1-1-2021

ELECTROMAGNETIC SURVEY DESIGN USING ARTIFICAL NEURAL NETWORKS

SHARIFUL ISLAM SHARIF
University of Mississippi

ELECTROMAGNETIC SURVEY DESIGN USING ARTIFICAL NEURAL NETWORKS

A Thesis

Presented for the partial fulfillment of requirements

for the Degree of Master of Science

in the Department of Civil Engineering

The University of Mississippi

Shariful Islam Sharif

August 2021

ABSTRACT

Acquiring geophysical information requires selection of the geophysical method based upon the defining physical property and a survey design with adequate resolution but cost effective based upon the size of the area to be surveyed. The objective of this study is to use artificial neural nets (ANNs) to design an optimal survey. The developed approach is tested for the case of soil pipe surveying using electromagnetics. Soil pipes are tortuous voids located within 1.5 m depth of the ground surface. They trend perpendicular to the slope and have cross-sectional dimensions on the order of millimeters to tens of centimeters. The contrast in electrical conductivity (EC) is significant especially if the soil pipe is filled with air. Based upon these characteristics an EM38B is chosen to survey the area. The EM38B is relatively fast and its maximum exploration depth is approximately 1.5m. The measured apparent electrical conductivity (ECa) is a dipole dependent weighted average over a soil volume of about $1m^3$. A benchmark high resolution survey was conducted having a 2D cross grid pattern with a 0.5m line spacing to ensure an overlap of soil volume being interrogated. The benchmark data set is then decimated (7 options) based upon orientation and line spacing options to simulate various surveying patterns. ANN models are developed using the various reduced datasets. The quantile method is used to generate a table to guide the choice of survey for a given ECa range. To validate the concept, an exercise is conducted starting with a reconnaissance survey consisting of a few lines based on surface features of soil

pipes. Using the table as a guide, a survey plan is proposed and the ANN models are created using this data set. The measured and model generated data are used to create the 2D ECa map using kriging interpolation. This map is in good agreement with the benchmark ECa map, although the second map required 60% less data.

DEDICATION

This work is dedicated to my father Mohammad Abdul Latif, to my mother Hosneyara Begum, my strong pillar, my source of knowledge and understanding.

To my dearest wife Fatema Tuz Johora and my daughter Sidrat Al Sharif,

Thank you for the love, support, and encouragement.

ACKNOWLEDGMENTS

I would like to express my sincere gratitude to my research advisor Dr. Craig Hickey, Director of the National Center for Physical Acoustics and Research Associate Professor of Physics and Geological Engineering, for his continuous encouragement and support over the years. His insight and guidance were invaluable in completing this research.

I would also like to express my sincere gratitude to my academic adviser and committee member Dr. Yacoub Najjar, Professor of Civil Engineering Department, for his valuable contribution, support, and encouragement. He has devoted his valuable time and energy in helping me navigate my graduate studies. I would also like to thank Dr. Hakan Irfan Yasarer, Assistant Professor of Civil Engineering Department, for his interest and valuable comments on my work.

Dr. Leti T. Wodajo, Md. Lal Mamud, and Md. Abdus Samad helped me during data collection for this study. Dr. Louis Zachos, Emeritus Associate Professor, Geology and Geological Engineering, helped me a lot in ArcGIS modeling. I would like to thank them for their cooperation.

This work was supported by the U.S. Department of Agriculture under Non-Assistance Cooperative Agreement 58-6060-6-009.

Thank you!

Shariful Islam Sharif

LIST OF ABBREVIATIONS AND SYMBOLS

ANNs	Artificial Neural Networks
EC	Electrical Conductivity
ECa	Apparent Electrical Conductivity
TR	Training
TS	Testing
ASE	Average Square Error
MARE	Mean Absolute Relative Error
R^2	Coefficient of Determination
GUI	Graphical User Interface
EM	Electromagnetic
MASW	Multichannel Analysis of Surface Waves
USDA	United States Department of Agriculture

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 Motivation of Research

Geophysical methods are cost effective methods for estimating spatial geological structure and composition of the earth. Furthermore, these methods can be repeated without destruction of the surface allowing for monitoring over time. Numerous standardized data-acquisition procedures are used to interpret desired subsurface parameters. Some major decisions that must be made before data acquisition are: suitability of instruments for the survey, selection of a traversing pattern, the size of the grid and its orientation, orientation and height of the instrument, and spacing between the measurements. Although less than a boring program, cost is still a major consideration in geophysical surveying. Finding a procedure that acquires the optimum amount of data is necessary to maintain a favorable cost/benefit ratio Appropriate survey design is therefore critical to the cost of the experiment in terms of the robustness, accuracy and precision of recovered geological information (Maurer et al., 2010). The acceptable survey design should ensure acquisition of required data that best resolve specific subsurface features or parameters of interest (Maurer and Boerner, 1998; Curtis and Maurer, 2000).

Artificial neural network (ANNs) are algorithms and mathematical models to reproduce the knowledge acquisition and information processing of the human brain (Zupan and Gasteiger,

1993). ANNs work as an alternative to classical mathematics and traditional techniques and have the capability to incorporate nonlinear issues. Feed forward-back propagation approaches in ANN modeling is the most common modeling procedure that is successfully used in engineering applications (Ghaboussi et al.,1991; Najjar et al.,1996; Yasarer and Najjar, 2010). Application of ANN modeling in geophysics is also reported in the literature (Poulton et al., 1992; Roth and Tarantola, 1994; Langer et al., 1996 and McCormack et al., 1993). Due to the continuous modification and improvement of ANN modeling, it's performance should be evaluated for challenging and complex problems. The motivation behind this research is the application of the ANN modeling approach to minimize the data collection effort for electromagnetic surveying of soil pipes.

1.2 Research Objectives

The geophysical method needs to be chosen based upon the properties of interest. Instrument selection and survey design depend upon the size of the area to be surveyed and the required resolution of the survey, which determine the acquisition time and therefore the primary cost of the survey. Application of ANNs have gained popularity in geophysics but no evidence has been found where ANNs are used for designing EM surveys. In this study, the advantage of using ANNs to obtain a cost-efficient EM data acquisition geometry for mapping subsurface soil erosion pipes is investigated. These soil pipes are a major contributor to the formation of gullies resulting in loss of agricultural farmland.

1.3 Work Scope

The thesis is composed of six chapters. The scope of each chapter is described as follows:

Chapter 1-Introduction: This chapter briefly presents the motivation of the research, research objective, and scope of work.

Chapter 2-Literaure Review: This chapter describes geophysical survey design parameters, measurement of apparent electrical conductivity (ECa) of soil using an EM38B, overview of ANNs, and application of ANN methods in geophysics.

Chapter 3- Electromagnetic Survey at Goodwin Creek, MS: Presents a general description of the site, selection of suitable EM equipment, data acquisition and processing, EM signatures of soil pipes and a plan view of soil pipes.

Chapter 4-Survey Design using ANNs: In this chapter, ANNs are used to choose appropriate survey orientation and line spacing.

Chapter 5-Survey Design from a Reconnaissance Survey: This chapter describes how the concept of designing and optimized EM survey beginning from an initial reconnaissance survey.

Chapter 6-Conclusion: Presents a summary of the work and provides recommendations for future research.

CHAPTER 2

LITERATURE RIVEW

2.1 Introduction

Geophysical surveys have become popular for near surface applications because they are cost and time effective and measurements can be easily repeated without altering the soil profile. Among the available geophysical methods, electromagnetic induction (EMI), ground penetrating radar (GPR), electrical resistivity (ER), seismic refraction, and multichannel analysis of surface waves (MASW) have been the most widely applied in soil studies. In this chapter, geophysical survey design parameters, measurement of apparent electrical conductivity (ECa) of soil using an EM38B, description of artificial neural network system, and application of neural network methods in geophysics are introduced.

2.2 Geophysical Survey Design Parameters

Major decisions must be made before conducting any geophysical survey. These are: choice of instruments for the survey, selection of a traversing pattern, determining the size of the grid and its orientation, in some cases the orientation and height of the instrument, and the spacing between the measurements. These preliminary decisions about geophysical surveys are described in this section.

2.2.1 Choice of Suitable Geophysical Instrument

A general introduction to the different geophysical instruments is shown in Table 2.1 (Clark, 1996).

Table 2.1: List of geophysical instruments

Type of Instrument	Application
Metal Detector	These instruments are inexpensive and fast. They are very efficient for determining the location of metallic objects.
Ground Penetration Radar	Ground Penetration Radar (GPR) is capable for providing estimates of the relative depths of features. 3D images of underground features can be created using GPR.
Conductivity Meter	This instrument is suitable for distinguishing different types of soil i.e. silt and sand. Sometimes it is used for distinguishing stone from soil. They are capable for detecting significant interference from lightning and electrical wires.
Resistivity Meter	Resistivity meter is excellent for distinguishing soils and also soils from rocks similar to conductivity meter. A resistivity meter is slower but less affected by noise than a conductivity meter.
Seismic	Speed of seismic refraction survey is slow but it shows excellent performance for the estimations of the depth to bedrock. It is also suitable for determining the location holes refilled with soil.

The decision about which instrument to use should be based on what types of features are of interest. Based on the description of survey site one instrument may be best, while another parameter may indicate that another instrument should be used. The decision should be made based on which physical property is most important and follow that recommendation. It is wise to evaluate several parameters rather than just one.

A geophysical survey may be done on a single site using different instruments. This allows a greater variety of features to be detected. The findings of such surveys will be more certain where

more than one instrument detects the same feature. But the multi-instrument survey increases the project cost. A survey with three different instruments may not triple the cost of a geophysical survey, but this survey will probably cost over twice that of a survey done with a single instrument (Bruce, 2004).

2.2.2 Selection of Traversing Pattern

During a geophysical survey, an instrument is carried, pulled, or pushed about the area of interest and measurements are made at many points. These measurements are often compiled into cross-sections or maps that show the distribution of values. There are many different methods for exploring an area and interpolation of data. Figure 2.1 represents typical traversing patterns used in geophysical surveying. Reconnaissance is done initially to look at the survey area. Most surveys are done with a rectangular pattern. A radial pattern is followed to explore around a known feature. A polygonal pattern is used when it is impossible to walk in the survey areas. A point-by-point pattern is used to measure specific selected points and a circumferential pattern is excellent for a tall conical mound.

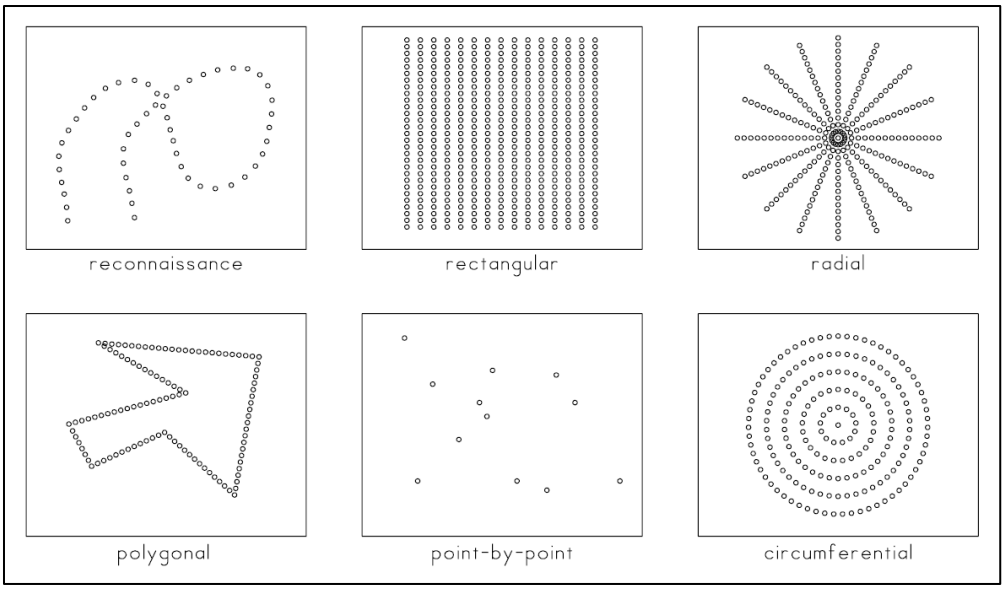


Figure 2.1: Typical traversing patterns of geophysical survey (Bruce, 2004).

2.2.3 Instrument Setup

Survey orientation and height of the instrument can change the readings of some geophysical instruments. Usually the instruments are kept close to the surface of the ground to get the greatest spatial resolution of underground features. In the presence of brush or grass at the surface, the instruments may need to be raised. Typical orientation and height of common geophysical instrument is shown in Table 2.2.

Table 2.2: Typical orientation and height of common geophysical instrument

Type of Instrument	Orientation	Height
Magnetometer	The sensors require a special orientation to allow the instruments to operate with minimum noise.	Heights for the lowest sensor range between 30 to 50 cm. The full range of heights is between 0 to 2 m. The greatest heights are best for deep features.
Ground Penetration Radar	Radars to be operated with the electrical dipoles of the antennas perpendicular to the line of traverse.	Usually, GPR should be placed 5 cm above the surface.
Conductivity Meter	There are two different orientations i.e. rotation about a vertical axis and rotation about a horizontal axis.	Directly on the surface or waist height based on instrument.
Resistivity Meter	The orientation is determined by the direction of the survey line of the moving electrodes. To detect thin and linear features, this orientation is very effective.	In general, spacing is equal to the anticipated depth of exploration. Deeper features must be detected by increasing the spacing of electrodes.
Seismic	Orientation has no major effect on survey.	The geophones are usually buried to reduce acoustic noise. The spacing between geophones is typically in the range of 0.25m to 2 m. The spacing affects spatial resolution. The depth of exploration is affected by distance between the impact point and the farthest geophone.

2.2.4 Size of Grid and Orientation

For covering a large survey site with a rectangular traversing pattern, the area needs to be broken up into smaller square, rectangular and parallelogram areas. These smaller square or rectangular areas are called grids. It is possible to locate points correctly in a parallelogram, but creating rectangular grids is much faster. Grids should be oriented in a way to reduce survey effort.

Typically, the side of any grid can be between 10 and 100 m. 20-m square grids are very common and can be measured in a fraction of a day. Lines should be short enough to see from one end of the grid to the other. But larger grids are more efficient because they reduce the number of survey lines. Extra time is not required to set up the instrument for each new line. For large survey sites, it is fastest to cross the length of it with a single long traverse if it is less than about 200 m. One directional survey along the grid lines is good practice. The standard orientation is to run the survey towards true north. Alignment with magnetic north is also acceptable.

2.2.5 Spacing Between Measurement

Selecting appropriate line spacing for geophysical measurements reduces cost and time. The smallest spacing is based upon the size of the smallest features. In general, the spacing between successive measurements for any geophysical survey is usually between 0.5 m to 2 m (Bruce, 2004). If the mobility of the survey equipment increases, the spacing between successive measurements on a survey line can be decreased. For example, GPR instruments typically make depth scans at 1 cm intervals. Measurements on excavated features require 0.2 - 5 cm intervals but for boreholes it is usually between 5 - 10 cm.

