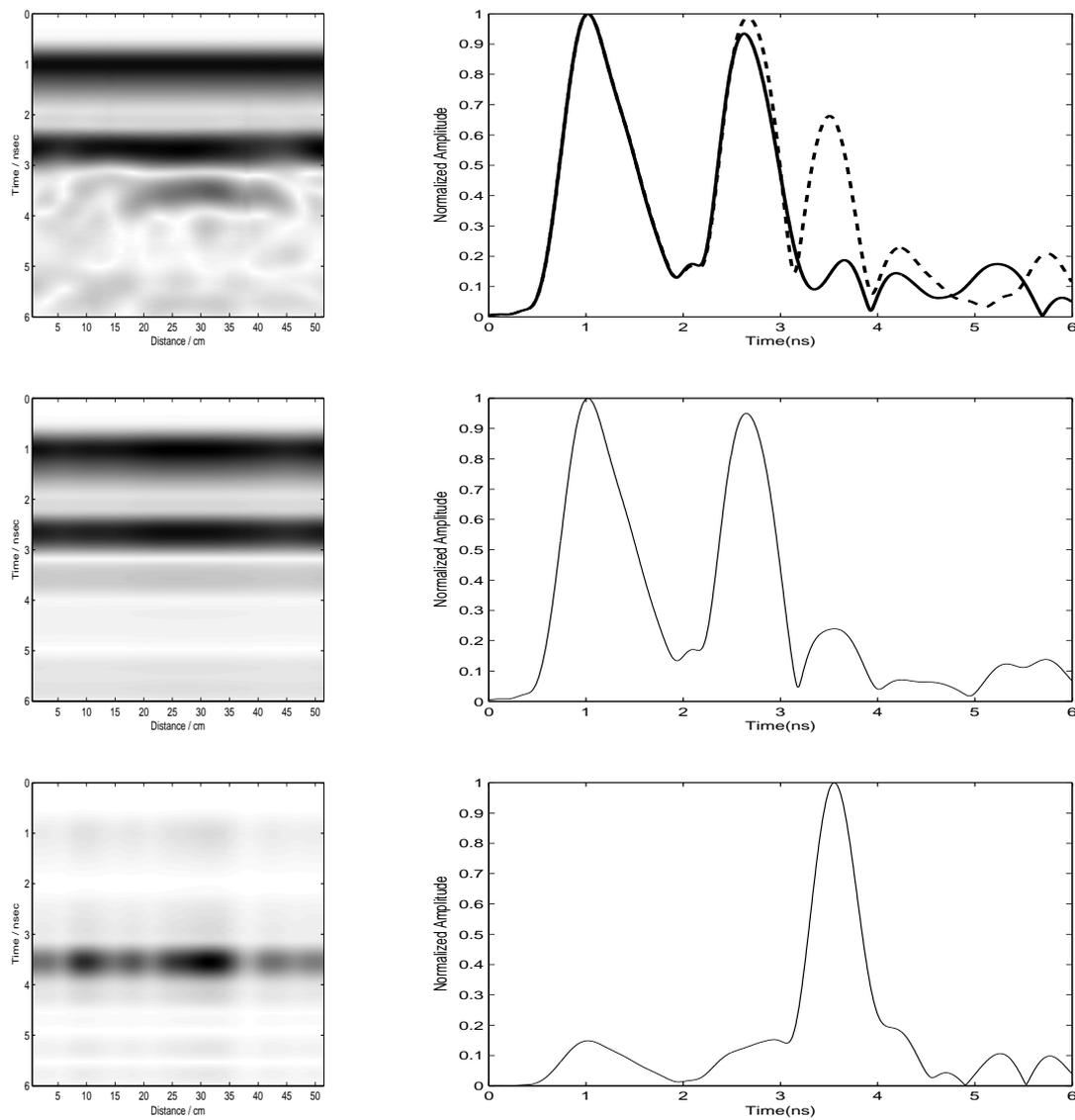
Figure 3.13: Eigenimages and associated source signal. First row is the raw data, second and third rows are the eigenimages and associated source signal.

Chapter 4

Independent Component Analysis (ICA)

In the previous chapter, three common techniques which were used to separate signal components based on blind source separation have been introduced. These techniques which have been discussed and applied to GPR data are the singular value decomposition, the factor analysis and the principle component analysis. These techniques decompose the signal sources using only the second order statistics. The objectives of those algorithms have been removing the second order correlations in the input data, while higher order dependencies remain . For Gaussian sources, these only solve the sources-separation.

In this chapter, a method called Independent Component Analysis (ICA) is reviewed (ICA enables the data to be represented by statistically independent components unlike principal component analysis or factor analysis, which lead to uncorrelated components). ICA is used aiming to solve the blind source separation problem. As in the previous chapter, we have found that many different uncorrelated representations of the signals would not be independent and would not separate the sources. Uncorrelatedness in itself is not enough to separate the components [30]. For this reason principal component analysis (PCA) or factor analysis (FA) cannot separate the signals: they give components that are uncorrelated, but little more [30]. The statistical independence takes into consideration higher order moments (because it is a stronger statistical property than decorrelation (the second order statistic used in PCA)). The usage of ICA to solve the BSS problem has been largely studied last years by researchers from the signal processing community. The idea of the algorithm was first introduced by Jeanny Hérault and Christian Jutten [36,61]. The solution of the BSS problem is done by assuming the sources statistically independent and have non-gaussian marginal distributions.[1]

---

[1]To be precise, at most one of the the sources can have Gaussian distribution.
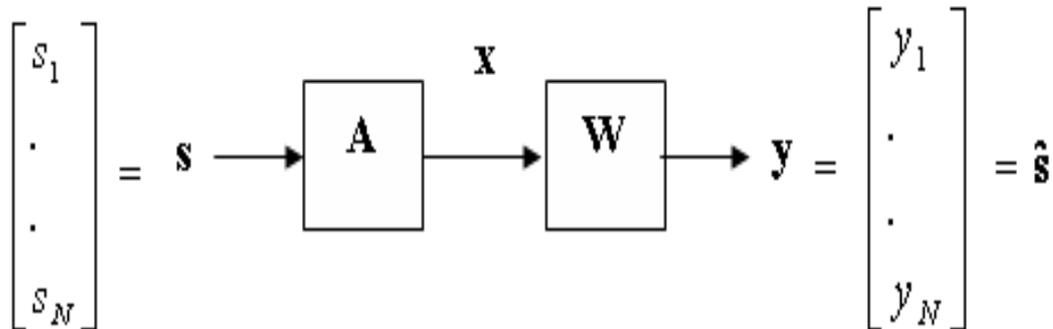
Figure 4.1: Independent Component Analysis (ICA) model. Unobserved signals: **s**; observation: **x**; estimated source signals: **y**, mixing and separating [64].

In this work we assumed that the sources are independent and linearly mixed using a stationary and instantaneous mixing. Furthermore, we assume that there exists at least as many nondegenerate mixtures as there are sources and that at most one of the sources has a Gaussian distribution. We as well assume that there exist an infinite amount of independent samples of the mixtures. ICA can be used either to solve the BSS problem [62], or as a feature extraction technique [63]. In BSS, the signals are separated assuming the original underlying source signals are mutually independently distributed. The other central application for ICA is feature extraction, where it provides a set of bases, which can be used to represent the observed data.

In this chapter the ICA is introduced in more details and several algorithms are reviewed for this implementation. Four of the most common ICA algorithms called: extended Infomax, the FastICA, Joint Approximate Diagonalization of Eigenmatrices (JADE), and the second order blind identification (SOBI) have been chosen for this purpose. We will apply these algorithms to GPR data, and evaluate their ability to separate the target and clutter signals.

4.1 Mathematical Notation

With the advent of the powerful computers, the ICA become feasible [65]. ICA is a set of mathematical techniques that are used to pull a little bit of information out of a large mass of data. It is a method used in the analysis and representation of multivariate data. In the most fundamental ICA model it is assumed that the observed data **x** have been generated from source data **s** through a linear process $\mathbf{x}=\mathbf{As}$, where both the sources **s** and the mixing matrix **A** are unknown. The sources **s** are assumed to be statistically independent of each other - hence the name Independent Component Analysis. ICA algorithms are able to estimate both the sources **s** and the mixing matrix **A** from the observed data **x** with very few assumptions. Let us

consider the GPR data being represented by the rectangular matrix $\mathbf{X}$, whose dimension is $M \times K$, $(i = 1, 2, .., M; k = 1, 2, ..., K)$, where i denotes the antenna position index, k denotes the time index. In general, ICA assumes that every $\mathbf{x}_i$ is a linear combination of each $\mathbf{s}_j$ as follows:

$$\mathbf{x}_i = \sum_{j}^{N} a_{ij}\mathbf{s}_j \tag{4.1}$$

$j = 1, 2, ..., N$ or in the matrix notation

$$\mathbf{X} = \mathbf{AS} \tag{4.2}$$

where $\mathbf{X}$ is the matrix holding the M A-scans in each row with k time samples, A is an $M \times N$ basis transformation or mixing matrix, and $\mathbf{S}$ is the matrix holding the N independent source signals in rows of K samples. The ICA of the vector $\mathbf{x} = (x_1, x_2, ..., x_M)^T$ is obtained by finding an $N \times M$, full rank, linear transformation (separating) matrix $\mathbf{W}$ such that the output signal vector $\mathbf{y} = (y_1, y_2, ..., y_N)^T$ defined by $\mathbf{y} = \mathbf{Wx}$ (see Figure 4.1), contains components that are as independent as possible, as measured by an information theoretic cost function. In other words, it is required to adapt the weights w_{ji} of the $N \times M$ matrix $\mathbf{W}$ of the linear system $\mathbf{y}(k) = \mathbf{Wx}(k)$ to combine of the observations $x_i(k)$ to generate estimations of the source signals

$$\hat{\mathbf{s}}_j(k) = \mathbf{y}_j(k) = \sum_{i=1}^{M} w_{ji}\mathbf{x}_i(k) \tag{4.3}$$

Or in matrix notation

$$\hat{\mathbf{S}} = \mathbf{Y} = \mathbf{WX} \tag{4.4}$$

Thus, the ICA solution for this problem is to find a linear transformation matrix $\mathbf{W}$, of dimension $N \times M$, which makes the outputs $\hat{\mathbf{s}} = (\hat{s}_1, \hat{s}_2, ..., \hat{s}_N)^T$ from the linear transformation of the dependent sensor signals $\mathbf{x}$ as independent as possible. The transformation $\hat{\mathbf{s}}$ is an estimate of $\mathbf{s}$, and if we for simplicity have $N = M$ and would have $\mathbf{W} = \mathbf{A}^{-1}$, then the independent components are exactly recovered up to a permutation and a scale. Hence, $\hat{\mathbf{s}} = \mathbf{A}^{-1} \mathbf{A} \mathbf{s} = \mathbf{Is}$, where $\mathbf{I}$ is the identity matrix. The starting point for ICA is the very simple assumption that the components $\mathbf{s}_j$ are statistically independent. For the reasons listed in [[65–67]], we must also assume that the independent components must have non-Gaussian distributions (except possibly one).

4.2 Statistical Independence

In the ICA method it is assumed that the output signals are independent. In this way if the set of N signals are originated from different physical sources then they

tend to be statistically independent of each other. The ICA method assumes that if a set of independent signals can be extracted from a mixture of signals then these extracted signals are likely to be the original source signals. ICA requires assumptions of independence that involve the cumulative distribution functions (cdfs) of source signals [68].

However, the principle of independence has been used for determining demixing matrix $\mathbf{W}$. The components $\mathbf{s}_j$ should be statistically independent. This means that the value of any one of the components gives no information on the values of the other components. Let us now consider a zero-mean N-dimensional random vector $\mathbf{s} = [s_1, s_2, ..., s_N]^T$, corresponding to N real valued, independent components $\mathbf{s}_j$. The probability density function (pdf) of the vector $\mathbf{s}$ can be written as the product of the marginal distributions of $\mathbf{s}_j$,

$$P(\mathbf{s}) = \prod_{j=1}^{N} P(\mathbf{s}_j) \tag{4.5}$$

To fulfill the requirements of statistical independence, the pdf product equality in Equation (4.5) must be satisfied. The covariance of two statistically independent variables is always zero. The opposite is not always true. However, in the special case of Gaussian variables, zero covariance does imply independence. This feature of Gaussian variables is used to find columns of $\mathbf{W}$ in $\mathbf{WX}{=}\mathbf{S}$ [69].

The basic idea of the ICA is to minimize the dependency among the output components. The dependency is measured by Kullback-Leibler (KL) divergence

$$KL = \int P(\mathbf{s}) \log \frac{P(\mathbf{s})}{\prod_{j=1}^{N} P(\mathbf{s}_j)} \; d(\mathbf{s}) \tag{4.6}$$

The KL divergence measures the distance between two probability distributions, and becomes zero when the distributions are equal. However it should be noted that the KL divergence is not symmetrical. Therefore the independent components can be found by minimizing the Kullback-Leibler divergence between the joint probability and marginal probabilities of the output signals [70]. Previously we stated that each measured signal in $\mathbf{X}$ is a linear combination of the independent signals in $\mathbf{S}$. The mixing matrix $\mathbf{A}$ is invertible such that $\mathbf{A}^{-1} = \mathbf{W}$. Each of the independent components in $\mathbf{S}$ can also be expressed as a linear combination of the measured signals in $\mathbf{X}$ ($\mathbf{S}{=}\mathbf{WX}$). The process of finding $\mathbf{W}$ is based on the Central Limit Theorem, which states that a sum of non-Gaussian random variables (such as those in $\mathbf{S}$) is more likely a Gaussian than are its individual components (e.g. if $\mathbf{x}_i$ is the sum of two components such that $\mathbf{x}_i = a_1 s_1 + a_2 s_2$, so $\mathbf{x}_i$ is more Gaussian than either s_1 or s_2). But the sought independent sources can be written as sums of the observed data. Thus, we can move toward independent sources by trying to find a sum of the observed data over vectors, which have maximum non-Gaussianity. This property can be measured in terms of higher order statistics, as kurtosis or negentropy. Once $\mathbf{W}$ is determined it is a simple matter to invert it to find $\mathbf{A}$. There are several algorithms were developed to measure the non-Gaussianity of $\mathbf{WX}$.

4.3 Measures of Non-Gaussianity

The measurers of non-Gaussianity are used to solve the ICA problem including higher-order moments kurtosis and negentropy, which measure the divination from the Gaussian distribution. Also minimizing mutual information and maximum likelihood estimation (Infomax) are used which are not direct measures of non-Gaussianity but lead to the same results. To use non-Gaussianity in ICA estimation, we must have a quantitative measure of non-Gaussianity of a random variable, say $\mathbf{s}$.

4.3.1 Kurtosis

The classical measure of non-Gaussianity is the kurtosis or the fourth-order cumulant. The kurtosis of a zero mean random variable $\mathbf{s}$ is defined by

$$\kappa(\mathbf{s}) = E\left\{\mathbf{s}^4\right\} - 3\left(E\left\{\mathbf{s}^2\right\}\right)^2 \tag{4.7}$$

When data is preprocessed to have unit variance, the right-hand side simplifies to $E\left\{\mathbf{s}^4\right\} - 3$, thus, the kurtosis is equal to the fourth moment of the data. This shows that kurtosis is simply a normalized version of the fourth moment $E\left\{\mathbf{s}^4\right\}$. For a gaussian $\mathbf{s}$, the fourth moment equals $3\left(E\left\{\mathbf{s}^2\right\}\right)^2$. Thus, kurtosis is zero for a gaussian random variable. For most (but not quite all) non-gaussian random variables, kurtosis is nonzero. Kurtosis can be both positive or negative. When the $\kappa(\mathbf{s}) > 0$ are considered supergaussian while those with $\kappa(\mathbf{s}) < 0$ are considered subgaussian. Thus, the non-gaussianity of a variable may be incremented simply by maximizing $|\kappa(\mathbf{s})|$; in this case the $\kappa(\mathbf{s}_i)$ is determined by the ith row vector $\mathbf{w}$ of $\mathbf{W}$.

4.3.2 Negentropy

negentropy is used as a measure of distance to normality. If the signal is Gaussian, the signal is said to have a normal distribution. Negentropy is always positive, is invariant by any linear invertible change of coordinates, and vanishes if and only if the signal is Gaussian. Negentropy is defined as

$$J(\mathbf{s}) = H(\mathbf{s}_{gauss}) - H(\mathbf{s}) \tag{4.8}$$

where $\mathbf{s}_{gauss}$ is a Gaussian random variable of the same covariance matrix as $\mathbf{s}$, and $H(\mathbf{s})$ is the differential entropy:

$$H(\mathbf{s}) = -\int P(\mathbf{s}) \log P(\mathbf{s}) d\mathbf{s} \tag{4.9}$$

where P is pdf. The gaussian variable has the largest entropy among all random variables of equal variance, therefore negentropy is always non-negative, and it is zero if and only if $\mathbf{s}$ has a Gaussian distribution. The Negentropy can be used to measure the non-gaussianity, but it is computationally very difficult. Therefore, simpler approximations of negentropy are very useful [67].

4.3.3 Approximating the Negentropy

The classical method of approximating the Negentropy is using higher order cumulants as [71]

$$J(\mathbf{s}) \approx \frac{1}{12} C_3(\mathbf{s})^2 + \frac{1}{48} C_4(\mathbf{s})^2 \qquad (4.10)$$

where $C_i(\mathbf{s})$ is the ith order cumulant of $\mathbf{s}$.

The above approximation is inaccurate due to finite sample size and expectation is not reliable, and the lack of robustness (sensitivity to outliers) is due the fact that samples contribute as their third or fourth power to the estimation of moments (such as skewness and kurtosis, first and second terms in the above equation) hence an outlier may strongly bias the estimated values. An alternative solution of the problems associated with Equation (4.10), which is still an approximation of negentropy. The polynomial functions $\mathbf{s}^3$ and $\mathbf{s}^4$ can be replaced by any two non-quadratic functions G_1 and G_2. In the case one non-quadratic function, a new approximation is proposed in [72] and can also be found in [66, 73]. It has the general form of the following

$$J(\mathbf{s}) \approx \left[E\left\{ G(\mathbf{s}) \right\} - E\left\{ G(\nu) \right\} \right]^2 \qquad (4.11)$$

In this equation $G(\mathbf{s})$ is some non-quadratic function, ν is a Gaussian variable of zero mean and unit variance. If $G(\mathbf{s}) = \mathbf{s}^4$ this equation becomes equal to kurtosis. There are functions that can be used for G that give a good approximation to the negentropy and are less sensitive to outliers than kurtosis [74].

4.3.4 Mutual Information

Mutual information (MI) is a major concept of information theory. MI is a quantity that measures the mutual dependence of the two random variables. It equals zero if and only if the variables are statistically independent, and is greater than zero otherwise. The mutual information of the components of the vector $\mathbf{s}_i$ $i = 1, 2, ..., N$ is defined as

$$I(s_1, s_2, ..., s_N) = \sum_{i=1}^{N} H(\mathbf{s}_i) - H(\mathbf{s}) = \int P(\mathbf{s}) \log \frac{P(\mathbf{s})}{\prod_{j=1}^{N} P(\mathbf{s}_j)} \, d(\mathbf{s}) \qquad (4.12)$$

Where $P(\mathbf{s})$ and $P(\mathbf{s}_i)$ denote the joint density and marginal density of $\mathbf{s}$ and $\mathbf{s}_i$. As $\mathbf{s}$ can be expressed as a linear combination of the measured signals in $\mathbf{x}$, $\mathbf{s}=\mathbf{W}\mathbf{x}$, it can be shown via a change of integration variable, that

$$H(\mathbf{s}) = H(\mathbf{x}) + \log \det \mathbf{w} \qquad (4.13)$$

therefore,

$$I(s_1, s_2, ..., s_N) = \sum_{i=1}^{N} H(\mathbf{s}_i) - \log \det \mathbf{w} - H(\mathbf{x}) \qquad (4.14)$$

The term $H(\mathbf{x})$ is a constant, therefore the criterion of the independency is

$$C(\mathbf{w}) = \sum_{i=1}^{N} H(\mathbf{s}_i) - \log \det \mathbf{w} \tag{4.15}$$

to be minimized with respect to $\mathbf{w}$. This (theoretical) criterion contains only the marginal entropies $H(\mathbf{s}_i)$ which have to be estimated [67, 75].

4.3.5 Maximum Likelihood Estimation

The maximum likelihood estimation is a very popular method to estimate the independent components [76]. If the source number are equal to independent components number $N = M$, therefore the mixing matrix is square, Also if it has full rank then the unmixing matrix $\mathbf{W} = \mathbf{A}^{-1}$. The log-likelihood can be written as

$$\log P(\mathbf{x}|\mathbf{A}) = N \log \det \mathbf{A}^{-1} + \sum_{i=1}^{N} \log P(\mathbf{s}_i) \tag{4.16}$$

differentiate above equation with respect to $\mathbf{W}$ we can obtain the gradient for updating the unmixing matrix in an iterative optimization method,

$$\frac{\partial}{\partial \mathbf{W}} \log P(\mathbf{x}|\mathbf{A}) = \frac{\partial}{\partial \mathbf{W}} N \log \det \mathbf{W} + \sum_{i=1}^{N} \frac{\partial \log P(\mathbf{s}_i)}{\partial \mathbf{s}_i} \frac{\partial \mathbf{s}_i^T}{\partial \mathbf{W}} \tag{4.17}$$

By substituting $\Phi(\mathbf{s}_i) = \frac{\partial}{\partial \mathbf{s}_i} \log P(\mathbf{s}_i)$ therefore

$$\frac{\partial}{\partial \mathbf{W}} \log P(\mathbf{x}|\mathbf{A}) = N \left(\mathbf{W}^T\right) + \Phi(\mathbf{s})\mathbf{X}^T, \tag{4.18}$$

if $\Phi = -\tanh$ this gives the infomax solution.

4.4 Preprocessing Data

Before applying an ICA algorithm on the data, it is usually useful to do some preprocessing. In this section, we discuss some preprocessing techniques that make the problem of ICA estimation simpler and better conditioned.

4.4.1 Centering

The most basic and necessary preprocessing is to center $\mathbf{X}$, i.e. subtract its mean vector $m = E\{\mathbf{X}\}$ so as to make $\mathbf{X}$ a zero-mean variable. This implies that $\mathbf{S}$ is zero-mean as well, as can be seen by taking expectations on both sides of Equation (4.2).

4.4.2 Whitening

Another useful preprocessing strategy in ICA is to first whiten the observed variables. This means that before the application of the ICA algorithm (and after centering), we transform the observed vector $\mathbf{X}$ linearly so that we obtain a new vector $\tilde{\mathbf{X}}$, which is white, i.e. its components are un-correlated and their variances equal unity. In other words, the covariance matrix of $\tilde{\mathbf{X}}$ equals the identity matrix:

$$E\left\{\tilde{\mathbf{X}}\tilde{\mathbf{X}}^T\right\} = \mathbf{I} \tag{4.19}$$

The whitening transformation is always possible. One popular method for whitening is to use the eigenvalue decomposition (EVD) of the covariance matrix $E\left\{\mathbf{X}\mathbf{X}^T\right\} = \mathbf{E}\mathbf{D}\mathbf{E}^T$, where $\mathbf{E}$ is the orthogonal matrix of eigenvectors of $E\left\{\mathbf{X}\mathbf{X}^T\right\}$ and $\mathbf{D}$ is the diagonal matrix of its eigenvalues, $\mathbf{D} = \mathrm{diag}(d_1, ..., d_n)$. Whitening can now be done by

$$\tilde{\mathbf{X}} = \mathbf{E}\mathbf{D}^{-1/2}\mathbf{E}^T\mathbf{X}, \tag{4.20}$$

where the matrix $\mathbf{D}^{-1/2}$ is computed by a simple component-wise operation as $\mathbf{D}^{-1/2} = \mathrm{diag}\left(d_1^{-1/2}, ..., d_n^{-1/2}\right)$ easy to check now that $E\left\{\tilde{\mathbf{X}}\tilde{\mathbf{X}}^T\right\} = \mathbf{I}$. Whitening transforms the mixing matrix into a new one $\tilde{\mathbf{A}}$. We have from Equations (4.2) and (4.20):

$$\tilde{\mathbf{X}} = \mathbf{E}\mathbf{D}^{-1/2}\mathbf{E}^T\mathbf{A}\mathbf{S} = \tilde{\mathbf{A}}\mathbf{S} \tag{4.21}$$

4.5 ICA Algorithms

Formulating a contrast function and then minimizing or maximizing it usually performs the estimation of the independent component. Let us express the difference between the formulation of the contrast function and the algorithm used to optimize it in the following equation:

$$\text{ICA method} = \text{Contrast function} + \text{Optimization algorithm} : \tag{4.22}$$

There are many algorithms that can extract a certain source signals from a linear mixture of different ones. These algorithms are based on the assumptions that the source signals are independent and non-Gaussian. There are two distinct approaches towards computing the ICA [77]. The first category employs high order cumulants and is found mainly in the statistical signal processing literature, as Sources Extraction Based on fourth order Cumulants [78], Joint Approximate Diagonalization of Eigenmatrices (JADE) algorithm [31,79], and MaxKurt/MinKurt Algorithm [31]. The second category uses the gradient-descent of non-linear activation functions and is mainly developed in the neural networks community, as an equivariant adaptive separation via independence (EASI) [31,80], Neural Approaches Based on Information Maximization [81] and Fixed-Point (Fast ICA) algorithm [82]. There are both batch

and adaptive versions of ICA algorithms depending on the particular application. In the adaptive case, the algorithms are obtained by stochastic gradient methods. In the case where all the independent components are estimated at the same time, the most popular algorithm in this category is natural gradient ascent of likelihood, or related contrast functions, like infomax. In the case where the computations are made in a batch-mode, much more efficient algorithms are available. The tensor-based methods are efficient in small dimensions and cannot be used in larger dimensions. The FastICA algorithm, based on a fixed-point iteration, is a very efficient batch algorithm that can be used to maximize both one-unit and multi-unit contrast functions, including likelihood [66].

