Material Type and Soil Moisture Determination Based on Radar Response

by

Enrico A. Mattei Lago

A thesis submitted in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE
in
ELECTRICAL ENGINEERING

UNIVERSITY OF PUERTO RICO
MAYAGÜEZ CAMPUS
2006

Approved by:

_____	_____
Miguel Velez Reyes, PhD Member, Graduate Committee	Date
_____	_____
Domingo Rodriguez, PhD Member, Graduate Committee	Date
_____	_____
Hamed Parsiani, PhD President, Graduate Committee	Date
_____	_____
Eric Harmsen, PhD Representative of Graduate Studies	Date
_____	_____
Isidoro Couvertier, PhD Chairperson of the Department	Date
_____	_____
José Mari Mutt, PhD Director of Graduate Studies	Date

ABSTRACT

Ground Penetrating Radar, *GPR*, is a widely used tool for probing the sub-surface in such applications as pavement monitoring, humanitarian de-mining, and exploratory geophysics, among others. Its advantage lies in that it provides information about the composition of the sub-surface in a fast, non-invasive fashion.

The work presented herein relies on the application of *GPR* to the problems of material type, e.g., soil types, and soil moisture determination. The task is attacked from a systems perspective to obtain a signature for each material at different moisture levels. This signature has discriminatory characteristics that are exploited with the use of a Neural Network, *NN*, which is able to distinguish between different materials and moisture levels with reasonable accuracy.

RESUMEN

Ground Penetrating Radar, *GPR*, se ha convertido en una herramiena muy utilizada para investigar la estructura del sub-suelo. Aplicaciones comunes son monitoreo de pavimentos, detectar minas que permanecen escondidas bajo tierra desde la segunda guerra mundial, en el campo de la geofísica, etc. La ventaja del *GPR* es su naturaleza para caracterizar la estructura del sub-suelo de una manera rapida y no-invasiva.

El trabajo aquí presentado, se basa en la aplicación del *GPR* en los problemas de identificación del tipo de suelo, y determinación de su contenido de humedad. El problema es atacado desde una perspectiva de sistemas, donde el objetivo es encontrar una caracteristica distinctiva del tipo de suelo, y su contenido de humedad. Esta característica es usada en un sistema basado en una red neural, la cual logra la discriminación deseada con una precisión razonable.

.

To my family . . .

ACKNOWLEDGEMENTS

This section is dedicated to those who helped in making this thesis a reality. During two and a half years, several people were instrumental to this work; Dr. Hamed Parsiani for your guidance, dedication, tolerance and counsel when I was distracted, and most importantly, for being a grand person, I thank you. Mairim, for developing a great GUI and being an even greater friend. Allen, Alex, and Adolfo for working hard in all the experiments.

I would also like to express my appreciation to Dr. Miguel Vélez, and Dr. Domingo Rodriguez for being part of my committee, and for being excellent *teachers*.

I would like to mention two people who witnessed, and helped me get through the tough times. Alexey, you have become a true friend. Marie, thank you for your advice and for making every day at LARSIP a fun day.

Finally, thanks to NASA TCESS grant no. NCC5-252 for providing the financial support for this project.

Table of Contents

Table List

Tables **Page**

Figure List

Figures **Page**

1 INTRODUCTION

1.1 Motivation

Knowledge of accurate soil moisture information is vital in the fields of agriculture, hydrology, and atmospheric sciences, civil engineering, among others. In atmospheric sciences, for instance, soil moisture information is an important factor that affects drastically climate predictions. In agriculture, soil moisture controls the scheduling of crop irrigation; poor moisture information is thus, in part, responsible for crop health and the success of precision agriculture programs.

Different methods have been used over the years to measure soil moisture. These methods, such as, gravimetric sampling, Time Domain Reflectometry (TDR), and Theta Probes are all invasive methods that require the insertion of probes or the removal of soil samples for laboratory analysis. Satellite remote sensing has been used for this purpose, but passive sensors provide spatial resolution, in the microwave region, that is too low for more spatial accuracy requirements. Active, microwave, sensors provide much better spatial resolution than passive sensors, but current algorithms are aimed at scenes with little or no vegetation cover. The remote sensing of soil moisture via satellite requires validation and calibration. The validation aspect depends on large moisture data from large tracks of land.

Current techniques to determine the soil moisture content using GPR are based on the calculation of the dielectric constant, which implies that a petrophysical relationship has to be

used to estimate the moisture given the dielectric constant. The problem with petropysical relationships is that they are site-specific, which means that they are only correct for the field and soil under which they were developed. So, there is a pending need for methods that do not rely on these models, and are equally applicable across any field. The word petrophysics, as defined by Schlumberger, refers to the study of the physical properties of rock types. These properties include, porosity, pressure, permeability, and saturation. With knowledge of these properties, the rock type can be inferred, as well as the presence of oil and gas. In the context of this thesis, the word petrophysical relationship is used to refer to an empirically derived equation that, with knowledge of certain properties of the soil, is used to infer its moisture content.

In this research, a Ground Penetrating Radar (*GPR*) is used to determine the soil moisture over land, with no vegetation cover. A *GPR* can measure electromagnetic reflections produced by variations in dielectric properties at the interface between different propagating mediums. It can provide high resolution, non-invasive, continuous spatial soil moisture measurements consistent with invasive measuring devices.

1.2 Literature Review

Several researchers have delved into the problem of soil moisture determination. Using invasive techniques (TDR, etc.) or non-invasive techniques (*GPR*, Satellite), this is still an open problem. Literature concerning, moisture determination with *GPR* is scarce, as most

publications concern moisture determination based on TDR measurements. Soil moisture determination using *GPR* has recently become a more active area of research.

Time Domain Reflectometry is a popular method to determine the moisture content thanks to the contributions of Topp, Davis, Annan, Dasberg, Dalton and others. The widespread use of TDR came after Topp developed a non-linear relationship between the volumetric water content and the dielectric constant [12]. This relationship is independent of soil type, salt content, density and temperature and has been used extensively to determine the moisture content of any field. TDR calculates the dielectric constant using the velocity of the electromagnetic wave through a waveguide. Soil moisture is obtained using the TDR calculated dielectric constant and equations such as those developed in [12].

Parallel transmission lines were found to be useful to measure moisture content in [1], [2] and [3], resulting in accurate measurements when compared against moisture content obtained by gravimetric sampling in a laboratory environment. These results were validated in [4] with experiments in a cornfield using three different variations of parallel transmission lines.

Chanzy, Lambot, Rubin and others have studied the use of GPR to measure soil moisture. The methods described in the literature are mainly:

- **Based on reflected wave velocity**

  The velocity of the reflected wave is calculated. This velocity is used to calculate the dielectric constant and then the equations in [12] are used to estimate the moisture content. This method can be divided into single offset reflection method and multi-offset reflection method. In the single-offset method, the distance between the transmitting and receiving antennas is not varied whereas in the multi-offset method, the distance between the antennas is varied.

- **Based on ground wave velocity**

  The ground wave is the part of the radiated energy that travels between the transmitter and receiver through the top layer of the soil [6]. This method involves a multi-offset GPR setup, in which the distance between the transmitter and receiver is continuously increased. Moisture content measured in this way is presented in [8], [9], and [10].

- **Based on Surface Reflection**

  This method relies on the amplitude of the reflected (relative to the amplitude of the wave reflected from a “perfect” reflector to obtain the normal incidence reflection coefficient. From the reflection coefficient, the dielectric constant of the soil is obtained. Later, dielectric constant is converted into soil moisture by means of a petrophysical relationship. As presented in [17], soil moisture can be adequately obtained using the reflection coefficient if appropriate calibration is performed.

- **Based on electromagnetic inversion**

  Model-based methods are becoming popular due to the accuracy of the estimates obtained. A forward model describing the physics of the problem is used to generate simulated responses. The forward model is subsequently inverted to get the estimates of the parameters, which most likely produced the observed response. This kind of method is more accurate because it is taking into consideration the actual physics of the problem in order to solve it. Its main drawbacks are that the model used must be very accurate, and requires lots of computing resources. If it is desired to obtain rapid soil moisture information, this method quickly becomes unfeasible. Recent work using this method can be found in [11].

1.3 Summary of Following Chapters

We first develop the necessary background theory, in Chapter 2. The third chapter presents the method of the Material Characteristics in Frequency Domain, MCFD. Chapter 4 gives a review of the basic theory of neural networks used in this thesis. The development of a Graphical User Interface is presented in chapter 5. The sixth chapter presents the experimental results and their discussion. The topic of clutter reduction is touched in chapter 7, conclusions and future work are given in Chapter 8.

2 THEORETICAL BACKGROUND

Abstract This chapter presents the underlying theory concerning electromagnetic wave propagation, and ground penetrating radar fundamentals. It also presents in some detail the current techniques that are used to estimate soil moisture content using *GPR*. We begin by introducing *GPR* concepts, and describing the equipment used in this research. We then go on to review the dielectric properties of materials, followed by a description, of the state of the art, in soil moisture determination using *GPR*.

2.1 Ground Penetrating Radar Concepts

Surface Penetrating Radar, commonly known as Ground Penetrating Radar (*GPR*), refers to a radar system that is specifically designed to probe the subsurface [21]. *GPR* works by transmitting electromagnetic energy into the ground and recording the portion of that energy which is reflected by the subsurface. Reflection of the electromagnetic wave occurs when there are sufficiently sharp differences in the dielectric properties of the propagating medium. For example, if the signal is transmitted into a multilayer medium consisting of N layers, with each layer being a different material (e.g., different dielectric properties), then the signal that was originally transmitted will be reflected N times. This phenomenon is what allows GPR to be used in many different applications, ranging from buried target detection to the determination of the material type and moisture content; the later two being the subject of this thesis.

2.1.1 *GPR System Characteristics*

A *GPR* system consists of a transmitter, a receiver, a processor, and a data display. The main characteristics of a *GPR* system are its range, depth resolution, and plan resolution. Also, the clutter plays an important factor in the processing of *GPR* data. A description of these characteristics follows [21],

The range of a *GPR* system is the distance that the electromagnetic wave travels in the subsurface. It depends mainly on the losses that the signal suffers when it travels from the

transmitter to the receiver. The most significant losses are material, spreading, and scattering losses [21].

Depth resolution refers to the ability of the *GPR* system to distinguish between different objects, which may be close to each other (at a different depth). Depth resolution depends on the bandwidth of the received signal in order to discriminate between the different targets. Convolution between the responses of different targets occurs when these are close to one another. In this situation, de-convolution is employed to separate the responses, to be subjected to further processing.

When targets are located at the same depth, and it is needed to distinguish between more than one, then plan resolution becomes a factor. Plan resolution depends on the antenna being used, as well as on the signal processing [21].

Clutter, as defined by Daniels in [21], are signals that are uncorrelated to the target scattering characteristics but occur in the same sample-time window and have similar spectral characteristics to the target wavelet.

2.1.2 *Electromagnetic Theory*

Natural materials such as soils have certain electric and magnetic properties, which control the speed of electromagnetic waves that travels through the material. The ability of a material to move charge and store energy is dictated by these properties. A few definitions are in place,

- Conductivity is a measure of the ability of a material to transport an electrical charge. Conductivity is a complex function of frequency, and is expressed as

$$\sigma(f) = \sigma' + j\sigma'' \qquad \textbf{2.1}$$

Where the frequency dependence is given by the imaginary part, σ'', which accounts for conduction losses.

- Permittivity describes how an applied electric field affects and is affected by the propagating medium. It is determined by the ability of a material to polarize when an electric field is applied. The reaction of a material to an applied electric field depends on the frequency of the applied field and hence, permittivity is expressed as a complex function of frequency,

$$\varepsilon(f) = \varepsilon' + j\varepsilon'' \qquad \textbf{2.2}$$

Where the real part, ε', represents energy storage, the imaginary part, ε'', represents energy loss, and $j = \sqrt{-1}$. In the limit, as the frequency tends to zero, the static permittivity is defined. The ratio of static permittivity to the permittivity of free space is called the dielectric constant, or relative permittivity, of the material.

$$\varepsilon_r = \frac{\varepsilon_s}{\varepsilon_0} \qquad \textbf{2.3}$$

Where ε_s is the static permittivity, and ε_0 is the permittivity of free space. Water content is the factor that most closely affects the dielectric constant of a soil sample. A sample of dry soil has a dielectric constant of about four, while the dielectric constant of free water is 81, [30]. It has been shown that as the soil's moisture content increases, its dielectric constant increases. The relationship proposed by Topp et. al., to obtain soil water content from dielectric constant is given by,

$$\theta = -5.3\times10^{-2} + 2.92\times10^{-2}\varepsilon - 5.5\times10^{-4}\varepsilon^2 + 4.3\times10^{-6}\varepsilon^3 \qquad \textbf{2.4}$$

It should be noted here that since the dielectric constant is a function of frequency, soil water content estimates derived from dielectric constant depend on the operating frequency of the instrument (i.e., TDR, GPR, etc.).

- Permeability is the degree of magnetization of a material that responds linearly to an applied magnetic field.

The velocity at which an electromagnetic wave propagates inside a material is given by

$$v_r = \frac{c}{\sqrt{\varepsilon_r}} \qquad \textbf{2.5}$$

Where c is the speed of light in a vacuum, equal to 3 x 10^8 m/s. As clearly seen from Equation 2.4, the velocity of the radar signal traveling in a material depends on the dielectric constant of the material. Most of the current methods of determining the moisture content of the soil using *GPR* estimate the dielectric constant. Dielectric constant can later be converted to moisture content by means of a petrophysical model.

2.1.3 GPR Data Description

Data generated by *GPR* is formed as images of the subsurface. Each column of the image is referred to as an A-scan. A B-scan is an ensemble of A-scans collected while the antenna is in motion. A sample A-scan, taken with the antenna directed towards the air is presented in Figure 2.1 while Figure 2.2 shows a sample B-scan.

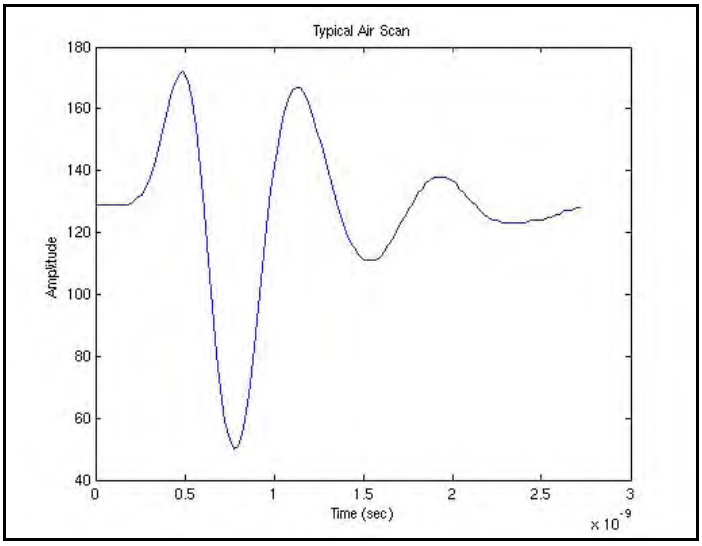


Figure 2.1 Typical air A-scan

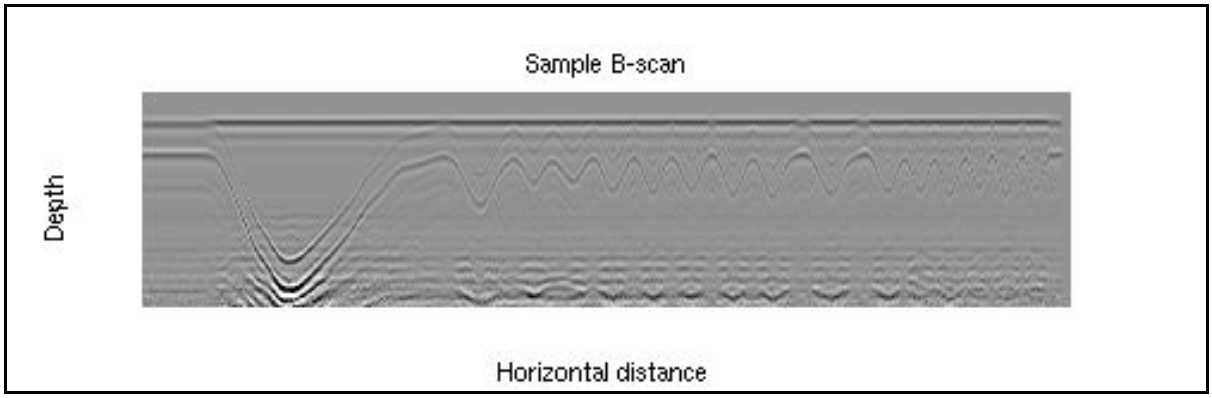


Figure 2.2 Example of GPR B-scan

2.1.4 *Equipment Description*

The *GPR* system used in this work is a GSSI SIRveyor SIR 20. It is a 2 hardware channel (antenna), 4 channel *GPR* system with a 1.5 GHz ground-coupled bowtie antenna. The system is shown in Figure 2.3. Antenna setup is presented in Table 2.1. The antenna setup given in Table 2.1 is a fixed setup given in [22] by GSSI inc. but it could be modified in any particular situation.

