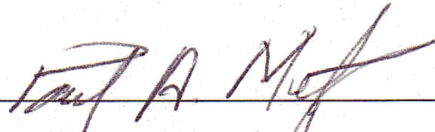


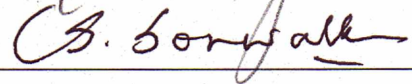
CHARACTERIZATION OF SUBSURFACE HYDRAULIC CONDUCTIVITY
ALONG THE PROPOSED ALASKA GAS LINE CORRIDOR USING
GEOPHYSICAL SIGNATURES

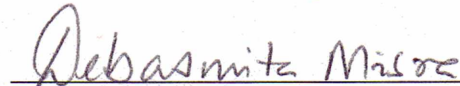
By

Peter A. Calvin


RECOMMENDED:








Advisory Committee Chair

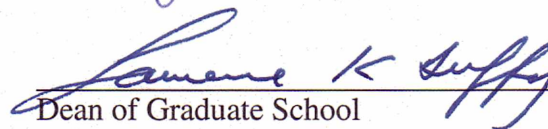


Chair, Department of Mining and Geological Engineering

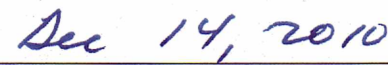
APPROVED:



Dean, College of Engineering and Mines



Dean of Graduate School



Date

CHARACTERIZATION OF SUBSURFACE HYDRAULIC CONDUCTIVITY
ALONG THE PROPOSED ALASKA GAS LINE CORRIDOR USING
GEOPHYSICAL SIGNATURES

A
THESIS

Presented to the Faculty
of the University of Alaska Fairbanks

in Partial Fulfillment of the Requirements
for the Degree of

MASTER OF SCIENCE

By
Peter Calvin, B.S.

Fairbanks, Alaska

December 2010

ABSTRACT

The objective of this research was to explore a cost-effective and non-invasive methodology to characterize spatial variability of hydraulic conductivity using airborne electromagnetic (AEM) signatures as an alternative to traditional techniques such as borehole sampling. The relationship of AEM measured apparent resistivity and magnetic field strength was explored using a small dataset that included 180 natural moisture (NM) content data and a total dataset of 546 grain size distributions that excluded the NM. The grain size distributions were used to develop soil indicator parameter and to estimate the hydraulic conductivity (K^*) using pedo-transfer functions. Predictive models were developed using three techniques; artificial neural network regression (ANNR), support vector regression (SVR), and artificial neural network classification (ANNC). The sole use of non-invasive parameters to characterize K^* proved insufficient. The inclusion of supplemental invasively collected parameters showed ANNR to best characterize the relationship ($R^2 = 0.64$) with the smaller dataset; while the SVR model performed best with the total dataset ($R^2 = 0.57$). ANNC was shown to be a viable alternative (overall accuracy = 88%) when broad characterization of K^* was sufficient. This study lays out a methodology that could be used for future K^* characterization using improved data set.

Table of Contents

	Page
Signature Page	i
Title Page	ii
Abstract	iii
Table of Contents	iv
List of Figures	vii
List of Tables	ix
Acknowledgements	x
1.0 INTRODUCTION	1
1.1 RESEARCH HYPOTHESES	3
1.2 OBJECTIVE	3
2.0 BACKGROUND	5
2.1 STUDY AREA	5
2.1.1 Hydrogeology of Upper Tanana River Valley	8
2.2 HYDRO-GEOPHYSICAL RELATIONSHIP	11
2.2.1 Electromagnetic Survey Measurements	12
2.2.2 Airborne Electromagnetic Surveys	15
2.3 ESTIMATION OF HYDRAULIC CONDUCTIVITY	18
3.0 DATA AND METHODS	21
3.1 DATA	21

3.1.1	Geophysical Data	21
3.1.2	Geologic Data	22
3.1.3	Hydraulic Data	22
3.2	METHODS.....	23
3.2.1	Artificial Neural Networks	24
3.2.2	Support Vector Machines	27
4.0	EXPLORATORY DATA ANALYSIS.....	31
4.1	GEOPHYSICAL DATA ANALYSIS	31
4.1.1	Electromagnetic Data Analysis.....	31
4.1.2	Magnetic Data Analysis.....	39
4.2	GEOLOGIC DATA ANALYSIS.....	42
4.2.1	Grain Size Distribution Analysis	42
4.2.2	Natural Moisture Content Analysis	48
4.3	HYDRAULIC DATA ANALYSIS	51
4.3.1	Estimation of Hydraulic Conductivity	51
4.3.2	Comparison of Hydraulic Conductivity Estimates	51
4.4	CORRELATION ANALYSIS.....	56
5.0	RESULTS AND DISCUSSION	60
5.1	MEASURES OF PERFROMANCE.....	60
5.2	ANALYSIS OF HYDRO-GEOPHYSICAL RELATIONSHIP	65