4.6 General Contrast Functions

Constructing an ICA algorithm requires firstly to formulate a contrast function G(y) that estimates the level of statistical independence between the components of **y**. Contrast functions (or contrasts) G serve as objective criteria for ICA. The contrast function means any function whose optimization enables the estimation of the independent components. For example, mutual information can be a contrast function for ICA [82] . Other contrast functions are related to the fact that mixtures of non-Gaussian sources signals are more Gaussian than the sources. Therefore, measures of non-Gaussianity like negentropy [72] , high-order cumulants [82], etc. However some authors reserve this term for a certain class of objective functions. Also the term loss function is used. Once a contrast function has been defined, the ICA algorithm has to maximize or minimize it. Usually, contrast functions are chosen in order to be differentiable because doing so allows using well known optimization tools like (natural) gradient ascent [80] or fixed-point techniques [82].

In particular, choosing G (whose derivative is g) is very important. The following are some examples functions with their derivatives.

$$G_1(y) = \frac{1}{a_1} \log \cosh a_1 y \qquad g_1(y) = \tanh(a_1 y) \qquad (4.23)$$

$$G_2(y) = -\frac{1}{a_2} \exp(-a_2 y^2/2) \qquad g_2(y) = y \exp(-a_2 y^2/2) \qquad (4.24)$$

$$G_3(y) = \frac{1}{4} y^4 \qquad g_3(y) = y^3 \qquad (4.25)$$

$$G_4(y) = \frac{1}{3} y^3 \qquad g_4(y) = y^2 \qquad (4.26)$$

where $1 \leq a_1 \leq 2$, $a_2 \approx 1$ [83].

4.7 The FastICA Algorithm

4.7.1 Formulation

The FastICA is an efficient and popular ICA algorithm which was introduced by Aapo Hyvrinen at Helsinki University of Technology. [1] FastICA is based on a fixed-point iteration scheme for finding a maximum of the non-Gaussianity of $\mathbf{w}^T \mathbf{x}$. This method equates $\mathbf{w}$ to the gradient of the measure of non-Gaussianity as

$$\mathbf{w} \longleftarrow E\big\{\mathbf{x}g(\mathbf{w}^T\mathbf{x})\big\}$$

where g is the derivative of the non-quadratic function G used in Equation (4.23) to Equation (4.26). The maxima of approximate negentropies of $\mathbf{w}^T\mathbf{x}$ are obtained at certain optima of $E\big\{G(\mathbf{w}^T\mathbf{x})\big\}$. According to the Kuhn-Tucker [84] conditions, the optima of $E\big\{G(\mathbf{w}^T\mathbf{x})\big\}$ under the constraint

$$E\big\{(\mathbf{w}^T\mathbf{x})^2\big\} = \|\mathbf{w}\|^2 = 1,$$

are obtained at points where

$$E\big\{\mathbf{x}g(\mathbf{w}^T\mathbf{x})\big\} - \beta\mathbf{w} = 0, \tag{4.27}$$

where β is some constant. Let us try to solve this equation by Newton's method. Denoting the function on the left-hand side of Equation (4.27) by F, we obtain its Jacobian matrix JF($\mathbf{w}$) as

$$JF(\mathbf{w}) = E\big\{\mathbf{x}\mathbf{x}^T g'(\mathbf{w}^T\mathbf{x})\big\} - \beta\mathbf{I} \tag{4.28}$$

where g' is the derivative of g. To simplify the inversion of this matrix, we approximate the first term in Equation (4.28). Since the data are sphered, a reasonable approximation can be taken as

$$E\big\{\mathbf{x}\mathbf{x}^T g'(\mathbf{w}^T\mathbf{x})\big\} \approx E\big\{\mathbf{x}\mathbf{x}^T\big\} E\big\{g'(\mathbf{w}^T\mathbf{x})\big\} = E\big\{g'(\mathbf{w}^T\mathbf{x})\big\}\mathbf{I} \tag{4.29}$$

The Jacobian matrix becomes then diagonal and can easily be inverted. Thus we obtain the following approximate Newton iteration:

$$\mathbf{w} \leftarrow \mathbf{w} - \big[E\big\{\mathbf{x}g(\mathbf{w}^T x)\big\} - \beta\mathbf{w}\big] / \big[E\big\{g'(\mathbf{w}^T\mathbf{x})\big\} - \beta\big], \tag{4.30}$$

where $\beta = E\big\{\big(\mathbf{w}^T\mathbf{x}g(\mathbf{w}^T\mathbf{x})\big)\big\}$. This algorithm can be further simplified by multiplying both sides of Equation (4.30) by $\beta - E\big\{g'(\mathbf{w}^T\mathbf{x})\big\}$, which gives, after algebraic simplification,

$$\mathbf{w} \leftarrow E\big\{\mathbf{x}g(\mathbf{w}^T\mathbf{x})\big\} - E\big\{g'(\mathbf{w}^T\mathbf{x})\big\}\mathbf{w} \tag{4.31}$$

The algorithm given above estimates only one independent component and is called a one-unit algorithm. The basic form of the one-unit FastICA algorithm is

---

[1]http://www.cis.hut.fi/projects/ica/fastica/fp.html

1. Choose an initial weight vector $\mathbf{w}$ (e.g. random).

2. Let $\mathbf{w}^+ = E\left\{\mathbf{x}g(\mathbf{w}^T\mathbf{x})\right\} - E\left\{g'(\mathbf{w}^T\mathbf{x})\right\}\mathbf{w}$

3. Take $\mathbf{w} = \mathbf{w}^+/\|\mathbf{w}^+\|$

4. Check the convergence, if not convergent, go back to 2.

Estimating the remaining independent component can be done using two approaches: The first approach is deflationary orthogonalization which has originally proposed in [85]. In this approach, one independent component is estimated at a time; when the estimate of the unmixing vector $\mathbf{w}_i$ is obtained, the vector is then orthogonalized with the previously estimated unmixing vectors $\mathbf{w}_j$ $(1 \leq j < i)$, by calculating the projection $p_{ij} = \mathbf{w}_i^T\mathbf{w}_j$ and subtracting it from the current vector:

$$\mathbf{w}_i \leftarrow \mathbf{w}_i - \sum_{j=1}^{i-1} p_{ij}\mathbf{w}_j$$

The second approach is a symmetric orthogonalization approach (all sources are recovered simultaneously). Symmetric orthogonalization achieves the same objective as deflationary orthogonalization by treating all vectors as a matrix and orthogonalizing them at once. In symmetric orthogonalization, each unmixing vector $\mathbf{w}_i$ is estimated independently and then the resulting matrix $\mathbf{W}$ is orthogonalized by using the matrix square root

$$\mathbf{W} \leftarrow \left(\mathbf{W}\mathbf{W}^T\right)^{-1/2}\mathbf{W}$$

where the eigenvalue decomposition of $\mathbf{W}\mathbf{W}^T = \mathbf{E}\,\mathrm{diag}(d_1, d_2, ..., d_m)\,\mathbf{E}^T$ defines the square root matrix as $\left(\mathbf{W}\mathbf{W}^T\right)^{-1/2} = \mathbf{E}\,\mathrm{diag}(d_1^{-1/2}, d_2^{-1/2}, ..., d_m^{-1/2})\,\mathbf{E}^T$. The convergence criterion then becomes $\mathbf{W}\mathbf{W}_{old}^T \sim \mathbf{I}$

The basic form to estimate $\mathbf{w}_i$ by the FastICA algorithm by using deflationary orthogonalization is as follows [83, 86]:

1. Center the data to make its mean zero and whiten the data to give $\mathbf{x}$.

2. Choose m, the number of ICs to estimate. Set counter $p \leftarrow 1$.

3. Choose an initial value of unit norm for $\mathbf{w}_p$ randomly

4. Let $\mathbf{w}_p = E\left\{\mathbf{x}g(\mathbf{w}_p^T\mathbf{x})\right\} - E\left\{g'(\mathbf{w}_p^T\mathbf{x})\right\}\mathbf{w}_p$

5. Do the following orthogonalization: $\mathbf{w}_p \leftarrow \mathbf{w}_p - \sum_{j=1}^{p-1}(\mathbf{w}_p^T\mathbf{w}_j)\mathbf{w}_j$

6. Let $\mathbf{w}_p = \mathbf{w}_p/\|\mathbf{w}_p\|$

7. If $\mathbf{w}_p$ has not converged, go back to step 4.

8. Set $p \leftarrow p + 1$. If $p \leq m$, go back to step 3.
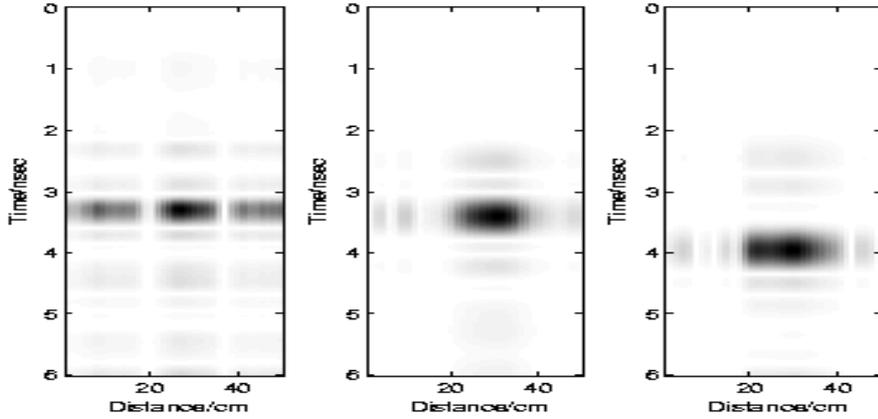
Figure 4.2: B-Scans of three targets at depth 3, 6 and 10cm, respectively.

where the convergence means that the old and new values of $\mathbf{w}$ point in the same direction [73], i.e. their dot-product is (almost) equal to 1 or

$$|\langle \mathbf{w}_{p+1}, \mathbf{w}_p \rangle| = 1$$

The symmetric orthogonalization is as follows

1. Center the data to make its mean zero and whiten the data to give $\mathbf{X}$.

2. Choose m, the number of ICs to estimate.

3. Choose initial value for $\mathbf{w}_j$, each of unit norm. Orthogonalize the matrix as $\mathbf{W} \leftarrow \left(\mathbf{W}\mathbf{W}^T\right)^{-1/2} \mathbf{W}$., where $\mathbf{W} = (\mathbf{w}_1, \mathbf{w}_2, ..., \mathbf{w}_m)^T$

4. Let $\mathbf{w}_j = E\left\{\mathbf{x}g(\mathbf{w}_j^T\mathbf{x})\right\} - E\left\{g'(\mathbf{w}_j^T\mathbf{x})\right\}\mathbf{w}_j$ for every $j = 1, 2..., m$

5. Do a symmetric orthogonalization of the matrix $\mathbf{W}$ by $\mathbf{W} \leftarrow \left(\mathbf{W}\mathbf{W}^T\right)^{-1/2} \mathbf{W}$

6. Check the convergence, if not convergent, go back to 4.

Here the convergence is $\mathbf{W}\mathbf{W}_{old}^T \approx \mathbf{I}$, and $\mathbf{W} = (\mathbf{w}_1, \mathbf{w}_2, ..., \mathbf{w}_m)^T$.

4.7.2 Experimental Results

To examine the ability of ICA technique for the elimination or reduction of the clutter in raw GPR data, the ICA technique has been applied to all measurements, see [87]. The IC's were estimated using FastICA algorithm and four different cost functions G_1, G_2, G_3 and G_4 (Equations (4.23), (4.24), 4.25 and (4.26)) have been used. The FastICA algorithm has been implemented in Matlab and is available on the web at http://www.cis.hut.fi/ pro-jects/ica/fastica/. Figure 4.2 shows the data after apply-
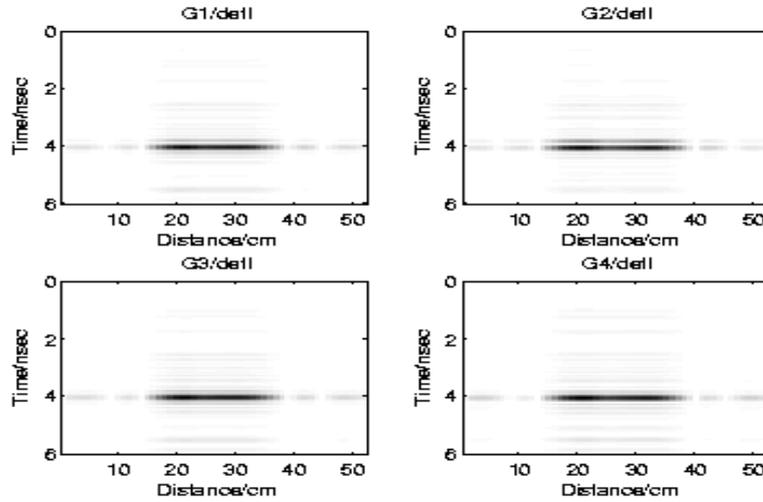
Figure 4.3: The data after applying the FastICA by using four different cost functions in deflationary technique.

ing the ICA algorithm to the data in Figure 2.8. From Figure 4.2, the target signals have been extracted from the GPR data, and most of the clutter signal (cross-talk and air-ground interface) has been removed for the targets at depth 6 and 10 cm. However for the target at depth 3 cm, the clutter has not been removed completely and the result of applying the ICA is the same as applying principle component analysis or singular value decomposition [48]. The results of applying ICA algorithm to the data of Figure 2.10 are shown in Figure 4.3 and Figure 4.4. The cost functions G_1, G_2, G_3 and G_4 have been used in Figure 4.3 and in Figure 4.4 in deflationary and symmetric techniques, respectively. The GPR images generated in Figure 4.3 are similar to those in Figure 4.4, which implies that both techniques (deflationary and symmetric) give the same results. Also in both techniques, G_1, G_3 and G_4 gave consistent result and have shown approximately small improvement over G_2.

4.8 The Infomax ICA

4.8.1 The Infomax Algorithm

The Infomax algorithm has been proposed by Bell and Sejnowski [81]. It is one of the popular ICA algorithms. It was first derived from the Infomax principle that maximises the entropy of outputs of a nonlinear network. The Infomax algorithm attempts to find a set of independent components using a neural network-based, mutual information minimization algorithm. As with the FastICA algorithm , the Infomax algorithm whitens the source data before generating the weight matrix $\mathbf{W}$ associated
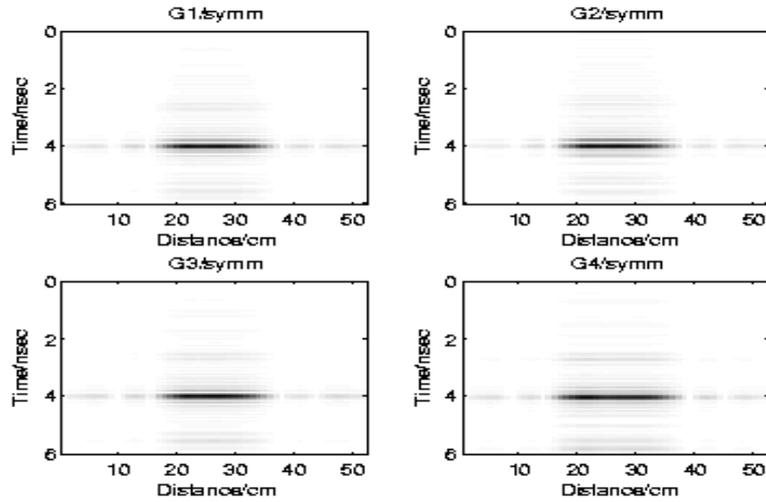
Figure 4.4: The data after applying the FastICA by using four different cost functions in symmetric technique.

with the independent components. The weight matrix is initialized to the identity matrix, which the algorithm iteratively refines using a learning algorithm. The learning algorithm random selects, without replacement, a subset of the whitened data $\mathbf{x}$ from which it computes an estimate of the unmixed signals $\mathbf{y} = \mathbf{W}\mathbf{x}$. The algorithm searches for $\mathbf{W}$ by estimating the likelihood that we can obtain the distribution of the observations from our estimate of $\mathbf{W}$. From the likelihood function, we can estimate a $\triangle\mathbf{W}$ that maximizes the likelihood that $\mathbf{W}$ produces a correct estimate.

Learning Rule

The learning algorithm can be derived using the maximum likelihood Estimation (MLE), which was firstly proposed in [88] and [89] for blind source separation. It has been pursued more recently in [90] and [91]. The probability density function (pdf) of the observations $\mathbf{x}$ can be expressed as [92]

$$p(\mathbf{x}) = |\det(\mathbf{W})|P(\mathbf{y}) \tag{4.32}$$

where the multivariate $P(\mathbf{y}) = \prod_{i=1}^{N} P_i(y_i)$ and $P_i(y_i)$ are the corresponding marginal pdf's.

The learning process tries to maximize $P(\mathbf{y})$ by maximizing the product of its marginal densities. If these marginals are not independent, then increasing one decreases the product of the others by the same amount, this leaves the value of the product unchanged.
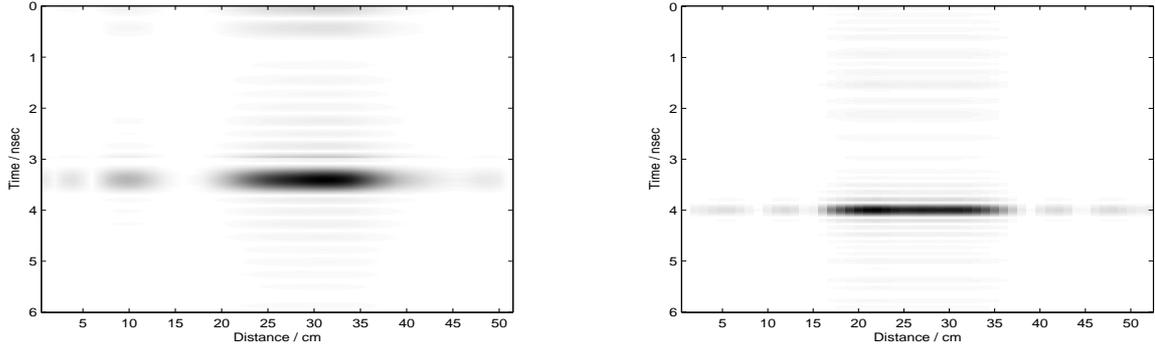
Figure 4.5: The target estimation by using infomax algorithm.

The log likelihood function is given by

$$L(\mathbf{y}, \mathbf{W}) = \log|\det(\mathbf{W})| + \sum_{i=1}^{N} \log p_i(y_i) \tag{4.33}$$

Maximizing the log-likelihood with respect to $\mathbf{W}$ gives a learning algorithm for $\mathbf{W}$ [81]:

$$\triangle\mathbf{W} \propto \left[\left(\mathbf{W}^T\right)^{-1} - \Phi(\mathbf{y})\mathbf{x}^T\right] \tag{4.34}$$

where

$$\Phi(\mathbf{y}) = -\frac{\frac{\partial p(\mathbf{y})}{\partial(\mathbf{y})}}{p(\mathbf{y})} = \left[-\frac{\frac{\partial p(y_1)}{\partial(y_1)}}{p(y_1)}, ..., -\frac{\frac{\partial p(y_N)}{\partial(y_N)}}{p(y_N)}\right]^T . \tag{4.35}$$

An additional and important feature was added to the Infomax ICA algorithm by Amari et.al. [93] who observed that a simpler learning rule, with a much faster and more stable convergence, could be obtained by multiplying the Infomax gradient of Equation (4.34) by $\mathbf{W}^T\mathbf{W}$. An efficient way to maximize the log-likelihood is to follow the "natural" gradient [93],

$$\triangle\mathbf{W} \propto \frac{\partial L(\mathbf{y}, \mathbf{W})}{\partial \mathbf{W}}\mathbf{W}^T\mathbf{W} = \left[\mathbf{I} - \Phi(\mathbf{y})\mathbf{y}^T\right]\mathbf{W} \tag{4.36}$$

Since $\mathbf{W}^T\mathbf{W}$ which scales the gradient, is positive-definite, it does not change the minima and maxima of the optimization. If we choose $\Phi(\mathbf{y}) = 2\tanh(\mathbf{y})$ the learning rule reduces to that in [81] with the natural gradient:

$$\triangle\mathbf{W} \propto \left[\mathbf{I} - 2\tanh(\mathbf{y})\mathbf{y}^T\right]\mathbf{W} \tag{4.37}$$

4.8.2 Experimental Results

To verify to theoretical results above, the infomax algorithm has been applied to all measurements, see [94]. Figure 4.5 shows the data after applying the ICA algorithm

to the data in Figure 2.8 and in Figure 2.10, respectively. From Figure 4.5, the target signals have been extracted from the GPR data, and most of the clutter signal (cross-talk and air-ground interface) has been removed for the targets at depth 6 and 10 cm.