2.3 Measurement of Apparent Electrical Conductivity (ECa) of Soil using an EM38B

Soil electrical conductivity is an indirect measurement which correlates with different soil physical and chemical properties. It is the ability of a material to transmit an electrical current and is commonly expressed in units of mS/m (milli-Siemens per meter). Figure 2.1 represents the conductivity values of different soil materials.

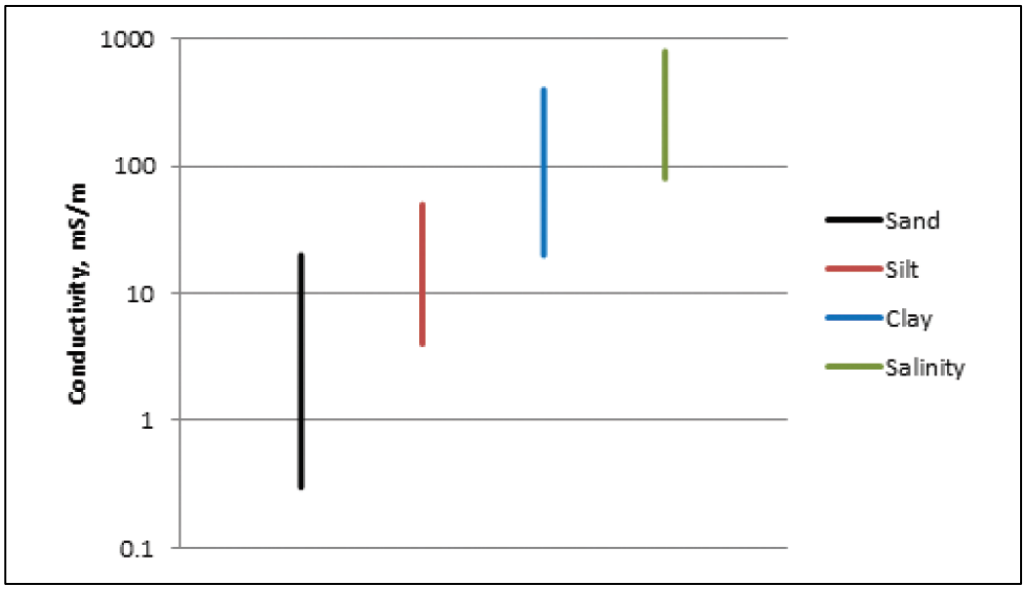


Figure 2.2: Conductivity values of different soil materials. (LSU AgCenter Pub. 3185 What Is Soil Electrical Conductivity?)

Soil EC is controlled by soil water content, clay content and mineralogy, and soil temperature (McNeill, 1980). These controlling factors have led to considerable interest in EM techniques to determine soil salinity (Nettleton et al., 1994; Johnston et al., 1997; Lesch et al., 1998). Other soils related uses of EM include measurement of clay content or depth to clay-rich layers (Williams et al., 1987; Doolittle et al. 1994), the depth of flood deposited sands (Kitchen et al., 1996) and the depth of splay deposits (Doolittle et al., 1995).

Electromagnetic induction-based measurements is a non-invasive and non-contact measurement technique. The measurement interrogates a finite sampling volume of the soil and

the resulting value is commonly referred to as the apparent conductivity (ECa). The EM-ECa sensor most often used in agriculture is the EM38, manufactured by Geonics Limited of Mississauga, Ontario, Canada (http://www.geonics.com).

2.3.1 Soil Water Content

Sheets and Hendrickx (1995) measured the ECa along a 1950 m transect in New Mexico over a 16-month period and found a linear relationship between conductivity and soil water content. Brevik et al. (1998) found that soil water content had a significant influence on soil ECa. Brevik and Fenton (2002) said that soil water content was the single most important factor influencing EC in central Iowa.

2.3.2 Drainage

Droughty areas typically have distinct textural differences from those with excess water. Soils in the middle range of the conductivity, which are both medium-textured and have medium water-holding capacity, may be the most productive. Since water holding capacity typically has the single greatest effect on crop yield, this is likely the most valuable use of EC measurements. Poor water infiltration can lead to poor drainage, waterlogging, and increased ECa. Jaynes et al., (1995) used ECa as an estimator of herbicide partition coefficients, theorizing that both were responding to changes in the soil drainage class.

2.3.3 Cation Exchange Capacity (CEC)

CEC is related to the percent of clay and organic matter (OM). Mineral soils enriched in organic matter, or with chemical fertilizers (e.g., NH_4OH) have higher CEC than non-enriched soils. As

the percent of clay and organic matter increase, the CEC also increases. Researchers have found a correlation between conductivity and CEC through its relationship to clay. Williams and Hoey (1987) used ECa to estimate within-field variations in soil clay content. McBride et al. (1990) related ECa measurements to CEC and exchangeable Ca and Mg.

2.3.4 Salinity

An excess of dissolved salts (Cl^-, SO_4^{2-}, NO_3^{2-}, PO_4^{3-}, Na^+, K^+, Ca^{2+}, Mg^{2+}, NH^{4+}, etc.) in the soil contributes to salinity which is readily detected by EM equipment. Inadequate leaching and excessive use of fertilizer will cause increases in the EC of soil.

2.3.5 Temperature

As temperature decreases to the freezing point of water, soil ECa decreases slightly. ECa decreases about 2.2% per degree Centigrade due to the increased viscosity of water and decreased mobility of ions. Electrical conductivity decreases sharply when the temperature of soil water is below the freezing point.

2.4 Operation of Electromagnetic (EM) Equipment

The induction of an electromotive force by the motion of a conductor across a magnetic field or by a change in magnetic flux in a magnetic field is called electromagnetic induction. This either happens when a conductor is set in a moving magnetic field (when utilizing AC power source) or when a conductor is moving in a stationary magnetic field. The general equation governing electromagnetic induction equipment is,

$$\frac{Hs}{Hp} = \frac{i\omega\mu ECs^2}{4} \quad \text{(Mcneill,1980)} \qquad \text{Equation 2.1}$$

Here, H_s= Secondary magnetic field at the receiver coil, H_p= Primary magnetic field at the receiver coil, ω= 2πf [f= frequency (Hz)], μ=Permeability of free space, EC= Soil Conductivity (mS/m), s= Inter coil spacing(m), i=√(-1).

Commonly used EM equipment includes EM31, EM34 and EM38. EM equipment has two operational modes - vertical dipole mode and horizontal dipole mode. The inter coil spacing or frequency can be varied to determine the variation of conductivity with depth. The effective exploration depth in a layered earth geometry is approximately 0.25 to 0.75 times the inter coil spacing for the horizontal dipole mode and 0.5 to 1.5 for the vertical dipole mode. The inter coil spacing of the EM 38 and EM 31 are 1m and 3.66m respectively. For the EM34, the inter coil spacing can be 10m, 20m or 40m.

The EM38B is composed of a transmitter coil and a receiver coil installed 1.0 m apart at opposite ends of a non-conductive bar. It operates at a frequency of 14.6 kHz. Figure 2.3 depicts the operation of the EM38B. The transmitter coil is energized with an alternating current from a 9 V battery, which generates a time-varying magnetic field in the earth. This primary magnetic field

(Hp) induces current to flow in the soil. These current loops induce their own magnetic field (Hi). The induced field is superimposed on the primary field and both Hp and Hi are measured in a receiving coil (Rx) at the other end of the instrument. The measurement is expressed by the ratio of the primary magnetic field (Hp) to the secondary magnetic field (Hs = Hp + Hi). It is a function of the different conductivities and the magnetic susceptibility in the subsoil as given by Equation 2.1.

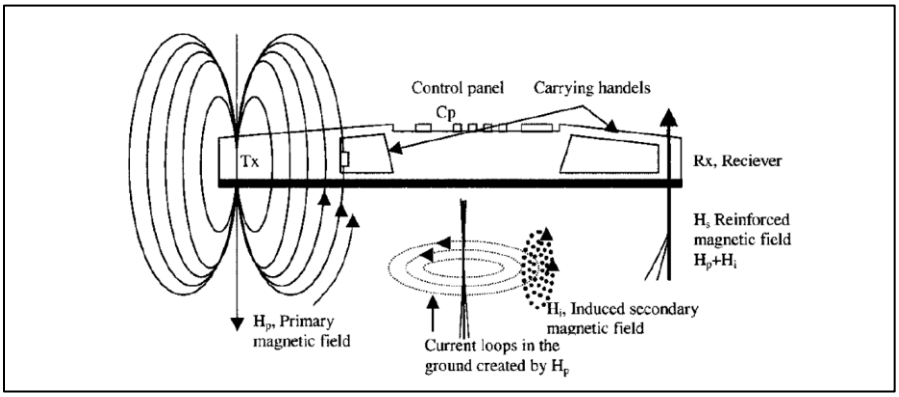


Figure 2.3: Operation of EM38B made by GEONICS (TN-6, GEONICS Limited)

Measurements of ground conductivity can be made with the instrument in either the vertical or horizontal dipole orientation (Figure 2.4). In the horizontal orientation the instrument measures to a depth of about 0.75 m with the greatest sensitivity just under the instrument. With the instrument in the vertical orientation it measures to a depth of about 1.5 m with the greatest sensitivity at about 0.4 m (TN-6, GEONICS Limited). Magnetic susceptibility is a measure of how much a material will become magnetized in an applied magnetic field. This additional parameter, important when searching for metallic objects, is determined using the in-phase part of the signal.

(a)

(b)

Figure 2.4: (a)Vertical dipole mode and (b) Horizontal dipole mode of EM instrument.

The EM38 requires initial calibration and zeroing before the start of data collection. Before the start of a measurement and at various times during the survey, apparent conductivity and in-phase readings should be obtained at a reference location outside the survey area. In EM surveys, drift represents differences in repeat measurements due to thermal distortion (change in temperature) of the coils. The EM38B provides good spatial resolution. It can collect 10 data points per second. For maximum resolution, the measurement stations should be close to a meter. Correct station spacing will be based on an estimate of the lateral dimensions of the anticipated

conductivity anomalies.

2.5 Accuracy Issues in Electromagnetic (EM) Survey

For accurate interpretation of the large amounts of ECa data, it is necessary to understand and consider issues related to how the data were collected and the intended application of the survey. It is well known that soil conditions, including temperature and moisture, influence ECa (McNeill, 1992). In non-saline soils, where the variation in ECa across a field will in general be much smaller, operational differences could be significant. Ambient conditions such as air temperature, humidity, and atmospheric electricity can also affect measurement of ECa with the EM38. Of these, air temperature generally has the largest effect (Geonics,1998).

With the advent of GPS technology, researchers have developed systems to mobilize the EM38 and synchronize its output with GPS positioning data (Carter et al., 1993; Jaynes et al., 1993; Cannon et al., 1994; Kitchen et al., 1996). Mobility of the EM38B system could potentially introduce error in ECa surveys. With the system it is impractical to mount a GPS antenna immediately above the EM38B. The distance between the GPS antenna and EM38B creates a position error, or offset, in the direction of travel. The output of the EM38B is designed for static operation, so operating speed can have an impact on data. Sudduth et al. (2001) found that in the vertical dipole mode ECa changed slightly with increasing operating speed (-0.4 mS/m per m/s). They also showed that the ECa changed about 1%/cm height of the sensor above the ground. They suggested using a uniform speed and minimum height for field surveying.

2.6 Artificial Neural Networks

Artificial neural networks (ANNs) are algorithms and mathematical models that attempt to reproduce the knowledge acquisition and information processing of the human brain (Zupan and Gasteiger, 1993). ANNs work on the principles of biological neural networks. In the human body, the neuron consists of three main elements i.e. soma, dendrites and axons. Neurons are interconnected using axons and dendrites to transfer information. The nervous system is a neural network composed of these interconnected neurons (Simson, 1990). Artificial neural networks (ANNs) systems typically consist of the following basic components (Agrawal and Daiutolo, 1992):

- A neuron or node
- An activation function associated with each node
- A real-valued bias associated with each node
- Transfer function
- Propagation rules
- Learning rules.

Similar to the human neural system, the ANN approach involves the gradual increase of acquired knowledge resulting from long-term experimentation.

2.6.1 Basic Elements of ANNs

Artificial neural networks (ANNs) consist of four main parts:

- Input layer
- Hidden layer(s)
- Connection weights
- Output layer

A schematic diagram of an ANN structure is depicted in Figure 2.5 and each part is described in this section. Usually a network performs three tasks sequentially (Najjar et al., 1996):

- Entry of input variables to input layer
- Information processing within hidden layer(s)
- Output generation in output layer.

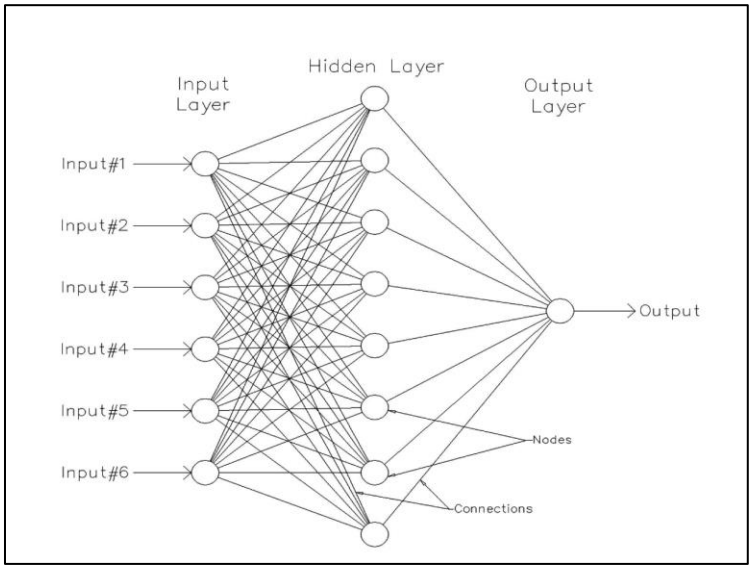


Figure 2.5: Schematic diagram of an ANN structure (Hakan, 2010)

2.6.1.1 Input Layer(s)

The input layer is the simplest layer of a network. It contains the input nodes and no mathematical operations are performed in this layer. The input layer is responsible for receiving, processing and forwarding information to the hidden nodes. Input variables depend on the number of input nodes which are assumed to influence the output. The performance of the network is influenced by the number of input variables.

2.6.1.2 Hidden Layer(s)

The hidden layer may contain one or more layers consisting of a set of nodes which processes information within the network body. The hidden layer processes the information passed on from the input layer and feeds it forward towards the output layer. In other words, it facilitates the flow of information between the input nodes and the output node via the connecting links. The accuracy of the developed models is considerably affected by the number of the hidden layers as well as the number of neurons involved within each layer.

2.6.1.3 Connection Weight(s)

Connection weights act as the interconnecting links between the neurons in the layers. Each neuron is connected to every other neuron in the next layer via individual links. The magnitude of connection weights is responsible for adjusting the output magnitude of the neuron. No side connections are used in this modeling procedure.