Figure 2.3 GSSI inc. SIR 20 GPR System

Table 2.1 Antenna Setup

Range	12 ns
Samples per Scan	512
Resolution	16 bits
Number of gain points	2
Vertical High Pass Filter	250 MHz
Vertical Low Pass Filter	3050 MHz
Scans per second	100
Vertical IIR High Pass	100 MHz
Transmit Rate	100 KHz

The SIR 20 can be operated in either one of three modes, namely point mode, free run, or survey wheel. In point mode, no data is collected until the user chooses to by clicking the computer mouse. The free run mode collects continuous data until stopped by the user. The rate at which data is collected is given in scans per second, a parameter that is user defined and controls the speed of the system. In survey wheel mode, data is collected based on the rotation of a survey wheel. In this mode the user is able to specify the sampling rate of the system. The survey wheel needs to be calibrated, and it allows the user to know the distance that has been traveled from the starting point.

3 Non-Parametric Transfer Function Estimation

Abstract Deconvolution is the processes of separating two signals that have been mixed by a process of convolution. Two *GPR* responses get convolved when the targets are not separated by more than one half wavelength. In this work, the signals that are convolved are the transmitted signal and the soil's response. It is desired to deconvolve the soil's response from the received signal. The response of the soil will be used as a signature for that particular soil type and moisture content. Two common methods are described in this chapter, namely the Empirical Transfer Function Estimate (*ETFE*), and one variation based on correlation analysis.

3.1 The Empirical Transfer Function Estimate (ETFE)

Suppose that for the brief time in which a *GPR* measurement is taken, the soil acts as a Linear Time-Invariant (LTI) system. Under this assumption, the signal at the receiver of the *GPR* can be modeled as the convolution of the transmitted signal and the response of the soil. Mathematically this is expressed as

$$y(t) = x(t) * h(t) \qquad \textbf{3.1}$$

Where *y(t)* is the signal recorded at the receiver, *x(t)* is the transmitted signal, *h(t)* is the soil's impulse response, and * is the convolution operator. Taking the Fourier Transform of Equation 3.1 we get,

$$Y(f) = X(f) \cdot H(f) \qquad \textbf{3.2}$$

Hence,

$$H(f) = \frac{Y(f)}{X(f)} \qquad \textbf{3.3}$$

In the system identification literature, Equation 3.3 is referred to as the Empirical Transfer Function Estimate, *ETFE*. The reason why it is called an empirical estimate is that the only assumption made on the system is that of linearity, [23]. Note that the *ETFE* is not defined at frequencies where the input is zero. If the input is zero, or close to zero at a certain frequency,

the *ETFE* is undefined at that specific frequency. The *ETFE* has the following properties, [23]:

- It is an unbiased estimate of the transfer function.
- Its variance, at a specific frequency, is given by the noise-to-signal ratio.
- At different frequencies, the *ETFE* gives asymptotically uncorrelated estimates.

We see that the *ETFE* is not a smooth estimate since its variance does not decrease with the length of the signal, but remains equal to the noise-to-signal ratio at that specific frequency. This comes from the fact that there is no relation between the different estimates (estimates at each frequency).

A way to smoothen the *ETFE* is based on correlation analysis, which is discussed in the following section. The magnitude of the quantity in equation 3.3 is what we call the Material Characteristics in Frequency Domain, *MCFD*

3.2 Correlation Analysis

Another way of estimating the frequency response of a LTI system is by means of the power spectrum of the input signal and the cross-power spectrum of the output and input signals. Again, suppose that the transmitted signal is input to a LTI system. The output signal is given by the convolution sum,

$$y[n] = \sum_{k=-\infty}^{\infty} h[k]x[n-k] \qquad \textbf{3.4}$$

Assume that the autocorrelation of the input signal is known. The cross-correlation of the input and output signals is given by,

$$R_{yx}[l] = \sum_{n=-\infty}^{\infty} y[n]x[n-l] \qquad \textbf{3.5}$$

Substitution of Equation 3.4 in Equation 3.5 yields,

$$R_{yx}[l] = \sum_{n=-\infty}^{\infty} \left(\sum_{k=-\infty}^{\infty} h[k]x[n-k] \right) x[n-l] \qquad \textbf{3.6}$$

After some simplification we obtain,

$$R_{yx}[l] = \sum_{k=-\infty}^{\infty} h[k]R_{xx}[l-k] \qquad \textbf{3.7}$$

Taking the Fourier Transform of Equation 3.7, the soil's frequency response becomes,

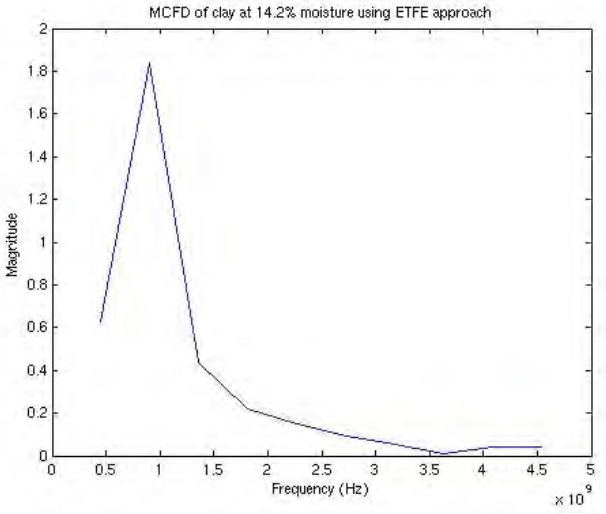
$$H(f) = \frac{S_{yx}(f)}{S_{xx}(f)} \qquad \textbf{3.8}$$

Equation 3.8 can be seen as a smoothed version of the *ETFE*. Here, $S_{yx}(f) = \Im\{R_{yx}(l)\}$ is an estimate of the cross power spectral density of the output and input, and S_{xx} $S_{xx}(f) = \Im\{R_{xx}(l)\}$ is an estimate of the power spectral density of the input. Note that, in its most simple case, that quantities $S_{yx}(f)$ and $S_{xx}(f)$ are estimated using the periodogram, which is an averaged Fourier Transform of the cross-correlation and auto-correlation sequences, respectively. Averaging has the effect of smoothing; this is why Equation 3.8 is a smoothened version of the *ETFE*. It can be shown that the variance of the *ETFE* is halved when it is calculated using Equation 3.8, [23].

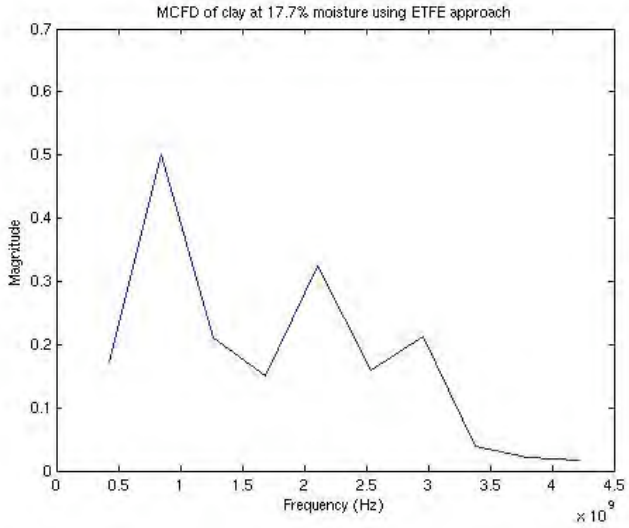
3.3 Example

Example calculations of the *MCFD* of a set of loam-type and clay-type soils *GPR* responses at different moisture content levels are presented in figures 3.1 and 3.2.

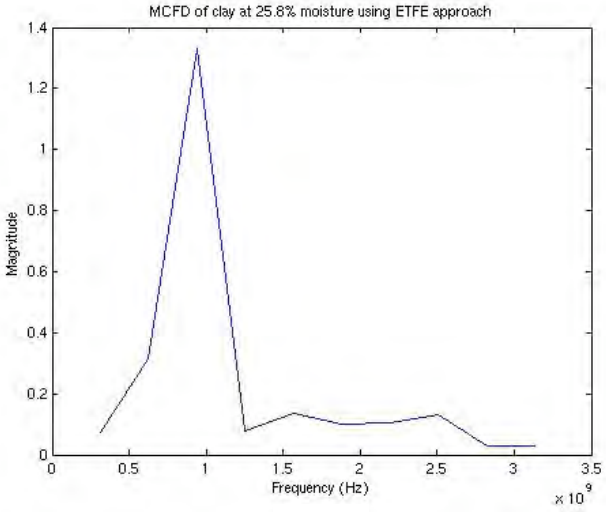
The example will show the differences in the characteristics of the materials with varying moisture levels, these differences is what makes the determination of the type of soil and its moisture possible.



(a)



(b)



(c)

Figure 3.1 MCFD calculation of clay soil type at three different moisture content levels using the ETFE approach. (a) 14.2% moisture, (b) 17.7% moisture, (c) 21.8% moisture

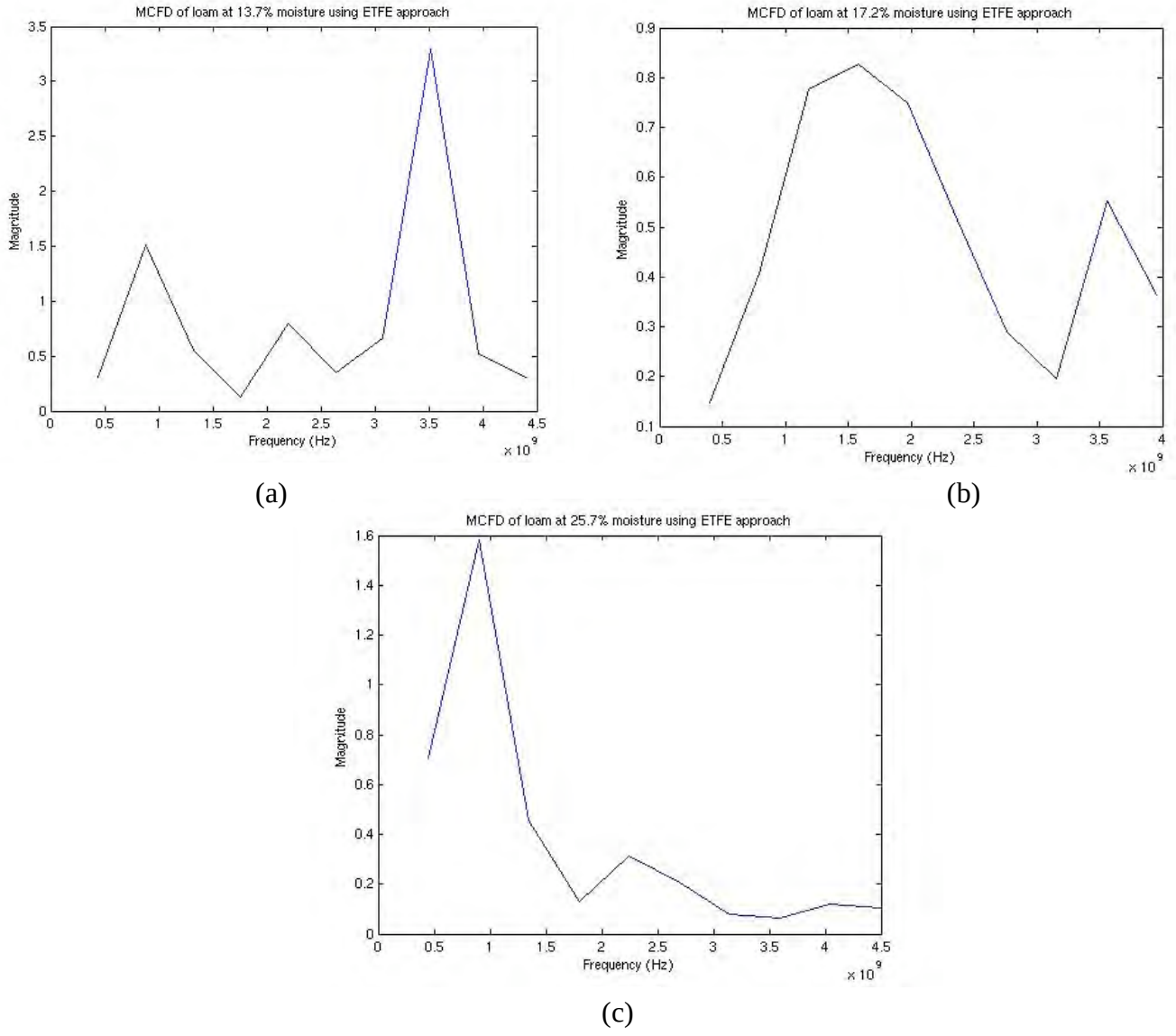


Figure 3.2 MCFD calculation of loam soil type at three different moisture content levels using the ETFE approach. (a) 13.7% moisture, (b) 17.2% moisture, (c) 22.6% moisture

Figures 3.1 and 3.2 show that the *MCFD,* could serve as the signature (or feature) needed to discriminate between soil types and moisture content levels. This is demonstrated for two distinct soil types, namely clay and loam. Note also that, for this application, we do not intend to estimate the true frequency response of the soil. Our goal instead is to obtain distinctive signatures that could serve as feature vectors. These signatures will be used as training features in a Neural Network (*NN*) based material type and moisture content determination system.

3.4 Summary and Conclusions

In this chapter, we have presented two non-parametric ways of obtaining the frequency response of a linear system, the *ETFE*, and its smoothed version using the power spectral densities. An example has been presented where it was shown that the *MCFD* could serve as a signature to discriminate between soil types and moisture content.

In the next chapter, the other building block of our system, *NNs,* will be discussed. A three-layer feed-forward *NN* will take MCFD vectors, as input, Theta Probe readings as target values, and the result will be a soil moisture content estimate.

4 Neural Networks

Abstract Neural Networks (*NNs*) are biologically inspired computational paradigms that seek to mimic the information processing nature of the nervous system. *NNs* have been applied in many application areas such as, buried target detection, speech recognition, stock market prediction, among others, with great success. In this chapter we review the fundamentals of *NNs* and the Back-Propagation learning algorithm is derived for the case of a 3-layer network. We review these fundamental concepts here since the work presented in this thesis is strongly dependent on the content of this chapter.