5.2.1	Neural Network Regression Analysis	65
5.2.2	Support Vector Regression Analysis	70
5.2.3	Neural Network Classification Analysis.....	76
5.2.4	Comparison of Model Predictions	79
6.0	CONCLUSIONS AND FUTURE WORK	83
	References.....	86

LIST OF FIGURES

	Page
Figure 2.1: Location and bounds of study area.....	7
Figure 2.2: Delineation of the hydrogeologic boundaries.	10
Figure 2.3: Illustration of apparent resistivity as a volume average.....	13
Figure 2.4: Basic electromagnetic system.	15
Figure 3.1: Feed forward multilayered artificial neural network.....	25
Figure 3.2: Nonlinear SVR using Vapnik's ϵ -insensitive loss function.....	28
Figure 4.1: Histogram of apparent resistivity at 140K and 40K Hz.	33
Figure 4.2: Contour plots of apparent resistivity	35
Figure 4.3: Contour plot of difference between ρ^* measured at 140K and 40K Hz.	36
Figure 4.4: Semi-variogram ρ^* at 140K Hz.....	38
Figure 4.5: Semi-variogram of ρ^* at 40K Hz.	38
Figure 4.6: Histogram plot of magnetic data.	40
Figure 4.7: Contour plot of magnetic data.	41
Figure 4.8: Semi-variogram magnetic data.....	41
Figure 4.9: Histograms of primary grain size distribution parameters.	44
Figure 4.10: Soil classifications description and frequency	46
Figure 4.11: Semi-variogram of soil indicator.....	47
Figure 4.12: Histogram natural moisture content percentage.....	49
Figure 4.13: Semi-variogram of natural moisture content percentage.	50
Figure 4.14: Histogram plots of natural-log transformed K^* distributions.	54

Figure 4.15: Semi-variogram of K^* obtained using (Cosby et al., 1984) PTF	55
Figure 4.16: Distribution of K^* with respect to soil class	58
Figure 4.17: Scatter plot of geophysical parameters ρ^* at 140K and H vs. K^*	59
Figure 5.1: Regression plot ANNR model results with an input of ρ^*	68
Figure 5.2: Regression plot of ANNR model results with inputs of ρ^* , SI, and XY	68
Figure 5.3: Regression plot of ANNR model results with inputs of ρ^* , XY, SI, NM.	69
Figure 5.4: Regression plot of SVR model results with an input of ρ^*	74
Figure 5.5: Regression plot of SVR model results with inputs of ρ^* , H, XY, and SI	74
Figure 5.6: Percentage of support vectors in the range of predictor variables.	75

LIST OF TABLES

	Page
Table 4.1: Summary statistics of the apparent resistivity data.....	32
Table 4.2: Statistical summary of magnetic data.	39
Table 4.3: Statistical summary of borehole sample grain size distributions.....	43
Table 4.4: Statistical summary of natural moisture content from borehole samples.....	48
Table 4.5: Statistical properties of estimated K^*	53
Table 4.6: Lognormal distribution parameters of K^*	53
Table 4.7: Comparison of population parameters between estimated and USDA measured K	53
Table 4.8: Coefficient of correlation between parameters.....	57
Table 5.1: Statistical summary of training and testing subsets from Total Dataset.....	64
Table 5.2: Statistical summary of the training and testing subsets from NM Dataset.....	64
Table 5.3: Performance measures for the ANNR model analysis.	67
Table 5.4: Performance measures and parameters for the SVR model analysis.	73
Table 5.5: Statistical summary of k-means clusters.....	76
Table 5.6: Neural network classification results.....	78
Table 5.7: Performance measures obtained using the best combination of input parameters from both the Total and the NM Dataset.....	82
Table 5.8: Accuracy of the best combination of input parameters from both the Total and NM Dataset for the classification model discussed in section 5.3.....	82

ACKNOWLEDGEMENTS

Throughout my graduate program, I have had the fortune of working and interacting with a number of exceptional people who contributed to both my education and research. First of all I would like to thank my advisor, Dr. Debasmita Misra. His expertise and guidance on hydrology and data analysis were invaluable throughout my research and study at the University of Alaska, Fairbanks. Our frequent discussions continually expanded my knowledge base while providing incite and encouragement. I would also like the Dr. Thomas Oommen for his assistance and expertise with respect to machine learning algorithms. His support provided me the knowledge and tools to delve deeper into analysis of my results while generously helping me work through issues as they arose.