2.6.1.4 Output Layer

One or multiple neurons can be found in the output layer of a network. The output neuron is responsible for computing a value for a certain parameter or variable.

2.6.2 Transfer Functions

While calculating the output of a neuron, a transfer function is required because the input could be very large or negative. To introduce nonlinearity in the model and to avoid large or negative values, an additional nonlinear transformation is required in the neuron's input to produce an output. Various types of transfer functions are used in ANN.

The sigmoidal function is one of most widely used functions in ANN modeling due to the nonlinear relation. It is represented by following equation:

$$f(\text{input}) = \frac{1}{1+e-(\text{input})} \qquad \text{Equation 2.2}$$

The sigmoidal function is preferred by users as it can accept any input in the range of (-∞,+∞) and map it into the ranges (0,+1).

The hard timer function can have two values: 0 and 1. This function is used only for ON/OFF or 1/0 outputs. This function is characterized by a threshold value of 9.

The output of the threshold logic function varies between 0 and 1 but the relation between these two values is linear. The interval width of the function is represented by α; the interval starts at 9 and has a width of 1/ α (Zupan et al.,1993).

2.6.3 Back Propagation Learning Algorithm

Back propagation learning algorithms are used in ANN modeling to improve its performance. A number of layers (i.e. input layer, hidden layer(s) and output layer, including a specified number of neurons) are connected together to create a back propagation neural network system. It is seen that one layer of hidden units can approximate any function with a finite number of discontinuities to arbitrary precision, provided the activation function of the hidden unit is non-linear (Hornik et al., 1989; Funahashi, 1989; Cybenko, 1989; Hartman et al., 1990). In most cases, a single layer of hidden nodes is used for a feed-forward network.

Once inputs pass through a network to calculate the output of a neuron in the output layer, the error is determined by comparing the calculated outputs to the actual values. This procedure is consequently used for error function determination. The error function works to adjust the error by modifying the connection weights linked to the output. Initially, the connection weights are typically assigned random or specified values. The obtained output value using the initial connection weights is usually not close to the actual output value. The error correction is done by propagating the error backwards. Using the adjusted connection weights, the new error is determined and is used to readjust the connection weights. This process is continuously repeated on all training datasets until the error is reduced to a predetermined minimum or an allowed tolerance (Najjar et al., 1997). The final connection weights which produce an allowable error are then stored to represent the network.

2.6.4 ANN Model Development

ANN models are usually developed following four sequential steps. In the first step, the database is divided into three different classes for training, testing, and validation. Usually 50% of the total data is selected randomly for training. The remaining 50% of the data is divided equally for testing and validation. In the second step, the network is trained and tested to determine the optimum hidden nodes and number of iterations. The three best-performing networks are selected based on their statistical accuracy measures. In the third step, the three best performing networks are validated using the validation data set. Finally, the selected networks are re-trained using all the data in order to increase the prediction accuracy of the network structures.

To compare the performance of networks, statistical accuracy measures such as the average square error (ASE), mean absolute relative error (MARE) and coefficient of determination (R^2) are evaluated. During the evaluation process, training, testing, validation and overall performance parameters should be considered. Minimum values of ASE and MARE and a maximum value of R^2 play key roles during performance evaluation. The ASE, MARE and R^2 value are expressed by the following equations:

$$ASE = \frac{\sum_{i=1}^{N}(X_i^A - X_i^P)^2}{N} \qquad \text{Equation 2.3}$$

$$MARE = \frac{\sum_{i=1(}^{N}|X_i^P - X_i^A|)/X_i^A}{N} \qquad \text{Equation 2.4}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(X_i^A - X_i^p)^2}{\sum_{i=1}^{N}(X_i^A - \overline{X_1})^2} \qquad \text{Equation 2.5}$$

where, X_i^A = Actual value, X_i^P = Predicted value, $\overline{X_1}$ = Mean of X_i^A ,N = Total number of data

More information about ANN modelling can be found in the following references: Rumelhart et al. (1986); Hopfield (1982); Haykin (1999); Fausett (1994); Herz (1991); Ghaboussi et al. (1991).

2.6.5 Development of Graphical User Interface (GUI)

At the end of an ANN modeling procedure, a file is generated which contains the biases and connection weights for determining outputs with the acceptable amount of error. A Graphical User Interface (GUI) can be built to predict outputs for any input within the range. Microsoft Visual Basic or any other suitable programming language can be used to generate a GUI from the biases and connection weights.

2.7 Application of ANNs in Geophysics

In the last decade, neural networks methods have gained popularity for geophysical applications. Neural networks have been used for: determining subsurface target location using electromagnetic surveying (Poulton et al., 1992); seismic inversion purposes (Roth and Tarantola, 1994; Langer et al., 1996); waveform recognition (Murat and Rudman, 1992); shear-wave splitting (Dai and MacBeth, 1994); seismic deconvolution (Wang and Mendal, 1992); event classification (Dowla et al., 1990); well log analysis (Huang et al., 1996); and trace editing (McCormack et al., 1993). However, there are no reported studies on the use of ANN for survey design.

CHAPTER 3

ELECTROMAGNETIC SURVEY IN GOODWIN CREEK, MS

3.1 Survey Site

Goodwin Creek Experimental Watershed (GCEW) is located in Panola County, Mississippi (Figure 3.1). The relatively flat cropland (slope <2%) currently occupies only 6% of the area, whereas the hilly forest and pasture lands occupy 39 and 55%, respectively (Kuhnle et al., 2008). The relatively flat (<2% slope) alluvial plains are typically Falaya silt loam and Collins silt loam soil series (Zhang et al., 2012). The surrounding hillslopes (2%–8% slope) are generally Grenada silt loam and Loring silt loam with some gullied land. The transition areas (2%–5% slope), which are in cropland, tend to be either Calloway silt loam or described as gullied land

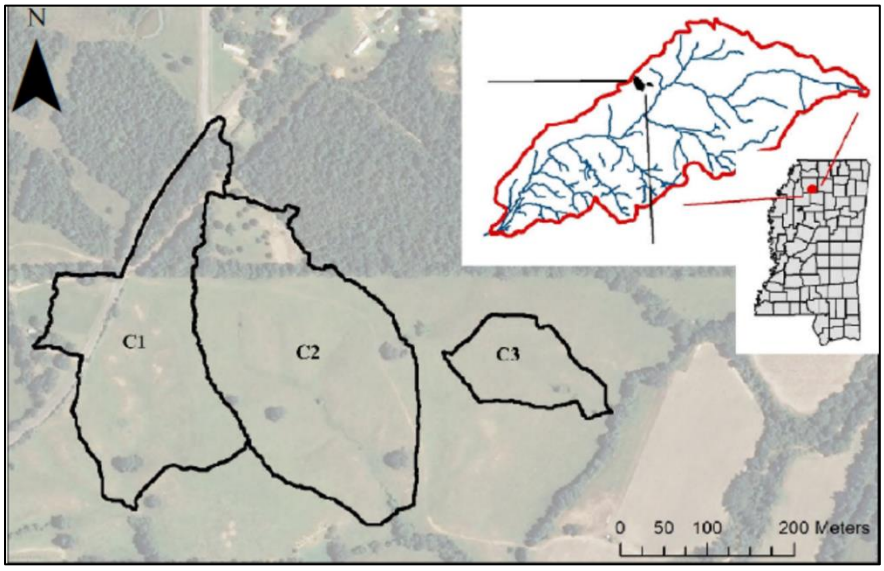
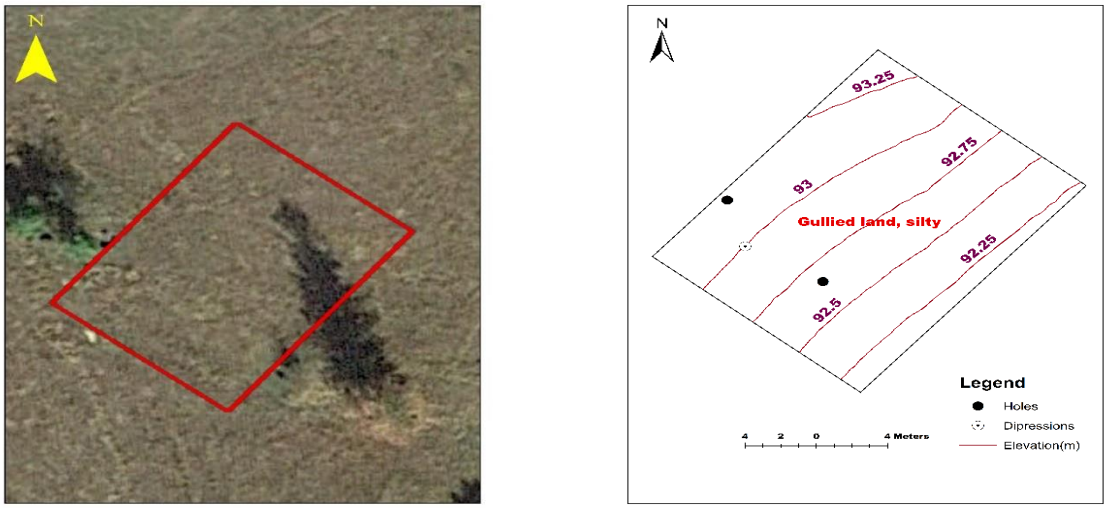


Figure 3.1: Location of Goodwin Creek, MS and the three catchments (Wilson et al., 2015).

Ephemeral gullies result from the junction of rills that form a branching or tree-like pattern of channels. They can also be formed by internal erosion that leads to pipe (tunnel) collapse. Five factors influence ephemeral gully formation: overland flow discharge & duration, slope and flow depth, planform curvature, soil characteristics, and vegetable characteristics. The research site (Figure 3.1) consists of three catchments based on soil pipe collapse features (Wilson et al.,2015). Among the three catchments, C2 and C3 contain soil pipe collapse features (i.e. sinkholes, flute holes and gully windows), but no pipe collapse features were identified in C1. In 2013, the main catchment C2 had 56 flutes holes while C3 had 14 flute holes. The flutes holes of catchment C2 averaged 29cm deep, 20cm wide and 25cm long, while the flute holes of catchment C3 averaged 56cm deep, 32cm wide and 38cm long.

For this study, electromagnetic surveying was conducted over a small section (20 m x 15 m) of C3 catchment (Figure 3.2a), having established collapse features that stem from internal soil

pipes. Two m) of C3 catchment (Figure 3.2a), having established collapse features that stem from internal soil pipes. Two gully windows, marked as holes formed due to soil pipe collapses, are located within the study area. A surface depression is located between the gully windows. Information related to typical soil features and elevation was collected from the USGS database. Using ArcGIS, a map was produced representing the soil features and elevation contour of the study area (Figure 3.2b).



(a) (b)

Figure 3.2: (a) Study area with collapse features and (b) elevation and soil information.

3.2 Selection of Suitable EM Equipment

The maximum depths of soil collapse features at C3 are between 30 cm and 115 cm, depending upon the type of feature (Wilson et al., 2015). Soil pipes are tortuous voids located within 1.5 m depth of the ground surface. They have cross-sectional dimensions from millimeters to meter. The contrast in apparent electrical conductivity (ECa) is significant especially if the soil

pipe is filled with air. Based upon these characteristics, an EM38B was used to conduct electromagnetic surveys in this study area. The EM38B has an inter coil spacing of 1m and provides measurements within two effective depth ranges: 0-1.5 m in the vertical dipole mode and 0-0.75 m in horizontal dipole mode. Apparent electrical conductivity (ECa) and in-phase measurements can be collected using a Geonics EM38B (Figure 3.3). The measured ECa is a dipole dependent weighted average over a soil volume of about $1m^3$. Therefore, the resolution limitation dictates that information can be used to infer locations having soil pipes but it can't provide information specific to the soil pipes themselves. The sensitivity in the horizontal mode is highest directly below the instrument, while sensitivity in the vertical dipole position reaches a maximum at approximately 30–40 cm below the instrument.

Figure 3.3: EM survey to detect soil collapse features.

3.3 Data Acquisition and Processing

The EM38B must be powered for thirty minutes before the start of measurements to allow

for the instrument to warm up and minimize instrument drift. All metal objects from the operator must be removed. Before the start of a measurement and at various times during the survey, apparent conductivity and in-phase readings were obtained at a reference location outside the survey area. The differences in these repeat measurements quantify the instrument drift due to the change of temperature (Robinson et al., 2004). Recalibration needs to be done between two successive measurements of ECa if the reference location changes more than 1 mS/m. Table 3.1 illustrates the calibration information of a reference location while conducting the survey. As the change in ECa was 1mS/m between the start and end of the survey for one direction, no calibration was performed.

Table 3.1: Calibration table for EM data collection on January 30, 2020.

Survey Orientation	Calibration	ECa (mS/m)	Start Time	End Time	Calibration Check	ECa (mS/m)
Perpendicular to soil pipes	11.09 am	7	11.16.30 am	11.53.07 am	11.55 am	8
Parallel to soil pipes	11.55 am	8	11.59.55 am	12.40.50 pm	12.42 pm	9

The survey was conducted by continuously walking and acquiring data at ten readings per second to ensure high spatial sampling. Data were collected and stored using an Archer2 with location information received from an Emlid Reach RS+ GPS. Real-time quality control (QC) was performed visually using the Archer2 display.

Geonics DAT38W software was used to download data from the data logger (Archer2). The software was also used to merge GPS and EM38B readings. Once the data were downloaded, readings from the reference location were used for drift correction using linear interpolation. The

timestamp of each data point was used to pick the correct encompassing reference reading for interpolation. Semi-variogram analysis using drift corrected data was then conducted to determine input parameters for kriging interpolation. ArcGIS software was used to generate ECa maps using ordinary kriging interpolation.

3.4 Response of EM38B on Soil Pipes

Four survey lines (Figure 3.4) using the EM38B were acquired in order to observe characteristic anomalies associated with internal soil pipes. Survey lines L1 and L2 were around 700 cm in length and were located between gully windows. Survey line L2 passed over the surface depression. Survey lines L3 and L4 were around 360 cm and 870 cm respectively and located on either side of a gully window. The EM38B provides position-dependent but not explicit depth-dependent information.

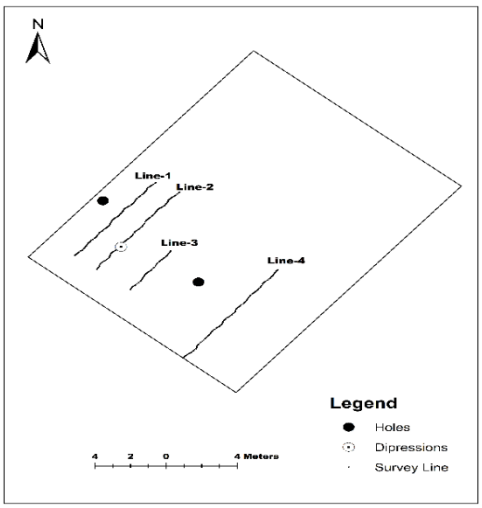


Figure 3.4: Plan-view of the survey area showing relative locations of the surface features and acquisition lines (L1, L2, L3, and L4).