4.1 Fundamentals

The fundamental processing unit of a *NN* is the neuron. The name neuron comes from its resemblance to the biological counterpart. Figure 4.1 shows a typical nerve cell. Information processing in biological neurons is as follows [http://en.wikipedia.org/wiki/Neuron]

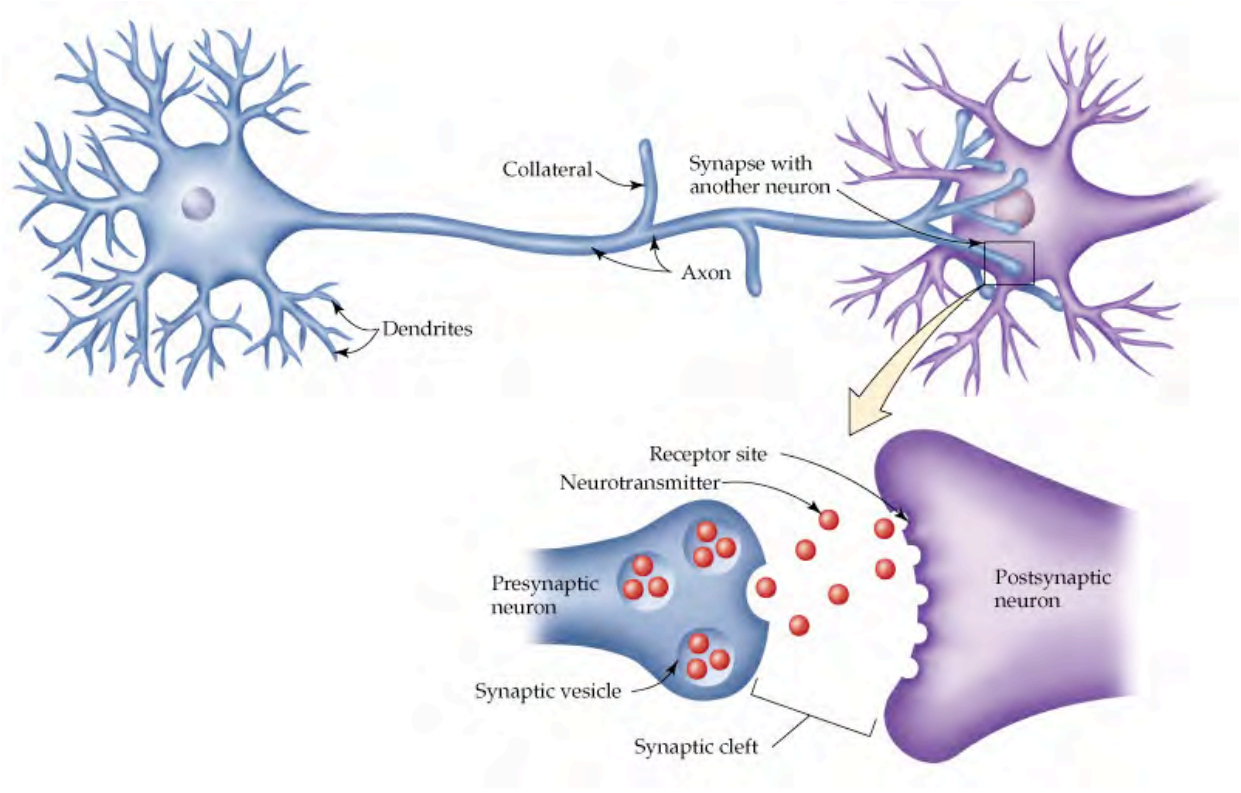


Figure 4.1 Information Processing in a biological neuron. *From cwx.prenhall.com*

Communication between neurons occurs through the synapse, where the axon terminal of one cell impinges upon a dendrite or soma of another. When an action potential reaches the

axon terminal, the wave of changing charges opens voltage gated calcium channels, thus allowing calcium ions to enter the pre-synaptic terminal. Calcium causes synaptic vesicles filled with neurotransmitter molecules to fuse with the membrane and release their contents into the synaptic cleft. The neurotransmitters diffuse across the synaptic cleft and activate receptors on the postsynaptic neuron. This process is illustrated in figure 4.1. A synapse can have excitatory or inhibitory behavior, which increases or decreases activity in the postsynaptic neuron, respectively.

McCulloch and Pitts introduced the first computational model of a neuron in 1943. They proposed a binary neuron model, in which the neuron was either on or off depending if the input exceeded a certain threshold, [12]. The work of McCulloch and Pitts fired an aggressive research effort by many people in order to obtain realistic neural models.

The A neuron is basically a linear combiner, which adds together the product of each individual input and its corresponding weight connection, see figure 4.2 below. Thus, the net input to the *ith* neuron is,

$$net_i = \sum_{j=1}^{N} x_j w_{ij} \qquad \textbf{4.1}$$

Where *N* is the number of inputs to the neuron.

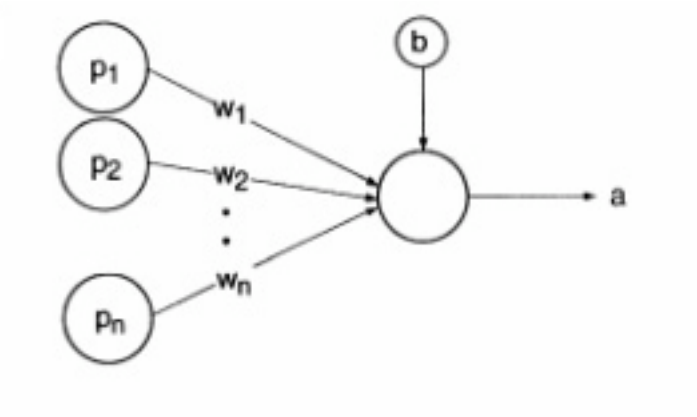


Figure 4.2 Linear Combiner

The net input to a neuron is then passed through a, generally, nonlinear function referred to as the activation function. There are different nonlinear activation functions that are used, the most popular being the sigmoid function given by,

$$f(x) = \frac{1}{1 + e^{-x}} \qquad \textbf{4.2}$$

A plot of the function in Equation 4.2 is shown in figure 4.3.

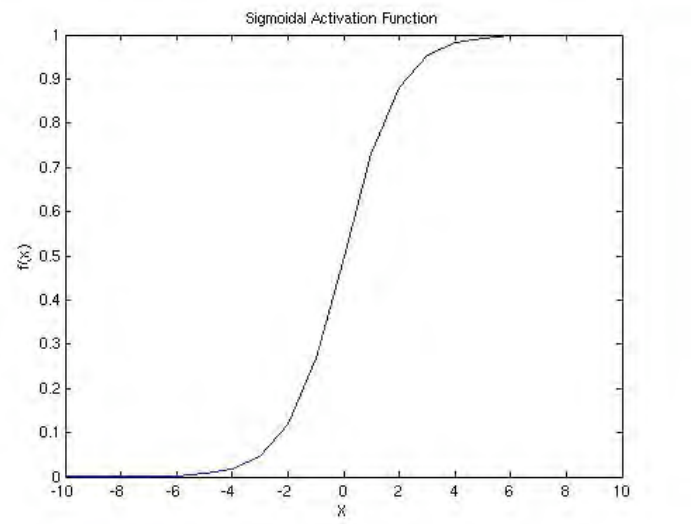


Figure 4.3 Plot of the sigmoidal activation function

There are many different topologies or architectures, which *NNs* can take. The most popular architecture is the multilayer perceptron (*MLP*), which consists of an input layer, one or more hidden (computational) layers, and an output layer. In the *MLP*, information is processed in the forward direction, from one layer to the next with no feedback connections. This type of network architecture is known as a feed-forward network. Figure 4.4 shows a graphical representation of this type of network with four input neurons, five neurons in the hidden layer, and one output neuron.

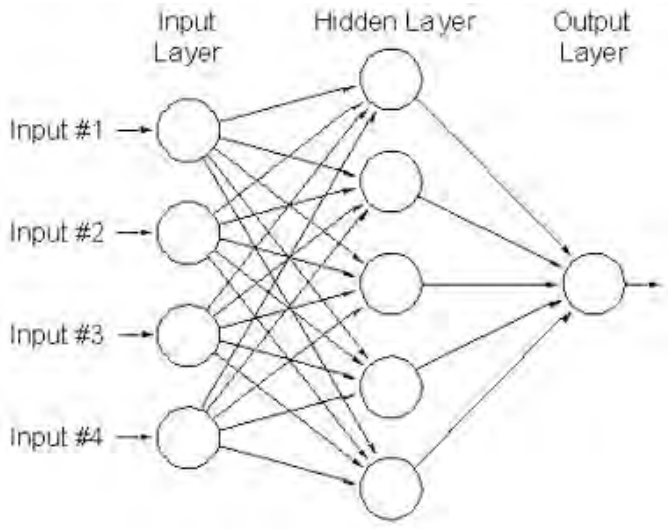


Figure 4.4 MLP Neural Network architecture

NNs work by learning an input-output mapping between a set of training features and desired outputs, or target values. They are called universal approximators since they can approximate any nonlinear mapping, by virtue of the universal approximation theorem [18].

In order for the *NN* to recognize an unknown pattern, it must be trained first. With training, it is sought to find the weight values that will minimize the difference between the actual output and the desired output. When the optimal weight value is found, the network would have learned the training patterns, and would be able to recognize patterns that were not present in the training phase. A popular *NN* training algorithm is the back-propagation algorithm, to be discussed next.

4.2 *NN* Training via Back-Propagation

The back-propagation algorithm is a learning law based on the steepest descent algorithm. Its name stems from that fact that in order to compute the weight change of a hidden layer, the error of the next layer must be computed first, thus this algorithm effectively back-propagates the errors, and is also known as back-propagation of errors. Next, the back-propagation algorithm is derived for a three-layer network.

The net input to a neuron in the hidden layer is the weighted sum of the input times the corresponding weight value plus a bias term, which is often treated as an additional weight.

$$net^{h}_{pj} = \sum_{i} w^{h}_{ji} \cdot x_{pi} + \theta^{h}_{j} \qquad \textbf{4.3}$$

Where x_{pi} is the *ith* input to that specific neuron, and w_{ij} is the weight value connecting the *ith* input to the *jth* hidden neuron, θ_j^h is the bias term, and the superscript "*h*" refers to connections on the hidden layer. The output of a neuron is given by,

$$i_{pj} = f^{h}_{j}(net_{pj}) \qquad \textbf{4.4}$$

Where f_j is the activation function of the *jth* neuron, and net^h_{pj} is the net input to that neuron. As stated earlier, the activation function can take different forms, depending on the application. Similarly, for the output layer we have,

$$net_{pk}{}^{o} = \sum_{j=1}^{L} w^{o}{}_{kj} \cdot i_{pj} + \theta^{o}{}_{k} \qquad \textbf{4.5}$$

Where i_{pj} is the *ith* input to that specific neuron, and w_{kj} is the weight value connecting the *kth* output to the *jth* hidden neuron, $\theta^{o}{}_{k}$ is the bias term mentioned earlier, and the superscript *"o"* indicates connection to the output layer . The output of the network is then given by,

$$o_{pk} = f^{o}{}_{k}(net^{o}{}_{pk}) \qquad \textbf{4.6}$$

In the back-propagation algorithm we wish to minimize the sum of squared errors with respect to the weights. This is done first for the output layer, and then for the hidden layer. The function to be minimized is,

$$E_p = \frac{1}{2}\sum_{k=1}^{M}(y_{pk} - o_{pk})^2 \qquad \textbf{4.7}$$

Where y_{pk} is the desired output of the network. Differentiating Equation 4.7 with respect to the output weight, we obtain the following update equation [12],

$$w_{kj}{}^{o}(t+1) = w_{kj}{}^{o}(t) + \eta(y_{pk} - o_{pk}) f_{k}{}^{o'}(net^{o}{}_{pk}) i_{pj} \qquad \textbf{4.8}$$

Where η is the learning rate parameter, which controls the size of the steps taken toward the minimum of equation 4.7. Following the same procedure, the weight update equation for the hidden weights becomes,

$$w^{h}{}_{ji}(t+1) = w^{h}{}_{ji}(t) + \eta\left\{\left(\sum_{k}(y_{pk} - o_{pk}) f_{k}^{o'}(net^{o}{}_{pk}) w_{kj}{}^{o}\right) f_{j}^{h'}(net_{pj}{}^{h}) x_{pi}\right\} \qquad \textbf{4.9}$$

The ground truth data to be used as target value in *NN* training is obtained with a Theta Probe, which gives actual soil moisture up to a depth of about 10 cm. For a neural network to be able to obtain the desired mapping, it must be trained with data that captures the essence of the general population since *NNs* are not able to extrapolate out of the training set. This fact makes *NNs*, as well as all other "black-box" models just as good as the training data.

It is of importance to discuss the relationship between the training sample size and the number of features when training the *NN*. It is shown in [31] that when the number of features increases linearly with the dimensionality, the error in the testing set does not depend on the dimensionality. Furthermore, it is also shown in [31] that features that carry little or no discriminatory information do not contribute much to the overall performance of *NNs*.

4.3 Summary and Conclusions

This chapter has presented the underlying *NN* background needed to understand the problem at hand. A review of *NNs* was presented making the analogy with the nervous system to make the transition into the computational framework. The basic elements of a *NN* were presented; finally, the popular back-propagation training algorithm was derived.

In the next chapter, we present the development of a Graphical User Interface (GUI) to facilitate user interaction, for the purpose of this project.

5 Graphical User Interface Development

Abstract Radar Remote Sensing using a Ground Penetrating Radar (*GPR*) has proved to be a useful method for determining soil moisture content. A software product that implements the proposed methodology to soil type and soil moisture determination has been developed in the C-language. Previously, the software was being used as a console application with little thought about future users. This chapter presents the development of a Graphical User Interface, GUI, which was written on top of the previously developed C-code.

This improvement uses the Tool Command Language and its graphical Toolkit (TCL-TK). It implements the proposed *MCFD/NN* method. First, it allows the user to process *GPR* images into *MCFD* data sets, which are used either for training the *NN* or as unknown moisture samples. Second, *NNs* can be created and trained from the *MCFD* data sets. Third, options to save and plot the results are provided. These capabilities make the *MCFD/NN-GUI* an easy to use interface for soil type and soil moisture determination using *GPR*.

5.1 Description

5.1.1 Create Data Sets

Once the GPR images are retrieved from the radar equipment, particular wavelets are selected, processed, and analyzed for soil moisture content. The first step is to calculate the wavelet's MCFD, returning a vector. Different collections of MCFD vectors are used for training, testing, and validation; which are called Data Sets. For each of the Data Sets, particular information about each wavelet is needed, for example: Filename, Scan Number, Target Value, etc. A spreadsheet-like window is used to enter and save the information for each Data Set. See Figures 5.1 and 5.2 for details.