4.9 Second Order Blind Identification (SOBI) Algorithm

4.9.1 Formulation

The second order blind identification (SOBI) algorithm uses only the second order statistic and is based on the joint diagonalization. Algorithms that operate explicitly on higher order statistics than second order statistic are classified as ICA. The SOBI algorithm has been developed first by Belouchrani et al. [95, 96]

Let the observed vector $\mathbf{x}$ be written as

$$\mathbf{x}(k) = \mathbf{A}\mathbf{s}(k)$$

The covariance matrix of the observed data is

$$\mathbf{R_x}(0) = E\{\mathbf{x}(k)\mathbf{x}^*(k)\} = \mathbf{A}\mathbf{R_s}(0)\mathbf{A}^H \tag{4.38}$$

and the time delay covariance matrix is

$$\mathbf{R_x}(\tau) = E\{\mathbf{x}(k+\tau)\mathbf{x}^*(k)\} = \mathbf{A}\mathbf{R_s}(\tau)\mathbf{A}^H \tag{4.39}$$

The superscripts $(*)$ denotes the conjugate transpose of the vector and H denotes the complex conjugate transpose of the matrix. The aim of the algorithm is to estimate the mixture matrix to recover the sources signals from observed data. This can be done by the following steps [97]

1. The first processing is to assume that the source signals ($\mathbf{s}$) have unit variance, this mean that

$$\mathbf{R_s}(0) = \mathbf{I} \quad \text{so that} \quad \mathbf{R_x}(0) = E\{\mathbf{x}(k)\mathbf{x}^*(k)\} = \mathbf{A}\mathbf{A}^H \tag{4.40}$$

2. The second preprocessing in SOBI is whitening the observed vector. The whitening technique can be done by using the matrix $\mathbf{Q}$, such that

$$E\{\mathbf{Q}\mathbf{x}\mathbf{x}^*\mathbf{Q}^H\} = \mathbf{Q}\mathbf{R_x}(0)\mathbf{Q}^H = \mathbf{Q}\mathbf{A}\mathbf{A}^H\mathbf{Q}^H = \mathbf{I} \tag{4.41}$$

From the above equation, if $\mathbf{Q}$ is the whitening matrix, then $\mathbf{Q}\mathbf{A}$ is a unitary matrix and

$$\mathbf{A} = \mathbf{Q}^{\sharp}\mathbf{U} \tag{4.42}$$

where $\sharp$ denotes the Moore-Penrose pseud inverse.
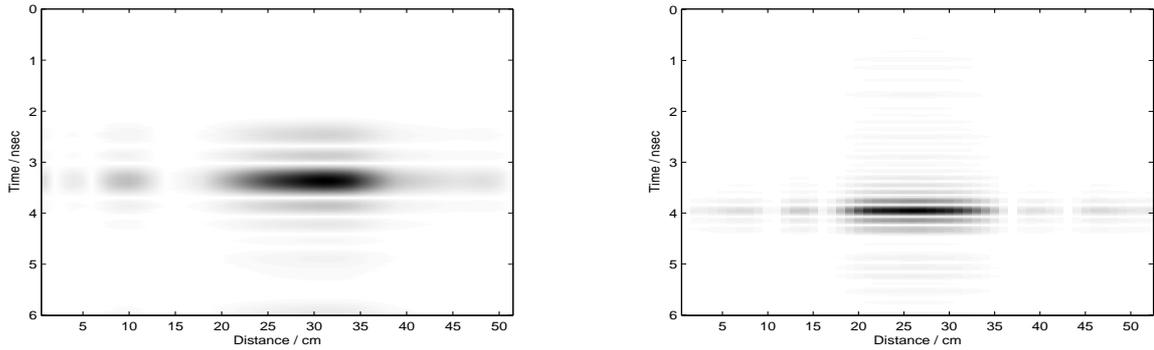
Figure 4.6: The target estimation by using SOBI algorithm.

3. Estimating the unitary matrix $\mathbf{U}$ by using one of the available numerical algorithms [98–104].

4. Estimating the source signals as [31, 96]

$$\mathbf{s}(k) = \mathbf{U}^H \mathbf{Q} \mathbf{x}(k) \tag{4.43}$$

4.9.2 Experimental Results

To verify to theoretical results above, the SOBI algorithm has been applied to all measurements. Figure 4.6 shows the data after applying the ICA algorithm to the data in Figure 2.8 and in Figure 2.10, respectively. From Figure 4.6, the target signals have been extracted from the GPR data, and most of the clutter signal (cross-talk and air-ground interface) has been removed for the targets at depth 6 and 10 cm.

4.10 The Joint Approximative Diagonalization of Eigenmatrices (JADE)

In this section, we will introduce the Joint Approximate Diagonalization of Eigenmatrices (JADE) algorithm. This algorithm performs the ICA decomposition of experimental data by computing the eigenvalue decomposition of their cumulant tensor. The eigenvalues of the cumulant tensor are vectors that describe the independent components of the data mixture corresponding to the different contributions (e.g., target and clutter). The JADE exploits the fourth-order cumulants of the data mixture that contains all fourth order information of the data. The cross-cumulants (off-diagonal elements of the cumulant matrix) characterize the statistical dependencies among the

different data components. The off-diagonal elements vanish (and hence the cumulant matrix is diagonal), if and only if all data components are statistically independent.

4.10.1 Cumulants

Under a linear transformation $\mathbf{Y} = \mathbf{WX}$, where the matrices $\mathbf{X}$ and $\mathbf{Y}$ are rewritting as column vectors form and hence $y_i = \sum_m w_{im} x_m$, the fourth order cumulants of $\mathbf{X}$ and $\mathbf{Y}$ are related according to

$$Cum\,(y_i, y_j, y_k, y_l) = \sum_{mnpq} w_{ip} w_{jq} w_{km} w_{ln} Cum\,(x_m, x_n, x_p, x_q) \tag{4.44}$$

Next , we apply the input data according to

$$\mathbf{Z} = \mathbf{QX} \tag{4.45}$$

where $E(\mathbf{ZZ} = \mathbf{I})$. The objective of JADE algoritm is to estimate the matrix $\mathbf{A}$, and the problem will hence reduce to finding an orthogonal matrix $\mathbf{U}$ such as

$$\mathbf{Z} = \mathbf{QX} = \mathbf{QAS} = \mathbf{US} \tag{4.46}$$

Now we can define the fourth order cumulants of the input data ($E(z_i) = 0$) as

$$Cum\,(z_i, z_j, z_k, z_l) = E\,(z_i z_j z_k z_l) - E\,(z_i z_j)\,E\,(z_k z_l) - E\,(z_i z_k)\,E\,(z_l x_j) - E\,(z_i z_l)\,E\,(z_j z_k) \tag{4.47}$$

On the other hand, the cumulants of $\mathbf{Z}$ and $\mathbf{S}$ (that are related by $\mathbf{X} = \mathbf{AS}$) will have the following simple relation when $\mathbf{S}$ has independent entries,

$$Cum\,(z_i, z_j, z_k, z_l) = \sum_q \kappa_4(s_q) a_{iq}, a_{jq}, a_{kq}, a_{lq} \tag{4.48}$$

where κ_4 is the kurtosis of $\mathbf{S}$, and a_{ij} denotes the $(ij)th$ entry of the matrix $\mathbf{A}$. Given a random $(n \times 1)$ vector z and any $(n \times n)$ matrix M, we define the associated cumulant matrix $C_z(M)$ as the $(n \times n)$ matrix whose elements are given by

$$[C_z(M)]_{ij} = \sum_{k,l=1}^{n} Cum\,(z_i, z_j, z_k, z_l)\,M_{kl} \tag{4.49}$$

or

$$C_z(M) = E\left\{\left(Z^T M Z\right) Z Z^T\right\} - R_z tr(M R_z) - R_z M R_z - R_z M^T R_z \tag{4.50}$$

Here $tr(.)$ denotes the trace and R_z denotes the covariance matrix of Z. It can be shown that such a matrix has the eigenvalue decomposition $C_z(M) = U\Lambda U^T$, with

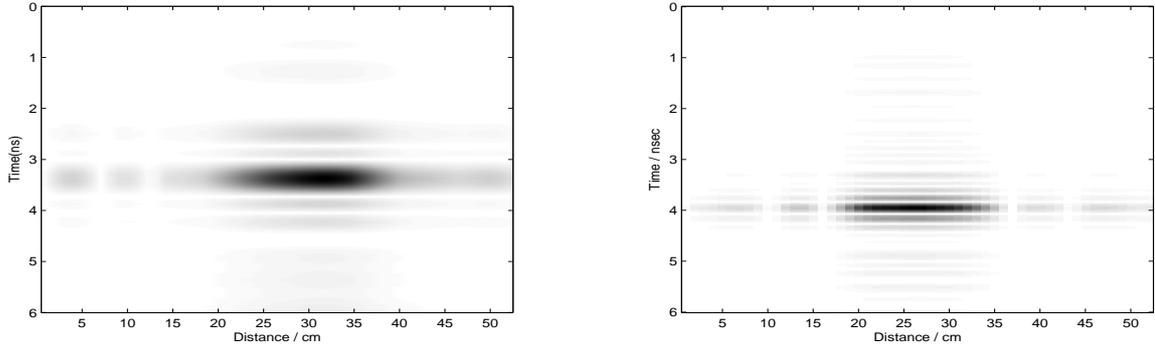$$\Lambda = diag\left\{\kappa_4(s_1) u_1^T M u_1, ..., \kappa_4(s_n) u_n^T M u_n\right\} \tag{4.51}$$

Figure 4.7: The target estimation by using JADE algorithm.

$u_i, i = 1, 2, ..., n$, denotes the *ith* column of U. Now, Let $M = M_1, ..., M_p$ be a set of P matrices of size $(n \times n)$ and denote the associated cumulant matrices for the sphered data by $Z = US$ by $C^i = C_z^i(M_i)$ with $1 \leq i \leq p$. Again, as above, for all i we have $C^i = U\Lambda_i U^T$ with being Λ_i a diagonal matrix given by Equation (4.51).

As will be summarized in the next section, the joint approximative diagonalization of eigenmatrices (JADE) is based on jointly diagonalizing a set of n^2 contracted quadri-covariance matrices $C_x(M_{pq})$, where M_{pq} are called the eigenmatrices. The eigenmatrices M_{pq} should satisfy the following conditions:

$$C_x(M_{pq}) = \lambda_{pq} M_{pq} \text{and} tr\left(M_{pq}M_{kl}^T = \delta(p, q, k, l)\right)$$

$(1 \leq p, q \leq n)$, where $\delta(p, q, k, l) = 1$ for $p = q = k = 1$ and 0 otherwise.

There are several techniques to select the eigenmatrices M_{pq} that satisfy the above relations. In the ideal case, we can choose $M_{pq} = e_p e_q^T$, where e_p denotes the n-dimensional vector with 1 at the pth position and 0 elsewhere.

4.10.2   JADE Algorithm

The joint approximative diagonalization of eigenmatrices (JADE) algorithm has been proposed by Cardoso and Sou-loumiac [79, 105, 106]. It is an exact algebraic approach to perform independent component analysis, and is essentially based on fourth-order cumulants of prewhitened input data [107]. The bottleneck with the JADE algorithm when dealing with high-dimensional problems is however the algebraic determination of the mixing matrices. In the following we will summarize a neural learning algorithm using higher-order neurons to adaptively determine these mixing matrices [57]:

1. Form the sample covariance R_x, compute a whitening matrix Q.

2. Apply a robust prewhitening or orthogonalization method to obtain the whitened (pre-processed) vector $Z(k) = QX(k)$.

3. Perform (PCA) of the sampled contracted quadri-covariance matrix with $M = \mathbf{I}$ and $\mathbf{I}$ is the identity matrix:

$$C_z = U\Lambda_I U^T$$

,

where $\Lambda_I = diag\left(\kappa_4(s_1), ..., \kappa_4(s_n)\right)$.

4. Select a fixed set $M = (M_1, ..., M_p)$ of $n \times n$ matrices and estimate the P sampled contracted quadrico-variance matrices $C_z(M)$.

5. Find an orthogonal joint diagonalizing matrix U for all n matrices $C_z(M)$.

6. Estimate the mixing matrix as:

$$\hat{A} = UQ^{-1}$$

and estimate the signal components as

$$\hat{S} = \hat{A}^{-1}X = U^T Z$$

The joint diagonalizer at step 5 is found by a Jacobi technique [106].

4.10.3   Experimental Results

The independent components in ICA have been estimating using the joint approximate diagonalization of eigenmatrices (JADE) algorithm which has been implemented in Matlab and is available on the home page of J. F. Cardoso at $http : //www.tsi.enst.fr/\ cardoso/$. The efficiency of JADE technique has been experimentally verified using two different sets of raw SFGPR [94]. In all measurements, the reflection from the surface has been removed or has been at least significantly reduced. Figures 4.7, show the data after applying the JADE algorithm to the data in Figure 2.8 and in Figure 2.10, respectively.

Chapter 5

Likelihood Method

In this section, we will apply the signal detection theory-namely, the likelihood ratio for detecting a landmine signal against the noise background. This technique is a rich history in both electrical engineering and mathematics, and its solution has involved, over the years, a wide variety of mathematical tools [108].

Signal detection theory states that the likelihood ratio is the basis for making the optimum decision between two hypotheses. The hypothesis testing firstly, introduced by [109], and became known statistical decision theory [110]. The first applications of statistical decision theory to the signal detection problems have been done by [111, 112].

The decision will be made between two hypotheses corresponding to presence of a target (H_1) or no target (H_0), respectively. This is a simple binary hypothesis testing where the noise alone hypothesis H_0 is referred to as the null hypothesis, and the signal present hypothesis H_1 as the alternative hypothesis. With respect to the hypothesis testing model, there are two types of errors. The first error is called false positive error, or false alarm. The second error, often called false negative error, or miss. The best decision region should give the best trade-off between two families of error probabilities under the Neyman-Pearson criterion. A common approach to the composite hypothesis testing is the generalized likelihood ratio test (GLRT) [113]. In the GLRT approach, the most probable individual hypothesis is computed to represent the likelihood of the composite hypothesis. The generalized likelihood ratio is defined as the ratio of the maximum value of the likelihood under H_0 to the maximum under H_1. For easy computation, the generalized log-likelihood ratio is used [114].

5.1 Signal Detection

Assume that for some physical measurements a GPR produces an output waveform $x = \{x(t) : t \in [0, T]\}$. Assume that this waveform has been produced by clutter and noise alone or by signal plus clutter and noise. These two possibilities are called the null hypothesis H_0 and the alternative hypothesis H_1, respectively, and are commonly written in the compact notation [115]:

$$H_0 : x = b + n \tag{5.1a}$$

$$H_1 : x = s + b + n, \tag{5.1b}$$

where x is the measured signal, b is the background clutter, s is the target signal and n is the noise. To decide between the null and alternative hypotheses one might apply a threshold, and make a decision that the signal is present or not. The question is where to set the threshold so as to ensure that the number of decision errors is small. There are two types of possible errors: the error of missing the signal (decide H_0 under H_1 (target signal is present)) and the error due to false alarm (decide H_1 under H_0 (no target signal is present)).

Let the threshold be denoted γ. Define the H_1 decision region $\Re_{H_1} = x : x(t) > \gamma$, for some $t \in [0, T]$. This region is called the critical region and simply specifies the conditions on x for which the detector declares the target signal to be present [116]. The probabilities of false alarm and miss are functions of γ, and they are given by $P_{FA} = P(\Re_{H_1}|H_0)$ and $P_M = P(\Re_{H_0}|H_1) = 1 - P(\Re_{H_1}|H_1)$ where $P(E|H_0)$ and $P(E|H_1)$ denote the probabilities of arbitrary event E under hypothesis H_0 and hypothesis H_1, respectively. The probability of correct detection $P_D = P(\Re_{H_1}|H_1)$ is commonly called the power of the detector and P_{FA} is called the level of the detector. For each value of $P_{FA} = \alpha$, we can calculate the value of threshold by knowing the distribution of the noise. The plot of the pair of points $P_{FA} = P_{FA}(\gamma)$ and $P_D = P_D(\gamma)$ over the range of threshold values $-\infty < \gamma < \infty$ produces a curve called the receiver operating characteristic (ROC) which completely describes the error rate of the detector as a function of γ.

5.2 Likelihood Ratio Test

If the probability density functions can be associated with the H_1 and H_0 hypotheses, then the Neyman-Pearson theorem [117] determines the optimal test statistic, yielding the highest probability of detection for a given probability of false alarm. The

Neyman-Pearson theorem yields the likelihood ratio test

$$\Lambda(x) = \frac{P(x|H_1)}{P(x|H_0)} \underset{<}{\overset{H_1}{\underset{H_0}{>}}} \gamma, \tag{5.2}$$

where x is the received signal from the sensor at the location under test, $P(x|H_1)$ and $P(x|H_0)$ are the probability density functions of the received signal under the target hypothesis, H_1, and under the clutter or no-mine hypothesis, H_0, respectively. When $\Lambda(x)$ is greater than the threshold γ, the decision is H_1, which is target (landmine) present. When $\Lambda(x)$ is less than γ, the decision is H_0, which means target not present, i.e. only clutter or nothing is there. When $\Lambda(x)$ is equal to γ, both H_1 and H_0 are equally likely. To use the likelihood ratio test in Equation (5.2), the target signal and the clutter signal have to be known completely. Because of this omniscient characteristic of the LRT, it is not a very practical approach to the detection problem. The probability density function (PDF) under the H_1 hypothesis (mine present), the PDF under the H_0 hypothesis (mine absent) are both illustrated in Figure 5.1 and they are obtained from real data which is shown in Figure 2.7 (the probability density function estimation was done using non-parametric "kernel smooth function"), where the solid and dashed traces represent the clutter and target plus clutter PDFs, respectively.

5.3   Maximum Likelihood Test

As has been stated in 3.2.1, the maximum likelihood method, is the procedure of finding the value of one or more parameters for a given statistic, which makes the known likelihood distribution a maximum. The maximum likelihood estimate for a parameter θ is denoted $\hat{\theta}$.

In Equation (5.2), the likelihood ratio test can be generated by knowing the probability density function of the noise n.

Let the received data x_i has been characterized as being Gaussian with mean μ and variance σ^2, where $i = 1, 2, ., N$. The probability density functions $P(x_i|H_1)$ and $P(x_i|H_0)$ can be written as

$$P(x_i|\mu, \sigma^2) = \frac{1}{(\sigma\sqrt{2\pi})} \exp\left\{-\frac{1}{2\sigma^2}[x_i - \mu]^2\right\} \tag{5.3}$$

Let $x_1, ..., x_N$ are stochastically independent, their joint probability density function

is the product of their marginal probability density functions.

$$P(x_1, ..., x_N | \mu, \sigma^2) = \prod_{i=1}^{N} P(x_i | \mu, \sigma^2)$$

$$P(x_1, ..., x_N | \mu, \sigma^2) = \frac{1}{(\sigma^2 2\pi)^{N/2}} exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^{N} [x_i - \mu]^2 \right\} \qquad (5.4)$$

If we substitute the received (observations) data $x_1, ..., x_N$ we obtain a function of the parameters μ and σ^2. The vector of parameters is denoted by θ

$$\theta = \begin{pmatrix} \mu \\ \sigma^2 \end{pmatrix} \qquad (5.5)$$

The joint probability density function of the received data x_i, seen as a function of the parameters θ is called the likelihood function, denoted by $L(\theta)$.

$$L(\theta) = \frac{1}{(\sigma^2 2\pi)^{N/2}} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^{N} [x_i - \mu]^2 \right\} \qquad (5.6)$$

The likelihood function can be used to obtain an estimator of the population parameters θ. A natural idea is, for a given set of observations, to maximize the probability of those observations, i.e. maximize the joint probability density function (likelihood function) w.r.t. θ. This gives the Maximum Likelihood estimator (MLE)

$$\hat{\theta} = arg \max_{\theta \in \Theta} L(\theta). \qquad (5.7)$$

where arg max stands for the argument of the maximum, thus Equation (5.7) holds if and only if $\hat{\theta}$ is the unique value of θ for which $L(\theta)$ is maximized.

Θ is the parameter space that has all possible values of the parameters. Equivalent definition of the above equation is

$$\hat{\theta} = arg \max_{\theta \in \Theta} \ln L(\theta). \qquad (5.8)$$

where ln is the natural logarithm. The function $\ln L(\theta)$ is called the *log likelihood function* .

5.4 Generalized Likelihood Ratio Test (GLRT)

The Probability density function of the measurement x can be defined as $P(x|\theta)$, where θ belongs to a parameter space Θ. We divide the parameter space Θ into two

regions Θ_{H_1}, and Θ_{H_0}. The Neyman-Pearson theorem can be defined as

$$\Lambda(x) = \frac{P(x|\theta_{H_1})}{P(x|\theta_{H_0})} \underset{H_0}{\overset{H_1}{\underset{<}{\overset{>}{}}}} \gamma \tag{5.9}$$

where x is a collection of N variables $x_1, ..., x_N$. The hypotheses in Equation (5.1) are simple when $\Theta = \{\theta_{H_0}, \theta_{H_1}\}$ consists of only two values and $\Theta_{H_0} = \{\theta_{H_0}\}$ and $\Theta_{H_1} = \{\theta_{H_1}\}$ are point sets.

When θ takes on more than two values at least one of the hypotheses, H_0 and H_1, is *composite* and the Neyman-Pearson theorem no longer applies. *A popular but ad hoc* alternative that enjoys some asymptotic optimality properties is to implement the generalized likelihood ratio test (GLRT):

Replace the unknown parameters by their maximum likelihood estimates in the above equation as

$$L_G(x) = \frac{P(x|\hat{\theta}_{H_1}; H_1)}{P(x|\hat{\theta}_{H_0}; H_0)} \underset{H_0}{\overset{H_1}{\underset{<}{\overset{>}{}}}} \gamma \tag{5.10}$$

$\hat{\theta}_{H_1}$ is the maximum likelihood of θ when H_1 is true, $\hat{\theta}_{H_0}$ is the maximum likelihood of θ when H_0 is true.

Equation (5.10) is known as the Generalized Likelihood Ratio Test (GLRT).

The Generalized Likelihood Ratio Test can be interpreted as a Likelihood Ratio Test which is based on the most likely values of the unknown parameters $\hat{\theta}_{H_1}$ and $\hat{\theta}_{H_0}$. The threshold γ is set to attain a specified level of P_{FA}.

5.5 Basic Analysis

To reduce the amount of clutter within the GPR data, we assume that the noise n has been characterized as being Gaussian with zero mean and variance σ^2. The probability density function of x under the hypothesis, H_1, is

$$P(x|H_1) = \frac{1}{(\sigma^2 2\pi)^{N/2}} \exp\left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^{N} [x_i - (b+s)]^2 \right\}$$

where x is a collection of N variables $x_1, ..., x_N$, and under the hypothesis H_0 the probability density function is

$$P(x|H_0) = \frac{1}{(\sigma^2 2\pi)^{N/2}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{N}[x_i - b]^2\right\}$$

Assume that the received data x_i are stochastically iid.
The generalized Likelihood ratio test Equation (5.9) is in the form

$$\Lambda(x) = \frac{\Lambda_1}{\Lambda_0} = \frac{P(x|\ b+s, \sigma^2, H_1)}{P(x|\ b, \sigma^2, H_0)} \overset{\overset{H_1}{>}}{\underset{\underset{H_0}{<}}{}} \gamma \qquad (5.11)$$

The generalized Likelihood ratio test Equation (5.10) is in the form

$$L_G(x) = \frac{L_1}{L_0} = \frac{P(x|\ \hat{b}+\hat{s}, \hat{\sigma}_1^2, H_1)}{P(x|\ \hat{b}, \hat{\sigma}_0^2, H_0)} \overset{\overset{H_1}{>}}{\underset{\underset{H_0}{<}}{}} \gamma \qquad (5.12)$$

Where the notation $\hat{\sigma}_0$ is the estimation of σ when the hypotheses H_0 is true, and $\hat{\sigma}_1$ is its estimation when the hypotheses H_1 is true. By taking the logarithm of the ratio, Equation (5.11) can be written as

$$\ln \Lambda(x) = \ln \frac{\Lambda_1}{\Lambda_0} = \ln \Lambda_1 - \ln \Lambda_0 \overset{\overset{H_1}{>}}{\underset{\underset{H_0}{<}}{}} \gamma \qquad (5.13)$$

We will start from $\ln \Lambda_0$

$$\ln \Lambda_0 = -\frac{N}{2}\left[\ln(2\pi) + \ln\sum_{i=1}^{N}\sigma_{0,i}^2\right] - \left[\sum_{i=1}^{N}\frac{1}{2\sigma_{0,i}^2}(x_i - b)^2\right] \qquad (5.14)$$

Where σ_i^2 is the diagonal elements of the covariance matrix. The log-likelihood function ($\ln \Lambda_0$) now is depending on the unknown parameters b and σ_i^2. These unknown parameters can be replaced by their maximum likelihood estimates to obtain the concentrated log-likelihood function.

For a likelihood function of more than one parameter, it is sometimes possible to write some parameters in terms of others, thereby reducing the number of independent

parameters. (The function is calculated using the parameter value, which maximizes the likelihood given the value of the other parameters.) This procedure is called concentration of the parameters and results in the concentrated likelihood function.