For the vertical dipole survey, changes in ECa along the line L1(Figure 3.5) was small. The larger values of vertical ECa on the end of the lines were associated with the harder soil while the lower ECa values were in areas having internal soil pipes. The apparent electrical conductivity also exhibits small fluctuations in signal strength at length scales much smaller than the spacing between antennae. Such behavior can be expected when one antenna traverses a small object located close to the surface. One would expect these types of signals to occur in pairs with a separation distance equal to the antennae separation distance. An air-filled soil pipe is expected to cause a drop in the apparent electrical conductivity but its response when it is smaller than the antennae spacing and close to the surface requires further investigation.

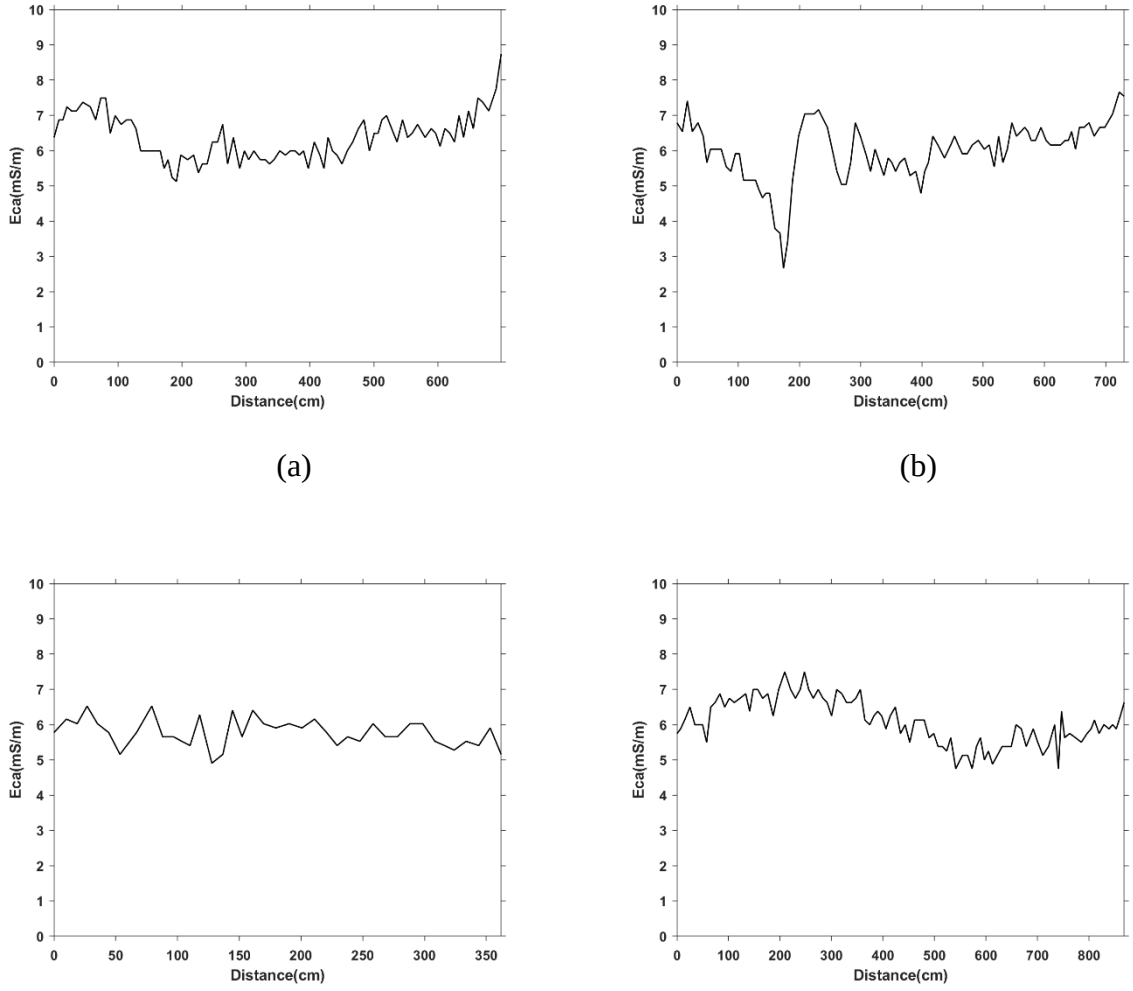


Figure 3.5: Vertical ECa for survey line: (a) L1, (b) L2, (c) L3 and (d) L4

The vertical dipole ECa of survey line L2 is shown in Figure. 3.5(b). There is a large change in ECa over the surface depression, located at a distance of 200 cm. The vertical ECa shows a signature expected when the system traverses a small metallic object close to the surface. When traversing a small, shallow, metallic object, a sudden decrease in ECa and drop in the in-phase component occurs as each antenna crosses the object. It is postulated that this depression is a partially filled-in sinkhole with a material of lower conductivity and a small metallic object very

close to the surface. If correct, this demonstrates that small metallic objects can easily dominate the signatures and would have to be accounted for if small scale features in EM data are used for soil pipe detection.

Changes in vertical ECa are not significant for survey line L3 (Figure 3.5c). Changes in vertical ECa for line L4 are also small (Figure 3.5d), but it is evident that the vertical ECa shows a drop in apparent conductivity starting at 500 cm and extending up to 750 cm, which could be an effect of the soil pipes.

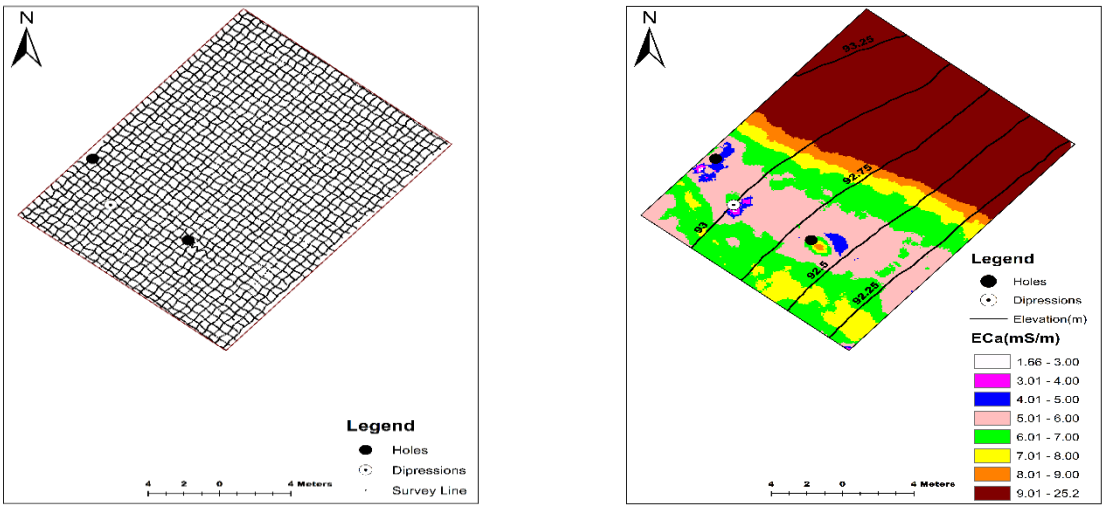
With the accepted state of interpretation of EMI data, the EM38B cannot detect individual pipes but appears to show reduced ECa in regions having soil pipes.

3.5 Mapping of Soil Pipes in EM Plan View

To ensure high spatial sampling, a benchmark survey (Figure 3.6a) was conducted on a 50 cm line grid oriented approximately parallel and perpendicular to the topography resulting in a total of 14676 data points. These measurements were conducted in order to produce a plan-view visualization (map) of the area. A semi-variogram analysis was conducted using ArcGIS and following parameters were used for ordinary kriging interpolation.

Table 3.2: Parameters used for kriging interpolation using ArcGIS

Type	Lag Size	Lag Number	Major Range	Nugget	Partial Sills
Gaussian	2.88	12	33.43	0.51	154.87



(a) (b)

Figure 3.6: (a) GPS grid at 50 cm x 50 cm interval used for generating vertical ECa map, (b) Vertical dipole ECa map using Kriging Interpolation

A vertical ECa map for the study site is shown in Figure 3.6(b). The vertical ECa has a range from about 3 mS/m to greater than 10 mS/m. The range of ECa at the study site is small and could imply a fairly uniform soil type based on traditional soil classifications. For example, clay-rich soil has an electrical conductivity of 2.5 – 10 mS/m, and topsoil has an electrical conductivity of 5 - 25 mS/m (Katsube et al., 2004). However, such conjecture makes it difficult to explain the trends in ECa. For a constant soil type, the distribution of soil moisture would be a controlling factor of the ECa. Due to drainage, it is expected that the soil moisture and therefore the ECa should decrease upslope. However, the trend in ECa does not correlate with elevation but increases transversely to the gully at a constant elevation. Another possibility for variations in ECa is the difference in soil structure between the upper soil layers and the underlying fragipan horizon. If the top granular layer is thicker, it could result in a higher ECa.

The ECa maps indicate that all the surface features (gully windows and depressions) indicative of soil piping are located within a low vertical ECa zone of 5.1 - 6 mS/m. This observation suggests that EMI can be used to define “zones” where soil piping networks are prevalent.

CHAPTER 4

SURVEY DESIGN USING ANN

4.1 Introduction

To detect probable locations of soil pipes, survey orientation is essential. The survey line spacing and directions are determined by the operator based on the problem at hand. In this chapter, ANN is used to choose the appropriate survey orientation and line spacing for an EM38B to determine the probable location of soil pipes with the minimum amount of effort.

4.2 Survey Orientation and Line Spacing

4.2.1 Data Separation for Developing ANNs Models

EM data was collected with 50 cm line spacing and the survey lines were parallel and perpendicular to the soil pipe features (i.e. holes (gully windows) and depression. A lot of effort was required to collect the data. This large data set is decimated to represent data collection efforts, with specific orientation and line spacing for ANN analysis. Figure 4.1 shows the process in data analysis for determining optimal surveys. Seven different surveys (referred to as patterns 1 to 7) are designed with different survey orientations and spacing as listed in Table 4.1 and displayed in Figure 4.2. Patterns 1-5 require the EM data to be collected in automatic mode, whereas patterns 6 and 7 would be in manual mode. Surveyors having no GPS connection would use manual mode.

All proposed surveys would require significantly less effort than the benchmark survey.

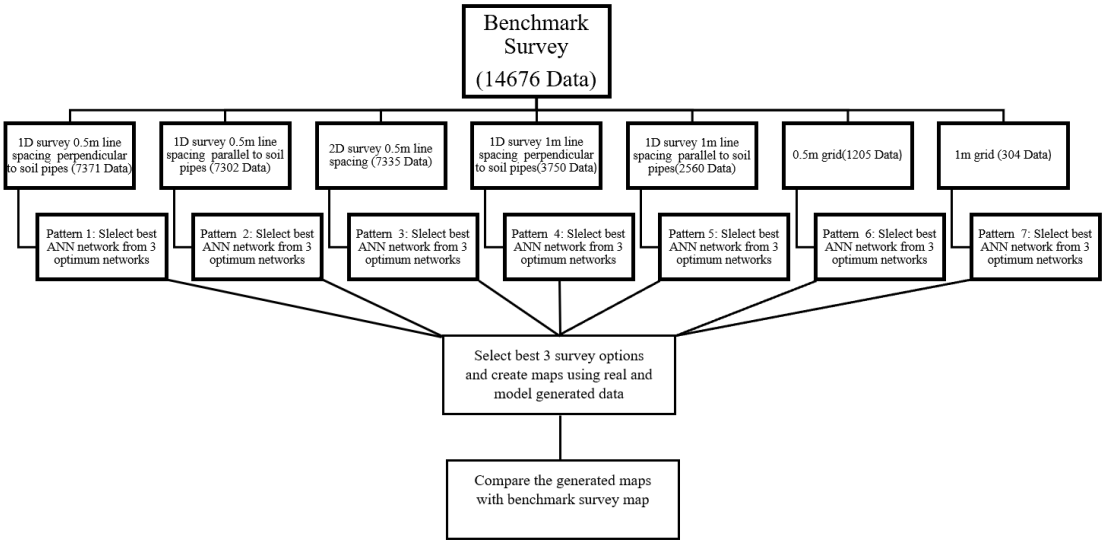
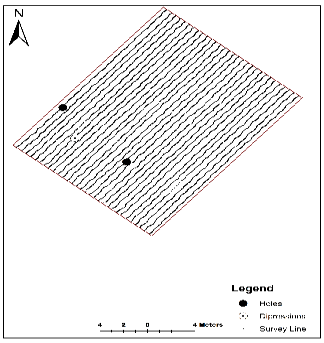


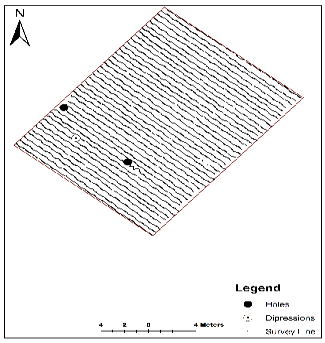
Figure 4.1: Flow chart for choosing optimal survey patterns

Table 4.1: Data separation for different survey orientation and spacing

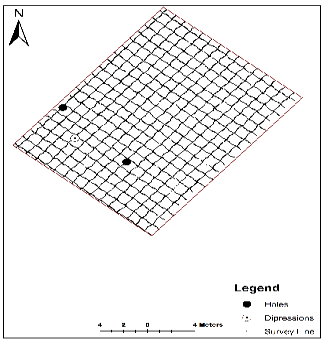
Pattern No	Orientation and Spacing	No of Data	ECa(mS/m) Max	ECa(mS/m) Min
1	1D survey with 50 cm line spacing perpendicular to soil pipe features	7371	26.19	0.25
2	1D survey with 50 cm line spacing parallel to soil pipe features	7302	25.17	0.75
3	2D survey with 1m line spacing	7335	25.77	0.75
4	1D survey with 1m line spacing perpendicular to soil pipe features	3750	25.77	2.66
5	1D survey with 1m line spacing parallel to soil pipe features	2560	25.17	0.75
6	0.5m grid spacing	1205	25.77	1.62
7	1m grid spacing	304	25.77	4.71



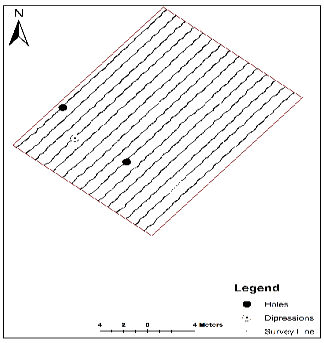
(a)



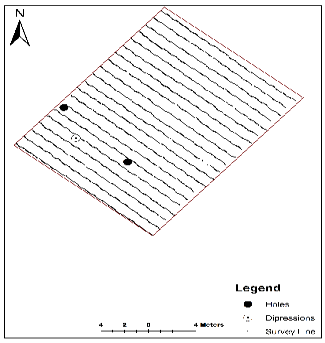
(b)