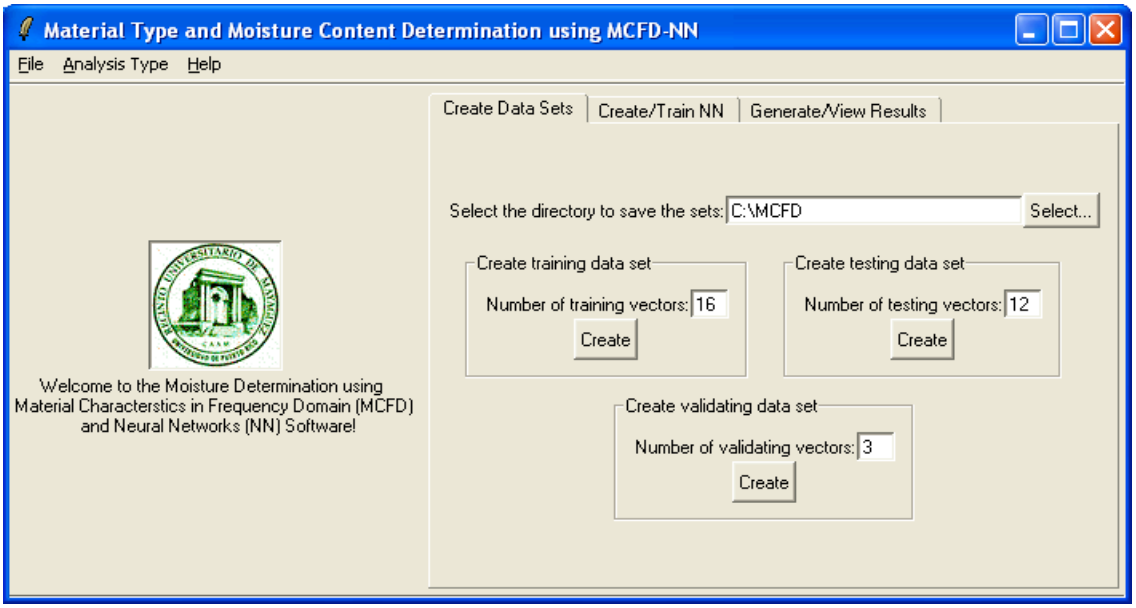


Figure 5.1 Create data sets window

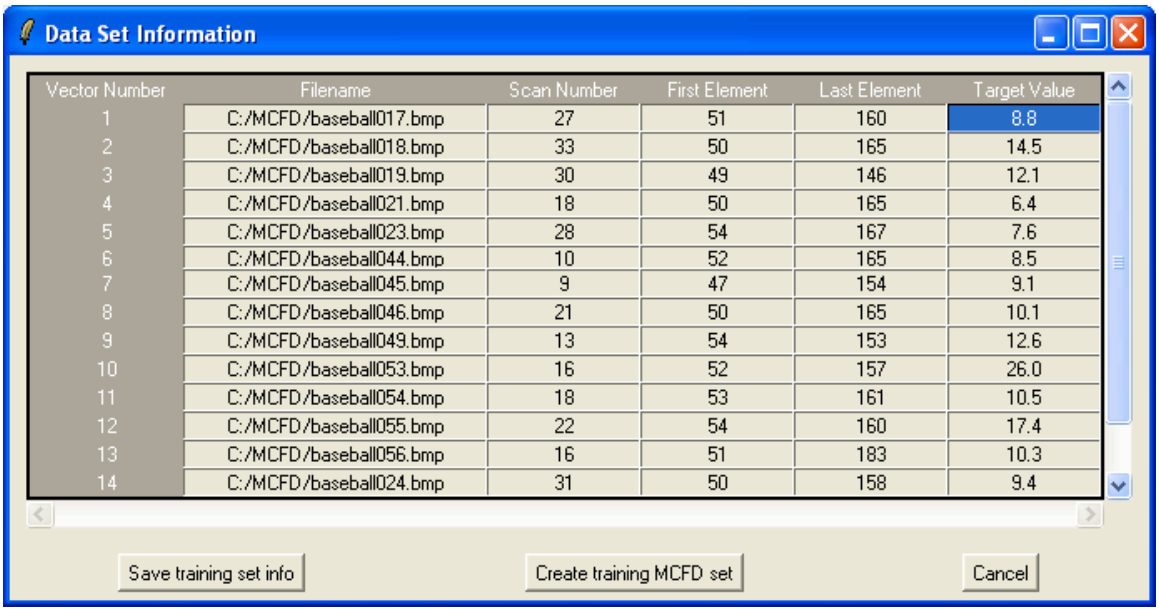


Figure 5.2 Enter data set information

The user has to enter the number of vectors he/she wants to create and then push the Create button for the Data Set Information window to display. There, all the necessary information should be entered. Afterwards, it may be stored, and the *MCFD* calculation may take place by clicking on the Create *MCFD* Data Set button. The output of this phase is a file containing the Data Set Information and also the files containing the *MCFDs* of the selected wavelets (in .txt and binary format).

5.1.2 Create and Train Neural Network

To obtain soil type or soil moisture, a neural network needs to be constructed and trained. The GUI allows the user to create a two-layer feed-forward network. After all fields have been entered, the *Train NN* button is pressed to begin the training phase. A progress bar appears indicating how training is progressing. When training ends, a binary output file is produced which contains the trained network parameters (weights). This window is shown in figure 5.3.

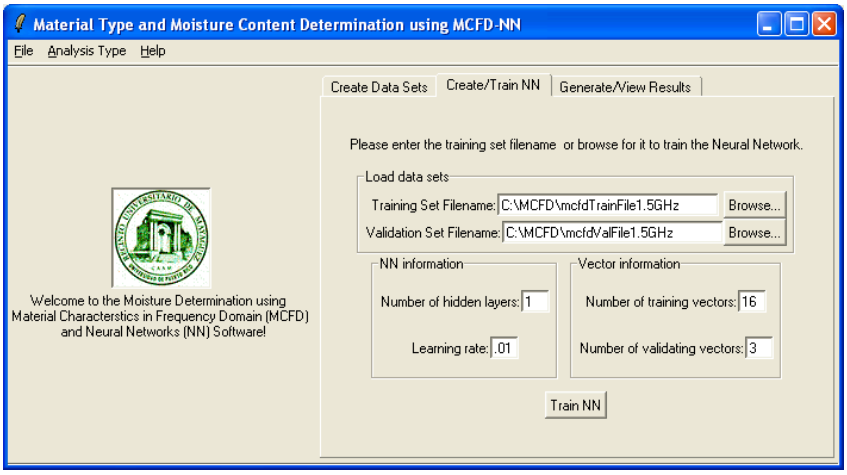


Figure 5.3 Create and train *NN*

5.1.3 Generate Results

Browsing for the files containing the testing set, and the trained NN is necessary to generate the results. The Analyze button is clicked to perform this task. The results are organized in a table and stored in a text file. Also, if desired, results may be viewed in form of a scatter plot of GPR moisture Vs. Theta Probe readings. See figures 5.4 & 5.5.

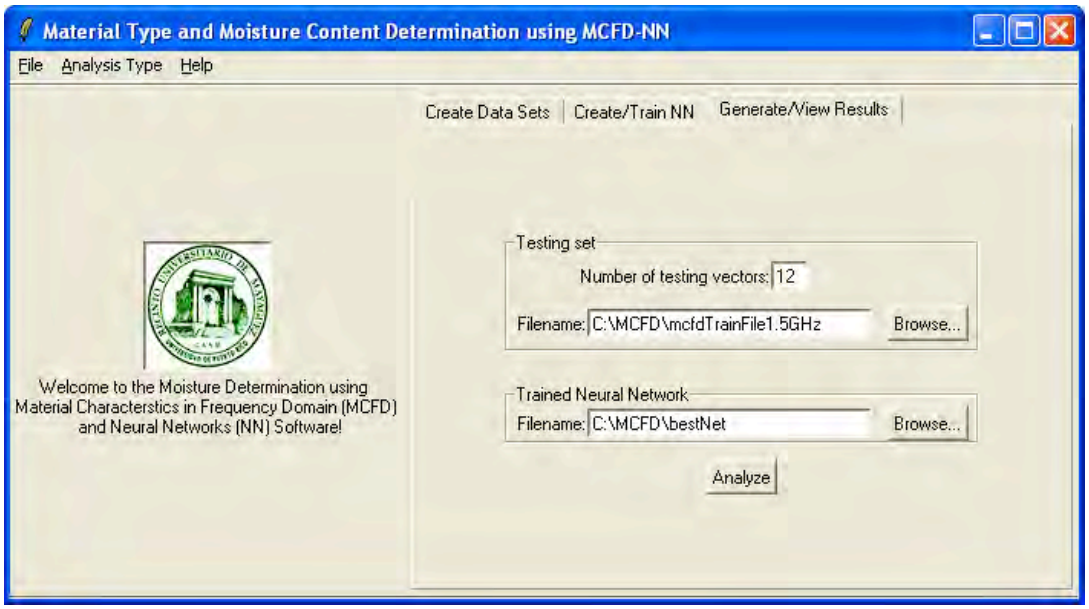


Figure 5.4 Test *NN* window

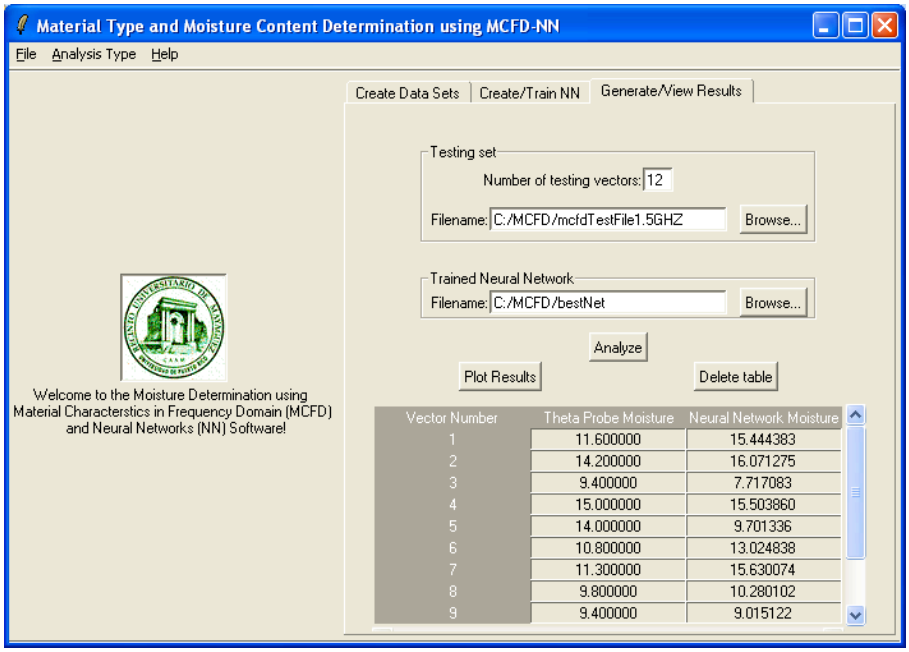


Figure 5.5 View results

5.2 Summary and Conclusions

The development of a GUI, which was designed to run on top of the C-language written software, has been presented and its capabilities described. This GUI represents an advantage for general users since it provides an easy to use, intuitive interface for soil type/moisture determination using *GPR* and the proposed *MCFD/NN* method.

In particular, the GUI provides options to create the necessary data sets, construct, train, and test a 2-layer feed-forward neural network. Finally, results can be viewed in tabular form or in the form of a scatter plot. The software is written in the C and TCL/TK languages, as such, the GUI is easily extendible, and platform independent.

In the following chapter, application of these methods to laboratory and field data will be presented. Of particular interest are the results obtained from a field campaign conducted at the baseball field located on the UPRM campus. Results show the performance of the *MCFD* approach for soil type/moisture determination in open field environment.

6 Experimental Description and Results

Abstract In this chapter, description of the experiments that were performed is presented and the results are discussed. First, experiments were performed in controlled laboratory conditions. For these experiments, three different soil types were used, namely, sand, loam, and clay. The *MCFD/NN* algorithm was employed as to determine the type of soil and its moisture content. Results show that soil type determination is possible using the proposed approach, and moisture content determination is successfully achieved, consistent with readings obtained using a Theta Probe.

The algorithm was then tested in uncontrolled, field conditions. A four-day campaign was conducted, during the months of February-March of 2005, and during the month of September of 2005. The *NN* was trained on individual days, and combined data. Acceptable results when the *NN* was trained on data from the first two days, and tested with data from the last two days (six months later) confirm the validity of the *MCFD/NN* approach.

6.1 Laboratory Experiments

6.1.1 Description

Experiments in controlled conditions helps us to analyze the strengths and weaknesses of the proposed approach, as well as determining its feasibility before testing it in more challenging scenarios. Material type and soil moisture experiments were performed using three different soil types, namely, sand, loam, and clay. The soils used in this work are the same as those used in [30]. As mentioned by Harmsen et al., the soils were analyzed as for their physical and chemical properties by Soilcon Laboratories, Ltd. of British Columbia, Canada, and are reproduced in Table 6.1. Specifically, construction sand, San Antón Loam, from Juana Díaz, PR, and Daguey Clay, from Finca Alzamora on the University of Puerto Rico, Mayaüez Campus were the types of soil that were used. A reasonably large amount of soil was put in a plastic bag, and placed inside a plastic container, which was then hung from a support about 5 feet above the ground, as depicted in figure 6.1.

Figure 6.1 Laboratory setup for data acquisition

Table 6.1 Chemical and Physical Properties of the three soil types used in the experiments

Analysis	Parameter	Units	Sand	Loam	Clay
	pH	pH units	7.27	6.44	4.65
	EC	(dS/m)	0.17	0.39	0.18
Total Carbon	TC	%	0.22	1.59	1.75
AmmoniumNitrogen	NH4	mg/kg	<1	5.31	2.31
Nitrate & Nitrite	Nitrogen	mg/kg	1.35	28.98	12.73
Total Nitrogen	TKN	%	0.12	0.14	0.30
Available P	Bray P	mg/kg	<0.6	11.73	7.02
Sulphate	S	mg/kg	7.90	18.00	21.00
Available Boron	B	mg/kg	<0.1	0.40	0.30
	Ca	mg/kg	2930.00	3150.00	869.57
Available Nutrients (NH4OAc Extractable)	K	mg/kg	<2	134.42	150.31
	Mg	mg/kg	160.64	450.72	263.58
	Na	mg/kg	8.85	64.64	15.71
	Cu	mg/kg	<0.03	3.61	3.42
Available Metals (0.1N HCl Extractable)	Fe	mg/kg	4.60	58.06	30.21
	Mn	mg/kg	23.79	76.59	19.88
	Zn	mg/kg	0.11	2.78	1.20
Cation Exchange Capacity		meq/100g	4.15	12.61	12.84
Exchangeable Cations					
	Ca	meq/100g	2.80	13.37	3.58
	Mg	meq/100g	0.50	3.55	2.55
	K	meq/100g	0.16	1.04	0.59
Total Organic Carbon	TOC	%	0.05	1.54	1.67
Texture	Sand	%	96.00	35.99	3.61
	Silt	%	1.56	39.53	29.06
	Clay	%	1.62	23.87	67.27
USDA Classification			Sand	Loam	Clay
Soil Characteristic Data	5 J/kg	% by vol	9.31	46.94	47.83
(Pressure vs. % Vol.)	10 J/kg	% by vol	6.86	39.52	46.12
	33 J/kg	% by vol	5.32	29.68	43.23
	70 J/kg	% by vol	4.77	26.41	41.56
	100 J/kg	% by vol	4.54	24.91	40.60
	300 J/kg	% by vol	3.95	20.92	37.20
	500 J/kg	% by vol	3.60	19.27	35.65
	800 J/kg	% by vol	3.27	17.90	34.14
	1200 J/kg	% by vol	2.93	16.73	32.52
	1500 J/kg	% by vol	2.71	15.80	31.43
Bulk Density		kg/m3	1387.80	1350.97	1092.38
Particle Density		kg/m3	2673.17	2533.14	2537.36
Total Porosity		% vol	48.08	46.66	56.95
Air Entry Tension		J/kg	0.00	3.87	0.72
Saturated Hydraulic Conductivity		cm/hr	65.70	0.24	2.77
Aeration Porosity	5 J/kg	% by vol	38.77	-0.27	9.12
Aeration Porosity	10 J/kg	% by vol	41.23	7.15	10.83
Available H20 Storage Capacity		% by vol	2.61	13.88	11.79

The experiments began with dry soil, and then water was added gradually. Every time that water was added, the soil was mixed as to obtain a homogeneous mixture. The *GPR* was placed on top of the soil and several scans were taken, then the actual moisture content was measured with a Theta Probe. This process was repeated until the soil was fully saturated.