To find the maximum likelihood estimates of b and σ_i^2, we start by differentiating $\ln \Lambda_0$ with respect to b,

$$\frac{\partial \ln \Lambda_0}{\partial b} = 0 \implies \hat{b} \qquad \hat{b} = \frac{1}{N} \sum_{i=1}^{N} x_i \tag{5.15}$$

In the same way differentiate $\ln \Lambda_0$ with respect to σ_0^2,

$$\frac{\partial \ln \Lambda_0}{\partial \sigma_0^2} = 0 \implies \hat{\sigma}_0^2 \qquad \hat{\sigma}_{0,i}^2 = \frac{1}{N} \sum_{i=1}^{N} \left(x_i - \hat{b} \right)^2 \tag{5.16}$$

Substitute $\hat{b}$ and $\hat{\sigma}_0^2$, $\ln L_0$ in Equation (5.12) can be written as

$$\ln L_0 = -\frac{N}{2} \left[\ln(2\pi) + \ln \sum_{i=1}^{N} \hat{\sigma}_{0,i}^2 \right] - \left(\frac{N}{2} \right) \tag{5.17}$$

Similarly $\ln L_1$ can be written as

$$\ln L_1 = -\frac{N}{2} \left[\ln(2\pi) + \ln \sum_{i=1}^{N} \hat{\sigma}_{1,i}^2 \right] - \left(\frac{N}{2} \right) \tag{5.18}$$

Substitute Equations (5.17) and (5.18) in Equation (5.12) , we get

$$\ln L_G = \frac{N}{2} \left[\ln \sum_{i=1}^{N} \hat{\sigma}_{0,i}^2 - \ln \sum_{i=1}^{N} \hat{\sigma}_{1,i}^2 \right] \begin{array}{c} H_1 \\ > \\ < \\ H_0 \end{array} \ln \gamma = \gamma' \tag{5.19}$$

Experimentally, it has been found that the optimal point of ROC curve is reached with certain γ value. This γ is found experimentally to be proportional with $\ln L_G$ and could be expressed by the equation

$$\gamma = \frac{N}{2} \left[ln \sum_{i=1}^{N} \hat{\sigma}_{0,i}^2 - ln \sum_{i=1}^{N} \hat{\sigma}_{1,i}^2 \right] \times \mu_0 \tag{5.20}$$

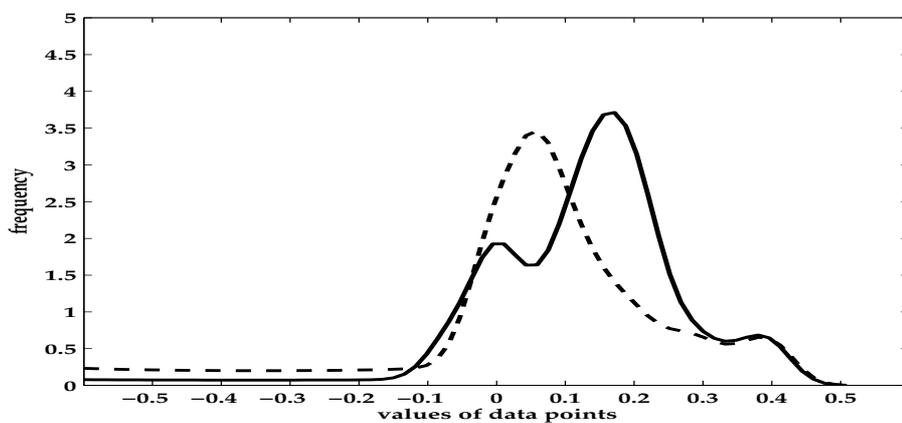Where μ_0 is the mean of A-scan data (a reference data) with no target.

Figure 5.1: Probability density functions for clutter (dashed) and buried PMN mine plus clutter (solid).

Figure 5.2: The data before and after applying the threshold equation (20), N=51 and depth of the landmine is 5 cm.
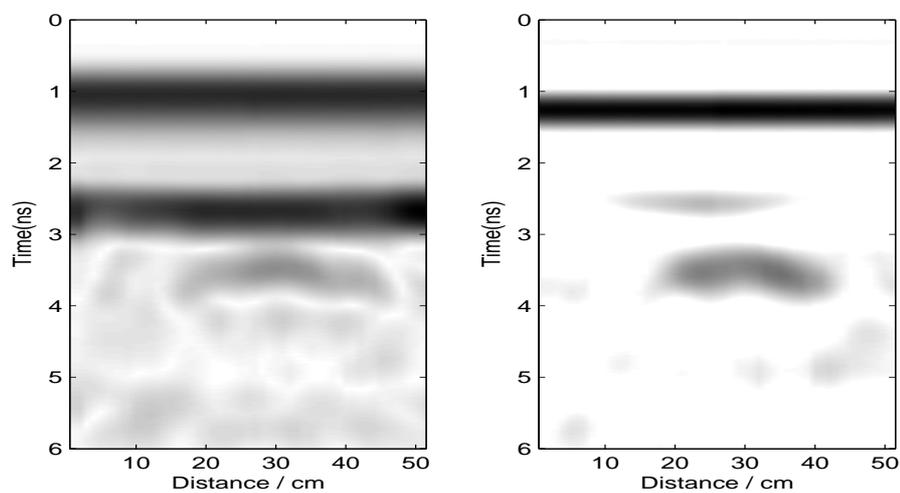
Figure 5.3: The data before and after applying the threshold equation (20), N=51 and depth of the landmine is 3 cm.
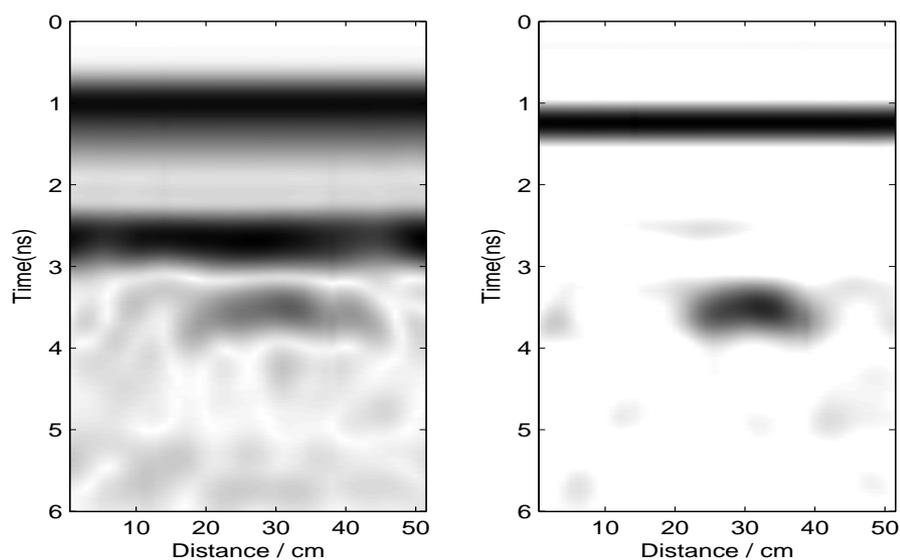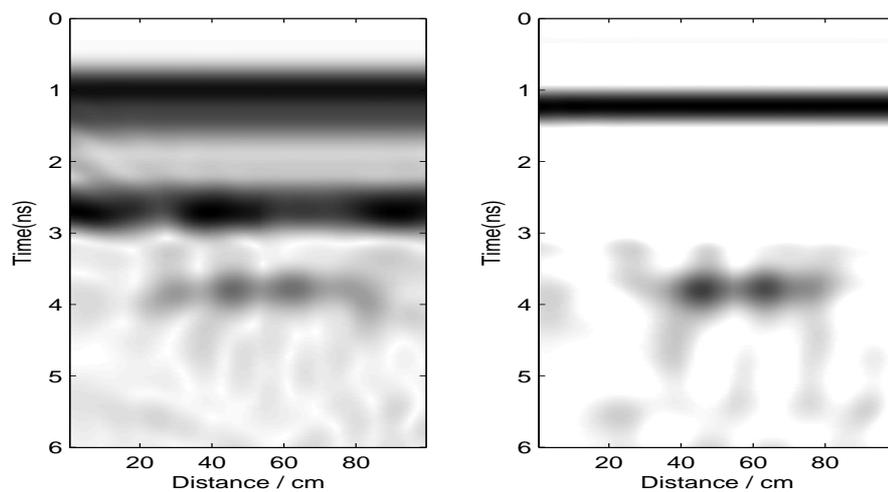

Figure 5.4: The data before and after applying the threshold equation (20), N=99 and depth of the landmine is 10 cm.

Figure 5.5: The data before and after applying the threshold equation (20), the left object is metallic cylinder with diameter 10cm and height 5cm, the right object is PMN landmine.
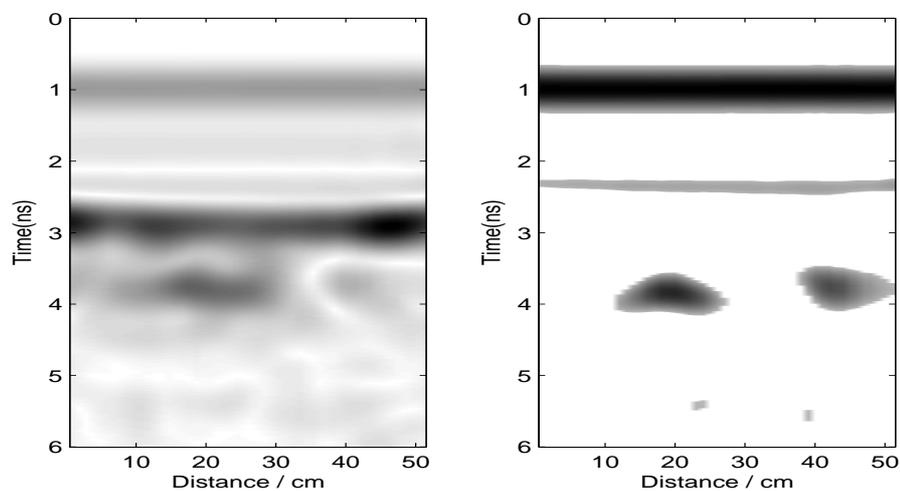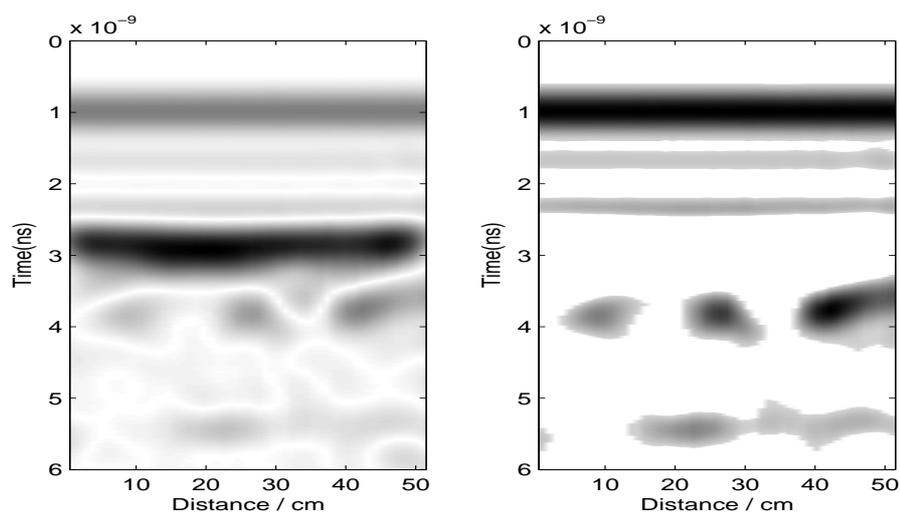

Figure 5.6: The left object is metallic cylinder with diameter 5cm and 3cm high, the middle object is PMN landmine and the right object is metallic cylinder with diameter 10cm and 5cm high.
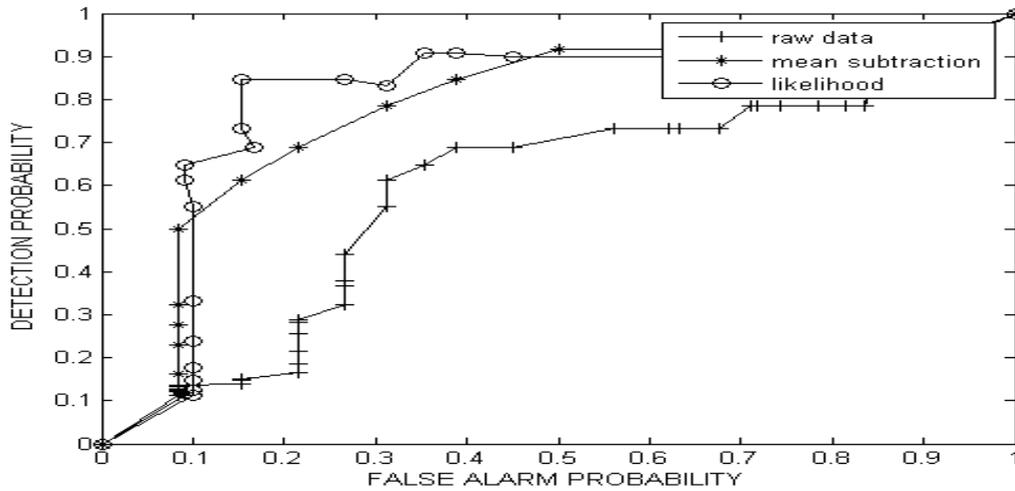
Figure 5.7: EPC curve after threshold, the EPC curve shows that the detection is improved.

5.6   Clutter Reduction

Our data consists of three main signals: The cross-talk between the transmitting and receiving antennas, the reflection from the air-ground interface, and the background resulting from scatterers within the soil.

The signal, which results from the cross-talk between the transmitting and receiving antennas, is constant because the distance between the two antennas is fixed for all measurements. This part can be removed by using timing filter. To reduce the reflection from the air-ground interface (surface reflection) we have applied the threshold to the each sample in the B-scan data. The results of applying the threshold to our data, which has been calculated according to Equation (5.20), are demonstrated in Figure 5.2 to 5.6. The parameters, which have been used in the experiments, are different targets, depths and N (the number of A-scans in a single B-scan). For example, in Figure 5.2 the target has been put at depth 5 cm, and N=51, in Figure 5.3 the target has been put at depth 3 cm, and N=51, and in Figure 5.4 the target has been put at depth 10 cm, and N=99.

In Figure 5.5, two different objects are used.

- A metal cylinder. 10 cm diameter, 5 cm high, 6 cm depth.

- A PMN landmine. 12 cm diameter, 5 cm high 7 cm depth.

In Figure 5.6, three different objects are used.

- A metal cylinder. 5 cm diameter, 3 cm high, 7 cm depth.

- A PMN landmine. 12 cm diameter, 5 cm high, 7 cm depth.

Mine	Depth	Area under EPC before processing	Area under EPC after processing
	3cm	0.69	0.74
PMN	5cm	0.66	0.74
	10cm	0.59	0.80

Table 5.1: The area under the EPC curve for different experiments.

- A metal cylinder. 10 cm diameter, 5 cm high, 6 cm depth.

5.7 Expected Performance Curve (EPC)

In several processing techniques concerned with classification tasks, curves like Receiver Operating Characteristic (ROC) are often used to compare two or more models with respect to various operating points. Instead of the ROC, an Expected Performance Curve (EPC) is introduces and used here for the comparion of different algorithms. It has the same meaning as the ROC.

For all the considered examples the area under the EPC curve is listed in Table 5.1. The closer to 1 the area under the EPC curve, the better the performance we have. In general, the EPC curve shows the performance of a particular detector or classifier. However, in this paper the EPC curve is considered a measure on the clutter reduction, i.e., the signal-to-clutter ratio. The EPC curve is shown in Figure 5.7, and the area under the EPC curves is calculated. The EPC curves are calculated according to reference [118]. Based on the knowledge of where the landmine is buried, and by applying a threshold (Equation (5.20)) on each reconstructed signal matrices we can classify the data into clutter and landmine. The EPC curve is thus looked at as a measure on how good the selected landmine signals are separated from the clutter, rather than a measure in general on classification or detection. Figure 5.7, shows the EPC curves before and after applying the threshold and for applying the mean subtraction method.

5.8 Conclusion

The present work deals with the problem of detecting buried landmine using stepped-frequency GPR. One problem that has to be dealt with is that the weak target signal will be hidden in the stronger background signal. In this work, the Likelihood processing technique is employed to reduce the clutter in GPR data. A new threshold formula has been derived which has been applied to the data. In all measurements, the reflection from the surface has been removed or has been at least reduced. The threshold

formula depends on all A-scans (B-scan). From EPC curves (Figure 5.7), the Likelihood processing technique showed significant improvement over mean subtraction method on real field GPR data.

Chapter 6

Combining Wavelet Packets with Higher-Order-Statistics

6.1 The Wavelet Transform

The idea of the wavelet is to decompose the data or signal f into a basis of functions ψ_i

$$f = \sum_i a_i \psi_i,$$

where the functions ψ_i should match the feature of data. One of the main features is decoorrelation.

Let $\psi : \mathbb{R} \to \mathbb{R}$ be a basis wavelet. Where ψ is a function of $L^2(\mathbb{R})$ with zero average

$$\int_{-\infty}^{\infty} \psi(x)dx = 0$$

Consider the family of wavelets obtained from ψ by scaling $a > 0$ and translating it by $b \in \mathbb{R}$

$$\psi_{a,b}(x) = \frac{1}{\sqrt{|a|}} \psi \left(\frac{x-b}{a} \right)$$

Let us define the Continuous Wavelet Transform (CWT) of a function $f(x) \in L^2(\mathbb{R})$ at the scale a and translation b as:

$$\left(\mathbf{W}_\psi f \right)(a,b) = \frac{1}{\sqrt{|a|}} \int_{-\infty}^{\infty} f(x)\psi \left(\frac{x-b}{a} \right) dx \tag{6.1}$$

In CWT, the signals are analyzed using a set of basis functions which relate to each other by simple scaling and translation. In the case of Discrete Wavelet (DWT), a time-scale representation of the digital signal is obtained using digital filtering techniques. The DWT is introduced in the concept of a multi-resolution analysis.

6.1.1 Multi-Resolution Analysis

Consider the space L^2, in multi-resolution analysis the space L^2 is decomposed in nested subspace V_j

$$... \subset V_{-2} \subset V_{-1} \subset V_0 \subset V_1 \subset V_2 \subset ...,$$

such that the closure of their union is L^2

$$\bigcup_{j=-\infty}^{\infty} V_j = L^2,$$

and the intersection

$$\bigcap_{j=-\infty}^{\infty} V_j = \{0\}$$

In dyadic case, we have the following properties

$$f(x) \in V_j \quad \Leftrightarrow \quad \text{dilation} \ \ f(2x) \in V_{j+1} \tag{6.2}$$

$$f(x) \in V_0 \quad \Leftrightarrow \quad \text{translation} \ \ f(x+1) \in V_0$$

Let us define a function $\varphi(x)$ such that the set of functions consisting of $\varphi(x)$ and its integer translates $\varphi(x-k)$ form a basis for the space V_0. For the other subspace V_j, we define

$$\varphi_{j,k}(x) = 2^{j/2}\varphi\left(2^j x - k\right)$$

Where φ is called *a scaling function* or *father function.*

6.1.2 Wavelet Function

Because the subspace V_j are nested, $V_j \subset V_{j+1}$, the V_{j+1} can be decomposed into V_j and W_j, where

$$V_j \oplus W_j = V_{j+1},$$

$$W_j \perp V_j,$$

and

$$\bigcup_{j=-\infty}^{\infty} V_j = \bigoplus_{j=-\infty}^{\infty} W_j = L^2$$

If the function $\psi(x) \in W_0$ obeys the translation 6.2 property where the set of functions consisting of $\psi(x)$ and its integer translates $\psi(x-k)$ form a basis for the space W_0. Then for the other subspace W_j, we define

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

where the ψ function is called *wavelet function* or *mother function.*

6.1.3 Wavelet Decomposition

Assume that V_0 and W_0 are subspace of V_1, then the the functions φ and ψ can be expressed in term of basis functions of V_1 as

$$\varphi(x) = 2\sum_k h_k \varphi(2x - k),$$

$$\psi(x) = 2\sum_k g_k \psi(2x - k),$$

where h_k and g_k are called the filter coefficients. Since $V_{j+1} = V_j \oplus W_j$, the function $f(x)$ which expressed in terms of the basis functions of V_{j+1} can be expressed in terms of the basis function V_j and W_j as

$$f(x) = \sum_k \lambda_{j+1,k} \varphi_{j+1,k}(x),$$

$$f(x) = \sum_l \lambda_{j,l} \varphi_{j,l}(x) + \sum_l \gamma_{j,l} \psi_{j,l}(x),$$

where $\lambda_{j,l}$ and $\gamma_{j,l}$ are the transform coefficients defined as

$$\lambda_{j,l} = \sqrt{2}\sum_k h_{k-2l} \lambda_{j+1,k},$$

$$\gamma_{j,l} = \sqrt{2}\sum_k h_{k-2l} \gamma_{j+1,k}$$

The filter h_k is a low-pass filter and g_k is a high-pass filter. The inverse of the wavelet transform can be obtained in the same way. Figure 6.1 shows that the original signal S, passes through two complementary filters (high-pass and low-pass) that splits the input signal into two decimated orthogonal components. The low-pass output is an approximation of the input signal. The highpass output contains the details of the input signal that are missing from the approximation. There is no information in the two outputs that overlaps, and nothing is lost. The input signal can be exactly reconstructed from the two outputs [119]. The decomposition process can be iterated, with successive approximations being decomposed in turn, so that one signal is broken down into many lower resolution components. This is called the wavelet decomposition tree, see Figure 6.2 The above shows that a signal can be decomposed by DWT into low-pass and high-pass components using the wavelet and scaling functions as shown in Figure 6.3. This is called the Mallat algorithm or Mallat-tree decomposition.

In the Figure 6.3, the signal is denoted by the sequence S. The low pass filter is denoted by h while the high pass filter is denoted by g. At each level, the high pass filter produces detail information, **D**, while the low pass filter associated with scaling function produces approximations, **A**. Figure 6.3 shows only one level decomposition, and reconstruction.
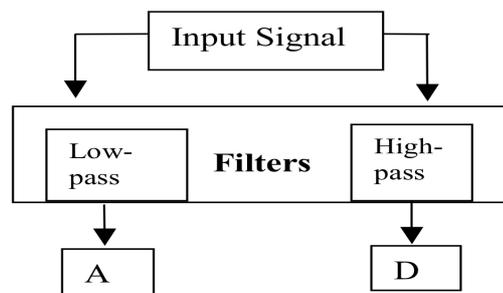
Figure 6.1: Implementation of the discrete wavelet transform for one-Level Decomposition.
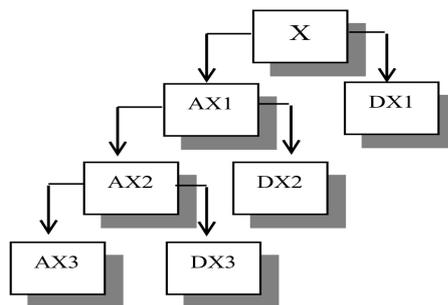

Figure 6.2: Implementation of the discrete wavelet transform for three-Level Decomposition.
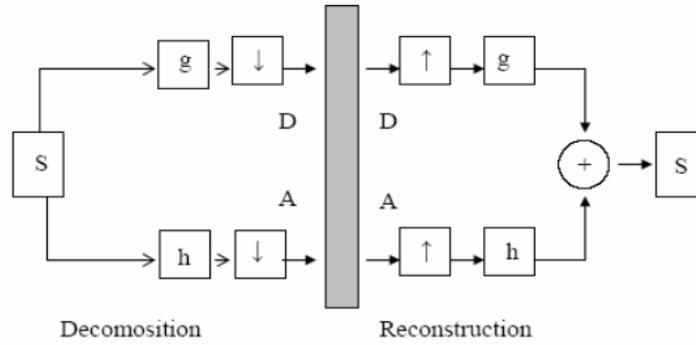
Figure 6.3: The filter bank: the signal S is filtered and downsampled. the high pass filter produces detail information, (D), while the low pass filter associated with scaling function produces approximations, (A). It can be reconstructed by upsampling and filtering with the correct filters.

6.1.4 De-Noising Using Wavelet Decomposition

Let us assume that the received B-scan data is X such that X = S + G, where S is the target signal and G is the clutter. Let C be the reference B-scan with no object. An estimate $\hat{S}$ for the signal S is obtained by denoising X using the wavelet transform. This can be expressed as: $Y = W(X), Z = D(Y, \lambda)$ and $\hat{S} = W^{-1}(Z)$ [120]. Where W and W^{-1} denote the forward and inverse wavelet transform, respectively, and $D(., \lambda)$ is the denoising operator using soft threshold. The threshold is selected as

$$\lambda = k\sigma\sqrt{\frac{j}{J}}$$

where σ is the 1D standard deviation, $J = \log_2(N)$, j is the scale index and k is a constant chosen to be 1.5. The data are transformed in the same way as for the reference scan C. The transformed C scan is subtracted from the data after transform. So the remaining clutter can be removed by passing the resulted image through the threshold. The final step is the inverse transform. The technique is performed using an A-scan. The result is given in Figure 6.4.