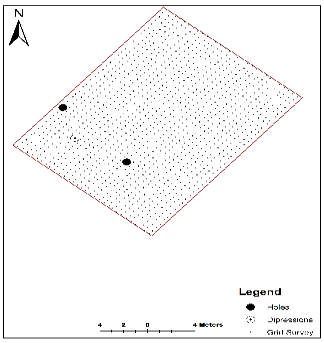
(c)



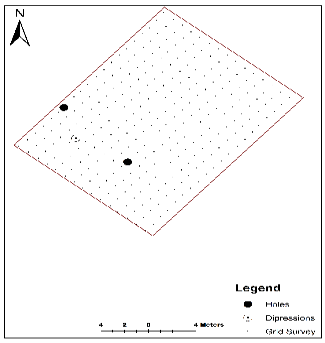
(d)



(e)



(f)



(g)

Figure 4.2: Different survey orientation and spacing: (a) pattern 1; (b) pattern 2; (c) pattern 3; (d) pattern 4; (e) pattern 5; (f) pattern 6; (g) pattern 7

4.2.2 Application of ANNs for Selecting Uniform Survey Pattern

For each proposed survey, the measurement location, easting and northing, are used as input parameters and the ECa is used as the output parameter. Before running ANN models, the output parameter is normalized between 0 and 37.16 mS/m. The data is then subdivided into 50% for training, 25% for testing, and the remaining 25% for validation. Training and testing are done together for a number of cases. After training and testing, 3 optimum networks are selected based on the minimum ASE and minimum MARE and maximum R^2 respectively. The performance of these selected networks is validated with the validation dataset and then retrained with the full data set. A single ANN network [A_(X_Y_Z) _B] is chosen based on its performance (minimum ASE, MARE and maximum R^2) during validation stage. The network is characterized by four parameters: A is the number of inputs, B is the number of outputs, X is the initial hidden node, Y is the final hidden node, and Z is the number of iterations. Table 4.2 presents the ANNs analysis for the 1D survey with a 0.5m survey line spacing. Network [2_(4_11_20000) _1] is chosen as the best ANN based on its performance in the validation stage. This procedure is repeated for all other surveying scenarios to obtain their respective statistical accuracies. Final ANNs networks, R^2 for linear regression, and statistical accuracy measures of the developed models for different survey patterns are presented in Table 4.3.

Table 4.2: Optimal ANNs selection for 1D survey 0.5m line spacing perpendicular to soil pipe features (pattern 1)

ANN Network		2_(4_11_20000)_1	2_(8_11_20000)_1	2_(9_11_20000)_1
TR	ASE	0.00023	0.00031	0.00029
	MARE	7.49	8.70	8.76
	R^2	0.99	0.99	0.98
TS	ASE	0.00023	0.000312	0.000288
	MARE	7.65	8.83	8.92
	R^2	0.99	0.99	0.99
VAL	ASE	0.00021	0.00028	0.00026
	MARE	7.33	8.47	8.55
	R^2	0.99	0.99	0.99
TR ALL	ASE	0.00022	0.00023	0.00018
	MARE	7.20	7.26	6.64
	R^2	0.99	0.99	0.99

Table 4.3: Performance of the "best" ANNs for each survey design pattern

Pattern No	ANNs Network	No of Data	R^2 (Linear Reg)	ASE	MARE	R^2 (ANNs)
1	2_(4_11_20000)_1	7371	0.80	0.00022	7.20	0.99
2	2_(1_5_20000)_1	7302	0.77	0.00029	9.53	0.98
3	2_(2_7_20000)_1	7335	0.79	0.00033	7.22	0.99
4	2_(8_11_20000)_1	3750	0.81	0.00026	5.68	0.99
5	2_(1_5_20000)_1	2560	0.76	0.00033	7.59	0.98
6	2_(9_11_20000)_1	1205	0.78	0.00028	6.07	0.99
7	2_(5_8_20000)_1	304	0.77	0.00044	6.25	0.98

After evaluating the performance of each model for TR ALL (Training ALL), they are ranked based on minimum ASE, MARE and maximum R^2. In Table 4.4, it is seen that pattern 1, pattern 4 and pattern 6 are the best 3 survey designs for conducting EM surveys for this site. It should be noted that although the number of data points for pattern 6 is much lower than pattern 1 and pattern 4, from a practical standpoint it takes more effort to conduct such a survey and should probably only be used if automatic mode is not available on the EM38B.

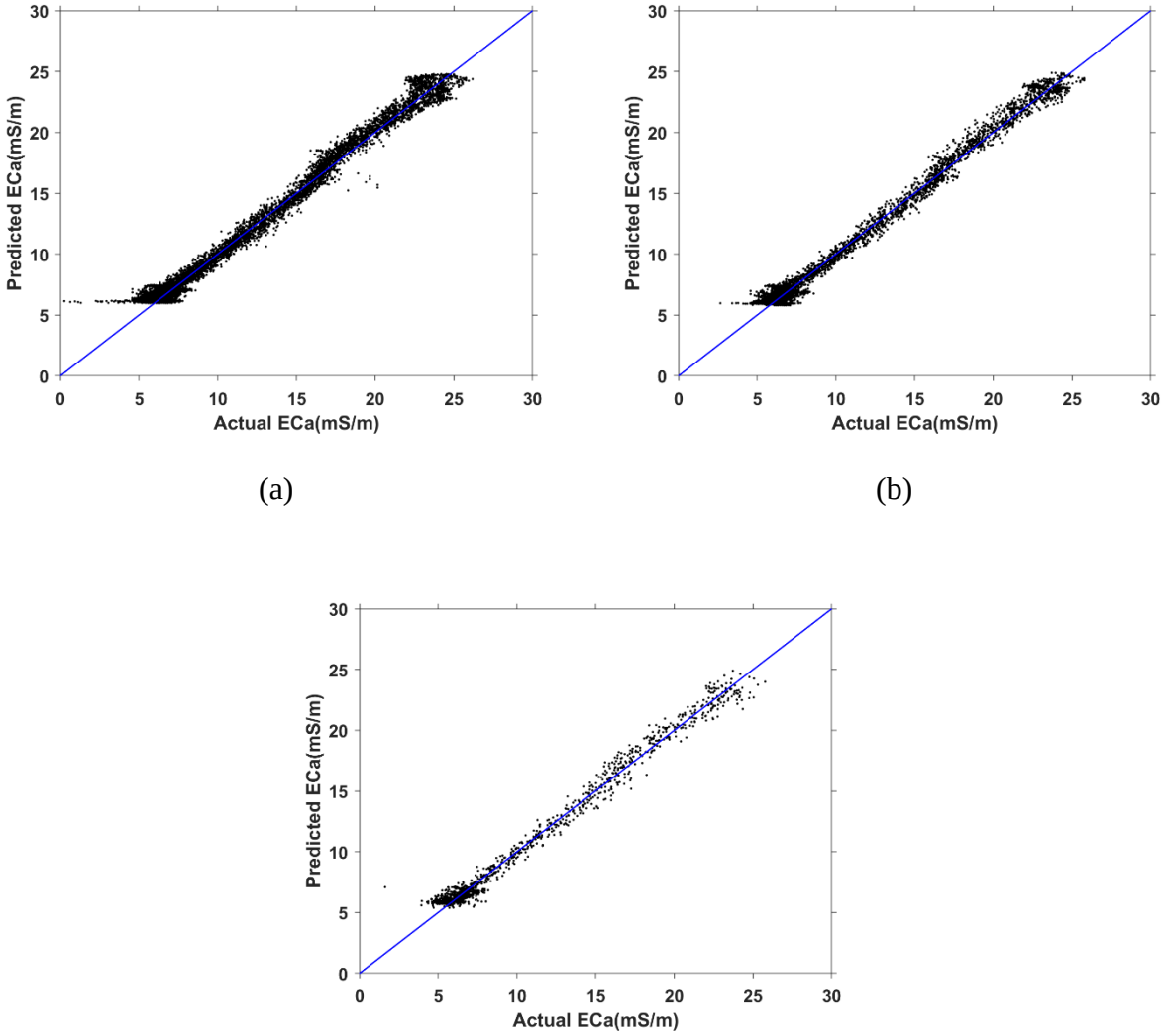
Table 4.4: Comparison of developed ANN patterns

Pattern No	No of Data	R^2 (Linear Reg)	ASE	MARE	R^2(ANNs)	Rank
1	7371	0.80	0.00022	7.20	0.99	1
4	3750	0.81	0.00026	5.68	0.99	2
6	1205	0.78	0.00028	6.07	0.99	3
2	7302	0.77	0.00029	9.53	0.98	4
3	7335	0.79	0.00033	7.22	0.99	5
5	2560	0.76	0.00033	7.59	0.98	6
7	304	0.77	0.00044	6.25	0.98	7

Plots of predicted vs. actual ECa for the best three survey orientations are shown in Figure 4.3. In spite of having high R^2 values, the developed ANN models cannot predict any ECa values less than 5 mS/m in Patterns 1 and 4 even though 2.65% of the total collected data had ECa values less than 5 mS/m. This is of concern because EM anomalies associated with air-filled soil pipes will have low ECa values.

The connection weights and biases of the developed ANN network are used to generate (predict) 7305 additional data points for pattern 1, 10926 for pattern 4 and 13471 for pattern 6. Then the subset of measured values and the generated data points are merged together to prepare maps using ArcGIS. The semi variogram analysis is conducted before applying kriging interpolation to the data sets. Figure 4.4 represents the interpolated maps for the benchmark survey and the three best models using Kriging interpolation. Only Figure 4.4b shows some evidence of soil pipe features. The soil pipe features (i.e. hole and depression) are located near the low vertical ECa zones of 3.1-4 mS/m and 4.1-5 mS/m. The difference between the ANN derived maps and the benchmark survey are shown in Figure 4.5. These ECa differences between ($\pm$1 mS/m) are assumed to be within the measurement error on the EM38. Measurements outside this range,the

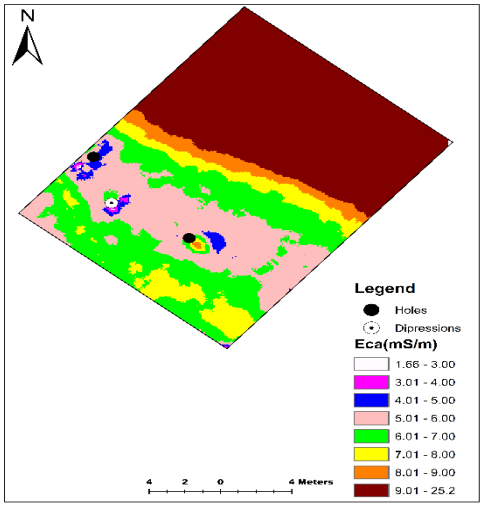
gray and red areas, are considered significant differences between the ANN derived maps and the benchmark map. As one might expect, the more data, the better the agreement. There are very few gray and red zones visible in Figure 4.5a. This is because the maximum number of real data was used to generate Model 1. As the number of real data decreases, more gray and red zones are observed for patterns 4 (Figure 4.5b) and 6 (Figure 4.5c).



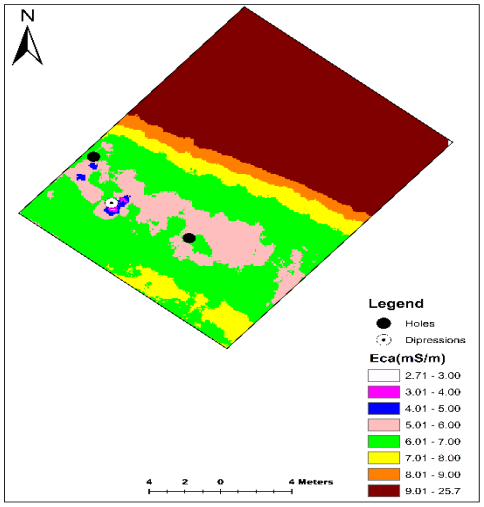
(a) (b)

(c)

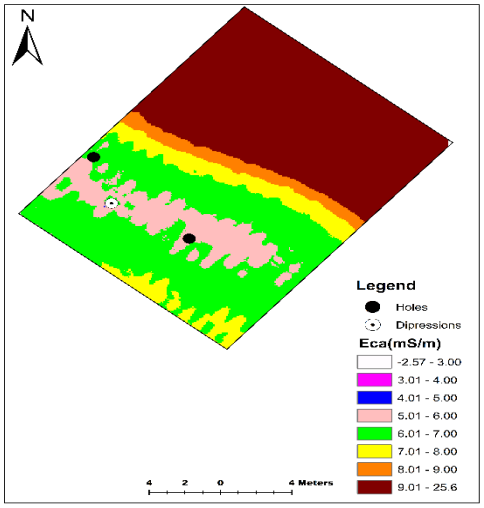
Figure 4.3: Predicted Vs. Actual Apparent Electrical Conductivity (ECa): (a) pattern 1, (b) pattern 4 and (c) pattern 6.



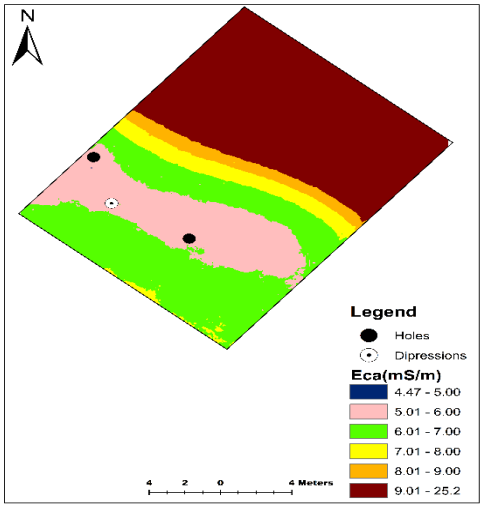
(a)



(b)

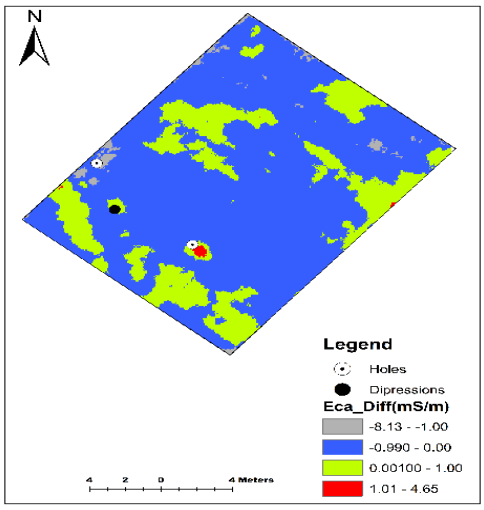


(c)

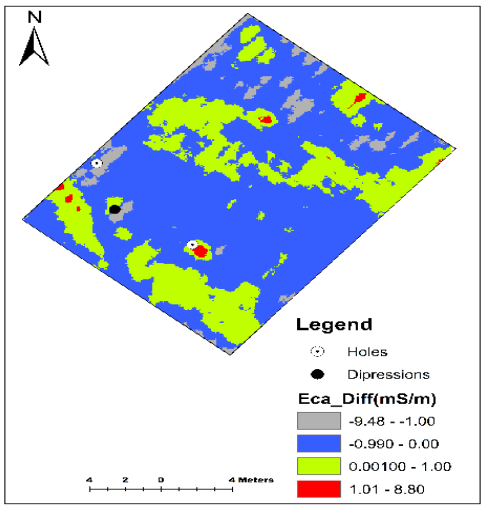


(d)

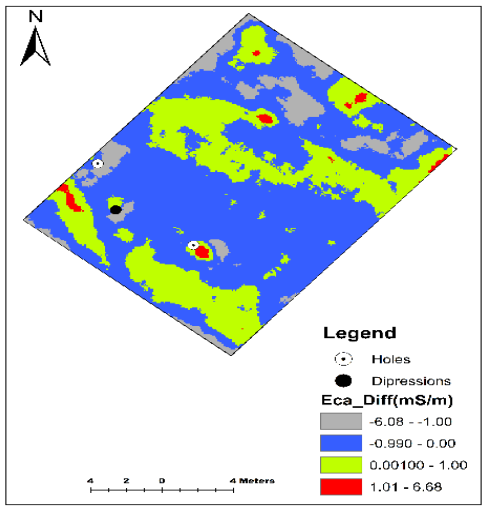
Figure 4.4: Maps generated using Kriging interpolation: (a) From the entire data set (14676) measured data), (b) from pattern 1 (7371 measured data and 7305 GUI generated data), (c) pattern 4 (3750 measured data and 10926 GUI generated data) and (d) pattern 6 (1205 measured data and 13471 GUI generated data).