I would like to thank my advisory committee members, Dr. Vikas Sonwalkar, and Dr. Paul Metz, for their guidance and constructive criticism of my work. In addition to the individuals above, I would like to thank Dr. Rajive Ganguli, Dr. Scott Huang and Dr. Margaret Darrow for their continual support and advice throughout the completion of my degree. I would also like to thank Dr. Diana Solie and Ms. Laurel Burns of the Alaska Department of Geological and Geophysical Surveys for providing the geophysical and geotechnical data. Without their assistance in the compilation, interpretation and processing of the data this work would not have been possible.

The funding that made this research and my study possible was generously provided by the Berry family through the Clarence Berry Fellowship and the University of Alaska Fairbanks Graduate School through the graduate school fellowship.

1.0 INTRODUCTION

Any design of facility or structure requires the knowledge of the subsurface hydraulic characteristics. These subsurface hydraulic characteristics are especially significant in northern latitudes where a disruption in the groundwater flow can alter the thermal energy balance leading to costly implications such as those caused by frost heave and excessive ice or hydraulic pressure. A parameter used to reflect the physical flow characteristics of the subsurface geologic medium is the hydraulic conductivity (K) measured in LT^{-1} . Hydraulic conductivity is known to have one of the highest degree of variability of all geotechnical and hydraulic properties with its magnitude varying over ten orders of magnitude ranging between 10^{-9} cm/s for clayey soils to 10^0 cm/s or higher for coarse gravel and crushed rock (Mbonimpa et al., 2002). Due to the large range of values and a high degree of natural spatial variability that is dependent on the deposition, highly clustered data sets are often required to accurately characterize the subsurface groundwater flow regime. Because areas of potential development especially in the arctic often involve vast regions of remote, rugged, and environmentally sensitive land, conventional methods of investigation such as borehole sampling and well tests can be economically prohibitive and environmentally invasive.

The Alaska Highway corridor between Delta Junction and Tok, Alaska, has long been of significant importance due to its role as a vital land-based transportation link between interior Alaska and the rest of the North America. More recently the significance of this corridor has been elevated due to a proposed buried chilled natural

gas pipeline extending from Alaska's North Slope to Alberta, Canada with an estimated cost of thirty five billion dollars that would travel through the corridor. Recent investigations into the area have shown that the majority of the corridor to be underlain by the pipeline falls under a discontinuous permafrost regime (Reger and Solie, 2008). A project of this type and scale will most certainly have a significant influence on the thermal balance of the subsurface. Due to the strong connection between energy and water balances (Woo, 1986), mitigating the danger of differential frost heave from the formation of ice lenses in the soil inducing potentially damaging loads to the pipeline in areas of discontinuous permafrost is critical (Kim et al., 2008). Because of the limited amount of available hydrologic data in the area, adequately addressing this issue in the design of the pipeline would require a significant amount of borehole sampling to characterize the hydraulic regime influencing the thermal energy balance of the subsurface. Such sampling will add considerably to both the cost and the potential for environmental damage.

A viable alternative to the large scale borehole sampling employs the relationship between hydraulic and electrical properties shown in lab and field scale research (Archie, 1942; Topp et al., 1980; Slater and Lesmes, 2002; Friedman, 2005; Doussan and Ruy, 2009). By employing the high density data coverage of the airborne electromagnetic (AEM) survey (less than 3 meter sample spacing) which measures the apparent resistivity and magnetic field strength, in conjunction with the relationship between electrical and subsurface hydraulic properties, an alternative method of subsurface hydraulic characterization using machine learning algorithms may be explored. Due to a limited

amount of hydraulic data the relationship between subsurface hydraulics and the AEM data will be explored using empirically derived hydraulic conductivity data estimated using grain size distributions. The method could offer potential improvements by characterizing the high degree of spatial variability while potentially providing cost savings and lowering the risk of environmental damage.

1.1 RESEARCH HYPOTHESES

Considering the complexity of quantifying the hydro-geophysical relationship within the Alaska Highway corridor between Delta Junction and Tok, Alaska the hypothesis of this research is bifold, as presented below:

- 1) That there exists a correlation between electrical properties and hydraulic conductivity of a porous media. On basis of this hypothesis, the research addresses the characterization of hydraulic conductivity of an example study area using AEM data.
- 2) That relationship between electrical properties and subsurface hydraulic conductivity is multifaceted and necessitates the use of machine learning algorithms to overcome the issues of complexity and multi-dimensional data.