6.2 Wavelet Packets

The wavelet packet method is a generalization of wavelet decomposition that offers a richer range of possibilities for signal analysis. In wavelet analysis, a signal is split into an approximation and a detail. The approximation is then itself split into a second-level approximation and detail, and the process is repeated. For an n-level decomposition, there are n+1 possible ways to decompose or encode the signal. In
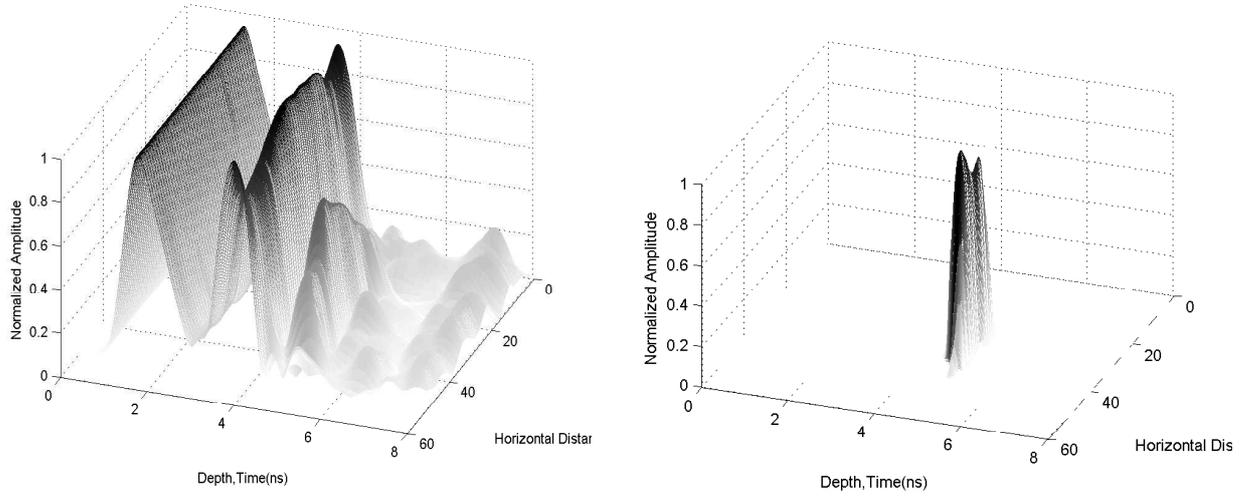
Figure 6.4: Three Dimension images before and after applied the wavelet transform. The target has been used at depth 6cm.

wavelet packet analysis, the details as well as the approximations can be split as shown in Figure 6.5. This yields more than $2^{2^{n-1}}$ different ways to encode the signal. This is the wavelet packet decomposition tree [121].

Let each wavelet coefficient in the above decomposition denoted by $W_{j,n,k}$ where $n \in N$, $(j,k) \in Z$ and $j = 1, 2, ..., J$ is the number of decomposition levels (scale parameter), $n = 1, 2, ..., 2^j$ is the frequency parameter, is consistent with the usual depth-position tree labelling (defines position in the tree), $k = 1, 2, ..., 2^{L-j}$, (where $N = 2^L$ i.e. N the sample number of the original signal) can be interpreted as a time-localization parameter.

De-Noising Using Wavelet Packet

The problem of the detection can be classically formulated as a binary hypothesis testing:

$$H_0 : \quad x(n) = e(n).$$

$$H_1 : \quad x(n) = t(n) + e(n) \tag{6.3}$$

Where $x(n)$ denotes the received GPR data, $t(n)$ is the target signal and $e(n)$ is the noise. We want to decide which of the two hypotheses is true: the noise alone H_0 hypothesis or the signal plus noise H_1 hypothesis. One solution concerned with finding a transform that decompose the original signal into two parts, noise and target signals.

Time-frequency analysis provides a powerful tool for the analysis of the signals. The wavelet transform is a subclass of the general class of time-frequency domain analysis.
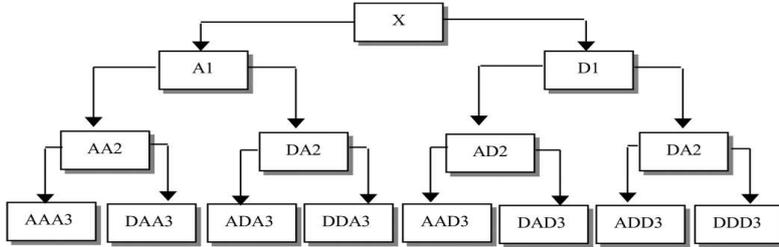
Figure 6.5: Implementation of the wavelet packet transform.

Using wavelet transform, one can decompose the original signal into a smooth part and a detail part. For most signals, energy mainly distributes in the smooth subband, and the energy in the details subband is clustered on a few large coefficients. In contrast, noise energy spreads over both the smooth subband and the detail subband. The time-frequency localization property of the WT can be further exploited in the high resolution analysis of local features, which is a distinct advantage in GPR applications. In noise suppression by wavelet transform, two factors are important for efficiency. The most important factor is a threshold level, and another factor is the number of shifted slices.

Determining a threshold, one should bear in mind that great computational effort must be avoided, and that the threshold must be adapted to the noise level of the data. This study shows how to remove the noise from the GPR data by using wavelet packet transform. The threshold level has been selected and used. Its value calculation is based on the Higher-order statistics. Assume that the observed data $x(n)$ are given by:

$$x(n) = t(n) + e(n) \tag{6.4}$$

Contains the target signal $t(n)$, $e(n)$ the noise , with $n = 1, 2, ..., N$ and $e(n)$ refers to all noise components, i.e., ground bounce, cross-talk and the noise. The two-dimension wavelet transform gives the two-dimensional wavelet packet coefficients in the form

$$WP_{j,s}^{x}(i) = WP_{j,s}^{t}(i) + WP_{j,s}^{e}(i) \tag{6.5}$$

where $WP_{j,s}^{x}(i)$, $WP_{j,s}^{t}(i)$ and $WP_{j,s}^{e}(i)$ are the wavelet packet coefficients of x, t, and e respectively, $j = 1, 2, ..., J$ is the number of decomposition levels, $s = 1, 2, ..., 2j$ is the number of scales, $i = 1, 2, ..., M$, where $M = N/2j$ and N is the length of the signal vector. When the noise is white Gaussian, the linear wavelet transform will

also produce white and Gaussian coefficients. Higher-order statistics test is used to determine the Gaussianity of the wavelet coefficients and then by setting all Gaussian coefficients to zero, the target signal, can be extracted.

The segmentation decision may then be formulated by the following binary hypothesis-testing problem:

$$H_0: \quad \text{set} \quad d(s) = 0 \quad \text{if the coefficient is Gaussian}$$

$$H_1: \quad \text{keep} \quad d(s) \quad \text{if the coefficients is non-Gaussian}, \tag{6.6}$$

where $d(s)$ is the wavelet coefficient.

6.3 Higher Order Statistics

Dealing with non-Gaussian random processes, the notions of higher order moments, cumulants, and their polyspectra called higher order statistics are of paramount importance in statistical signal processing. If $x(n), n = 0, \pm 1, \pm 2, \pm 3, ..$ is a real stationary discrete- time signal and its moments up to order p exist, then its p-th order moment function is given by

$$m_p(\tau_1, \tau_2, ..., \tau_{p-1}) = E\{x(n)x(n + \tau_1), ..., x(n + \tau_{p-1})\}, \tag{6.7}$$

and depends only on the time differences $\tau_1, \tau_2, ..., \tau_{p-1}$ $\tau_i = 0, \pm 1, \pm 2, \pm 3, ...,$ for all i. Here $E\{.\}$ denotes statistical expectation and for a deterministic signal, it is replaced by a time summation over all time samples (for energy signals) or time averaging (for power signals). If in addition the signal has zero mean, then its cumulant functions (up to order four) are given by [122]
second-order cumulant:

$$C_2(\tau_1) = m_2(\tau_1) \tag{6.8}$$

third-order cumulant:

$$C_3(\tau_1, \tau_2) = m_3(\tau_1, \tau_2) \tag{6.9}$$

fourth-order cumulant:

$$c_4(\tau_1, \tau_2, \tau_3) = m_4(\tau_1, \tau_2, \tau_3) - m_2(\tau_1)m_2(\tau_3 - \tau_2) - m_2(\tau_2)m_2(\tau_3 - \tau_1) - m_2(\tau_3)m_2(\tau_2 - \tau_1). \tag{6.10}$$

By setting all the lags to zero in the above cumulant expressions, we obtain the variance, skewness and kurtosis, respectively,
Variance:

$$\gamma_2 = C_2(0) = E\{x^2(n)\} \tag{6.11}$$

Skewness:

$$\gamma_3 = C_3(0, 0) = E\{x^3(n)\} \tag{6.12}$$

Kurtosis:

$$\gamma_4 = C_4(0,0,0) = E\{x^4(n)\} - 3\left[E\{x^2(n)\}\right]^2 \tag{6.13}$$

When estimating higher-order statistics from finite data records, the variance of the estimators is reduced by normalizing the input data to have a unity variance, prior to computing the estimators. Equivalently, the third order and fourth-order statistics are normalized by the appropriate powers of the data variance, thus we define the Normalized skewness:

$$S = \frac{C_3(0,0)}{[C_2(0)]^{1.5}} = \frac{E\{x^3(n)\}}{[Ex^2(n)]^{1.5}}, \tag{6.14}$$

and the
Normalized kurtosis:

$$K = \frac{C_4(0,0,0)}{[C_2(0)]^2} = \frac{E\{x^4(n)\}}{[Ex^2(n)]^2} - 3, \tag{6.15}$$

6.4 Denoising by Combining Wavelet Packets with Higher-Order-Statistics

We have assumed that the received data consists of two parts, noise, and target signals and that, $e(n)$ is a white Gaussian noise and the two signals $t(n)$ and $e(n)$ are stochastically independent. Then the higher-order statistics of the receiving data (x) can be written as [123]

$$H(x) = H(t) + H(e) \tag{6.16}$$

Where H(t) and H(e) are the higher-order statistics of the target, and the noise signals, respectively. We also assume that some of the wavelet coefficients of the receiving data belong to the target signal, and some of it belongs to the noise signal. So we have target, and noise coefficients. Figure 6.6 shows the wavelet packet coefficients at each scale to the data in Figure 2.8. The problem now, is how can we separate the target coefficients from all the data coefficients. Higher-order statistics are traditionally used to accomplish this task. Our candidate are skewness and kurtosis, which are the normalized version of third-order cumulant and the fourth-order cumulant. For Gaussian process, theoretically the process has a skewness or a kurtosis value that equals zero. But in practice, the skewness and the kurtosis are estimated and their values are authorized to exist in a confidence interval, which is conditioned by the probability properties of the estimator. By using the Bienayme-chebyshev inequality, given a desired confidence percentage, the estimator can be framed between two values depending on the first statistics of the estimator.

6.4.1 Gaussianity test

The test we present here are based on the normalized both forth and third-order cumulant, Equations (6.15) and (6.14). In the Gaussian test the noise is assumed to be white Gaussian, and the noise and the target signals are stochastically independent. Firstly, let us discuss the kurtosis processing after that we will discuss the skewness.

Kurtosis Test

The kurtosis of the received data (x) can be written as [123]

$$K(x) = K(b) + K(t) \tag{6.17}$$

Where $K(b)$ and $K(t)$ are the kurtosis of the clutter and target signals, respectively. We also assume that some of the wavelet coefficients of the received data belong to the target signal, and some of it belong to the noise signal. So we have target, and noise coefficients. The problem now, how can we separate the target coefficients. Instead, determination of the wavelet coefficients of the noise by checking its Gaussianity using the higher-order statistics. When the noise is white stationary, the coefficients remain white stationary. Our candidate is the kurtosis, which is the normalized version of the forth-order cumulant. The Gaussian process has a kurtosis value that equals zero. To perform a de-noising procedure on the wavelet coefficients, all Gaussian coefficients are set to be zero. The forth-order cumulant, kurtosis Equation (6.15), has been computed by the method of moments. In this method, while fitting a probability distribution to a sample, the parameters are estimated by equating the sample moments to those of the theoretical moments of the distribution. Even though this method is conceptually simple, and the computations are straight-forward, it is found that the numerical values of the sample moments can be very different from those of the population from which the sample has been drawn, especially when the sample size is small [124]. In our work we have a limited number of data sample, so we are not able to obtain an exact value of the kurtosis. Instead, we have an estimate value using sample averages. The estimation of the kurtosis can be calculated as [125]

$$\hat{K} = \frac{1}{N} \sum_{i=1}^{N} \left(\frac{x_i - \hat{\mu}}{\hat{\sigma}} \right)^4 - 3, \tag{6.18}$$

where $\hat{\mu} = \frac{1}{N} \sum_{i=1}^{N} x_i$ and $\hat{\sigma} = \frac{1}{N} \sum_{i=1}^{N} (x_i - \hat{\mu})$. The estimated value of the kurtosis is allowed to exist in a confidence interval. Normally, the confidence interval is calculated when the probability density function is known. But the probability density function of the fourth-order cumulant of a Gaussian sequence is not known analytically. A partial solution for this problem is to use the Bienayme-chebyshev inequality. The inequality makes it possible to frame our estimate and reads for the $\hat{K}$ estimator [126]

$$Prob \left(|\hat{K} - E(\hat{K})| \leq a \sqrt{var(\hat{K})} \geq 1 - \frac{1}{a^2} \right) \tag{6.19}$$

A fixes confidence percentage corresponds to a value of the factor $a = 1/\sqrt{1-\alpha}$. Now, the kurtosis estimator varies between

$$\pm \frac{1}{\sqrt{1-\alpha}}\sqrt{var(\hat{K})}, \tag{6.20}$$

with α is the authorized confidence percentage value. The variance of the kurtosis is [127]

$$Var(\hat{K}) = \frac{1}{N\mu_2^3}[\frac{\mu_8}{\mu_2} - 4\frac{\mu_6\mu_4}{\mu_2^2} + 4\frac{\mu_4^3}{\mu_2^3} - \frac{\mu_4^2}{\mu_2} + 16\frac{\mu_4\mu_3^2}{\mu_2^2} - 8\frac{\mu_3\mu_5}{\mu_2} + 16\mu_3^2] \tag{6.21}$$

Where μ_i denotes the i-th central moment. In the case where the N coefficients $WP_{j,s}^e(i)$ are white and Gaussian, the variance of the kurtosis estimator $\hat{K}$ and the bias are evaluated as: [128]

$$Var(\hat{K}) \approx 24\sigma^8/N$$

$$B(\hat{K}) = -6\sigma^4/N$$

where σ^2 is the variance of the noise. The bias has been neglected, when the confidence interval has been computed. So the confidence interval is based only on the variance. The kurtosis estimator (for $\sigma=1$) varies between

$$\pm \frac{1}{\sqrt{1-\alpha}}\sqrt{24/N} \tag{6.22}$$

The kurtosis has been framed with 97.5 % of confidence by the following inequality [129]:

$$-6.3\sqrt{\frac{24}{N}} \leq \hat{K} \leq 6.3\sqrt{\frac{24}{N}} \tag{6.23}$$

Skewness Test

The proceeding of skewness test is the the same of the kurtosis test, except of small differences. The skewness of the received data (x) can be written as [123]

$$S(x) = S(b) + S(t) \tag{6.24}$$

Where $S(b)$ and $S(t)$ are the skewness of the clutter and target signals, respectively. The estimation of the skewness can be calculated as:

$$\hat{S} = \frac{1}{N}\sum_{N=1}^{N}\left(\frac{x_i - \hat{\mu}}{\hat{\sigma}}\right)^3, \tag{6.25}$$

where $\hat{\mu} = \frac{1}{N}\sum_{N=1}^{N}x_i$ and $\hat{\sigma} = \frac{1}{N}\sum_{N=1}^{N}(x_i - \hat{\mu})$. The estimated value of the skewness is allowed to exist in a confidence interval. The Bienayme-chebyshev inequality, which makes it possible to frame our estimates for the estimator and is expressed as [126]:
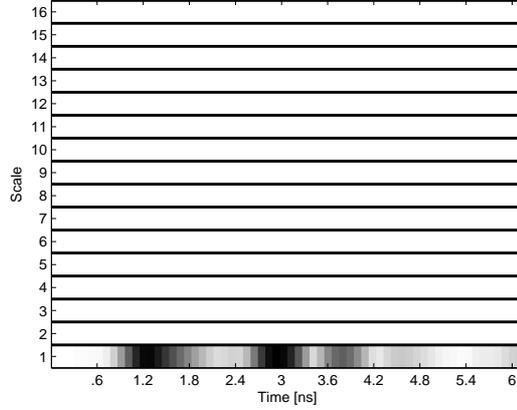
Figure 6.6: The wavelet packet of GPR data.

$$Prob\left(|\hat{S} - E(\hat{S})| \le a\sqrt{var(\hat{S})}\right) \ge 1 - \frac{1}{a^2}$$

A fixed confidence percentage corresponds to a value of the factor $a = 1/\sqrt{1 - \alpha}$, where α is the authorized confidence percentage value.

Now, the skewness estimator varies between

$$\pm \frac{1}{\sqrt{1 - \alpha}}\sqrt{var(\hat{S})}, \tag{6.26}$$

The variance of the skewness is [127]

$$var(\hat{S}) = \frac{1}{n}\left\{\frac{\mu_6}{\mu_2^3} - 6\beta_2 + 9 + \frac{\beta_1}{4}(9\beta_2 + 35) - \frac{3\mu_5\mu_3}{\mu_2^4}\right\} \tag{6.27}$$

where μ is the central moment, $\beta_1 = \mu_3^2/\mu_2^3$ and $\beta_2 = \mu_4/\mu_2^2$

In the case where the n coefficients $WP_{j,s}^e(i)$ are white and Gaussian, the variance of the third-order cumulant is evaluated as [127]

$$var(\hat{S}) = 6\sigma^6/n$$

where σ^2 is the variance of the n noise. The variance of skewness estimator (for $\sigma=1$) is

$$var(\hat{S}) = 6/n \tag{6.28}$$

The simple test for Gaussianity measure is that the skewness varies between

$$\pm \frac{1}{\sqrt{1 - \alpha}}\sqrt{6/n} \tag{6.29}$$

The skewness has been framed with 90 % of confidence by the following inequality [130]:
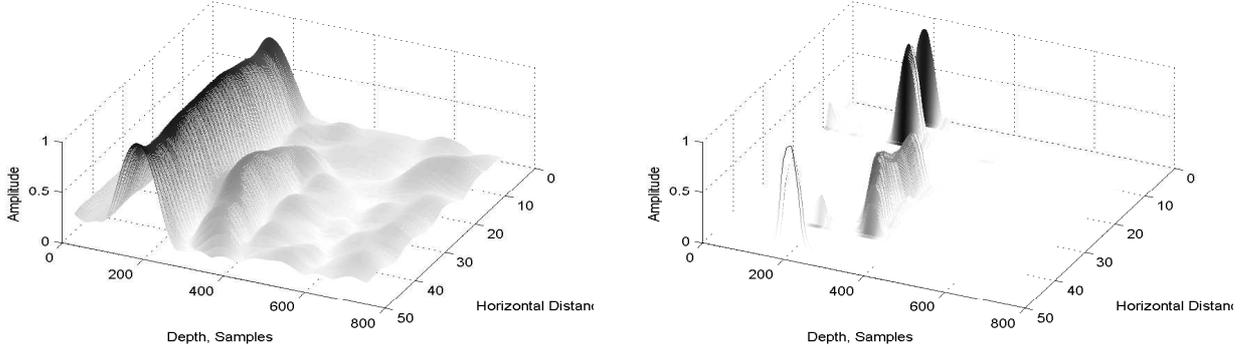
Figure 6.7: Three Dimension images before and after applying the algorithm for GPR data. The target is at depth 3cm.

$$-\frac{3}{10}\sqrt{\frac{6}{n}} \leq \hat{S} \leq \frac{3}{10}\sqrt{\frac{6}{n}} \tag{6.30}$$

Now, the decision rule is applied to each coefficients and becomes:

H_0 : HOS of the coefficient have a value within the above interval.

H_1 : otherwise leave the coefficient.

In other words, our decision depends on the value of the skewness and kurtosis (HOS) of the coefficient. If the HOS of the coefficient takes a value within the above intervals, the coefficient will be Gaussian. Otherwise, the coefficient will be non-Gaussian. The next step, we applied this segmentation algorithm to the GPR data.

6.4.2 Clutter removal

To remove the clutter from GPR data, we have used kurtosis or skewness test algorithm. The algorithm has been applied to remove the noise. The steps of the algorithm are

1. Remove the cross-talk signal by a time gating technique.

2. Compute the wavelet packet coefficients of the received data $WP_{j,s}^{x}(i)$ at level j, scale $s = 1, 2, ..., 2j$.

3. Estimate the kurtosis or the skewness for the wavelet packet coefficients of each scale using Equation (6.25) or (6.18).

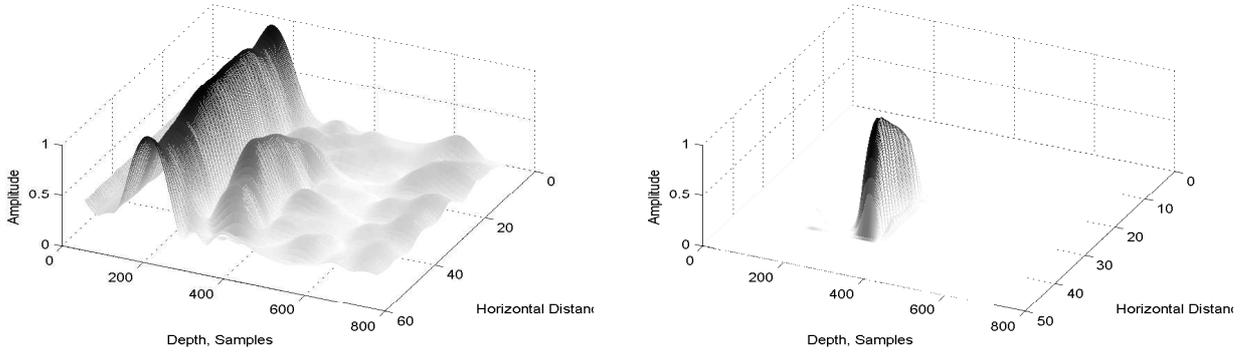4. Apply the Gaussianity test Equation (6.23) or (6.30).

Figure 6.8: Three Dimension images before and after applied the algorithm for GPR data. The target has been used at depth 5cm.
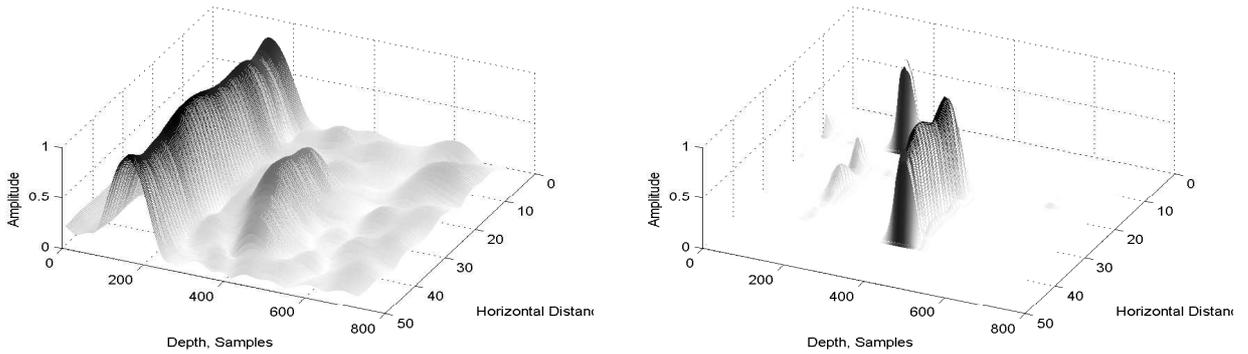

Figure 6.9: Three Dimension images before and after applied the algorithm for GPR data. The target has been used at depth 10 cm.