(a)



(b)



(c)

Figure 4.5: ECa difference maps for: (a) pattern 1, (b) pattern 4 and (c) pattern 6.

4.2.3 Application of Quantile Method in ANNs Modeling

ANNs constructed using the entire site data cannot predict low ECa(<5mS/m) and these low values are representative of air-filled soil pipes. The inability to predict the low ECa values might be associated with the large range of ECa. The quantile method is incorporated into the analysis to further divide the study area based on ECa. Figure 4.6 represents how the quantile method is used with ANN for selecting survey patterns.

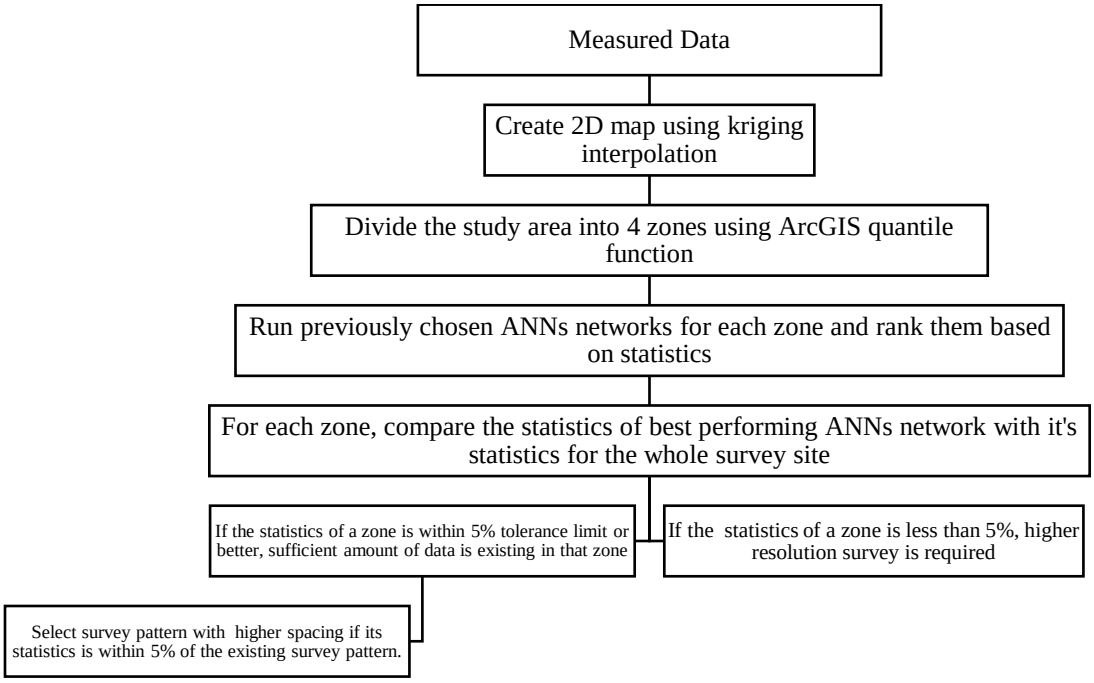


Figure 4.6: Application of quantile method in ANN modeling

The measured data is used to generate a 2D map using kriging interpolation. Then the entire survey area is subdivided into 4 zones based on ECa ranges using the quantile function of ArcGIS (Figure 4.7). This process is used for the 4 different data sets to get a uniform ECa range which satisfies most of the survey patterns (Figure 4.7). The zones are chosen to be rectangular shapes as the survey is usually conducted following regular geometric shapes (i.e. rectangular and square shapes).

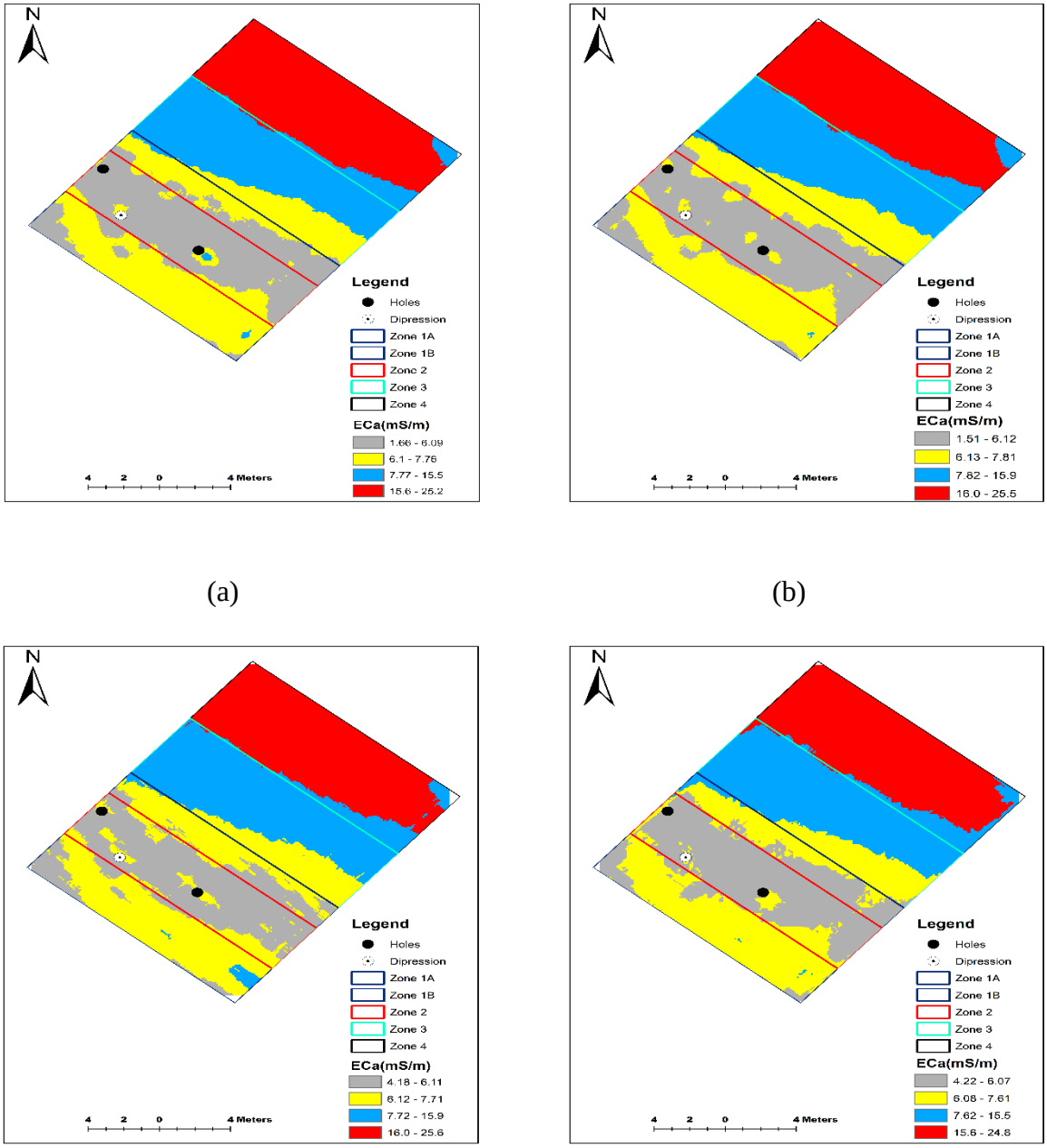


Figure 4.7: Zoning the study area using quantile the method: (a) Benchmark survey, (b) pattern 1 and (c) pattern 4 and (d) pattern 6

Table 4.5 represents the ECa ranges of the respective zones. Zone 2 is the most critical zone having the lowest ECa values (1.51- 6.12 mS/m). The ECa of zone 1 is higher than the ECa of zone 2. It is seen that zone 1 has two parts which can be denoted as zone 1A and zone 1B. Although zone 1 could be analyzed in total, for better accuracy the two parts are analyzed separately. Zone 3 and 4 have much higher ECa values than zone 1 and 2.

Table 4.5: ECa ranges for respective zones

ECa(mS/m)	Description of ECa(mS/m)	Zone
1.51-6.12	Very low	2
6.13-7.81	Low	1A and 1B
7.82-15.9	Medium	3
16-25.6	High	4

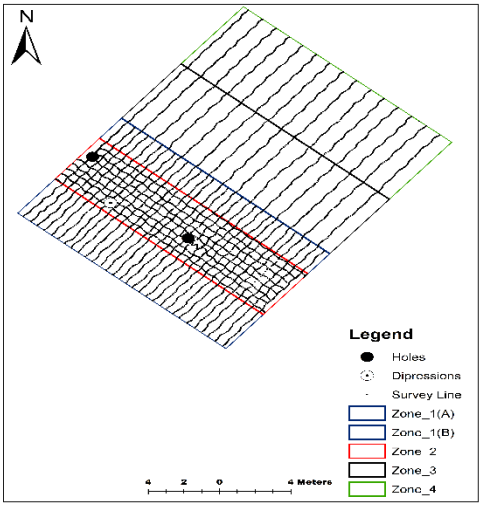
The next step is to perform ANN analysis on each zone to find the statistical measures of the different zones for a specific survey pattern. For example, 7371 data were collected over the entire site using pattern 1. These 7371 data are divided into four zones based on their ECa range and renormalized to run the ANN analysis. Similarly, the ANN networks for patterns 4 and 6 provide statistics for every zone. For a specific zone, the statistics for each pattern are ranked based on minimum ASE and MARE and maximum R^2. From Table 4.6, it is seen that pattern 1 gives the best statistics for zones 1, 2, 3 and 4.

Table 4.6: Statistical measures of developed ANN models for different zones for different survey patterns

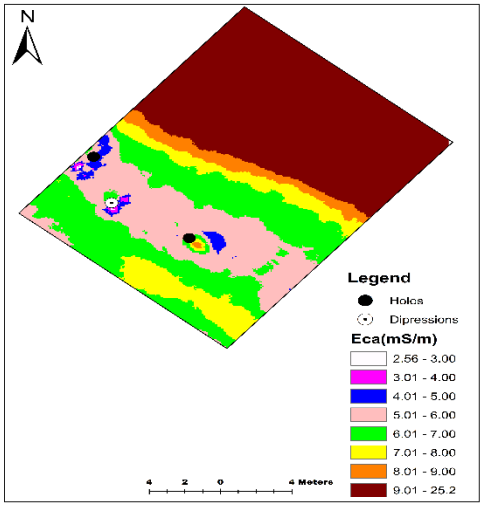
Zone	Pattern No	No. of Data	R^2 (Linear Reg)	ASE	MARE	R^2 (ANNs)	Rank
1A	1	1155	0.17	0.00030	5.84	0.51	1
	4	584	0.19	0.00081	7.75	0.52	2
	6	212	0.10	0.00247	6.99	0.47	3
1B	1	695	0.50	0.00021	4.87	0.66	1
	4	354	0.44	0.00035	4.61	0.65	2
	6	118	0.45	0.00596	5.35	0.60	3
2	1	1381	0.01	0.00185	13.03	0.09	1
	4	707	0.02	0.00358	6.75	0.26	2
	6	228	0.08	0.00641	7.63	0.21	3
3	1	1882	0.94	0.00038	3.44	0.97	1
	4	951	0.95	0.00051	3.42	0.97	2
	6	318	0.94	0.00055	3.49	0.97	3
4	1	2260	0.75	0.00049	2.05	0.98	1
	4	1156	0.98	0.00052	1.99	0.97	2
	6	330	0.73	0.00087	2.56	0.96	3

The statistics of each zone, Table 4.6, are compared to the statistics of the entire site, Table 4.4, to determine an optimal survey pattern for each zone. For zone 1A, MARE is lower than the MARE of the entire site for pattern 1. For zone 1B, ASE and MARE is lower than the ASE and MARE of the entire site. R^2 of zone 1A and 1B is moderate considering the number of data points. Therefore, survey pattern 1 should be sufficient for zone 1. For zone 2, the statistics provided by the ANN network is significantly lower than the statistics of entire area. This indicates that a higher resolution survey of this zone is required. The MARE and R^2 of zone 3 and 4 are very similar to the MARE and R^2 of entire site for pattern 1. It is also seen that the statistics (i.e. MARE and $R^{2)}$ provided by pattern 4 is within the 5% tolerance limit of the statistics of pattern 1 for zone 3 and 4. Therefore, pattern 4 could be used for zone 3 and 4 to minimize cost. Following the survey design in Figure 4.8a, this would require 6765 data to be collected. Using this "measured data"

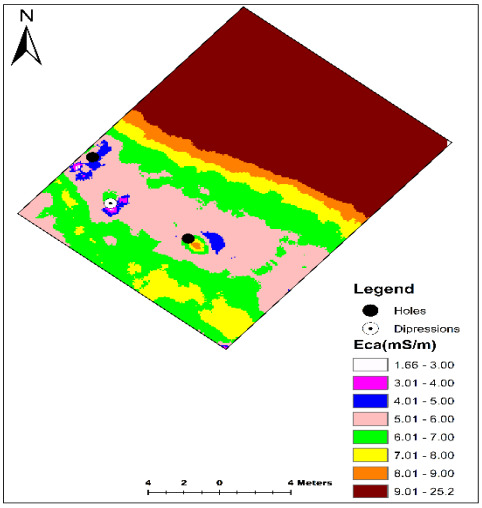
and 7911 data generated using the ANN models results in a database containing 14676 data. This database is used to create a 2D map by kriging interpolation (Figure 4.8b). Compared with the 2D map of the benchmark survey (Figure 4.8c), no significant difference of ECa is found (Figure 4.8d). The soil pipe signatures are visible in Figure 4.8b and Figure 4.8c.