To quantify the error between the actual soil moisture, measured by Theta Probe, and the soil moisture obtained using our approach, the absolute error is used.

$$E = \left| T.\mathrm{Pr}obe - GPR \right| \qquad \textbf{6.1}$$

Where *T.Probe* is the moisture content measured by the Theta Probe, *GPR* is the moisture content obtained using the *MCFD/NN* method.

6.1.2 Results and Discussions

The method was first tested with the purpose of soil type determination. Training, validation and testing sets were created. The validation set is required in order to see how well the model obtained by the *NN* fits data that was not present in the training set. Towards this goal, the training procedure shown in Figure 6.2 was implemented. Assume that the training has converged to the desired threshold. Present the validation set and let α be the *MSE* on this set, and β the desired validation threshold. If $\alpha \geq \beta$ then a neuron is added to the hidden layer.

All the weights in the network are re-initialized and training is resumed with this new configuration.

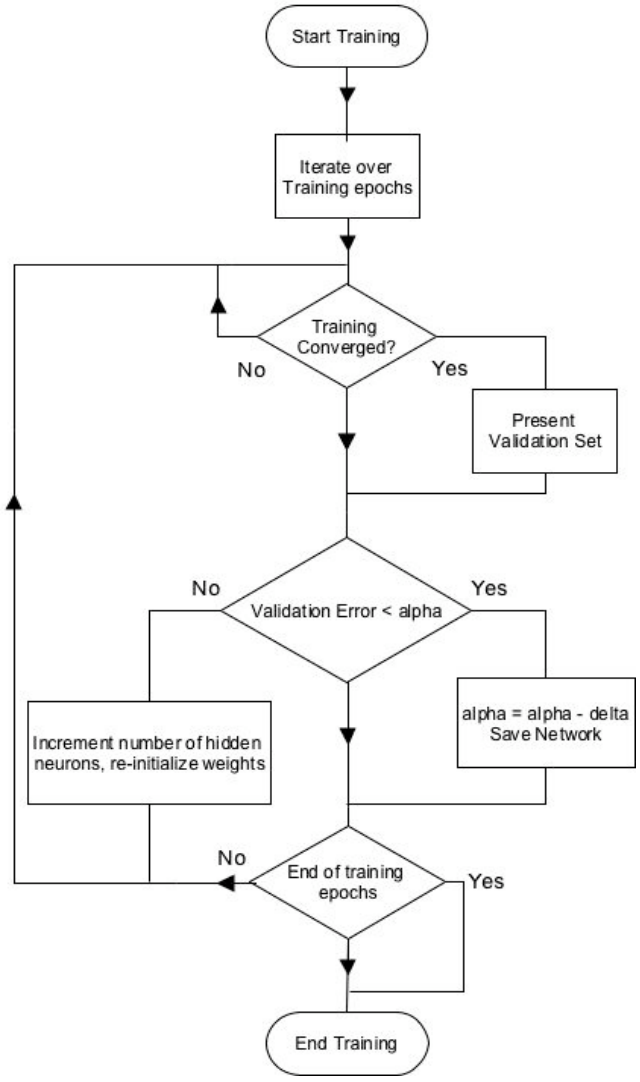


Figure 6.2 Neural Network training procedure

When $\alpha < \beta$, β is recomputed as $\beta_{new} = \beta_{old} - \delta$, where δ is an arbitrarily chosen positive constant, and $0 < \delta \leq 1$. The network is saved as the best so far in the training. With this

training procedure it is assured that when training has finished, the best possible network would have been saved.

Results of soil type determination are presented in Figure 6.3.

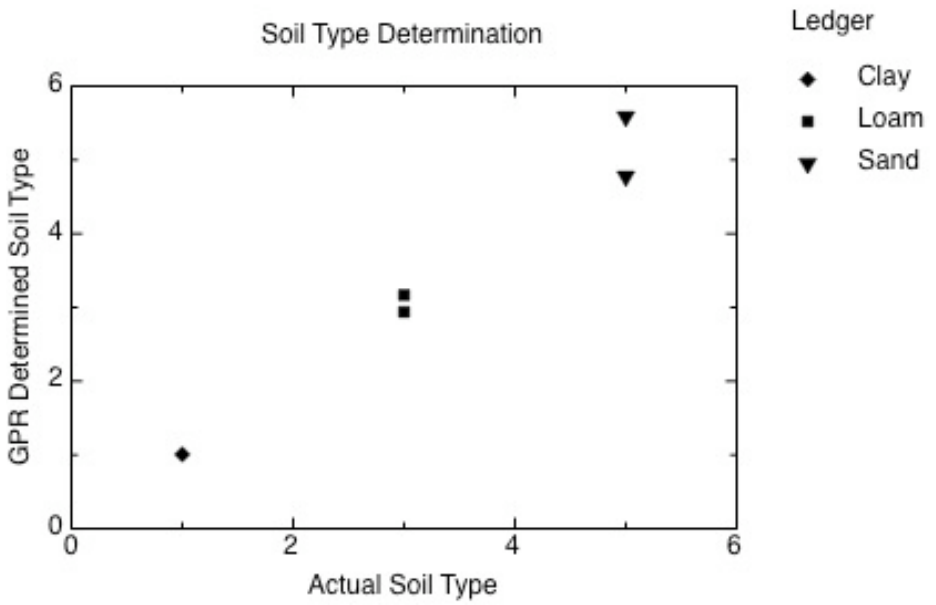


Figure 6.3 Soil Type determination

As it is shown in the figure, there is one sample of clay, and two samples each of loam and sand. Clearly soil type of each of the five soil samples in the testing set was determined with good accuracy. The samples used in this experiment ranged from dry to wet, so the moisture that was being introduced may have become a source of error. Nevertheless, the proposed approach seems effective, and its usefulness will be reinforced with laboratory experiments

for soil moisture determination, and later with more challenging experiments in uncontrolled environments.

Now, we turn our attention to soil moisture determination. There were three different *NNs* trained in this case, one for each of the three distinct soil types. Results show good accuracy as presented in figures 6.4 through 6.6.

Figure 6.4 shows the results for the clay experiment. The training samples are presented in red, and the testing samples are in green. Although the amount of data is not very large, a constraint that became inevitable due to the nature of the experiments, good results were obtained. As seen in figure 6.4, only one of the samples fell way off the line. This may be due in part to a small training set, and in part due to the soil not being homogeneously mixed to obtain a uniform moisture distribution, which may have caused the Theta Probe to give inaccurate readings.

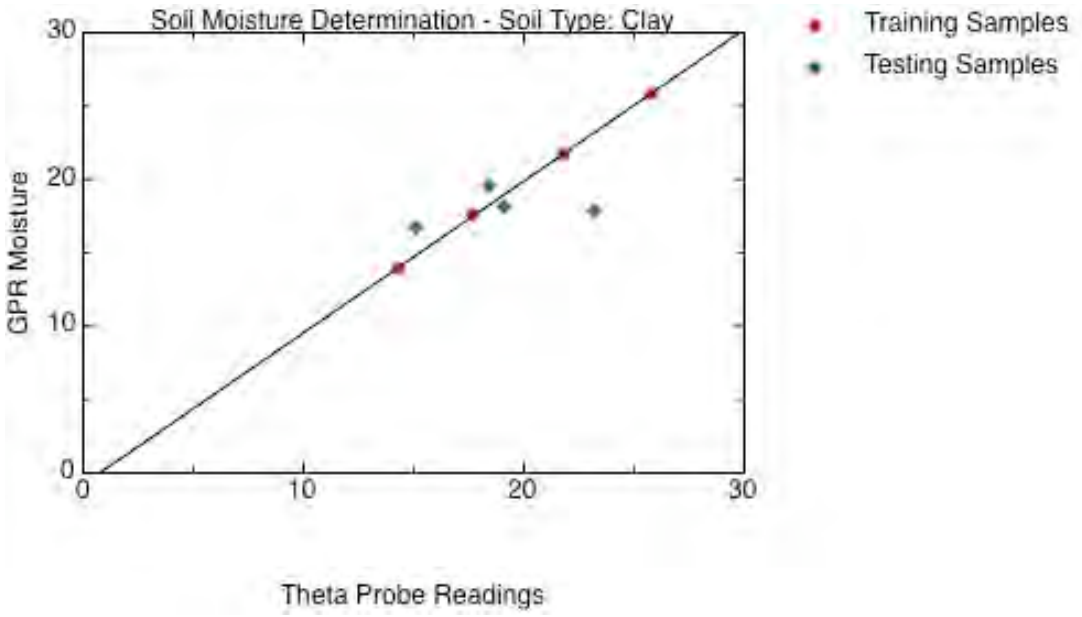


Figure 6.4 Results of soil % moisture determination for clay soil

Results for the loam are shown in Figure 6.5. We can see that in this case, greater accuracy was obtained, with all testing samples falling close to the line. The loam soil is easier to manage than the clay, so the errors due to poor mixing were reduced.

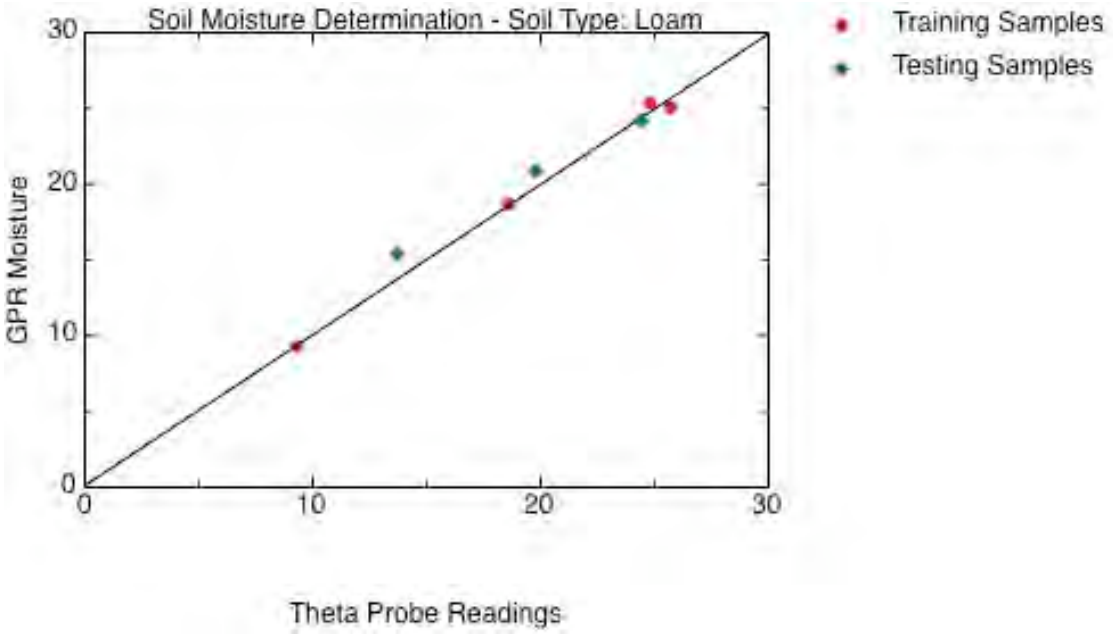


Figure 6.5 Results of soil % moisture determination for loam soil

Figure 6.6 depicts the results when the soil being used was sand. Again, better results were obtained as compared to Figure 6.4. Table 6.3 shows the same results in tabular form, with the relative error specified for each case.

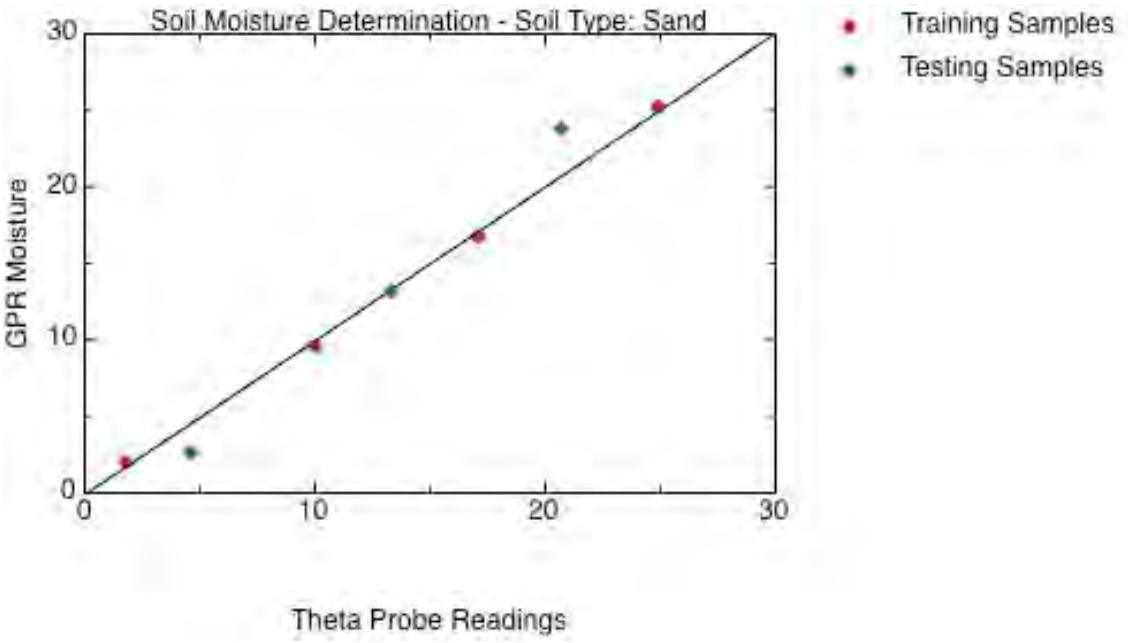


Figure 6.6 Results of soil moisture determination for sand soil

Table 6.2 Error calculation for soil moisture laboratory experiments

	Theta Probe reading (%)	*GPR soil moisture (%)*	*E (%)*
Clay	15.10	16.73	1.63
	18.40	19.56	1.16
	19.10	18.17	0.93
	23.18	17.87	5.31

Loam	13.70	15.38	1.68
	19.80	20.88	1.08
	24.40	24.21	0.19
Sand	4.60	2.67	1.93
	13.30	13.18	0.12
	20.30	23.82	3.52

6.2 Field Experiments

6.2.1 Description

A four-day field campaign extending over a period of six months (two mini-campaigns, two days each and six months apart) was conducted at the baseball field inside the University of Puerto Rico Mayagüez campus with the objective of obtaining open-field soil moisture information.

Soil moisture determination in true open field conditions with *GPR* is a difficult task that has not been tackled by many due to the fact that an open field is full of hardships that are beyond the control of the researcher, such as inhomogeneities, the field being cluttered with

grass, rocks, and other non-soil items, field saturation due to extreme precipitation, etc. These conditions constantly change the characteristics of the field. The effect of some of these factors can be reduced with efficient signal processing techniques. Some work can be found in [7] using ground wave techniques, and more recently in [24] based on electromagnetic inversion.

The first part of the campaign dates back to February-March of 2005 while the second part took place during the month of September of 2005. Field conditions were subjected to noticeable changes in the time between March and September, going from having no grass to an almost outgrown field that showed scattered wetness due to the continuous precipitation characteristic of the area. These changes in field conditions produced soil moisture data in the range of 6% to 37%, as measured by a Theta Probe. The exact distribution is shown in Table 6.3.