5. Apply the segmentation rule to the coefficients.

6. For the remaining coefficients a hard threshold has been applied to further improve the SNR if needed. The threshold value is calculated using the following relation:

$$\delta_S = \bar{\sigma}_s \sqrt{2log(N)} \tag{6.31}$$

where $\bar{\sigma} = median(|WP_{j,s}^t|)/0.6745$

7. Reconstruct the target signal from the remaining coefficients.

6.5 Conclusions

The main purpose of this communication was to study the effect of combining two powerful tools, namely, the wavelet packet transform and the higher order statistics in detecting AP mines using GPR in a severe noisy environment. The skewness and kurtosis test were selected as the HOS measurements to be used in this algorithm. The skewness or kurtosis algorithm test consists of two parts, and has been applied to GPR data. The first part of the algorithm is to remove the Gaussian noise, and the second part is used for further improvement in SNR if necessary. For kurtosis, the threshold value is calculated using confidence percentage value $\alpha = 97.5\%$ using Equation (6.30). And for skewness, the threshold value is calculated using confidence percentage value $\alpha = 90\%$ by using Equation (6.23). The Daube-chies wavelet has been used of order 4 and number of levels $J = 4$. The algorithms give good result, which proves its validity. The three dimension images of the GPR data and the result of applying the algorithm are depicted in Figures 6.7, 6.8, and 6.9 for target at depth 3cm, 5cm and 10cm, respectively. The surface reflection and the reflection within the earth have been approximately removed. However, the algorithm prove to be successful in denoising GPR data. When the clutter has been assumed to be white Gaussian noise, and it has been shown that most of the clutter has been removed using the described algorithm.

Chapter 7

Further Experimental Results

This section summarizes the results of applying the signal processing techniques described in the preceding chapters to GPR data.

7.1 Results of Clutter-Reduction Techniques

7.1.1 Singular Value Decomposition (SVD)

Two techniques of SVD algorithms have been deployed. Ranking of SVD components shows that the background information and the target information contained in the second components respectively. On the other hand, when mean subtraction method has been used; the first component contains the target information. The threshold has been employed to improve the SNR, where a new formula to calculate the threshold has been developed. Good results for automatic selection of SVD components have been achieved for some landmines as PMN, PMN2, PPM, and PFM. Moreover, for another landmines as M14 or cylindrical wood , SVD is not sufficient technique to separate the mine signal from received data.   Figure 7.1 shows that the data of PMN2 landmine at depth 5 cm under the surface, and the result of applying SVD technique for First algorithm (Section 3.1.2) and second algorithm (Section 3.1.2), respectively. Figure 7.2 shows the result of applying both algorithm to the same data. From Figure 7.2, the detection technique has been proven and approximately the clutter has been eliminated.

7.1.2 Factor Analysis (FA)

Factor analysis which has been described in Section 3.2 was applied to GPR data. Only the first two factors have been chosen. In almost measurements, the strong
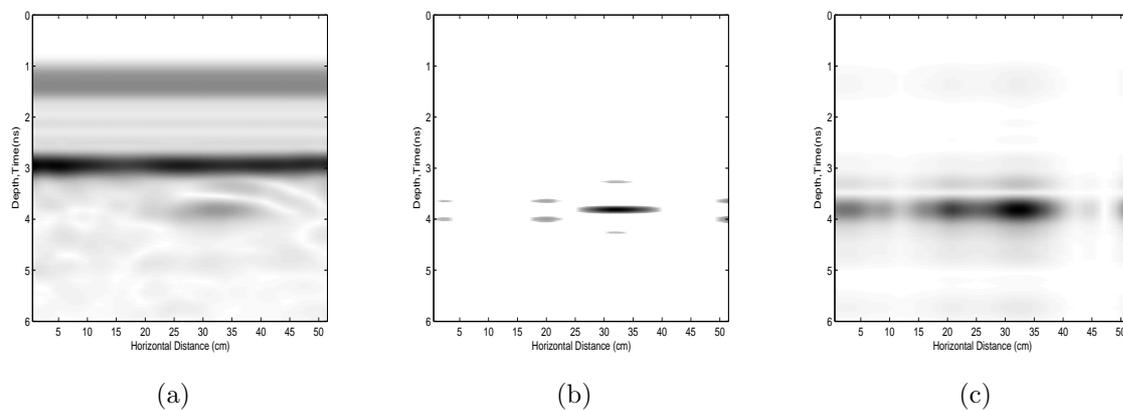
(a) (b) (c)

Figure 7.1: (a) Anti-personnel land mines at depth 5 cm PMN2 (b) data after applying first algorithm Section 3.1.2(c) data after applying second algorithm Section 3.1.2
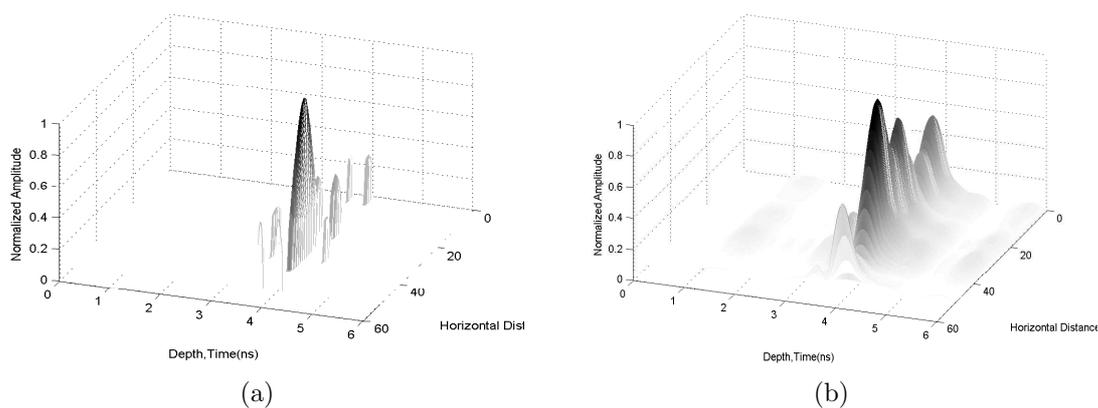

(a) (b)

Figure 7.2: Three dimension images of applying SVD technique for First algorithm (Section 3.1.2) and second algorithm (Section 3.1.2), respectively.

clutter was almost eliminated and target signals get prominent clearly. Figure 7.3 demonstrates samples of PPM landmine B-scans at depth 1, 5, 10 cm, respectively. In the other hand, factor analysis could not be estimated the landmine signals as M14 or wood target. In case of some targets as PFM, the factor analysis gives good result when the depth of the target is more than or equal 5 cm. However, targets just below the surface ($\approx 1cm$), factor analysis gives bad result.
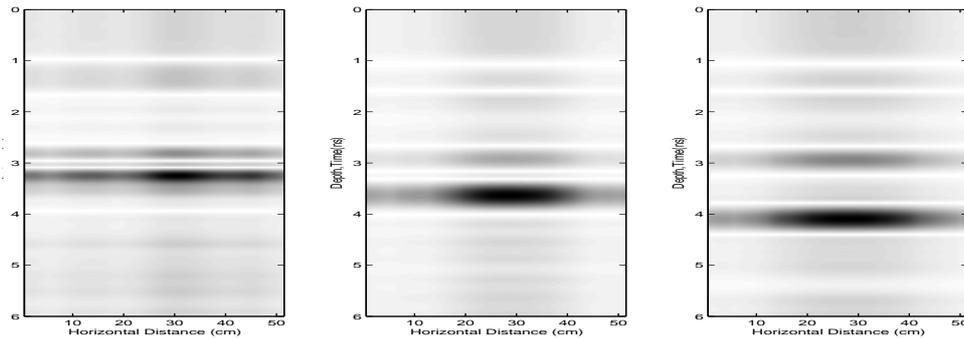

Figure 7.3: The data after applying the factor analysis. A PPM landmine has been used at depths 1, 5, 10 cm below the sand surface.

7.1.3 Principal Component Analysis (PCA)

The results of Principal Component Analysis (Section 3.3) are shown in Figure 7.4. In Figure 7.4 the data after applying PCA for PPM landmine at different depths 1, 5, 10 cm, respectively. The reflection from the surface has been removed or has been at least reduced. PCA has given the same results as SVD and factor analysis when the target signal is very low (M14 and wood). The PCA cannot be used to detect landmine made of wood, or small landmine as M14.

7.1.4 FastICA

To examine the ability of ICA technique for the elimination or reduction of the clutter in raw GPR data, the FastICA technique has been applied to all measurements.

The IC's were estimating using FastICA algorithm (Section 4.7) and four different cost functions G1, G2, G3 and G4 in Equations (4.23),(4.24),(4.25) and (4.26) have been used in deflationary and symmetric techniques. Figure 7.5 shows the data after applying the ICA algorithm to the PPM landmine data in depths 1, 5 and 10 cm, respectively. From Figure 7.5, the target signals have been extracted from the GPR data, and most of the clutter signal (cross-talk and air-ground interface) has been
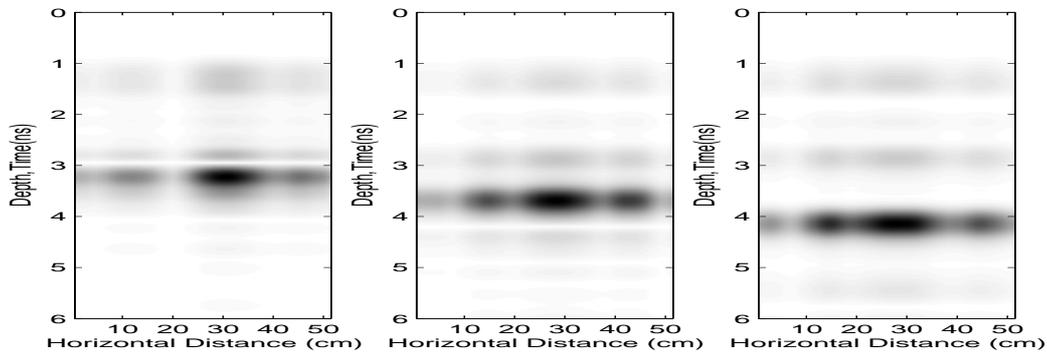
Figure 7.4: The data after applying the PCA. A PPM landmine has been used at depths 1, 5, 10 cm below the sand surface.
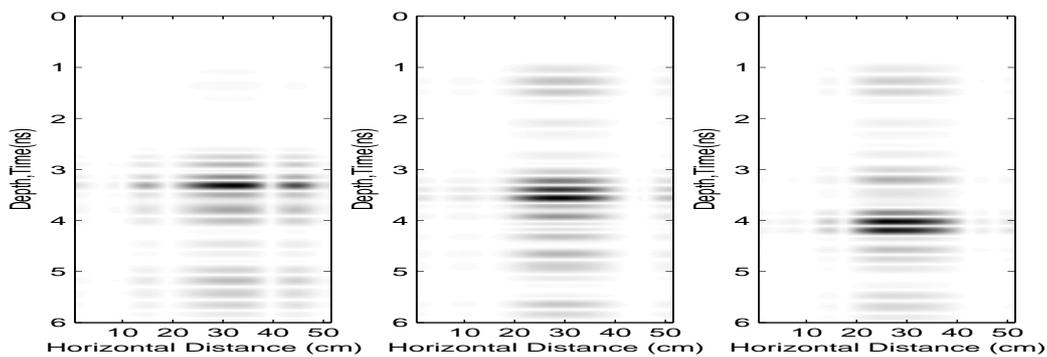

Figure 7.5: The data after applying the FastICA. A PPM landmine has been used at depths 1, 5, 10 cm below the sand surface.

removed for the targets at depth 1, 5 and 10 cm. The deflationary and symmetric techniques give the same results. Moreover in both techniques, G1, G3 and G4 gave consistent result and have shown approximately small improvement over G2.

Although the FastICA gives good result for almost measurements, it gives bad result for M14 and wood.

7.1.5 Infomax

Good results have been obtained by applying infomax algorithm almost in all measurements. Figure 7.6 shows the A-scans before and after applying infomax algorithm. The PMN2 landmine has been used at depth 10 cm. While Figure 7.7 demonstrates the result in three dimension. From Figures 7.6 and 7.7, the target signals have been extracted from the GPR data, and most of the clutter signal (cross-talk and
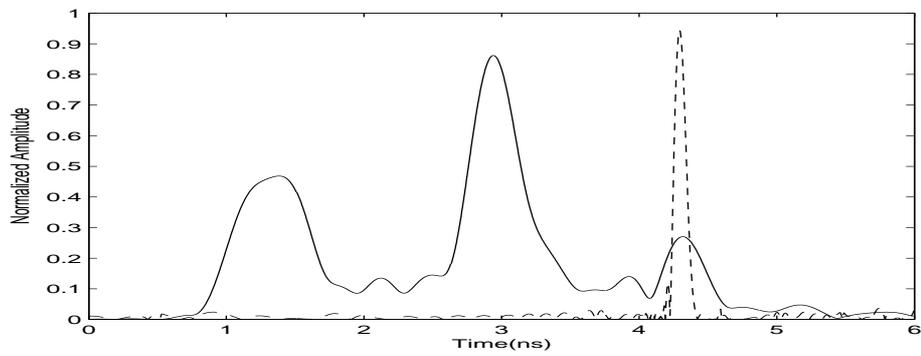
Figure 7.6: A-scan of the received data (solid) and the data after applying infomax technique. A PMN2 landmine has been used at depths 10 cm below the sand surface.
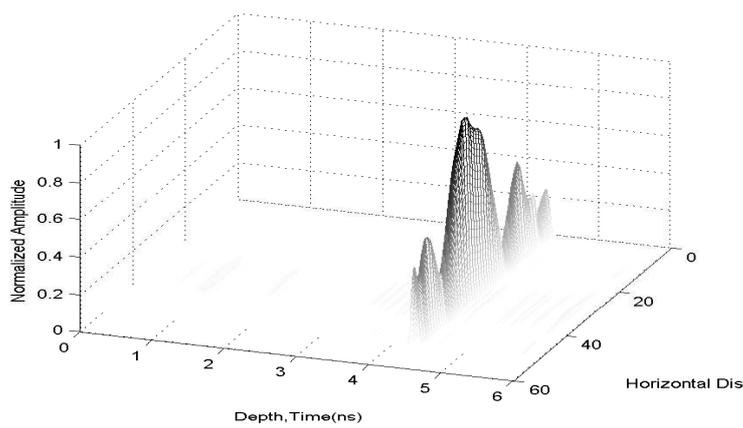

Figure 7.7: The data after applying the infomax technique. A PMN2 landmine has been used at depths 10 cm below the sand surface.

Figure 7.8: A-scan of the received data (solid) and the data after applying SOBI technique. A PMN2 landmine has been used at depths 10 cm below the sand surface.

Figure 7.9: The data after applying the SOBI technique. A PMN2 landmine has been used at depths 5 cm below the sand surface.

air-ground interface) has been removed.

7.1.6 SOBI

The SOBI algorithm has been applied to all measurements. Figure 7.8 and Figure 7.9 show the one dimension and three dimension after applying the SOBI technique. The PPM2 landmine has been used at depth 5 cm. From Figures 7.8 and 7.9 the target signal has been extracted and most of clutter has been eliminated. The results indicate that the target signal can be extracted by using SOBI technique.
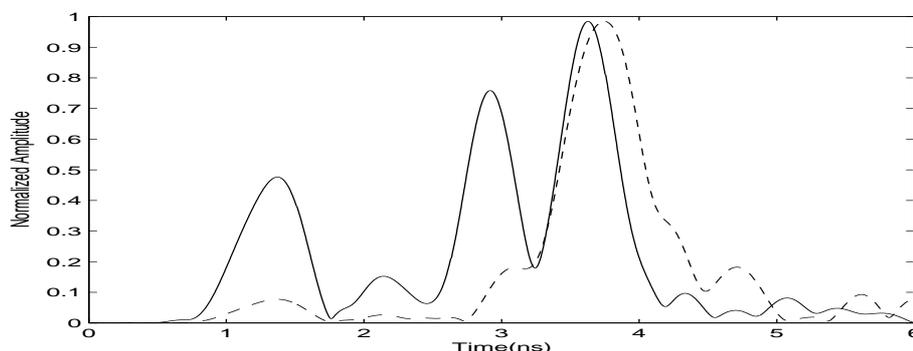
Figure 7.10: A-scan of the received data (solid) and the data after applying JADE technique. A PMN landmine has been used at depths 5 cm below the sand surface.
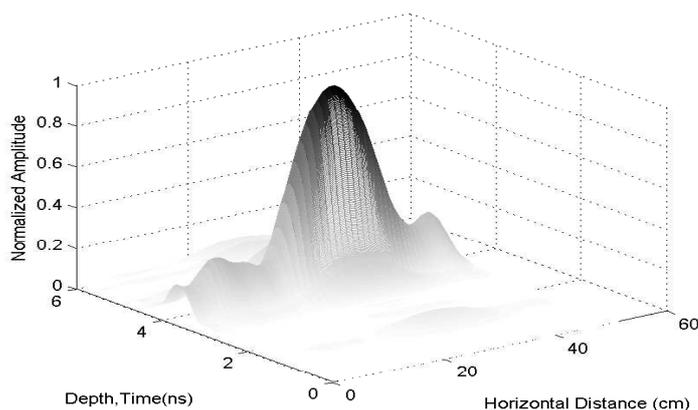

Figure 7.11: The data after applying the JADE technique. A PMN landmine has been used at depths 5 cm below the sand surface.

7.1.7   JADE

The Jade algorithms gives good results for estimating both clutter and target signals. Figure 7.10 and Figure 7.11 show the one dimension and three dimension images of the data after applying the Jade algorithm to the data in Figure 2.8. The PPM2 landmine has been used at depth 5cm.

7.1.8   Wavelet Transform

In this work the wavelet transform has been applied to all measurements. The reference B-scan **b** which the measurement with no target has been used. The data are transformed in the same way as for the reference scan **b** (B-scan with no target). The

Figure 7.12: A-scan of the received data (solid) and the data after applying wavelet transform technique. A PMN2 landmine has been used at depths 10 cm below the sand surface.

transformed **b** scan is subtracted from the data after transform. So the remaining clutter can be removed by passing the resulted image through the threshold. The final step is the inverse transform. The technique is performed using an A-scan. The results are given in Figure 7.12 and in Figure 7.13. From Figure 7.12 and Figure 7.13 the cross-talk, air-ground interface and reflection within the soil approximately have been removed.

7.1.9   Wavelet Packets Combined with Higher-Order-Statistics

An algorithm has been developed for this purpose which combines two powerful tools: The wavelet packet analysis and the higher-order-statistics (HOS).

The combination of wavelet packet analysis and the HOS have been applied to all measurements. This algorithm divided into two algorithms, namely skewness and kurtosis algorithm. Each algorithm consists of two parts, the first part of the algorithm is to remove the Gaussian noise, and the second part is used for further improvement in SNR if necessary. The threshold value is calculated using Gaussianity test Equation (6.30) for kurtosis and Equation (6.23) for skewness.

The Daube-chies wavelet has been used of order 4 and number of levels $J = 4$. The distance between two antennas is constant, thus the cross-talk has been removed before applying this algorithm by a time gating technique. The one and three dimension result images of applying the skewness test are depicted in Figures 7.14, and 7.15 to the data in Figure 2.8. While Figures 7.16 and 7.17 are the result of applying the Kurtosis test to the same data.

The surface reflection and the reflection within the earth have been approximately removed.

Figure 7.13: The data after applying the wavelet transform technique. A PMN2 landmine has been used at depths 10 cm below the sand surface.
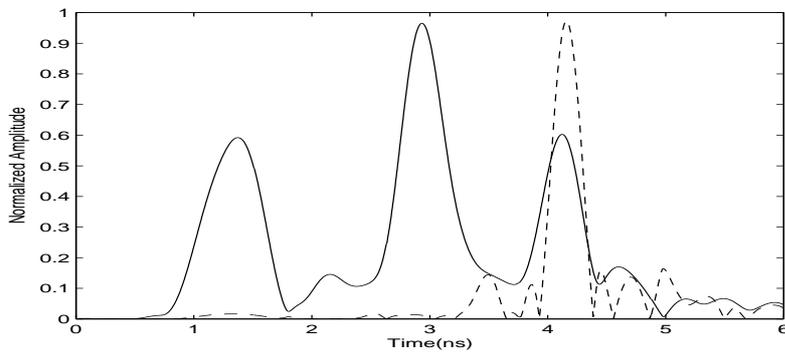

Figure 7.14: A-scan of the received data (solid) and the data after applying skewness test. A PMN landmine has been used at depths 5 cm below the sand surface.
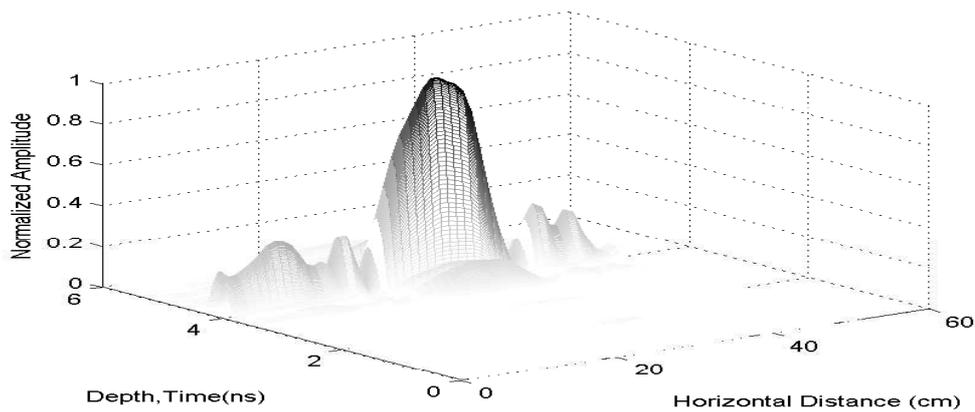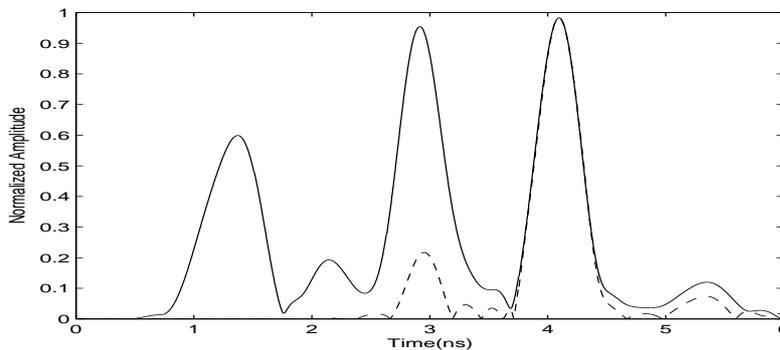
Figure 7.15: The data after applying the skewness test. A PMN landmine has been used at depths 5 cm below the sand surface.
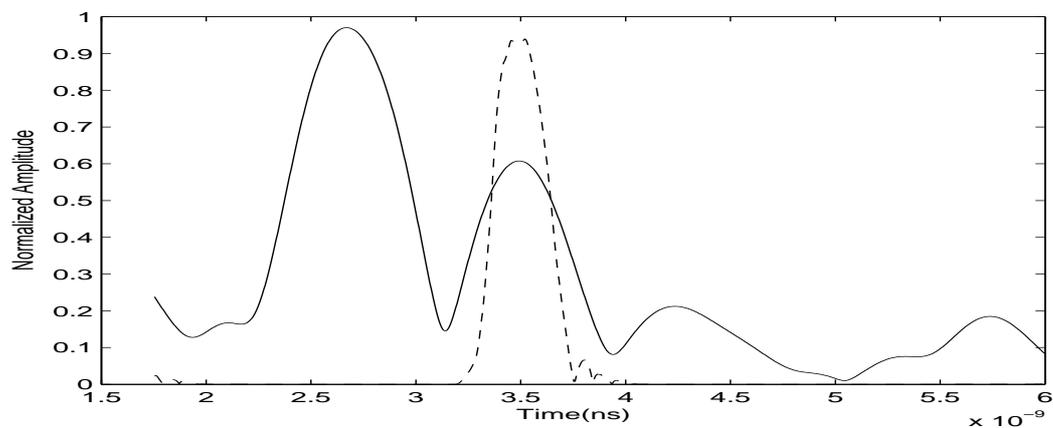
Figure 7.16: A-scan of the received data (solid) and the data after applying the kurtosis test. A PMN landmine has been used at depths 5 cm below the sand surface.
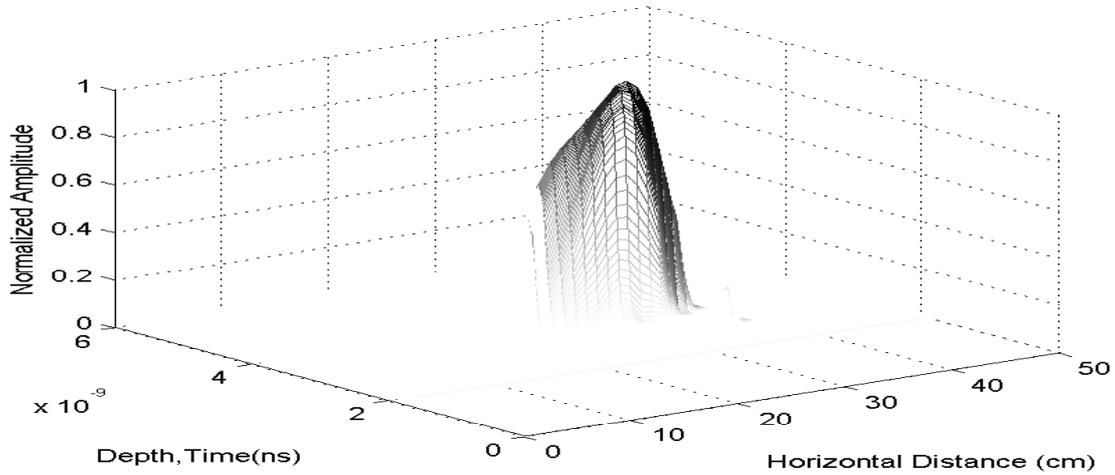
Figure 7.17: The data after applying the kurtosis test. A PMN landmine has been used at depths 5 cm below the sand surface.