1.2 OBJECTIVE

The objective of this research is to explore the relationship between the geophysical data acquired through an AEM survey and the estimated hydraulic conductivity values by using machine learning algorithms (MLA) so that it may be used

in predictive modeling or characterization of subsurface heterogeneity. Due to a limited amount of hydraulic data within the area empirically derived hydraulic conductivity data estimated using grain size distributions will be used in place of field measured values to explore the relationship. In order to accomplish this, the following items are included in the scope of this research:

- 1) Review of past studies and research outlining the general hydrogeologic conditions of the study area.
- 2) Determination of applicable machine learning algorithms for the development of predictive models from which hydraulic conductivity can be estimated.
- 3) Refinement and analysis of AEM data.
- 4) Compilation, refinement, and analysis of geologic data.
- 5) Development of hydraulic data set using the available grain size distributions in conjunction with empirically derived functions.
- 6) Analysis of the statistical relationships between the hydraulic, geophysical and geologic data.
- 7) Development and analysis of predictive models estimating hydraulic conductivity using machine learning algorithms.

2.0 BACKGROUND

The characterization of subsurface hydraulic conductivity using AEM surveys cannot be complete without first understanding of the topics and theory that this research is based upon. This chapter provided a brief description of the topics, theory, and literature pertaining to the items listed below:

- Overview of study area detailing its location and the general hydrologic and geologic conditions.
- Summary of the theory behind the relationship between the electrical and hydraulic properties of a porous geologic medium.
- Overview of AEM surveys.
- Evaluation of the estimation of hydraulic conductivity using grain size distributions.

2.1 STUDY AREA

The area of interest for this study lies along the Alaska Highway corridor in interior Alaska bounded by the city of Delta Junction (65°02'16" N, 145°43'56" W) at the North West extent and the city of Tok (63°20'12" N, 142°59'8" W) at the South East extent (**Figure 2.1**). The corridor is approximately 108 miles long in the northwest direction lying primarily in the Yukon-Tanana lowlands. The lowlands consist primarily of unconsolidated glacial and alluvial deposits by the Alaska Mountain Range to the Southwest and the Yukon-Tanana upland to the northeast. The peaks of the Alaska Range adjacent to the Tanana highway corridor rise to elevations of over 6,500 ft above

sea level while the hills of the Yukon-Tanana upland north of the corridor rise to over 3,300 ft above sea level. The main axial stream within the area is the Tanana River, a large, meandering, braided stream traveling in the northwest direction along the lowland (Reger et al., 2008). The stream descends from the approximate elevations of 1,600 to 1,000 ft above sea level between Tok and Delta Junction, respectively.



Figure 2.1: Location and bounds of study area [digital elevation model of study area adapted from Solie and Burns, (2007) and image of Alaska adapted from Saskal, (2008)].

2.1.1 Hydrogeology of Upper Tanana River Valley

The first hydrologic investigations within the upper Tanana River Valley were conducted by Pewe (1955) investigating the groundwater and springs in the vicinity of Big Delta in his study of permafrost and groundwater in the Middle Tanana Valley. Pewe (1955) found the groundwater gradient to slope in the northwest direction at three to five feet per mile. Subsequently, further investigations confirmed the observations noted in Pewe (1955) through surveys of piezometric levels, ground water gradients, and soils (U.S. Army Corps of Engineers, 1959; Wallner et al., 1961; Holmes, 1965).

The general hydrogeology, water quality and surface water conditions were first mapped by Anderson (1970) in his investigations into the general hydrologic conditions in the Tanana Basin. He found that the main supply to the aquifers came from seepage from the surrounding glacial streams indicating that the water table and natural water content of the surrounding drainages are likely to be higher than the plateaus and valley floors. Wilcox (1980) later reported on the hydrogeology of the Delta-Clearwater area, describing the ground and surface water conditions of the area as well as the influence of permafrost. It was noted that the main rivers and creeks in the area including the Tanana River are perched above the groundwater table and provide a significant amount of recharge to the aquifer. The loss of water from the rivers to the aquifer was shown through the significant increase in water table depth measured from wells as the distance from surface water corridors increased. The water table depth was shown ranged from near surface to roughly up to 400 ft below the surface one mile away from local wells sampled.

More recently detailed surficial and engineering geologic investigations and maps have been provided by Reger et al., (2008). The maps and report detail the distribution of unconsolidated geologic material and their depositional characteristics. These delineations provide a base from which significant alterations in subsurface hydraulic characteristics can be related back to the changes in physical properties of the geologic media.