Table 6.3 Distribution of Measurements

Moisture Percent range	*Measurements*
5% - 9%	15
10% - 15%	20
16% - 20%	7
21% - 25%	0
26% - 30%	2

31% - 35%	5
36% - 40%	1

Due to the inhomogeneous (even at the scale of the antenna footprint) nature of the soil in an uncontrolled environment, there can be sharp dielectric contrasts within a small area. Such contrast is observed as variations in the readings obtained by the Theta meter when probing the area under the footprint of the *GPR*. In order to measure true soil moisture that is to be used as target values in the *NN* training. To reduce the effect of soil inhomogeneity, three Theta Probe readings were obtained at each measurement point along the main diagonal of the antenna footprint. The median value was used as target value for NN training. Several scans were taken in order to select the best available scan by visual inspection.

To measure the accuracy to which soil moisture is obtained, as compared to actual soil moisture, absolute errors are used. The definition of the absolute error given in equation 6.1 is modified here to take into account the variations in the Theta Probe readings at the footprint of the *GPR*, as described above. The absolute error is given by,

$$E_D = \frac{|T.\mathrm{Pr}obe - GPR|}{D} \qquad \textbf{6.2}$$

Where *T.Probe* is the actual soil moisture as obtained by the Theta meter, *GPR* is the soil moisture obtained using the *MCFD/NN* method, and *D* is defined as the range of the variations in the Theta Probe readings,

$$D = \frac{T.\mathrm{Pr}obe_{\max}}{T.\mathrm{Pr}obe_{\min}} \qquad \textbf{6.3}$$

6.2.2 Results and Discussion

The data collected following the procedure given in section 6.2.1 was analyzed, and data points where the three Theta Probe readings differed by less than 3% were kept, otherwise it was rejected. The remaining data points were divided in training, validation, and testing sets. The training set consisted of data points that showed more consistency in the Theta Probe readings.

Training of the NN was done following the procedure presented in Figure 6.2. Four *NNs* were created. The first NN was trained and tested on data taken from the first day of measurements, back in February 2005. The second NN was trained and tested on data taken from the second day (March). Combined data from the first two days was used to train and test the third *NN*. Finally, a fourth *NN* was created. Training data was taken from the first two days (February-March); this *NN* was tested on data taken from the September leg of the

campaign, six months later. With this test, we will be able to assess the capabilities of the *MCFD/NN* approach to obtain accurate soil moisture information, over time.

The results are shown in figures 6.7 through 6.10, along with relative error calculations for the testing set in each of the four cases.

As the results show, soil moisture was determined in an open field, uncontrolled, environment with good accuracy. The obtained accuracy can be seen from tables 6.4 through 6.7. Specially, table 6.7 presents the results when the *NN* was trained and tested on data taken six months apart. Average errors are presented in table 6.8 for each of the four *NNs.* The smallest error, on average, resulted when the *NN* was trained on data from the first two days of the campaign, and tested on data from the last two days; a six month difference.

An average error of 1.63% is not a prohibitive large price to pay when determining soil moisture over large tracks of land. The proposed approach is well suited for on-line implementation, allowing the user to obtain soil moisture maps for a given field. Also, the method has an important application when it is sought to obtain data for satellite validation.

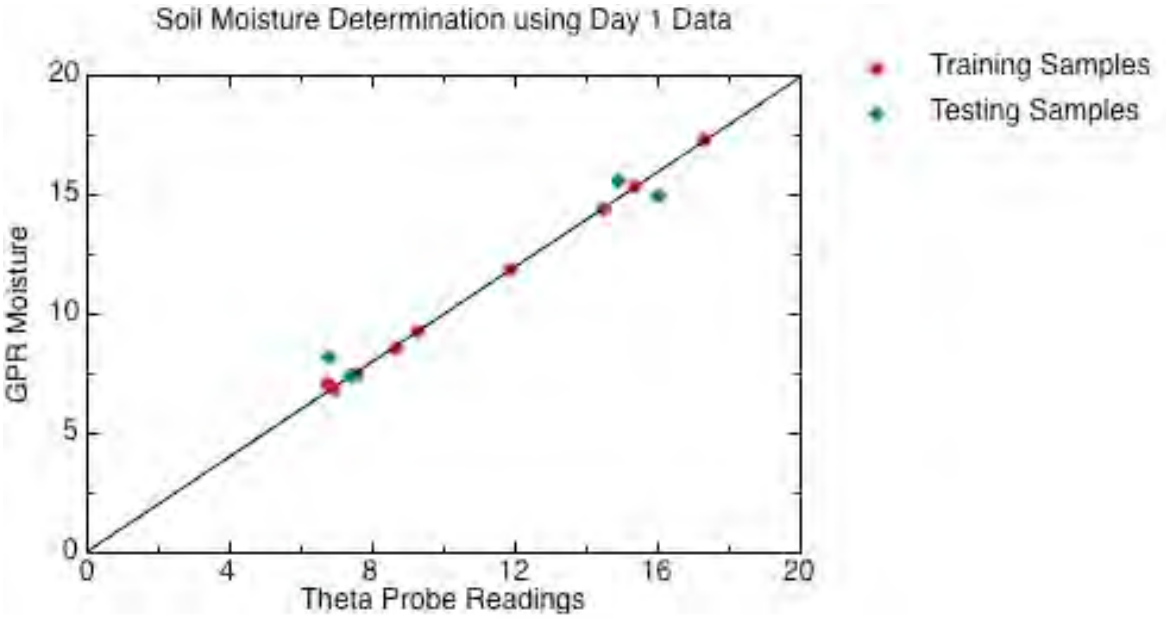


Figure 6.7 Soil moisture determination. Day one data

Table 6.4 Error calculation for data in figure 6.7

T.Probe reading (%)	*GPR moisture (%)*	*D*	E_D *(%)*
6.77	8.20	1.09	1.31
7.40	7.43	1.11	0.03
14.88	15.60	1.11	0.65
16.0	14.96	1.06	0.98

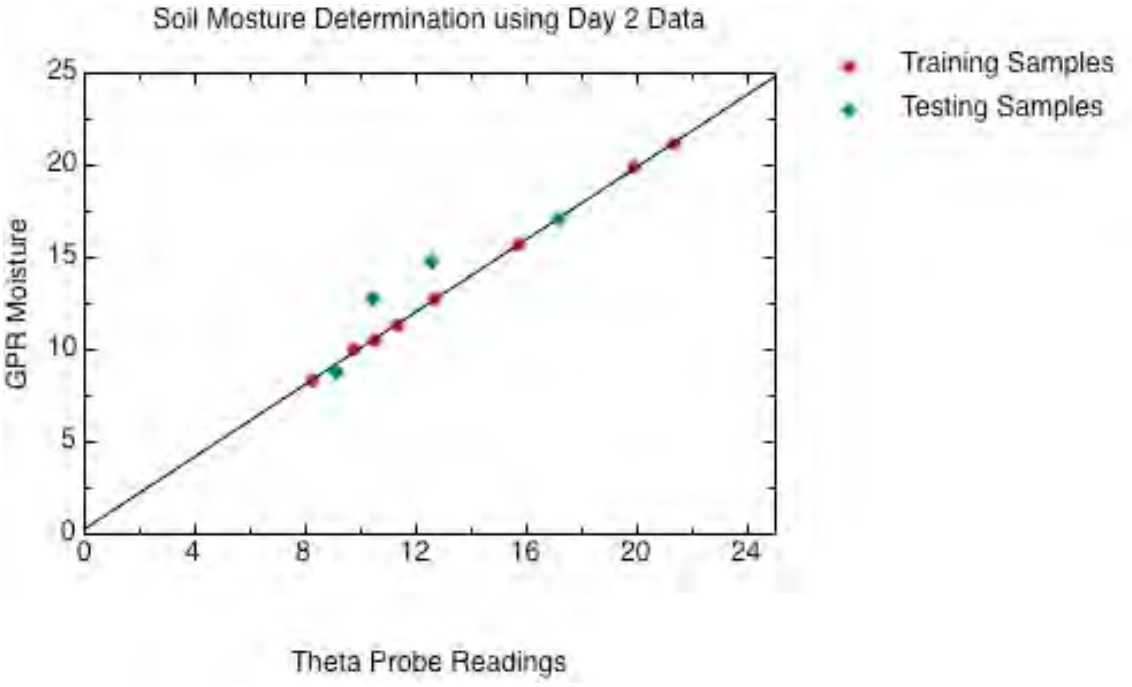


Figure 6.8 Soil moisture determination. Day two data

Table 6.5 Error calculation for data in figure 6.8

T.Probe reading (%)	*GPR moisture (%)*	*D*	E_D *(%)*
9.10	8.80	1.07	0.28
10.43	12.78	1.02	2.30
12.56	14.82	1.10	2.05
17.16	17.10	1.13	0.05

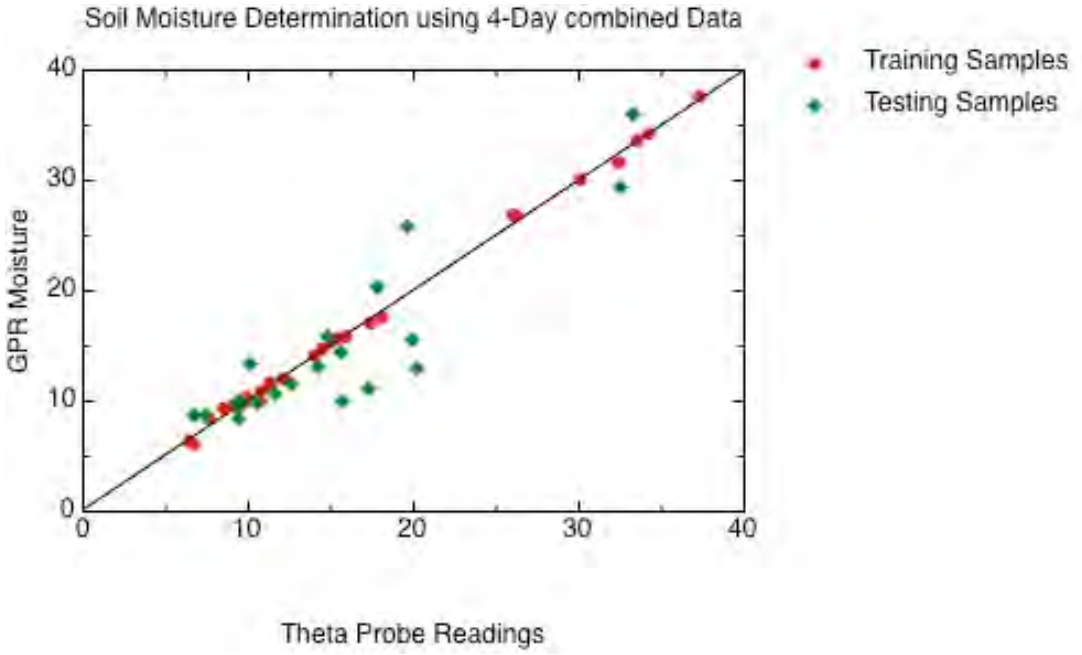


Figure 6.9 Soil moisture determination. 4-day combined data

Table 6.6 Error calculation for data in figure 6.9

T.Probe reading (%)	*GPR moisture (%)*	*D*	E_D *(%)*
11.6	10.7	1.28	0.70
17.8	20.4	1.18	2.20
32.5	29.4	1.12	2.77

17.3	11.1	1.26	4.92
15.7	10.0	1.21	4.71
33.3	36.0	1.08	2.5
19.9	15.6	1.19	3.61
14.2	13.1	1.19	0.92
9.40	8.44	1.40	0.69
7.40	8.72	1.11	1.19
6.70	8.74	1.09	1.87
9.40	9.63	1.32	0.17
20.2	12.9	1.22	5.98
14.8	15.9	1.11	0.99
10.1	13.4	1.14	2.89
15.6	14.4	1.17	1.03
12.6	11.5	1.08	1.02
9.4	10.0	1.33	0.45
19.6	25.8	1.17	5.30
10.5	9.87	1.02	0.62

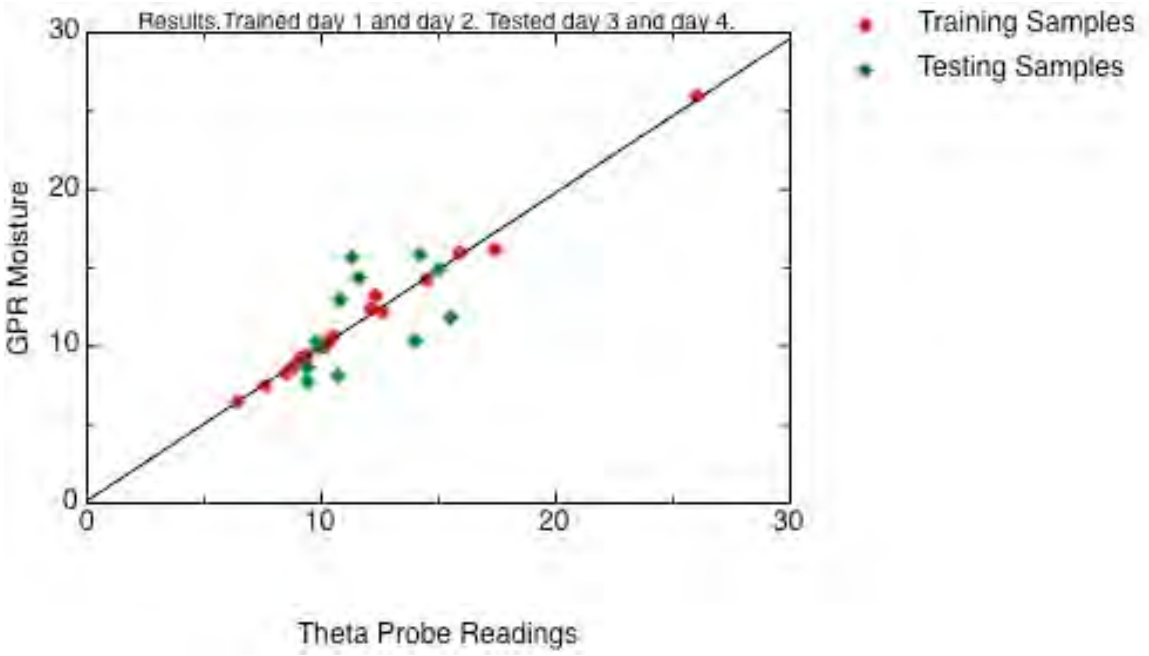


Figure 6.10 Soil moisture determination. Trained day 1 and day 2. Tested day 3 and day 4

Table 6.7 Error calculation for data in figure 6.10

T.Probe reading (%)	*GPR moisture (%)*	*D*	E_D *(%)*
11.6	14.4	2.70	1.03
14.2	15.9	1.19	1.43
9.4	7.76	1.40	1.17
15.0	14.9	1.07	0.09
14.0	10.3	1.03	3.59

10.8	12.9	1.17	1.79
11.3	15.7	1.13	3.89
9.80	10.3	1.17	0.43
9.40	8.67	1.11	0.66
15.5	11.9	1.13	3.19
10.7	8.15	1.12	2.28
9.90	9.94	1.12	0.03

Table 6.8 Average Errors of the four *NNs* for the open field experiment

Data	*Average E_D (%)*	*E_D Standard Deviation*
Trained and Tested from day 1	0.74	0.54
Trained and Tested from day 2	1.17	1.17
Trained and Tested from all 4 days	2.23	1.80
Trained from days 1 & 2 Tested from days 3 & 4	1.63	1.34

6.3 Summary and Conclusions

Soil type and soil moisture determination has been successfully accomplished with good accuracy, as has been reported. The soil type has been determined for sand, loam, and clay types of soil in laboratory conditions. In the case of soil moisture, laboratory experiments were performed on the previously mentioned soil types. Also, results of experiments in open field conditions have been presented showing the effectiveness of the *MCFD/NN* approach. It was shown that the method of *MCFD* was able to capture the relevant characteristics of the field under study under varying moisture conditions, allowing the NN to obtain an adequate mapping for the field as shown in figure 6.10. The average moisture difference obtained was 1.63 %.