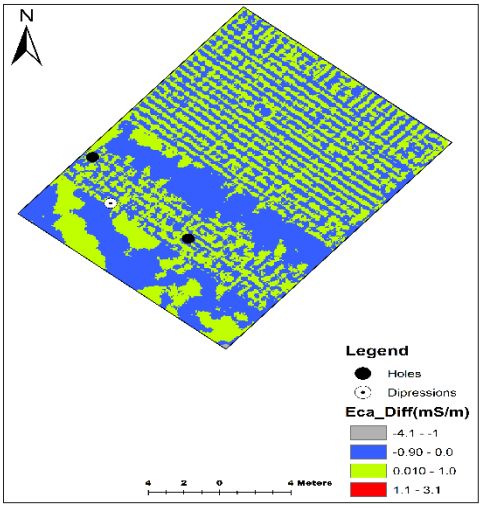
(a)



(b)



(c)



(d)

Figure 4.8: (a) Survey plan based on ECa, (b) 2D ECa map created following proposed survey plan, (c) 2D ECa map of benchmark survey and (d) ECa difference map.

4.2.4 Selection of Survey Pattern Based on ECa

The statistical measures (i.e. ASE, MARE and $R^{2)}$ change significantly for the zones having low ECa based on the number of collected data. The maximum amount of data is required for the lowest range of conductivity. Spacing should be increased gradually observing the increment of ECa. Based on the analysis, a survey plan can be proposed for any survey site with similar ranges of conductivity (Table 4.7). Reconnaissance data must be acquired to establish the probable ECa ranges of the survey site. After doing ANNs modeling using that reconnaissance data, an effective survey plan can be designed using Table 4.7. If the reconnaissance data has very good ASE, MARE and R^2 for a zone having high ECa, the surveyor can select a sparser design spacing for those zones.

Table 4.7: EM survey plan based on apparent conductivity of soil (ECa)

ECa(mS/m)	Survey Plan
1.51-6.12	2D survey with 50 cm line spacing
6.13-7.81	1D survey with 50 cm line spacing perpendicular to soil pipes.
7.82-15.9	1D survey with 1m line spacing perpendicular to soil pipes.
16-25.6	1D survey with 1m line spacing perpendicular to soil pipes.

CHAPTER 5

SURVEY DESIGN FROM RECONNAISSANCE

5.1 Introduction

An approach for designing EM survey for soil pipe mapping was developed in the previous chapter based on decimating benchmark data from a high-resolution survey. In reality, the survey design will start with minimum data collected using a reconnaissance survey. In this chapter, the approach outlined in Figure 5.1 is followed to map the field site.

5.2 Reconnaissance

The first step is to collect reconnaissance data using apriority information. For this example, the reconnaissance survey consists of linear acquisition lines constrained to be close to visually observed surface features (gully windows) and oriented approximately parallel and perpendicular to the natural elevation contours. Additional survey lines around the perimeter of the area and within the area to maintain a 4m line spacing constrained to a rectangular grid are also collected. The data is collected using the EM38B in automatic mode so the data along lines will be dense.

Figure 5.2 represents the reconnaissance survey lines of the study area. Information related to typical soil type and elevation are obtained from the USGS database. A total of 2370 data are collected and ECa ranges between 2.66 mS/m and 25.17mS/m. For linear regression, the coefficient of determination, R^2 of the collected data is 0.81.

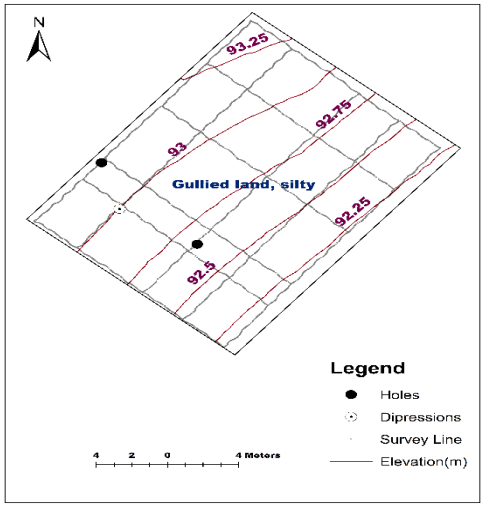


Figure 5.1: Reconnaissance in the study area of Goodwin Creek, MS

5.3 ANNs Modeling using Reconnaissance Data

Before running ANN models with reconnaissance data, normalization of ECa output is conducted. The data is subdivided into 3 sets - 50% for training, 25% for testing, and the remaining 25% for validation. Models 2_(3_11_20000)_1, 2_(5_11_20000)_1 and 2_(6_11_20000)_1 networks are selected as optimum networks based on minimum ASE and minimum MARE and maximum R^2 , respectively. The statistical accuracy measures of the ANNs analysis are shown in Table 5.1. Network 2_ (5_11_20000) _1 is chosen as the best ANN network from the three optimum networks

based on their performance in VAL stage.

Table 5.1: Statistical accuracy measures of the developed model using reconnaissance data.

ANN Network		2_(3_11_20000)_1	2_(5_11_20000)_1	2_(6_11_20000)_1
TR	ASE	0.00032	0.00031	0.00033
	MARE	6.46	6.09	6.19
	R^2	0.99	0.99	0.99
TS	ASE	0.00031	0.00031	0.00033
	MARE	6.34	6.09	6.24
	R^2	0.9908	0.9911	0.9905
VAL	ASE	0.00032	0.00031	0.00033
	MARE	6.19	6.34	6.36
	R^2	0.99	0.99	0.99
TR ALL	ASE	0.00029	0.00026	0.00031
	MARE	5.64	5.73	6.22
	R^2	0.99	0.99	0.99

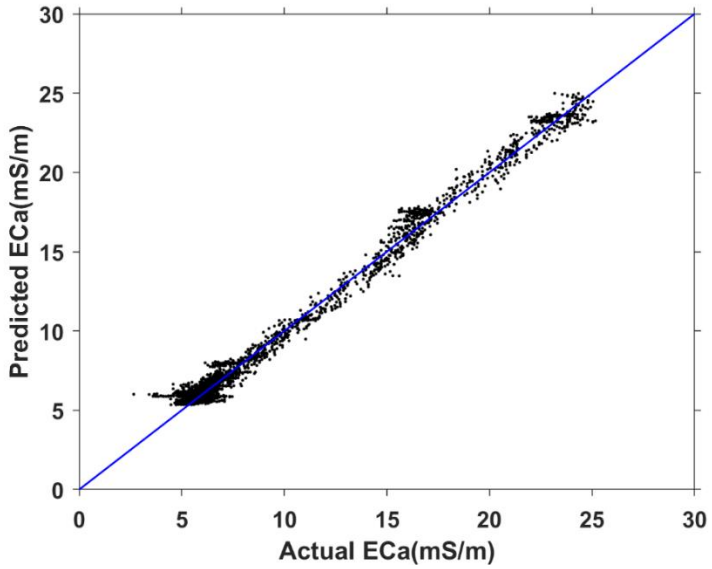


Figure 5.2: Predicted Vs. Actual ECa for reconnaissance data

From Figure 5.2, it is seen that the developed ANN model is not capable of predicting ECa less than 5 mS/m in spite of having a high R^2 value. The study area needs to be divided into several

zones and ANN analysis must be conducted for specific zones. If the ANN model could predict low ECa values, no zoning would be required and a uniform survey pattern could be chosen from Table 4.1 considering the desired accuracy.

5.4 Survey Design from Reconnaissance Data using ANNs and Quantile Method

Based on quantile method the study area is divided into 4 different zones (Figure 5.3). Zone 1(ECa: 6.13-6.743 mS/m) is divided into 2 parts:1A and 1B. Zone 2 (ECa: 4.24-6.12 mS/m) is the most critical zone because of low ECa values. Zone 3 (ECa: 7.44-15.1 mS/m) and 4 (ECa: 15.2-25.1 mS/m) have very high ECa values and the probability of getting soil pipe features is very low. To improve the performance of ANNs modeling, the chosen ANN network 2_ (5_11_20000) _1 is run for each zone. The statistical measures of the zones are listed in Table 5.2.

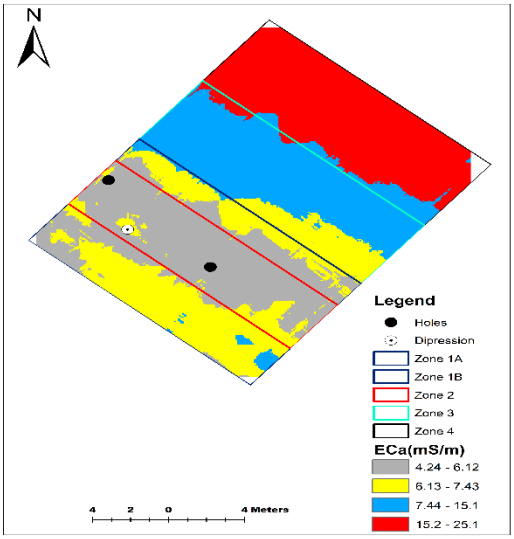


Figure 5.3: Zoning the study area using quantile method

Table 5.2 : Statistical measures of developed ANN model for different zones using reconnaissance data

Zone	No of Data	R^2 (Linear Reg)	ASE	MARE	R^2 (ANNs)
Entire Site	2370	0.81	0.00026	5.73	0.99
1A	374	0.23	0.00285	4.42	0.77
1B	122	0.41	0.00267	4.05	0.79
2	567	0.11	0.00435	7.52	0.34
3	477	0.93	0.00050	3.59	0.97
4	832	0.77	0.00068	2.30	0.98

The statistics of a specific zone are compared to the statistics of the entire site. ASE is higher but MARE is lower and R^2 is similar for the data of zone 3 and 4 compared to the ASE, MARE and R^2 of the data of entire site. But for lowest ECa zone (zone 2), the statistics are very poor (i.e. ASE and MARE are high but R^2 is very low). For zone 1A and 1B, the statistics are not satisfactory.

The survey plans listed in Table 4.7 can be used to choose a survey pattern from the reconnaissance data because the ECa ranges of the reconnaissance data are similar to the ECa ranges mentioned in Table 4.7. From Table 4.7, the recommended survey pattern for a zone having ECa between 1.51 and 6.12 mS/m is a 2D survey with 50 cm line spacing. This pattern should be used for zones having ECa between 4.24 and 6.12 mS/m to ensure maximum amount of data collection. For zone 1, a 1D survey with 50 cm line spacing should be sufficient. For zone 3 and 4, the statistics are excellent and similar to the entire site. So, no additional data should be required for zones 3 and 4.

Table 5.3: Proposed survey plan based on reconnaissance data

Zone	ECa (mS/m)	Survey orientation and spacing
1	6.13 -7.43	1D Survey 50 cm spacing perpendicular to soil pipe features
2	4.24 - 6.12	2D Survey 50 cm spacing
3	7.44 -15.1	No additional Survey is required
4	15.2 - 25.1	No additional Survey is required

Table 5.4: Statistical measures of ANN models using zone 1A and 1B data for 1D survey with 50 cm line spacing perpendicular to soil pipes.

Zone	ANNs Network	No of Data	R^2 (Linear Reg)	ASE	MARE	R^2 (ANNs)
1A	2_(4_11_20000)_1	1155	0.17	0.00030	5.84	0.51
1B	2_(4_11_20000)_1	695	0.49	0.00021	4.87	0.66

The EM survey design for the entire site based upon the reconnaissance data and ECa ranges is shown in Figure 5.4a. A total 5976 data are required. Two GUIs are created for zone 1A and 1B from developed ANN models. The statistical measures of the developed models are shown in Table 5.4. Another two GUIs are developed from ANN analysis of reconnaissance data of zone 3 and zone 4. 8700 data are generated using the four developed GUIs. 5967 measured data and 8709 GUI generated data are merged together to create a database of 14676 data. The semi variogram analysis is conducted before applying kriging interpolation to the data set. Then a 2D map of the study area is created using an 8cm cell size (Figure: 5.4b). This map is compared with the 2D map of benchmark survey (Figure 5.4c). Both maps have similar structures near the actual soil pipe features. Figure 5.4d shows the difference between the benchmark survey map and the

map produced using the optimal survey design. Assuming a ±1mS/m accuracy in measurements (blue and green), the ECa difference is acceptable for most of site with a reduced effort of 60%.

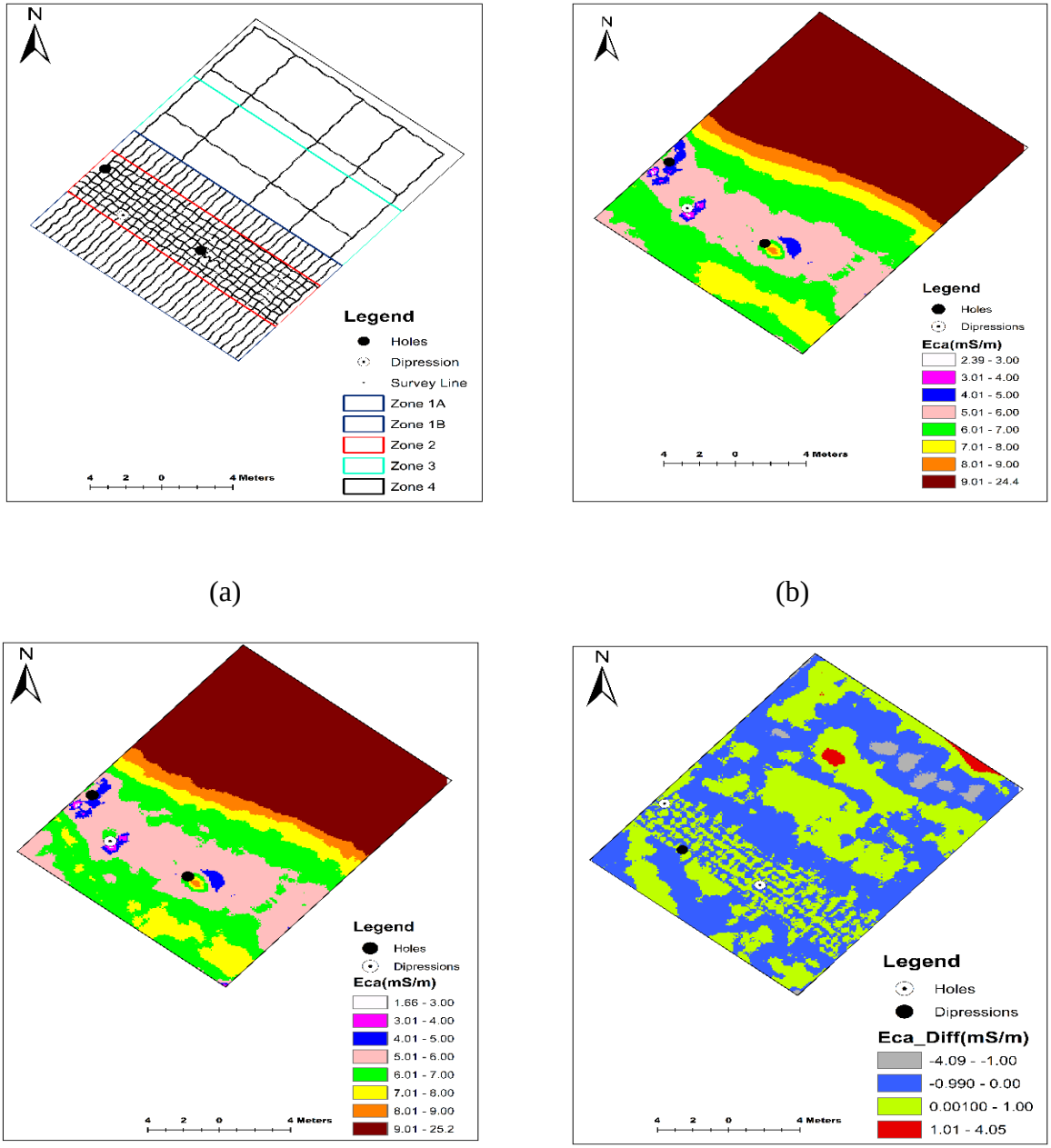


Figure 5.4: (a) Survey plan, (b) 2D map using 5967 measured data and 8709 GUI generated data, (c) 2D map using 14676 measured data; (d) map with ECa difference