7.1.10 Likelihood Method

In this work, the Likelihood processing technique is employed to reduce the clutter in GPR data. A new threshold formula has been derived which has been applied to the data. In all measurements, the reflection from the surface has been removed or has been at least reduced. The one and three dimension result images of applying the algorithm are depicted in Figure 7.18 and in Figure 7.19 for PMN landmine at depth 5 cm. Figures 7.18 and 7.19 demonstrate one and three dimension images of the data before and after applying the algorithm.

7.2 Performance Measures of Clutter-Reduction Methods

In this thesis we use different signal processing techniques. The comparison of these techniques will be based on signal-to-noise ratio (SNR) and Expected Performance Curves (EPC).

7.2.1 Signal-to-Noise Ratio (SNR)

For effective detection and discrimination of the APL's, the calculation of the signal-to-noise ratio (SNR) is needed [131]. This is done in this thesis according to the empirical formula explained in the technical report [132].

Figure 7.18: A-scan of the received data (solid) and the data after applying Likelihood technique. A PMN landmine has been used at depths 5 cm below the sand surface.
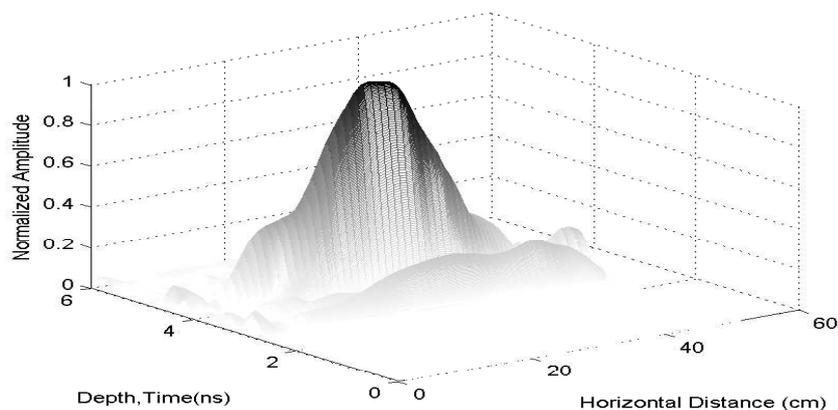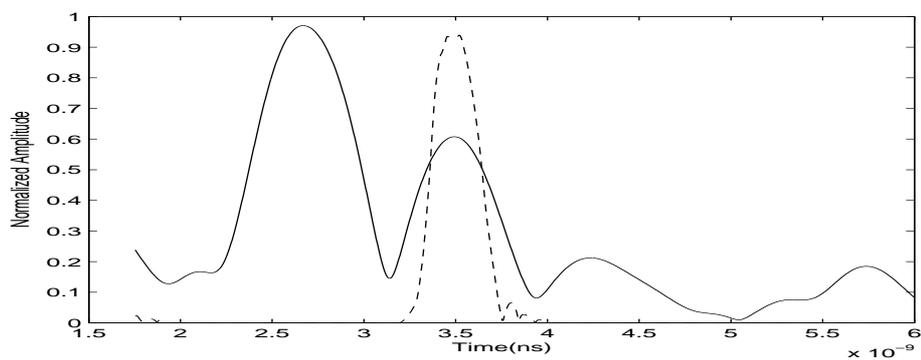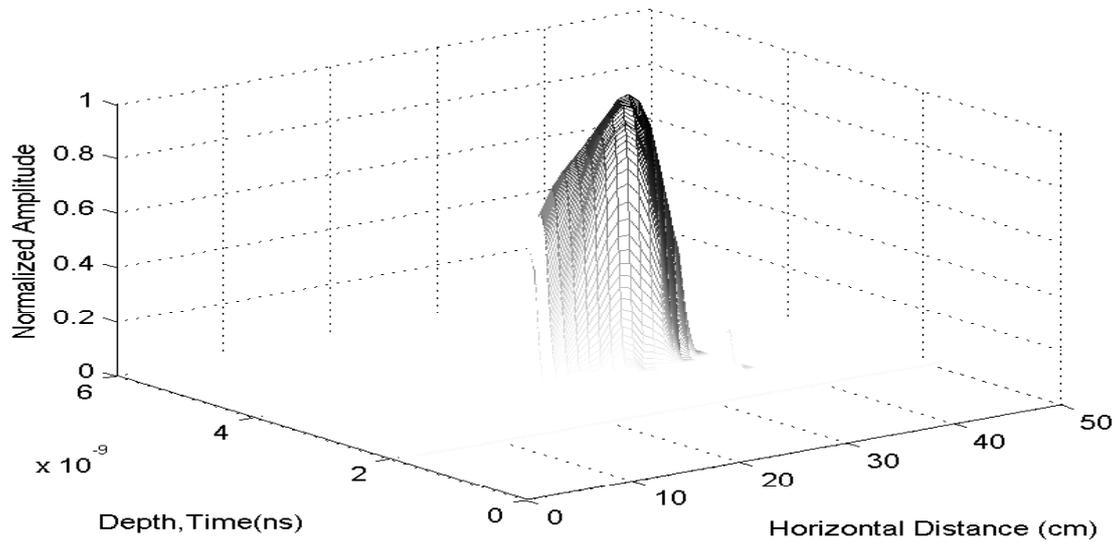

Figure 7.19: The data after applying the Likelihood technique. A PMN landmine has been used at depths 5 cm below the sand surface.
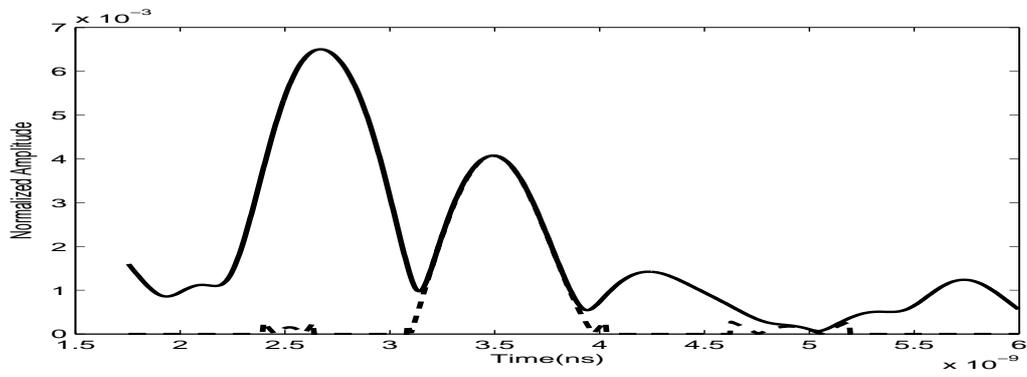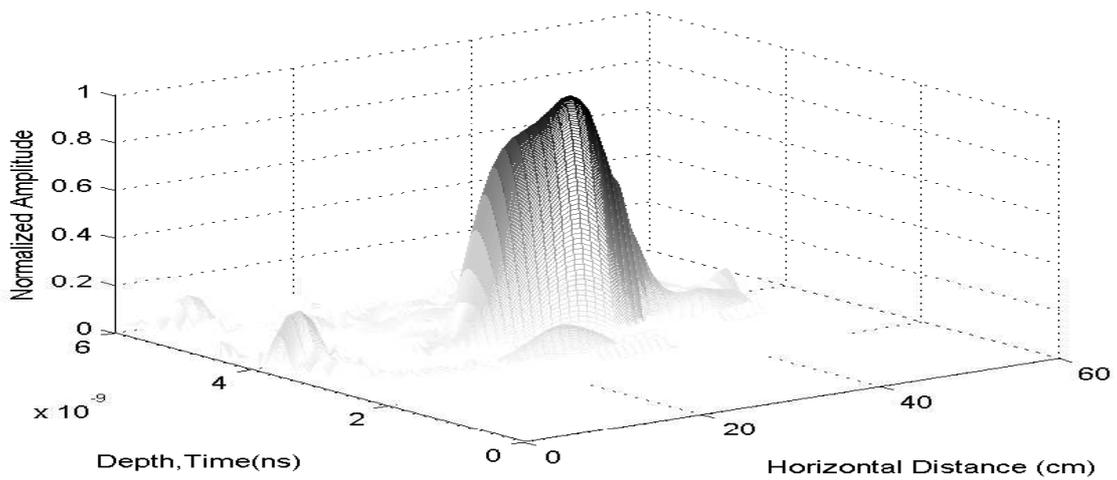
To calculate the SNR, let us assume that the received data in the region where $r < r_1$ consist of a target signal, a background and noise For region where $r > r_2$, we expect that only background and noise are received. The set of coordinates (x, y) for $r < r_1$ and $r > r_2$ are denoted by R_1, and R_2, respectively.

An empirical SNR can now defined as

$$SNR = \frac{\text{Average energy of signals in } R_1}{\text{Average energy of signals in } R_2} - 1$$

Let us assume that N_1 and N_2 denote the number of measurements in R_1 and R_2, respectively, we have then

$$\text{SNR} = \frac{\frac{1}{N_1} \sum_{(x,y) \in R_1} \|\mathbf{x}\|^2}{\frac{1}{N_2} \sum_{(x,y) \in R_2} \|\mathbf{x}\|^2} - 1$$

$$= \frac{\frac{1}{N_1} \sum_{(x,y) \in R_1} \|\mathbf{s} + \mathbf{b} + \mathbf{e}\|^2}{\frac{1}{N_2} \sum_{(x,y) \in R_1} \|\mathbf{b} + \mathbf{e}\|^2} - 1$$

$$= \frac{\frac{1}{N_1} \sum_{(x,y) \in R_1} \|\mathbf{s}\|^2 + \|\mathbf{b}\|^2 + \|\mathbf{e}\|^2}{\frac{1}{N_2} \sum_{(x,y) \in R_1} \|\mathbf{b}\|^2 + \|\mathbf{e}\|^2} - 1$$

$$\text{SNR} \approx \frac{\frac{1}{N_1} \sum_{(x,y) \in R_1} \|\mathbf{s}\|^2}{\frac{1}{N_2} \sum_{(x,y) \in R_1} \|\mathbf{b}\|^2 + \|\mathbf{e}\|^2} \tag{7.1}$$

In the above formula, $\mathbf{x}$, $\mathbf{s}$, $\mathbf{b}$ and $\mathbf{e}$ are the received data, the target signal, the background signal and the noise, respectively. The derivation of the above equation is based on the assumption that the target signal, background signal and the noise are independent, and the background and the noise are stationary over all measurements. The expression of SNR in Equation (7.1) has the same standard definition of SNR as given in [6].

7.2.2   Expected Performance Curves (EPC)

In several processing techniques concerned with classification tasks, curves like Receiver Operating Characteristic (ROC) are often used to compare two or more models with respect to various operating points. Instead of the ROC, an Expected Performance Curve (EPC) is used here for the comparion of different algorithms. It has the same meaning as the ROC.

The area under the EPC curve is used to perform the detection and the reduction of clutter. The closer to 1 the area under the EPC curve, the better the performance we have. In general, the ROC curve shows the performance of a particular detector or classifier. However, in this work the EPC curve is considered as measurement on the clutter reduction, i.e., the signal-to-clutter ratio. The EPC curves are calculated according to [118].

The following steps EPC curve can be calculated:
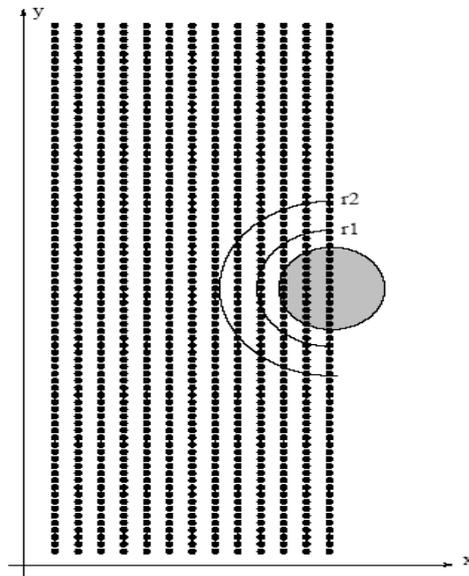
Figure 7.20: Definition of SNR.

1. Estimating the background $\hat{\mathbf{b}}$, and finding test statistics of A-scans in the no target region, and use a threshold of ∞.

   The estimation of the background can be done averaging, or median across all A-scans in the absence of the target. Moreover, choosing moving window, the estimation of background can be formed.

2. Finding test statistics T of A-scans in the test region. If T is less than threshold, this mean that there is no target. "where $T_{bs} = max \left| x - \hat{b} \right|$ the decision that there is target if T_{bs} is larger than a threshold, T_α, where T_α is obtained empirically."

3. Calculating the ratio of A-scans where the target is declared to the total of A-scans tested in the above step. This is P_{fa} or P_d depending on weather the target is present or not.

7.3 Comparison and Discussion

Different signal processing techniques to reduce clutter in GPR data have been presented in this thesis. Our comparison of above techniques are based on SNR and EPC curve. Equation 7.1 has been used to calculate SNR. Table 7.1 gives SNR for all signal processing techniques. By using SNR as criterion to compare above techniques, it can be stated that the JADE based technique appears to have the best

Table 7.1: SNR for different signal processing techniques

Method	SNR
Data	0.039
Mean subtraction method	2.31
SVD (Second algorithm)	5.1
SVD (First algorithm by using threshold)	8.2
Factor Analysis	0.29
Principal Component Analysis	7.3
FastICA	3.8
JADE	18.9
SOBI	10.7
Infomax	4.8
Wavelet Transform (WT)	2.1225
WT with HOS (Kurtosis test)	9.0259
WT with HOS (Skewness test)	6.9523
maximum likelihood algorithm	0.1857

overall performance by a significant margin. EPC curve has been used as the second criterion for the Comparison. The area under EPC curve has been calculated to measure the performance of each technique. Table 7.2 gives the area under EPC, where we conclude that the JADE algorithm provides the best detection performance.

For the SVD technique (Section 3.1), good results for automatic selection of SVD components have been achieved. Ranking of SVD components shows that the background information and the target information contained in the second components respectively. SNR has been improved by using threshold technique, where a new formula to calculate the threshold has been developed. In Sections 3.2 and 3.3 factor analysis and principle component analysis have been studied and applied to GPR data. In all measurements, and in both techniques, the reflection from the surface has been removed or has been at least reduced. Unfortunately, both techniques are insignificant algorithms to eliminate the clutter from GPR data. In other hand, both techniques can be used as pre-processing for measured data if necessary.

For the Likelihood technique the application of a new threshold has deleted the air-ground interface for targets at depths of 5 cm and 10 cm.

Also Independent Component Analysis (ICA) which is one of the most existing new topics in field of signal processing has been applied to the GPR data. Four of the most common ICA methods have been chosen to perform the estimation of the independent component. The main difference between different algorithms consists in the choice of the dependence measure (mutual information, marginal entropies, cumulants, etc.) and the optimization technique (gradient-based or algebraic methods). These algorithms are the extended Infomax, the FastICA, the JADE, and the SOBI. The four algorithms have been studied and compared to each other.

Table 7.2: Area under EPC for different signal processing techniques

Method	Area
Data	0.51
Mean subtraction method	0.77
SVD (Second algorithm)	0.79
SVD (First algorithm by using threshold)	0.81
Factor Analysis	0.61
Principal Component Analysis	0.79
FastICA	0.53
JADE	0.82
SOBI	0.75
Infomax	0.83
Wavelet Transform (WT)	0.51
WT with HOS (Kurtosis test)	0.80
WT with HOS (Skewness test)	0.81
maximum likelihood algorithm	0.79

In all cases, the target signals have been extracted from the GPR data, and most of the clutter signal (cross-talk and air-ground interface) has been removed. SNR and the area under the EPC curve have been used to measure a performance of the reduction of clutter and a comparison between the four algorithm. As can be seen from Table 7.1, both JADE and SOBI methods have shown better performances over Infomax and FastICA. Also the author found the same results by using EPC technique in another data, see [94]. In other hand, the results from Figure 7.21, which shown the EPC curves of the considered algorithms and Table 7.2, indicated that JADE and Infomax methods have shown significant performances over FastICA, but insignificant margin performances over SOBI.

A simple algorithm using wavelet transform to reduce the clutter has been applied. The wavelet transform has been improved by using Higher Order Statistic. In this technique skewness and kurtosis have been used. However kurtosis has shown better performance over skewness.
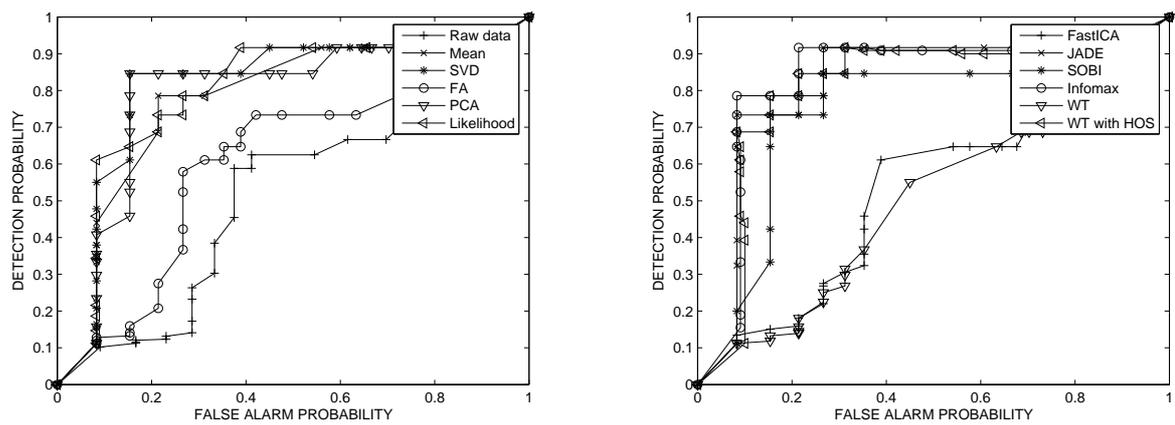
Figure 7.21: The EPC curves for all signal processing techniques.

Chapter 8

Conclusions and Suggestions for Future Work

8.1　Conclusions

This thesis has dealt with the problem of detecting buried non-metallic anti-personnel (AP) land mines using stepped-frequency ground penetrating radar.

Weak reflection signals obtained from buried objects are usually blurred by strong clutter, which mainly comes from flat or rough ground surfaces, underground inhomogeneities, and coupling between the transmitting and receiving antennas. Therefore, reducing or eliminating the clutter signal is of fundamental importance.

The objectives of this work have been to study, develop and compare signal processing techniques as regarding their ability to extract landmine signal from GPR measurements.

In the preceding chapters, several denoising methods for discriminating landmines from clutter using GPR measurements have been discussed in detail. These methods have been divided into three categories: First one are linear data analysis methods such as factor analysis (FA), independent component analysis (ICA) and blind source separation (BSS). The second one is the maximum likelihood detection algorithm. The third and last one is Combining Wavelet Packets with Higher-Order-Statistic.

In Chapter 3, singular value decomposition (SVD), factor analysis (FA) and principle component analysis (PCA) have been applied to GPR measurements. Thresholding has been employed to improve the SVD performance, where a new threshold formula has been developed. In Chapter 5, clutter-reduction algorithms for GPR, based on a maximum likelihood algorithm has been improved. In Chapter 6, adaptive clutter-reduction algorithm has been developed. This algorithm is based on combining the wavelet packets with higher-order-statistic. The algorithm shows promising performance when applied to raw GPR data. The pre-processing is not needed for applying the algorithm. On the other hand, the pre-processing of measured data is absolutely necessary for many other clutter reduction algorithm, as e.g. ICA. The

Independent Component Analysis (ICA) algorithm has proved to show a better performance over all signal processing.

8.2   Suggestions for Future Work

The research presented in this work has advanced the effectiveness of GPR techniques for APL detection. Further investigations are however still needed. The BSS problem, for example has been studied using a linear data model, when the noiseless observation data are assumed to be linear mixtures of the source signals. A non-linear model is however realistic and should be considered in future. On the other hand, most of the presented methods were developed in case of noiseless data and differ from one another in the way they enforce independence. Therefore, further investigations are needed to establish algorithms that perform noisy ICA. Recently, solutions based on higher order statistics (HOS) only, quasi whitening, and mixture of Gaussian models have been proposed.

Linear ICA is suitable when it is reasonable to assume that the observations have been generated by a linear mixing from some independent source signals. However, in general and for many practical problems, mixtures are more likely to be nonlinear or subject to some kind of nonlinear distortions due to sensory or environmental limitations. Extending the existing validity of the existing algorithms to nonlinear ICA (NLICA) is not straightforward. There have been few initial attempts in the field. Feasible solutions have been disappointing due to problems of uniqueness and/or computational complexities. ICA and NLICA may be approached through non-linear PCA (NLPCA) and mixture models, and some work has demonstrated tentative links.

In blind separation of dependent sources, the independence assumption alone is not valid anymore. For achieving blind separation in this case, one must model the dependences between the sources in one way or another. Modeling the dependences between the sources is often useful because the dependence information can be utilized to achieve better estimation results. If the dependences between the sources can be modeled sufficiently well, such a model can sometimes be used to remove or suppress the dependences so that the remaining signals become easier to separate adequately using standard BSS methods. There does not yet exist any uniform, generally accepted theoretical basis for handling statistical dependence. Instead, diverse approaches have been developed during the last years for handling this problem.

Combining wavelet packets with higher-order-statistics algorithms seems to be promising when applied to raw GPR data which have not been pre-processed for clutter removal. The thresholding method described here depends on the skewness or kurtosis. The sensitivity of this algorithm to various wavelet families should be further investigated in future.

Bibliography

[1] Magnus Lundberg. Electro-optical land mine detection and contributions to iterative decoding and power estimation. Technical report no. 373l., licenciate thesis, Department of Signals and Systems,School of Electrical and Computer Engineering, chalmers university of technology, Sweden, 2001.

[2] Anti-personnel landmines - friend or foe?. *ICRC Publications,Geneva*, March 1996.

[3] Ilaria Bottigliero. *120 Million Landmines Deployed Worldwide: Fact or Fiction.* Pen and Sword Books Ltd, Barnsley, South Yorkshire, UK, 2000.

[4] J.W.Brooks. *The Detection of Buried Non-Metallic Anti-Personnel Land Mines.* Ph. d. dissertation,, The University of Alabama in Huntsville, August 2000.