It is concluded from this result that after a suitable model has been obtained using the *MCFD/NN* approach, that model is valid over an undetermined period of time. This is particularly useful in the area of agriculture where fast, accurate soil moisture needs to be obtained at a low cost.

In the following chapter, a methodology to reduce the effects of clutter (non-soil items) will be presented. The method is based on the fuzzy c-means clustering algorithm and it is intended to be employed when continuous measurements are to be taken with the GPR connected at the back of a moving vehicle.

7 Clutter Reduction using Fuzzy C-Means Clustering

Abstract The acquired Ground Penetrating Radar (*GPR*) signal is a combined reflection of the soil and clutter responses. In this research, the interest is to capture the actual soil response, reducing the clutter contribution within the soil, such as rocks, grass, roots, etc. This separation is accomplished using Fuzzy C-Means clustering method.

The proposed clutter reduction algorithm works by creating clusters of the many backscattered wavelets over an area of land, size adjustable by the user. From this pool of signals, a predetermined number of clusters are produced, out of which only one is selected as a *clean* signal, based on its shape characteristics. The remaining clusters are considered to be *clutter clusters*. The centroid of the cluster containing the clean signal is used as a benchmark to extract the clutter signal from other centroids. The result is a clutter characteristic response (*CCR*). The *CCR* is used to clean the received signal. This algorithm makes no prior assumptions on the clutter or soil response. The computation of the clusters, centroids, and *CCRs* are renewed as often as the user deems necessary, based on the nature of the field whose moisture is to be determined.

7.1 Review of Fuzzy Logic

In this section, a review of the fundamentals of fuzzy logic is given. Although this material can be found in the literature, such as [26], it is reproduced here to make the chapter self-contained. Fuzzy logic grows from the fact that when making decisions, we rarely have hard-limited decisions. Instead there's a certain degree of imprecision or smooth transitions when deciding between two options. Fuzzy logic tries to mathematically describe reasoning by accounting for that imprecision with the use of fuzzy rules, which are a combination of fuzzy sets. Fuzzy sets are described next.

7.1.1 Fuzzy Sets

A set is defined as a collection of objects which can be treated as a whole, such that an item from a given universe is either a member or not [26]. A set is completely characterized by its members. A fuzzy set is a modification of an ordinary set that allows partial membership. A grade of membership is specified such that the transition between membership and non-membership is gradual, not abrupt. The degree of membership of all the elements is what describes a fuzzy set.

Mathematically, a fuzzy set S of a set X is a mapping $S: X \rightarrow [0,1]$. Thus, *x* belongs to S with *S(x)* degree of membership. The function *S(x)* is referred to as the membership function. The region of support of the function $S: X \rightarrow [0,1]$ is given by the set $R(S) = \{x \in X \mid S(x) \neq 0\}$. A *normal* fuzzy set is a set $S: X \rightarrow [0,1]$ where there is an $x \in X$ such that $S(x) = 1$. A

convex fuzzy set is a set $S:\Re \to [0,1]$ where it happens that given $x \le y \le z$ it must be true that $f(y) \ge f(x) \wedge f(z)$. A *fuzzy number* is a fuzzy set $S:X \to [0,1]$ that is normal and convex, with finite support.

7.1.2 Membership Functions

The membership function of a fuzzy set is a function $S:X \to [0,1]$. The membership function gives a number from zero to one to every element in *X*, indicating to what extent that element belongs to *S*. Some common membership functions are the triangular, trapezoidal, Gaussian, and sigmoidal Z- and S- functions.

The triangular function with endpoints (*a*,0) and (*b*,0), and high point (*c*,α) is given by,

$$S(x) = \begin{cases} \alpha\left(\frac{x-a}{c-a}\right) \text{ if } a \le x \le c \\ \alpha\left(\frac{x-b}{c-b}\right) \text{ if } c \le x \le b \\ 0 \qquad \text{otherwise} \end{cases} \qquad \textbf{7.1}$$

The trapezoidal function with endpoints (*a*,0) and (*b*,0), and high points (*c*,α) and (*d*,α) is given by,

$$S(x) = \begin{cases} \alpha\left(\frac{x-a}{c-a}\right) & \text{if } a \le x \le c \\ \alpha & \text{if } c \le x \le d \\ \alpha\left(\frac{x-b}{d-b}\right) & \text{if } d \le x \le b \\ 0 & \text{otherwise} \end{cases} \qquad \textbf{7.2}$$

The Gaussian function with mean c and variance σ^2 is given by,

$$S(x) = e^{-\frac{(x-c)^2}{2\sigma^2}} \qquad \textbf{7.3}$$

The sigmoidal functions (S- and Z-) are of the following form,

$$S(x) = \frac{1}{1 + e^{-(x-m)\sigma}} \qquad \textbf{7.4}$$

An increasing or decreasing function is determined by the parameter σ, the parameter m determines if the curve is shifted to the right or to the left.

7.1.3 Fuzzy Partitions

Suppose we want to divide de set X into n mutually exclusive partitions $\{S_1, S_2, ..., S_n\}$. Then the following conditions must hold,

- $S_1 \bigcup S_2 \bigcup ... \bigcup S_n = X$
- $S_i \bigcap S_j = 0$ if $i \neq j$

In words, the union of all n partitions must be equal to the original set that was partitioned. Also, as mentioned before, partitions must be mutually exclusive. Complying with the above two conditions gives us *crisp* partitions, meaning that an element of *X* either belongs to partition S_i or it does not belong. Now, when we talk about fuzzy partitions, we are allowing all elements of the set *X* to belong to all partitions with a certain degree of membership. A fuzzy partition is defined as follows,

Definition *A finite set of normal fuzzy subsets* $\{S_1, S_2, ..., S_n\}$ *is a fuzzy partition of a set X if*

- $\sum_{i=1}^{n} S_i(x) = 1 \ \forall \ x \in X$
- Each S_i is normal; that is, for each *i*, $S_i(x_i) = 1$ for some x_i

The first condition assures that the degree of membership, across all partitions, of each element of *X* does not exceed unity. The second condition states that in every partition there must be an element of *X* with full membership to that partition.

Having reviewed the necessary fuzzy logic background, we now move on to discuss the fuzzy c-means clustering algorithm, which is the backbone of the proposed technique for clutter identification and reduction from open field GPR images of land.

7.1.4 Fuzzy C-Means Clustering

The fuzzy c-means (*FCM*) algorithm has its roots in the work of Dunn and Bezdeck It is a fuzzy version of the popular k-means clustering algorithm. In *FCM* clustering each data point is allowed to belong to all clusters with a certain degree of membership, this is accomplished by minimizing the following fuzzy objective function,

$$J_m = \sum_{i=1}^{N}\sum_{j=1}^{C} u_{ij}{}^{m}\left\|x_i - c_j\right\|^2,\ 1 \le m < \infty \qquad \textbf{7.5}$$

Where $m \in \Re$ and $m > 1$, x_i is the *ith* data vector of dimension *d*, u_{ij} is the degree of membership of x_i in the *jth* cluster, c_j is the center of cluster *j* and it is of dimension *d*, and $\|*\|$ is any vector norm. The goal is to find the cluster centers and membership grades, which minimize J_m. It can be shown that the update equation for the cluster centers is given by [26],

$$c_j = \frac{\sum_{i=1}^{N} u_{ij}^{\ m} x_i}{\sum_{i=1}^{N} u_{ij}^{\ m}} \qquad \textbf{7.6}$$

and for the membership grades the update equation becomes,

$$u_{ij} = \frac{1}{\sum_{k=1}^{C} \left(\frac{\|x_i - c_j\|}{\|x_i - c_k\|} \right)^{\frac{2}{m-1}}} \qquad \textbf{7.7}$$

The *FCM* algorithm goes as follows,

FCM Algorithm

- Initialize the membership matrix M with random numbers in the range [0,1].
- *Calculate the cluster centers according to Equation 7.6.*
- *Compute the value of the objective function in Equation 7.5. Stop if the objective function is below a threshold, ε, or if its improvement over the previous iteration is below a tolerance, δ.*
- *Update the membership matrix, M, according to Equation 7.7.*
- *Go to step 2.*

7.2 Fuzzy Clutter reduction Method

Obtaining soil moisture content information in uncontrolled, open field conditions is very important in the fields of agriculture, where soil moisture is critical in order to schedule crop irrigation. Also, the growing enthusiasm, and need for precision agriculture demands soil moisture information to be obtained in a fast, accurate fashion. *GPR* has emerged as a leading technology for this application due to its potential to give accurate soil moisture estimates in a non-invasive manner, thus saving lots of time and manual labor typical of commonly used techniques such as TDR.

Inherent in all open field scenarios is the presence of undesired signals, referred to as clutter. In the problem of determining the moisture content of the soil using *GPR*, clutter is identified as all non-soil reflections, such as reflections coming from patches of grass, its roots, rocks lying on top of the ground as well as underground, and any other such items. Since our interest is in obtaining the moisture content from the top layer of the soil, our signal of interest is the so-called *ground bounce* signal. Note how clutter is application specific, for example, in the problem of land-mine detection, the ground-bounce signal would be considered as clutter.

Consider the scenario where the *GPR* is connected to the back of a moving vehicle, collecting data as the vehicle moves over a track of land, as shown in figure 7.1

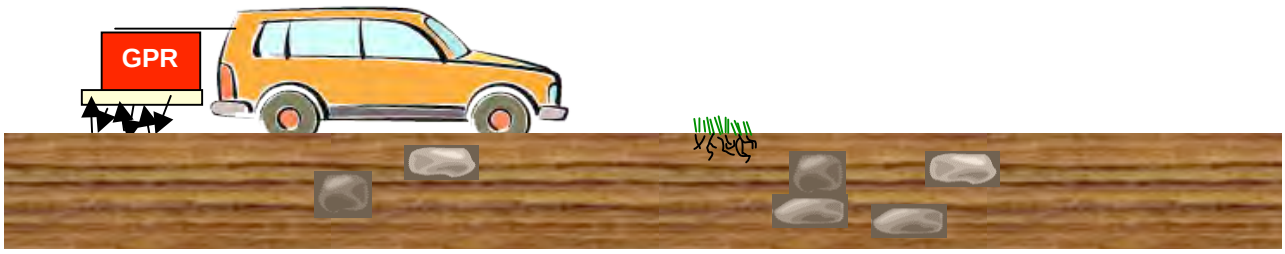


Figure 7.1 Vehicle mounted GPR unit scanning a track of land

As the vehicle moves forward, the *GPR* scans the field and receives data both from uncluttered (clean), and cluttered soil. Our goal is to identify and remove the clutter contribution.

The proposed approach uses *FCM* to create clusters of the many backscattered signals over an area of land, size adjustable by the user. From this pool of signals, a predetermined number of clusters are produced, out of which only one is selected as *clean* signal, based on its shape characteristics. The remaining clusters are considered to be *clutter clusters*. The centroid of the cluster containing the *clean* signal is used as a benchmark to extract the clutter signal from other centroids. The result is a clutter characteristic response (*CCR*). The *CCR* is used to clean the received signal. This method makes no assumptions on the clutter signal. Note that as the vehicle travels along the field, the clutter will change randomly, so this procedure is renewed as often as necessary, genuine *CCRs* representative of the area that is being scanned.

The received GPR signal can be expressed as

$$y(t) = x(t) * \left(h_s(t) + h_c(t)\right) \qquad \textbf{7.8}$$

Where *x(t)* is the input signal, $h_s(t)$ is the soil response, and $h_c(t)$ is the clutter response, * represents a convolution operation. A *clean* signal is written as $s(t) = x(t) * h_s(t)$; hence the *CCR* is found by subtraction,

$$CCR = y(t) - s(t) \qquad \textbf{7.9}$$

The proposed algorithm is as follows,

Fuzzy Clutter Reduction Algorithm

- *Collect data by scanning the field for a specified distance. This distance should be small enough so that it has approximately constant moisture content.*
- *Use FCM to compute a feasible number of clusters*
- *Choose the cluster center, which has the highest correlation (positive or negative) with the reflection obtained from the air, as the clean signal.*
- *For each cluster center, except the "clean" cluster center, calculate the CCR using Equation 7.9*
- *For each cluster, use the CCR calculated in step 4 to remove the clutter subtracting the corresponding CCR from each signal in the cluster.*

7.3 Results

An experiment was performed at the UPRM baseball field to make preliminary tests of the proposed algorithm. Measurements were taken at adjacent locations ranging from clean to cluttered conditions. This test was done manually moving the antenna from one location to the next at discrete steps. One GPR scan was taken at each location.

Figure 7.2 presents three wavelets collected during the experiment. The signals shown in figure 7.2 are representative of the type of signals observed during the experiment. Notice that cluttered signals are not just cluttered, they can be slightly cluttered, as in (c), moderately cluttered, as in (b), all the way to very cluttered, as in (d). This is the reason why the FCM algorithm is employed. A total of twenty signals were used in this test. As an example, two clusters were specified and the proposed algorithm was run. The decision to select only two clusters was based on the partition coefficient, which is a cluster validity index proposed by Bezdeck [25], $PC = \frac{\sum_{k=1}^{n}\sum_{i=1}^{c} u_{ik}^{\ 2}}{n}$. Where n is the number of vectors and c is the number of clusters. This validity index measures the amount of overlap between clusters. A valid number of clusters is found by solving $\max_{c}\left\{\max_{\Omega_c}\{F\}\right\}$, $c = 2,3,...,n-1$. In essence, this index measures the amount of fuzziness.