CHAPTER 6

CONCLUSIONS

6.1 Summary

The goal of this research was to select appropriate survey orientations and line spacings for using electromagnetic induction equipment to detect soil pipes in a cost-effective manner. Surveying with the EM38B is relatively fast and the maximum exploration depth is approximately 1.5m, which covers the range of depth of soil pipes. EM surveys show lower apparent electrical conductivity (ECa) on soil pipe features than their surroundings. After data acquisition and processing, feed forward back propagation algorithms of ANNs were used to determine survey orientation and line spacing and optimum data acquisition speed for detecting soil pipes. It was found that if a uniform line spacing was chosen for the whole survey site, ANNs could not predict the low apparent electrical conductivity (ECa) due to the wide range of values. This issue could be solved by dividing the data using the quantile method during ANNs modeling.

Using ANN modeling for each zone, a table was developed that suggests an optimum survey design for a given ECa range. For the zone having the lowest ECa (1.51-6.12 mS/m), a 2D survey with 50 cm line spacing is required for maximum data. For zones having an ECa between 6.13-7.81 mS/m, a 1D survey with a 50 cm line spacing is required. For the zones with higher ECa

(7.82-15.9 mS/m and 16-25.6 mS/s), a wider line spacing can be used.

To validate the concept, an exercise was conducted starting with a reconnaissance survey consisting of a few lines based on surface features of soil pipes. Using the table as a guide, a survey plan was proposed and the ANN models were created using this data set. The measured and model generated data were used to create the 2D ECa map using kriging interpolation. The map was in good agreement with the benchmark ECa map, although the second map required 60% less data.

6.2 Recommendation for Future Research

The use of ANNs in geophysical survey design is a new concept. In this research, the aim was to design EM surveys to detect soil pipe features. The proposed survey design concept was validated using a reconnaissance survey at the same site. However, the approach needs to be validated by conducting surveys at multiple sites having different ECa ranges.

More data needs to be collected and included in the developed ANN models to improve model accuracies. These models can be used in future investigations to train and test with new data.

During model development, easting and northing are used as inputs to predict ECa. But other data (i.e., elevation, infrared signatures, hyperspectral signatures, height of the instrument, data acquisition speed and temperature, etc.) should be collected for use as inputs for future ANNs modeling. These models will predict ECa more accurately than the existing models. For a new survey site, the origin should be the lowest elevation point of the study area and the distance of the other data points should be calculated from that origin.

BIBLIOGRAPHY

Agrawal, S. K. and Daiutolo, H., (1992), "Reflex-percussive grooves for runways: Alternative to saw-cutting", *Transportation Research Board,* No. 836, pp. 55-60.

Bevan, B. W., (2004), "Design of Geophysical Surveys," *Technical Report: 10.13140/RG.2.2.25151.92326.*

Brevik E.C. and Fenton, T.E., (2002), "The relative influence of soil water content, clay, temperature, and carbonate minerals on soil electrical conductivity readings taken with an EM-38 along a mollisolcatena in central Iowa," *Soil Survey Horizons, No.* 43, pp. 9–13.

Brevik E.C., Fenton, T.E. and Lazarai, A., (2006), "Soil electrical conductivity as a function of soil water content and implications for soil mapping," *Precision Agric Precision Agric (2006) 7,* pp. 393–404.

Cannon, M.E., McKenzie, R.C. and Lachapelle, G., (1994), "Soil salinity mapping with electromagnetic induction and satellite-based navigation methods," *Can. J. Soil Sci. 74,* pp.335–343.

Carter, L.M., Rhoades, J.D. and Chesson, J.H., (1993), "Mechanization of soil salinity assessment for mapping," *ASAE Paper 931557, ASAE, St. Joseph, MI.*

Clark, A. J., (1986), "Archaeological geophysics in Britain", *Geophysics, 51(7),* pp. 1404 - 1413.

Curtis, A., and Maurer, H. R., (2000), "Optimizing the design of geophysical experiments—Is it worthwhile?" *Eos, Transactions, American Geophysical Union, 81,* pp.224–225.

Cybenko, G., (1989), "Approximation by superposition of a sigmoid function. Mathematics of Control", *Signals and Systems,* 2(4), pp. 303-314.

Dai, H., and MacBeth, C., (1994), "Split shear-wave analysis using an artificial neural network," *First Break, 12,* pp.605–613.

Doolittle, J.A., Sudduth, K.A., Kitchen, N.R., Indorante, S.J., (1994), "Estimating depths to claypans using electromagnetic induction methods*."J. Soil Water Cons. 49 (6*), pp. 572–575.

Dowla, F. U., Taylor, S. R., and Anderson, R. W., (1990), "Seismic dis- crimination with artificial neural networks: Preliminary results with regional spectral data," Bull. Seis. Soc. Am., 80, pp.1346–1373.

Fausett, L., (1994), “Fundamentals of Neural Networks: Architectures, Algorithms, and Applications,” Prentice-Hall, Inc., Upper Saddle River, NJ.

Funahashi, K. I., (1989), “On the approximate realization of continuous mapping by neural networks,” Neural Networks, 2(3), pp. 183-192.

Geonics (1998), “EM38 ground conductivity meter operating manual,” Geonics Ltd, Mississauga, Ont., Canada.

Ghaboussi, J., Garett, J. H. and Wu, X., (1991), “Knowledge-based modeling of material behavior with neural networks”, Journal of Engineering Mechanics, 117 (1), pp. 132-153.

Wilson, G. V., Rigby, J.R. and Dabney, S.M., (2015), “Soil pipe collapses in a loess pasture of Goodwin Creek watershed, Mississippi: Role of soil properties and past land use,” Earth

Surf. Process. Landforms.

Hartman, E. J., Keeler, J. D., and Kowalski, J. M., (1990), “Layered neural networks with Gaussian hidden units as universal approximations,” Neural Computation, 2(2), pp. 210-215.

Haykin, S., (1999), “Neural networks: A Comprehensive Foundation”, Second Edition, Prentice Hall, Upper Saddle River, New Jersey.

Hopfield, J.J., (1982), “Neural networks and physical systems with emergent collective computational abilities,” Proc. National Academy of Science, USA, Vol. 79, pp. 2554-2558.

Hornik, K., Stinchcombe, M. and White, H., (1989), “Multilayer feedforward networks are universal approximators,” Neural Networks, 2 (5), pp. 359-366.

Huang, Z., Shimeld, J., Williamson, M., and Katsube, J., (1996), “Permeability prediction with artificial neural network modeling in the Ventura gas field, offshore eastern Canada,” Geophysics, 61, pp.422–436.

Jaynes, D.B., Colvin, T.S. and Ambuel, J., (1993), “Soil type and crop yield determinations from ground conductivity surveys,” ASAE Paper 933552, ASAE, St. Joseph, MI.

Jaynes, D.B., Novak, J.M. and Moorman, T.B., Cambardella, C.A., (1995), “Estimating herbicide partition coefficients from electromagnetic induction measurements,” J. Environ. Qual. 24, pp. 36–41.

Katsube, T.J., Keating, P.K., McNairn, H., Das, Y., DiLabio, R., Singhroy, V., Connell-Madore, S., Best, M.E., Hunter, J., Klassen, R. and Dyke, L., (2004), “Soil moisture and electrical conductivity prediction and their implication for landmine detection technologies, in,” *Detection and Remediation Technologies for Mines and Minelike Targets IX*.

Kitchen, N.R., Sudduth, K.A. and Drummond, S.T., (1996), “Mapping of sand deposition from 1993 midwest floods with electromagnetic induction measurements,” J. Soil Water Cons. 51, pp.336–340.

Kitchen, N.R., Sudduth, K.A., Drummond, S.T., (1996), "Mapping of sand deposition from 1993 midwest floods with electromagnetic induction measurements," *J. Soil Water Cons. 51*, pp. 336–340.

Kuhnle, R.A., Bingner, R.L., Alonso, C.V., Wilson, C.G. and Simon, A., (2008), "Conservation practice effects on sediment load in the Goodwin Creek Experimental Watershed," J. Soil Water Conserv.

Langer, H., Nunnari, G., and Occhipinti, L., (1996), "Estimation of seismic waveform governing parameters with neural networks," *J. Geophys. Res*., 101, 20109–20 118.

Lesch, S.M., Strauss, D.J., Rhoades, J.D., (1995), "Spatial prediction of soil salinity using electromagnetic induction techniques: II. An efficient spatial sampling algorithm suitable for multiple linear regression model identification and estimation*," Water Resources Research 31*, pp. 87–398.

Maurer, H. C., Curtis, A., and Boerner, D.E., (2010), "Recent advances in optimized geophysical survey design," Geophysics, vol 75, no. 5, pp. A177-A194.

Maurer, H. R., and Boerner, D. E., (1998b), "Optimized design of geophysical experiments," The Leading Edge, 17, pp.111-119.

McBride, R.A., Gordon, A.M. and Shrive, S.C., (1990), "Estimating forest soil quality from terrain measurements of apparent electrical conductivity," Soil Sci. Soc. Am. J. 54, pp. 290–293.

McCormack, M. D., Zaucha, D. E., and Dushek, D. W., (1993), "First-break refraction event picking and seismic data trace editing using neural networks," Geophysics, 58, 67–78.

Mcneill, J.D., (1980), "Electromagnetic Terrain Conductivity Measurement at Low Induction Number," Technical Note TN 6. Geonics Limited, Mississauga, ON, Canada.

McNeill, J.D., (1992), "Rapid, accurate mapping of soil salinity by electromagnetic ground conductivity meters. In: Advances in Measurement of Soil Physical Properties: Bringing Theory Into Practice," *Spec. Publ. 30, SSSA, Madison, WI,* pp. 209–229.

Murat, M. E., and Rudman, A. J., (1992), "Automated first arrival picking: A neural network approach," Geophys. Prosp., 40, 587–604.

Najjar, Y. M., and Basheer, I. A., (1996), "Stress-strain modeling of sands using artificial neural networks," *ASCE, Journal of Geotechnical Engineering, 122(11),* pp. 949-951.

Najjar, Y. M., Basheer, I. A., and McReynolds, R. L., (1997), "Modeling the durability of aggregate used in concrete pavement construction: A neuro-reliability-based approach," Final Report no.: KS-97-3, Kansas Department of Transportation.

Poulton, M. M., Sternberg, B. K., and Glass, C. E., (1992), "Location of subsurface targets in geophysical data using neural networks," *Geo-physics, 57,* 1534–1544.

Robinson, D.A., Lebron, I., Lesch, S.M. and Shouse, P., (2004), "Minimizing Drift in Electrical Conductivity Measurements in High Temperature Environments using the EM-38," Soil Sci. Soc. Am. J.

Roth, G., and Tarantola, A., (1994), "Neural networks and inversion of seismic data," *J. Geophys. Res., 99,* pp. 6753–6768.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J., (1986), "Learning representation by back-propagating errors," *Nature, 323(6088),* pp. 533-536.

Sheets, K.R. and Hendrickx, J.M.H., (1995), "Noninvasive soil water content measurement using electromagnetic induction," *Water Resources Research 31.*. issn: 0043-1397.

Simson, Patrick K., (1990), "Artificial neural systems," Pergamon Press, New York, NY.

Sudduth, K.A., Drummond, S.T. and Kitchen, N.R., (2001), "Accuracy issues in electromagnetic induction sensing of soil electrical conductivity for precision agriculture," *ELSEVIER, Computers and Electronics in Agriculture, 31 (2001),* pp. 239–264

Wang, L.-X., and Mendel, J. M., (1992), "Adaptive minimum prediction- error deconvolution and source wavelet estimation using Hopfield neural networks," Geophysics, 57, 670–679.

Williams, B.G. and Hoey, D., (1987), "The use of electromagnetic induction to detect the spatial variability of the salt and clay contents of soils," *Aust. J. Soil Res*. 25, pp. 21–27.

Wilson, G. V., Rigby, J.R. and Dabney, S.M., (2015), "Soil pipe collapses in a loess pasture of Goodwin Creek watershed, Mississippi: Role of soil properties and past land use," Earth Surf. Process. Landforms.

Yasarer, H., (2010), "Characterizing the permeability of concrete mixes used in transportation applications: A neuronet approach, "MS Thesis, Kansas State University.

Yasarer, H., and Najjar, Y., (2010), "Permeability prediction model for concrete mixes used in Kansas PCC pavements," Transportation Research Board 91st Annual Meeting, Report# 12-3457.

Zajicova, K. and Chuman, T. (2019), "Application of ground penetrating radar methods in soil studies: A review," *ELSEVIER, The Global Journal of Soil Science,* pp. 116-129.

Zhang, T., and Wilson, G. V, (2012), "Spatial distribution of pipe collapses in Goodwin Creek Watershed, Mississippi," Hydrol. Process. 27, 2032–2040.

Zupan, J. and Gasteiger, J., (1993), "Neural Networks for Chemists: An Introduction," VCH Publishers, New York, NY.

VITA

Shariful Islam Sharif is from Dhaka, Bangladesh. He completed his BSc in Civil Engineering from Bangladesh University of Engineering and Technology in September, 2015. From October, 2015 to April, 2019 he served Bashundhara Group, Bangladesh Building Systems Ltd. and Corps of Engineers, Bangladesh Army. In May, 2019 he joined National Center for Physical Acoustics as a graduate research assistant. He has expertise in ArcGIS, ANNs Modeling, designing earth retaining structure and pile foundation and conducting geophysical survey for subsoil investigation. After graduating from University of Mississippi, his plan is to get involved a commercial organization to practice geophysics and geotechnical engineering.