[5] MacDonald J., Lockwood J.R., McFee J.E., Altshuler T., Broach J.T., L. Carin, Harmon R.S., Rappaport C., Scott W.R., and Weaver R. Alternatives for landmine detection. Technical report, RAND (http://www.rand.org/ publications/MR/MR1608/MR1608.appg.pdf), 2003.

[6] J. J. Daniels. *Surface Penetrating Radar.* Radar sonar navigation and avionics series, The Institution of Electrical Engineers, London, UK, 1996.

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

[8] C. P. Gooneratne, S. C. Mukhopahyay, and G. Sen Gupta, editors. *A review of sensing technologies for landmine detection : unmanned vehicle based approach.*

2nd International Conference on Autonomous Robots and Agents,Palmerston North, New Zealand, December 13-15, 2004.

[9] International campaign to ban landmines. http://www.icbl.org/.

[10] United States Commitment to Humanitarian Demining. http://www.state.gov/ t/pm/rls/rpt/hk/2001/6961.htm.

[11] Ackenhusen J. G., Quentin A. H., C. King, and Wright J. A. Detection of mines and minefields. *Technical Report 440000-217-T, Infrared Information Analysis Center, Veridian Systems, Ann Arbor, MI.*, 2001.

[12] Hakan Brunzell. *Signal processing techniques for detection of buried landmines using ground penetrating radar*. PhD thesis, Department of Signals and Systems, Chalmers University of Technology, 1998.

[13] Don C.G. Using acoustic impulses to identify a buried nonmetallic object. *in 127th Meeting of The Acoustical Society of America, (M.I.T)*.

[14] W.R. Scott, Jr. Schroeder, C.T. Martin J.S., and Larson G.D. Investigation of a technique that uses both elastic and electromagnetic waves to detect buried land mines. *Proceedings of the AP2000 Millennium Conference an Antennas and Propagation, Davos, Switzerland*, April 2000.

[15] Defence Research and Development Canada. Landmine detection, http://www.dres.dnd.ca/researchtech/products/mileng_products/rd20001_ildp/ index_e.html.

[16] Mikael Georgson, Lars Pettersson, Stefan Sjökvist, and Magnus Uppsäll. Mine detection using infrared imaging technique. *Int. Conf. MINE'99, Firenze, Proceedings*, 1999.

[17] Moody KA and LeVasseur JP. Current and emerging technologies for use in a hand-held mine detector. *The. Royal Military College of Canada*, 2000. http://www.rmc.ca/academic/gradrech/minedetector.pdf.

[18] M. Nolte, A. Privalov, J. Altmann, V. Anferov, and F. Fujara. ^{1}H -^{14}N cross-relaxation in trinitrotoluene: A step toward improved landmine detection. *J. Phys. D: Appl. Phys*, (35):939–942, 2002.

[19] M.D. Rowe and Smith J.A.S. Mine detection by nuclear quadrupole resonance. *In Proceedings of the EUREL International Conference on The Detection of Abandoned Land Mines. Edinburgh, UK,*, (431):62–66, October 1996.

[20] Buan Sen Tan. TEM horn antennas for landmine detection applications. thesis report, Queensland University, May 2002.

[21] Benson R. C., R. A. Glaccum, and M. R. Noel. Geophysical techniques for sensing buried wastes and waste migration. *National Water Well Association.*, 1983. http://www.hanford.gov/dqo/project/level5/burwm/burwm.pdf.

[22] Imse J.P. and E.N. Levine. Conventional and state-of-the-art geophysical techniques for fracture detection. *Proceedings Second Annual Eastern Groundwater Conference*, pages 261–278, 1985.

[23] Wright D.L., Olhoeft G.R., and Watts R.D. Ground-penetrating radar studies on cape cod, in conference on surface and borehole geophysical methods in ground water investigations. *Proceedings: U.S. Environmental Protection Agency and National Water Well Association, San Antonio, TX,*, pages 666–680, 1984.

[24] Pradip Kumar Mukhopadhyay. *Three-dimensional Borehole Radar Imaging*. PhD thesis, University of Cape Town, Department of Electrical Engineering, 2005.

[25] Sheriff R.E. and L.P. Geldart. *Exploration Seismology*. Cambridge University Press (New York), 1982.

[26] Wilkinson A.J., R.T. Lord, and M.R. Inggs. Stepped-frequency processing by reconstruction of target reflectivity spectrum. *Proceedings of the 1998 IEEE South African Symposium on Communications and Signal Processing*, page 101104, 1998.

[27] Langman A. and M.R. Inggs. Pulse versus stepped frequency continuous wave modulation for ground penetrating radar. *Proceedings of IGARSS*, 2001.

[28] Reynolds J.M. *An introduction to applied and environmental geophysics.* John Wiley & Sons Ltd, Chichester & London, 796 pp, 1997.

[29] Daniels J. J. Fundamentals of ground penetrating radar. *Proceedings of the Symposium on the Application of Geophysics to Engineering and Environmental Problems. Colorado School of Mines, Golden, Colorado*, pages 62–142, 1989.

[30] Aapo Hyvarinen, J. Karhunen, and E. Oja. *Independent Component Analysis.* John Wiley, New York, 2001.

[31] Cichocki A. and Amari S. *Adaptive Blind Signal and Image Processing: Learning Algorithms and Applications.* John Wiley & Sons, West Sussex, UK, 2003.

[32] N. Thirion, J. MARS, and J. L. BOELLE. Separation of seismic signals: A new concept based on a blind algorithm. In *in Signal Processing VIII, Theories and Application*, pages 85– 88. Elsevier,Triest, Italy, September 1996.

[33] G. H. Golub and C. F. Van Loan. *Matrix Computations.* Baltimore, MD, USA, second edition, 1989.

[34] H. H. Harman. *Modern Factor Analysis.* University of Chicago Press, 2nd edition edition, 1967.

[35] E. Oja. Principal components,minor components, and linear neural networks. *Neural Networks*, 5:927–935, 1992.

[36] C. Jutten and J. Hérault. Blind separation of sources, part i: An adaptive algorithm based on neuromimetic architecture. *Signal Processing*, 24:1–10, 1991.

[37] A. K. Barros. The independence assumption: Dependent component analysis. In M. Girolami, editor, *in Advances in Independent Component Analysis.* Springer, 2000.

[38] A. Cichocki, R. Zdunek, and S. Amari. New algorithms for non-negative matrix factorization in applications to blind source separation. *2006 IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP2006*, pages 14–19, Toulouse, France, May 2006.

[39] P.J. Huber. Exploratory projection pursuit. *J. of the American Statistical Association*, 82(397):249–266, 1987.

[40] V. D. Vrabie, J. I. Mars, and J.-L. Lacoume. Modified singular value decomposition by means of independent component analysis. *Signal Process.*, 84(3):645–652, 2004.

[41] P. Dewilde and E.F. Deprettere. Singular value decomposition: An introduction. In SVD and signal processing, editors, *In E.F. Deprettere*. North Holland, New York, 1988.

[42] Daniel P.Berrar, Werner Dubitzky, and Martin Granzow. *A Practical Approach to Microarray Data Analysis*, chapter 5, pages 91–109. Kluwer Academic Publishers, 2003.

[43] P. Comon and T. Kailath. An array processing technique using the first principal component. In *First International Workshop on SVD and Signal Processing*, September 1987. Extended version published in: SVD and Signal Processing, E.F. Deprettere editor, North Holland, 1988, 301–316.

[44] Richard J. Vaccaro, editor. *SVD and Signal Processing, II. Algorithms, Analysis and Applications*, 1991.

[45] William K. Pratt. *Digital Image Processing.* John Wiley & Sons Inc., second edition, 1991.

[46] Uri Barkan. Dynamic structures of neuronal networks. Master's thesis, The Department of Physics, Tel-Aviv University, February 2005.

[47] Michael E. Wall, Andreas Rechtsteiner, and Luis M. Rocha. *Singular value decomposition and principal component analysis In: A Practical Approach to Microarray Data Analysis*, chapter 5, pages 91–109. Kluwer Academic Publishers, Boston, MA, 2003.

[48] Fawzy Abujarad, Galal Nadimy, and Abbas Omar. Clutter reduction and detection of landmine objects in ground penetrating radar data using singular value decomposition (SVD). *Proc. 3rd international workshop on advanced ground-penetrating radar*, pages 37–41, May 2005.

[49] Richard A. Reyment and K. G. Joreskog. *Applied Factor Analysis in the Natural Sciences.* 1993.

[50] D. B. Rubin and D. T. Thayer. EM Algorithms for ML Factor Analysis. *Psychometrika*, 47(1):69–76, 1982.

[51] Z. Ghahramani and G. Hinton. The em algorithm for mixtures of factor analyzers. Technical report CRG-TR-96-1, Department of Computer Science, University of Toronto, 1997.

[52] L. Saul and M. Rahim. Maximum likelihood and minimum classification error factor analysis for automatic speech recognition. *IEEE Transactions on Speech and Audio Processing*, 8(2):115–125, 1999.

[53] F. B. Nielsen. Variational approach to factor analysis and related models. Master's thesis, Informatics and Mathematical Modelling, Technical University of Denmark, DTU, Richard Petersens Plads, Building 321, DK-2800 Kgs. Lyngby, 2004. Supervised by Professor Lars Kai Hansen and Associate Professor Ole Winther.

[54] Michael E. Tipping and Christopher M. Bishop. Mixtures of probabilistic principal component analysers. *Neural Computation*, 11(2):443–482, 1999.

[55] S.-H. Yu, R. K. Mehra, and T. R. Witten. Automatic mine detection based on ground penetrating radar. In J. Broach A. C. Dubey, J. F. Harvey and R. E. Duncan, editors, *in Proc. SPIE, on Detection and Remediation Technologies for Mines and Mine-Like Targets IV*, volume 3710, pages 961–972, 1999.

[56] B. Karlsen, J. Larsen, H.B.D. Sørensen, and K.B. Jakobsen. Comparison of PCA and ICA based clutter reduction in GPR systems for anti-personal landmine detection. In *Proceedings of the 11th IEEE Signal Processing Workshop*

on Statistical Signal Processing, Orchid Country Club, Singapore, pages 146 – 149, Aug 2001. P 1664.

[57] Fawzy Abujarad and Abbas Omar. GPR data processing using the component-separation methods PCA and ICA. *IEEE International Workshop on Imaging Systems and Techniques, Minori, Italy*, April 2006.

[58] J. Shlens. A tutorial on principal component analysis: Derivation, discussion and singular value decomposition, 2001.

[59] C. Fancourt and J. Principe. Competitive principal component analysis for locally stationary time series, 1998.

[60] Fawzy Abujarad and Abbas Omar. Factor and principle component analysis for automatic landmine detection based on ground penetrating radar. *German Microwave Conference (GeMiC) 2006, Karlsruhe, Germany*, March 2006.

[61] C. Jutten, J. Hérault, P. Comon, and E.Sorouchiary. Blind separation of sources, part I, II and III. *Signal Processing*, 24:1–29, 1991.

[62] P. Comon and P. Chevalier. Source separation: Models, concepts, algorithms and performance. In S. Haykin, editor, *Unsupervised Adaptive Filtering, Vol. I, Blind Source Separation*, Series on Adaptive and learning systems for communications signal processing and control, pages 191–236. Wiley, 2000.

[63] Jaakko Särelä. *Exploratory Source Separation in Biomedical Systems*. PhD thesis, Department of Computer Science and Engineering, Helsinki University of Technology, Espoo,Finland, 2004.

[64] J. Cardoso. Blind signal separation: statistical principles. In R.-W. Liu and L. Tong, editors, *Proceedings of the IEEE, Special issue on blind identification and estimation*, volume 9, pages 2009–2025, 1998.

[65] James V. Stone. *Independent Component Analysis: A Tutorial Introduction.* MIT Press/Bradford Books, Cambridge, MA, 1st edition, September 2004.

[66] Aapo Hyvarinen. Survey on independent component analysis. *Neural Computing Surveys*, 2:94–128, 1999e.

[67] Aapo Hyvarinen and Erki Oja. Independent component analysis: A tutorial. Technical report, Laboratory of Computer and Information Science, Helsinki University of Technology,1999.http://www.cis.hut.fi/aapo/papers/ IJCNN99_tutorialweb/.

[68] James V. Stone and John Porrill. Independent component analysis and projection pursuit: A tutorial introduction, 1998.

[69] A. Morgenstjerne, B. Karlsen, H. B. Søren-sen J. Larsen, and K. B. Jakobsen. A comparative and combined study of EMIS and GPR detectors by the use of independent component analysis. *Proceedings of the 2005 Detection and Remediation Technologies for Mines and Mine-Like Targets, AeroSense 2005,SPIE*, 2005.

[70] S. Amari, A. Cichocki, and H. H. Yang. A new learning algorithm for blind signal separation. In David S. Touretzky, Michael C. Mozer, and Michael E. Hasselmo, editors, *Advances in Neural Information Processing Systems*, volume 8, pages 757–763. The MIT Press, 1996.

[71] M.C. Jones and R. Sibson. What is projection pursuit? *Journal of the Royal Statistical Society, A*, 150:1–36, 1987.

[72] Aapo Hyvarinen. New approximations of differential entropy for independent component analysis and projection pursuit. *Advances in Neural Information Processing Systems*, 10:273–279, 1998. MIT Press.

[73] Aapo Hyvarinen and E. Oja. Independent component analysis: Algorithms and applications. *Neural Networks*, 13:(4–5):411–430, 2000.

[74] Alan Oursland, Judah De Paula, and Nasim Mahmood. Case studies of independent component analysis.available at http://www.oursland.net/ tutorials/ica/ica-report.pdf.

[75] THU Hoang, DINH-TUAN Pham, and Claudette AH-SOONL. Mutual information based independent component analysis of array cgh data,2005. http://www.math-info.univ-paris5.fr/map5/publis/PUBLIS05/2005-17.pdf.

[76] D.-T. Pham, P. Garrat, and C. Jutten. Separation of a mixture of independent sources through a maximum likelihood approach. *In Proc. EUSIPCO*, pages 771–774, 1992.

[77] Bogdan Matei. A review of independent component analysis techniques. *Electrical and Computer Engineering Department Rutgers University, available at.* http://www.caip.rutgers.edu/riul/research/tutorials/tutorialica.pdf.

[78] P. Comon. Independent component analysis, a new concept?. *Signal Processing, Elsevier*, 6(3):287–314, April 1994.

[79] J. F. Cardoso and A. Souloumiac. Blind beamforming for non-gaussian signals. *IEE Proceeding part F*, 140(6):362–370, 1993.

[80] J. Cardoso and B. Labeld. Equivariant adaptative source separation. *IEEE Transactions on Signal Processing*, 44(12):3017–3030, 1996.

[81] Anthony J. Bell and Terrence J. Sejnowski. An information-maximization approach to blind separation and blind deconvolution. *Neural Computation*, 7(6):1129–1159, 1995.

[82] Aapo Hyvarinen and Erki Oja. A fast fixed-point algorithm for independent component analysis. *Neural Computation*, pages 9:1483–1492, 1997.

[83] A. Hyvärinen. Fast and robust fixed-point algorithms for independent component analysis. *IEEE Transactions on Neural Networks*, 10(3):626–634, 1999.

[84] D. G. Luenberger. *Optimization by Vector Space Methods.* John Wiley & Sons, 1969.

[85] N. Delfosse and P. Loubaton. Adaptive blind separation of independent sources: a deflation approach. *Signal Processing*, 45(I):59–83, 1995.

[86] An xing Zhao, Yan sheng Jiang, and Wen bing Wang. Exploring independent component analysis for GPR signal process. *Progress In Electromagnetics Research Symposium, Hangzhou, China,*, pages 22–26, August 2005.

[87] Fawzy Abujarad and Abbas Omar. GPR detection of landmine by FastICA. *Proc. 11th International Conference on Ground Penetrating Radar, Columbus,Ohio, USA*, June 19-22, 2006.

[88] M. Gaeta and J.-L. Lacoume. Source separation without prior knowledge: The maximum likelihood solution. *Proc. EUSIPO*, pages 621–624, 1990.

[89] Pham D. T. and Garrat P. Blind separation of mixture of independent sources through a quasi-maximum likelihood approach. *IEEE Trans. on Signal Processing*, 45(7):1712–1725, 1997.

[90] Pearlmutter BA and Parra LC. A context-sensitive generalization of ICA. *International Conference on Neural Information Processing,Hong Kong*, September 1996.

[91] J.F. Cardoso. Infomax and maximum likelihood for blind source separation. *IEEE Signal Processing Letters*, 4(4):112–114, April 1997.

[92] S. Amari and J. Cardoso. Blind source separation semiparametric statistical approach. *IEEE Trans. Signal Processing*, 45(11):2692–2700, 1997.

[93] Shun ichi Amari. Natural gradient works efficiently in learning. *Neural Computation*, 10(2):251–276, 1998.

[94] Fawzy Abujarad and Abbas Omar. Comparison of independent-component-analysis (ICA) algorithms for GPR detection of non-metallic land mines. *In Proc. SPIE Image and Signal Processing for Remote Sensing XI, Stockholm, Sweden*, 6365, September 11-14, 2006.

[95] A. Belouchrani, K. A. Meraim, J. F. Cardoso, and E. Moulines. Second-order blind separation of correlated sources. *In Proc. Int. Conference on Digital Sig. Processing,Cyprus*, pages 346–351, 1993.

[96] A. Belouchrani, M.G. Amin, and K. Abed-Meraim. Direction finding in correlated noise fields based on joint block-diagonalization of spatio-temporal correlation matrices. *IEEE Signal Processing Letters*, 4(9), September 1997.

[97] A. Belouchrani, K. A. Meraim, J. F. Cardoso, and E. Moulines. A blind source separation technique using second-order statistics. *IEEE Trans. Sig. Proc.*, 45(2):434–444, 1997.

[98] Jean-François Cardoso and Antoine Souloumiac. Jacobi angles for simultaneous diagonalization. *SIAM Journal on Matrix Analysis and Applications*, 17(1):161–164, 1996.

[99] P. Comon. Independent component analysis, and the diagonalization of symmetric tensors. In H. Dedieu, editor, *European Conference on Circuit Theory and Design – ECCTD*, pages 185–190, Davos, Aug 30-Sept 3, 1993. Elsevier.

[100] P. Comon. Tensor diagonalization, a useful tool in signal processing. In M. Blanke and T. Soderstrom, editors, *IFAC-SYSID, 10th IFAC Symposium on System Identification*, volume 1, pages 77–82, Copenhagen, Denmark, July 4-6 1994. invited session.

[101] B.D. Flury and B.E. Neuenschwander. Simultaneous diagonalization algorithms with applications in multivariate statistics. In In R.V.M. Zahar, editor, *Approximation and Computation,BirkhAuser, Basel*, volume 9, page 179205, 1995.

[102] G. Hori. A new approach to joint diagonalization. In *Proceedings of Second International Workshop on Independent Component Analysis and Blind Signal Separation (ICA2000)*, pages 151–155, Helsinki, Finland, June 2000. Helsinki University of Technology, IEEE Signal Processing Society.

[103] A.J. van der Veen. Joint diagonalization via subspace fitting techniques. In *In Proc. of IEEE ICASSP 2001*, Salk Lake, USA, 2001.

[104] M. Wax and J. Sheinvald. A least-squares approach to joint diagonalization. *IEEE Signal Processing Letters*, 4(2):5253, February 1997.

[105] J. F. Cardoso and B. H. Laheld. Equivariant adaptive source separation. *IEEE Transactions on Signal Processing*, 44(12):3017–3030, Dec 1996.

[106] J.F. Cardoso. High-order contrasts for independent component analysis. *Neural Computation*, 11(1):157–192, 1999.

[107] Ch. Ziegaus and E.W. Lang. A neural implementation of the JADE algorithm using higher-order neurons. *Neurocomputing 56c*, pages 79–100, 2004.

[108] Kailath and Poor. Detection of stochastic processes. *IEEETIT: IEEE Transactions on Information Theory*, 44, 1998.

[109] A. Wald. *Sequential Analysis*. New York.Wiley, 1947.

[110] A. Wald. *Statistical Decision Functions*. Wiley, 1950.

[111] D. Middleton. Statistical theory of signal detection. *IRE Trans. Prof. Group Inform. Theory*, vol. PGIT-3:26–52, 1954.

[112] W. W. Peterson, T. G.Birdsall, and W. C. Fox. The theory of signal detectability. *Trans. Prof. Group Inform. Theory*, PGIT-4, 1954.

[113] Steven M. Kay. *Fundamentals of Statistical Signal Processing: Estimation Theory*. Prentice Hall, NJ, 1993.

[114] Chuhong Fei, D. Kundur, and R. H. Kwong. A hypothesis testing approach for achieving semi-fragility in multimedia authentication. *Proc. SPIE*, 6072, 2006.

[115] Brunzell Hakan. Clutter reduction and object detection in surface penetrating radar. *IEE Proceedings*, pages 688–691, Oct 1997.

[116] A.O.Hero. *CRC Handbook on Signal Processing*, chapter Detection and Classification. CRC Press, 1997.

[117] Steven M. Kay. *Fundamentals of Statistical Signal Processing, Detection Theory*. Prentice-Hall, 1998.

[118] A. M. Zoubir, I. J. Chant, C. L. Brown, B. Barkat, and C. Abeynayake. Signal processing techniques for landmine detection using impulse ground penetrating radar. *IEEE Sensors Journal*, 2(1):41–51, February 2002.

[119] Stephen W. Kercel, Marvin B. Klein, and Bruno Pouet. Wavelet and wavelet-packet analysis of lamb wave signatures in real-time instrumentation. *IEEE Mountain Workshop on Soft Computing in Industrial Applications*, pages 25–27, June 2001.

[120] Fawzy Abujarad, Jöstingmeier, and Abbas Omar. Clutter removal for landmine using different signal processing techniques. *Proc. tenth international conference on ground penetrating radar*, pages 697–700, June 2004.

[121] Matlab wavelet toolbox, available online at:http://www.mathworks.com/access/helpdesk/help/pdf_doc/wavelet/wavelet_ug.pdf.

[122] C. Nikias and J. Mendel. Signal processing with higher-order statistics. *IEEE Trans. Signal Processing*, 41:10–38, Jan 1993.

[123] C. L. Nikias and A. P. Petropulu. *Higher-Order Spectra Analysis: A Nonlinear Signal Processing Framework.* Prentice-Hall, 1993.

[124] A. Sankarasubramanian and K. Srinivasan. Investigation and comparison of sampling properties of l-moments and conventional moments. *Journal of Hydrology*, 218(1-3):13–34, 1999.

[125] Tae Hwan Kim and Halbert White. On more robust estimation of skewness and kurtosis. Technical report, Department of Economics, UC San Diego, September 2004.

[126] Corinna Cortes and Mehryar Mohri. Confidence intervals for the area under the ROC curve. In Lawrence K. Saul, Yair Weiss, and Léon Bottou, editors, *Advances in Neural Information Processing Systems 17*, pages 305–312. MIT Press, Cambridge, MA, 2005.

[127] Kendall and M.G.and Stuart A.and Ord J.K. *Kendall's Advanced Theory of Statistics*, volume 1 of *Distribution Theory*. New York: Oxford University Press, 6th edition, 1987.

[128] Philippe Ravier and Pierre-Olivier Amblard. Combining an adapted wavelet analysis with fourth-order statistics for transient detection. *Signal Process.*, 70(2):115–128, 1998.

[129] Fawzy Abujarad, Galal Nadimy, and Abbas Omar. Combining wavelet packets with higher-order-statistic for GPR detection of non-metallic anti-personnel land mines. *Proc. SPIE Image and Signal Processing for Remote Sensing XI, Brugge, Belgium*, 5982:380–390, September 2005.

[130] Fawzy Abujarad, Galal Nadimy, and Abbas Omar. Wavelet packets for GPR detection of non-metallic anti-personnel land mines based on higher-order-statistic. *Proc. 3rd international workshop on advanced ground-penetrating radar*, pages 21–24, May 2005.

[131] G. Alli, A. Bicci, G. De Pasquale, G. Manacorda, G. Pinelli, and S. Sensani. Fully polarised 3D-SAR GPR for the recognition of hidden mines. *Proceedings of the Mine Identification Novelties Euroconference 1999, Florence*, October 1999.

[132] H. Brunzell. Detection of shallowly buried objects using impulse radar. Technical report, Department of Applied Electronics, Chalmers University of Technology, S-412 96 Goteborg, Sweden, May,13 1996.