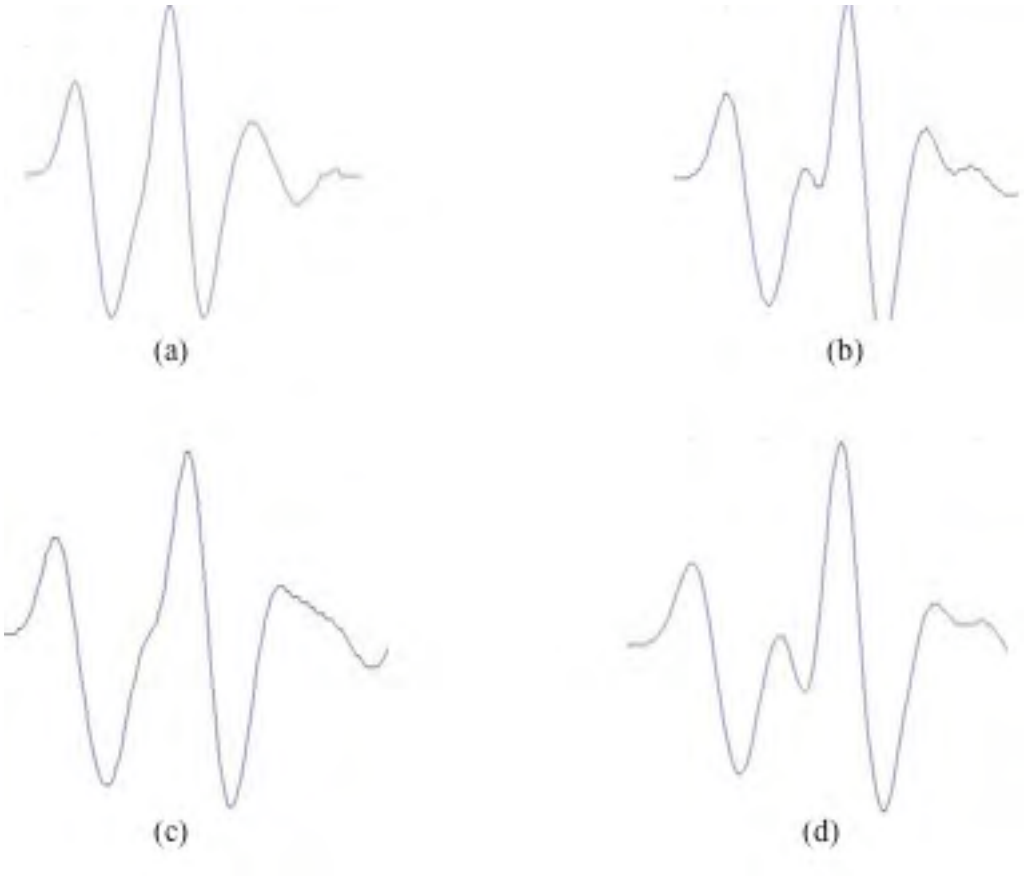
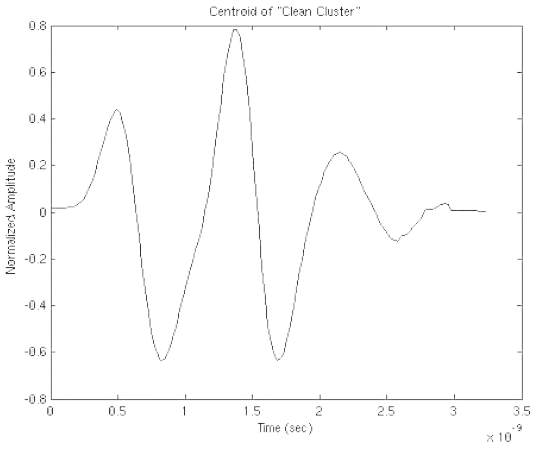
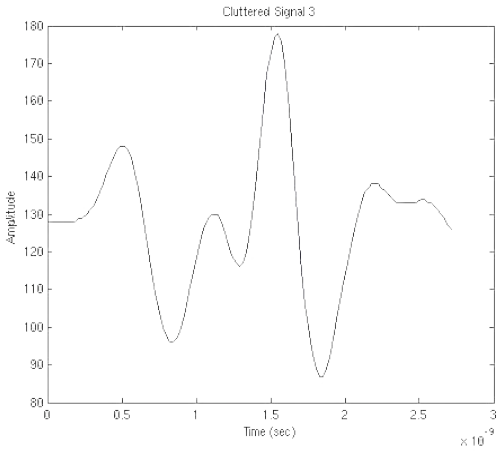


Figure 7.2 Representative wavelets. (a) clean wavelet. (b) cluttered wavelet. (c) semi-clean wavelet. (d) cluttered wavelet

Applying the proposed methodology to the data that was collected during the experiment resulted in the two cluster centers depicted in figure 7.3. The plot to the right represents the center of the *clean* cluster, while the plot on the right represents the center of the *cluttered* cluster. As shown, we were able to separate *clean* signals from *cluttered* signals. Next, we present the extracted *CCR*, computed using Equation 7.9, along with the signal obtained after eliminating the *CCR.* These are shown in figure 7.4.

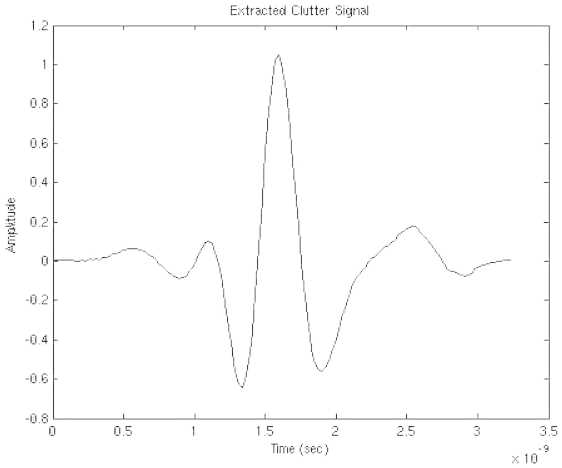


(a)

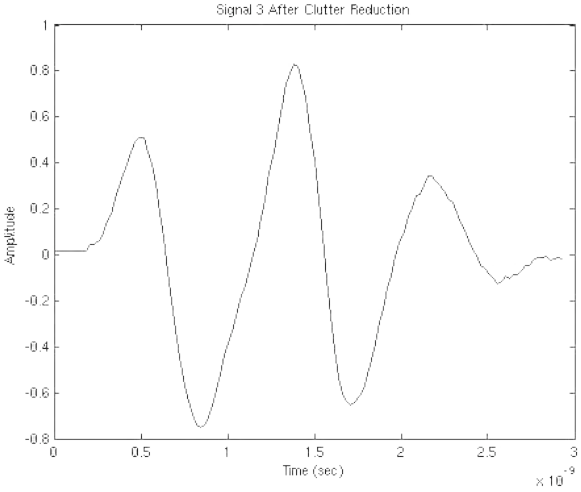


(b)

Figure 7.3 Cluster centers. (a) "Clean" cluster center. (b) "Cluttered" cluster center.



(a)



(b)

Figure 7.4 (a) *CCR*. (b) Signal after eliminating the *CCR*.

Comparing figure 7.3.a with figure 7.4.b we see that the proposed method was able to successfully identify and eliminate the clutter contribution from the signal shown in figure 7.3.b. Note that this algorithm is renewed as often as necessary to account for clutter variations, i.e. amount of clutter, and the type of clutter that is present in the area that's being surveyed.

7.4 Summary and Conclusions

In this chapter we have reviewed some necessary background on fuzzy logic, en route to presenting a methodology for clutter extraction and removal from open field *GPR* images of land with the purpose of obtaining soil moisture content information. The proposed algorithm is based on *FCM* clustering. Albeit the proposed algorithm has only been preliminary tested, it shows promise to obtain genuine clutter characteristics. Further tests must be performed before making conclusions on the feasibility of the algorithm, such as surveying a small portion of open land with the *GPR* actually connected to the back of a vehicle, and employing the *MCFD/NN* method to obtain the soil moisture content.

8 Conclusions and Future Work

8.1 Conclusions

This thesis has presented an approach to determine the soil type and moisture content using *GPR*. The approach, *MCFD/NN,* differs of typical methods in that it does not assume a prior model for the data, rather, we employ a neural network to obtain the desired information. Also, we operate the *GPR* in monostatic mode, thus reducing the time needed for data acquisition as compared to common methods such as the multi-offset method presented in [6].

It has been shown that a distinctive signature obtained from the *ETFE* can be used as feature vector for the *NN* to determine its soil type or moisture content. Some of the advantages of the new approach over traditional methods are,

- After a *NN* has been trained, the output to a given testing set is obtained almost immediately. Since *NNs* can be trained off-line, this type of system is well suited for on-line implementations. This fact is advantageous for agricultural applications since the user would be able to obtain a near real-time moisture map of the desired field. Physical approaches, in contrast, are more computationally demanding.
- It is a well-known fact that *NNs* are less sensitive to noise than other methods, since they learn the underlying pattern in the data. This is an advantage over

model-based methods, which requires accurate forward modeling to obtain the improved estimates that they are known for. For rapid soil moisture determination in open field conditions, these model-based methods have not been proven. Even with the state of computing power, they require much computation time. Some work is being done by Lambot, [24], in this area.

Some disadvantages of *NN* based systems include:

- Since *NNs* are "black-box" models their performance depends on the quality, and completeness of the training data. A *NN* cannot extrapolate out of the training set. This becomes a disadvantage when compared with physical approaches, since the latter are mathematical equations specifically derived to solve the problem at hand.
- If greater accuracy is demanded, *NNs* may fall behind the physical approach. This stems from the fact that, as the word implies, a physical model tries to mimic, as closely as possible, the actual physical phenomena that's taking place. If the model is realistic enough, then the results are surely to be good.

Our results can be summarized as follows, in the area of soil type determination it was shown that the *MCFD/NN* is able to discriminate between soil types. Experiments were done in laboratory conditions to distinguish between sand, loam, and clay type of soils.

In the area of soil moisture content determination, laboratory, and open field experiments were presented and discussed. Results show good agreement between the MCFD/NN moisture and Theta Probe readings. In particular, the most interesting results were obtained from the open field experiment, when the NN was trained and then tested with data taken six months later. The results showed acceptable consistency with the Theta meter, with an average moisture difference of 1.34%. This gives us a clue on the validity of the *MCFD/NN* method to obtain soil moisture information with good temporal accuracy.

The last chapter of this thesis presented a preliminary methodology to obtain the genuinely clutter signal that's inevitable in uncontrolled environments. The method based on the *FCM* algorithm shows potential, but needs to be tested in a real situation with moisture content estimates based on the output signals of this algorithm.

8.2 Future Work

Future work in the area of soil moisture determination based on the *MCFD/NN* approach needs to concentrate on the following,

- Clutter identification and suppression.
- Further evaluation of its open field performance i.e., more testing data for an already trained *NN*.
- Extension to a vehicle-mounted, off-ground, experimental setup.
- The creation, and validation of a moisture map for a chosen field.

REFERENCES

[1] G.C. Topp, J.L. Davis, and A.P. Annan "Electromagnetic determination of soil water content: Measurements in coaxial transmission lines." Water Resour. Res., 16:574-582, 1980.

[2] G.C. Topp, J.L. Davis, and A.P. Annan "Electromagnetic determination of soil water content using TDR I: Application to wetting fronts and steep gradients." Soil Sci. Soc. Am. J., 46:672-678, 1982.

[3] G.C. Topp, J.L. Davis, and A.P. Annan "Electromagnetic determination of soil water content using TDR II: Evaluation of installation and configuration of parallel transmission lines." Soil Sci. Soc. Am. J., 46:678-684, 1982.

[4] G.C. Topp, and J.L. Davis "Measurements of soil water content using time-domain reflectrometry(TDR) a field evaluation". Soil Sci. Soc. Am. J., 49:19-24, 1985.

[5] Dasberg, S. and F.N. Dalton "Time domain reflectrometry field measurements of soil water content and electrical conductivity." Soil Sci. Soc. Am. J., 49:293-297, 1985.

[6] Huisman, J. A., Hubbard, S. S., Redman, J. D., Annan, A. P. "Measuring Soil Water Content with Ground Penetrating Radar: A Review." Vadose Zone Journal 2:476-491, 2003.

[7] S.S. Hubbard ,Grote,K., and Y. Rubin, "Mapping the volumetric water content of a California vineyard using high-frequency GPR ground wave data." Leading Edge Explor. 21:552-559,2002.

[8] S. Hubbard, K. Grote, M.B. Kowalsky, and Y. Rubin. "High Resolution estimation of near-subsurface water content using surface GPR ground wave information."

[9] A. Chanzy, A. Tarussov, A. Judge, and F. Bonn. "Soil water content determination using a digital ground penetrating radar." Soil Sci. Soc. Am. J., 60:1318-1326, 1996.

[10] K. Grote, S. Hubbard, and Y. Rubin. Field-scale estimation of volumetric water content using ground-penetrating radar ground wave techniques." Water Resources Research, 39, 2003.

[11] Lambot, S., "Hydrogeophysical Characterization of Soil Using Ground Penetrating Radar." PhD Thesis, Université catholique de Louvain, November 2003.

[12] Freeman, J. A., Skapura, D. M, "Neural Networks: Algorithms, Applications and Programming Techniques", Addison-Wesley Publishing Company, 1991.

[13] Dalton, F.N, "Measurement of soil water content and electrical conductivity using time-domain reflectrometry." International Conference on Measurement of Soil and Plant Water Status. Centennial of Utah State Univ., pp. 95-98, 1987.

[14] Dasberg, S. and A. Nadler, "Field sampling of soil water content and electrical conductivity with time domain reflectometry." International Conference on Measurement of Soil and Plant Water Status. Centennial of Utah State Univ., pp. 99-102, 1987.

[15] Topp, G.C., "Electromagnetic determination of soil water content: measurements in coaxial transmission lines." Water Resour. Res., 16:574-582, 1980.

[16] Topp, G.C., "The application of time-domain reflectrometry (TDR) to soil water content measurement." International Conference on Measurement of Soil and Plant Water Status. Centennial of Utah State Univ., pp. 85-94, 1987.

[17] Guy Serbin and Dani Or. "Ground-penetrating radar measurement of soil water content dynamics using a suspended horn antenna." IEEE Trans. Geoscience and Remote Sensing, 42:1695-1705, 2003.

[18] Simon Haykin. "Adaptive Filter Theory." Prentice Hall, 2002.

[19] Leon Peters Jr., Jeffrey J. Daniels, and Jonathan D. Young. "Ground-penetrating radar as a subsurface environmental sensing tool." Proceedings of the IEEE, 82:1802-1822, 1994.

[20] Hamed Parsiani, Pedro Torres, and Maritza Torres. "Material characteristics in Fourier domain (*MCFD*) formulation, a signature to determine soil type, moisture, and vegetation health based on multilayer ground penetrating radar reflection." Proceedings of IASTED SIP, 2005.

[21] D.J. Daniels. "Surface-Penetrating Radar." The Institution of Electrical Engineers, 1996.

[22] "SIRveyor SIR-20." Geophysical Survey Systems, 2002.

[23] Lennard Ljung. "System Identification: Theory for the User." 2nd Ed. Prentice Hall, 1998.

[24] S. Lambot, L. Weihermller, I. van den Bosch, M. Vanclooster, and E.C. Slob. "Full-wave inversion of off-ground monostatic GPR signal focused on the surface reflection for identifying surface dielectric permittivity." Proceedings of the 3rd International Workshop on Advanced Ground Penetrating Radar, 2005.

[25] J. C. Bezdeck, "Cluster Validity with fuzzy sets." J. Cybernet. vol. 3, 3:58-72, 1974.

[26] Jan Jantzen, "Tutorial on fuzzy logic" Tech Report, U. of Denmark, Dpt. Of Automation, 1998.

[27] Enrico Mattei and Hamed Parsiani, "Soil Moisture Determination based on Radar Response and Artificial Neural Networks" WSEAS Trans. Signal Processing, Issue 2, Vol 1, pp 276, November 2005.

[28] Hamed Parsiani, Enrico Mattei, "Open Field Soil Moisture Measurements with Radar", Proceedings of the 1st WSEAS International Conference on Remote Sensing, November 2-4 2005, Venice, Italy.

[29] Hamed Parsiani, Enrico Mattei, Allen Lizarraga, Mairim Ramos, "Soil Moisture Determination based on MCFD/NN-GUI Algorithm using Wideband Radar Images of Land", Proceedings of the 7th IASTED Internacional Conference on Signal and Image Processing, August 15-17 2005, Honolulu, Hawai.

[30] Eric Harmsen, Hamed Parsiani, Maritza Torres, “Evaluation of Several Dielectric Mixing Models for Estimating Soil Moisture Content in Sand, Loam, and Clay Soils”, ASAE Annual Internacional Meeting, July 26-30 2003, Las Vegas, Nevada.

[31] Yoshihiko Hamamoto, Shunji Uchimura, Shingo Tomita, “On the Behavior of Artificial Neural Network Classifiers in High-Dimensional Spaces”, IEEE Trans. On Pattern Análisis and Machine Intelligence, Vol. 18, No. 5, May 1996.