The principal factors observed in these reports (Wilcox, 1980; Reger et al., 2008) include that the aquifer system of the highway corridor is primarily composed of thick unconsolidated alluvial and glacial deposits of gravel, sand, and silt at depths of over 2,500 feet. The water table within these aquifers ranges from near surface to over 400 ft below the ground surface indicating that the subsurface is relatively permeable and well drained. These large relatively permeable aquifers are resultant of the terminal moraines, broad piedmont fans, and coarse outwash of the Alaska Mountain Range. The stratification of the deposits varies widely with lenticular deposits of gravel, sand, and silt forming an interconnected system of aquifers separated by leaky confining layers. This system of aquifers is bounded to the North and South by the bedrock of the Alaska Mountain Range and Yukon-Tanana uplands, respectively. These formations provide the hydro-geologic boundaries controlling regional groundwater flow defining the Tanana River drainage basin shown in **Figure 2.2**.

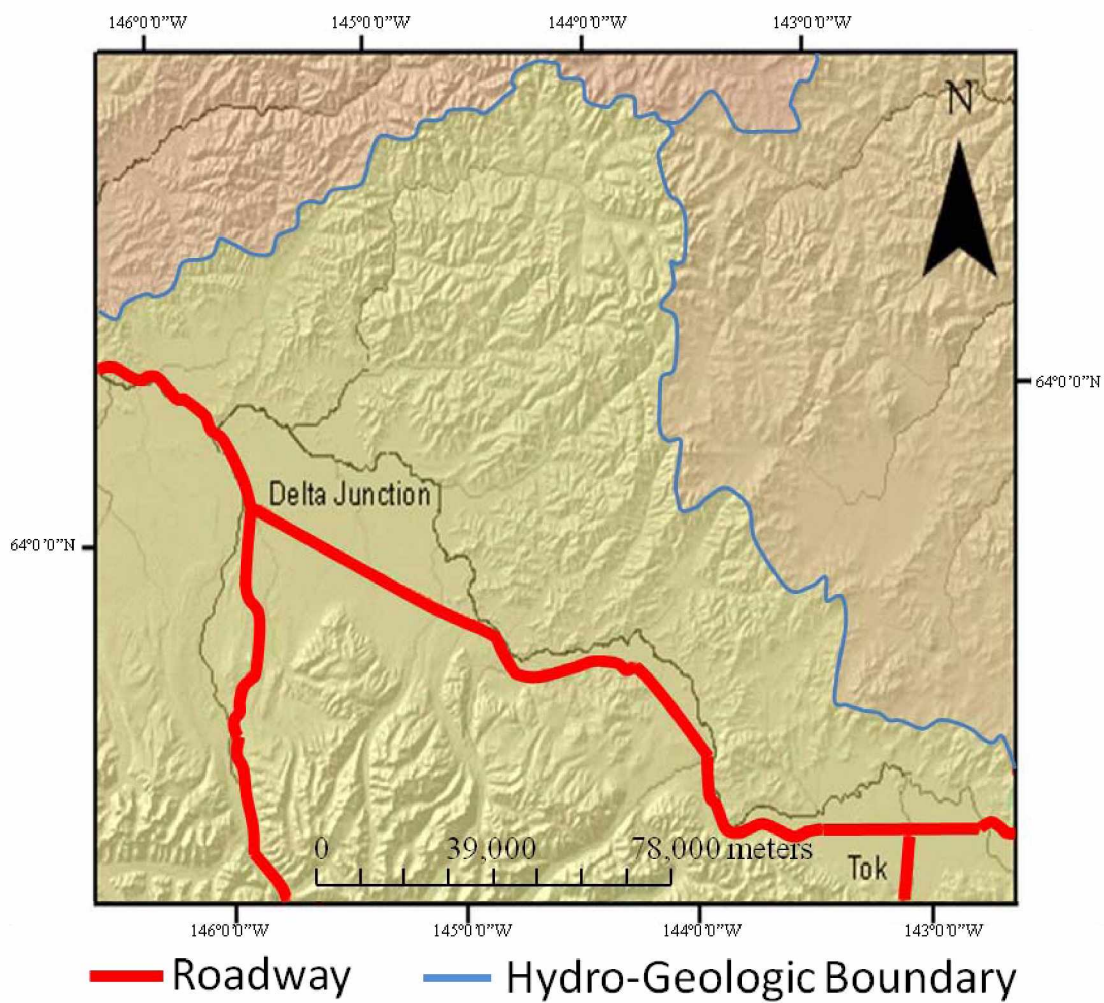


Figure 2.2: Delineation of the hydrogeologic boundaries.

2.2 HYDRO-GEOPHYSICAL RELATIONSHIP

Water and the accompanying electrolytes that are found in the pore spaces of much of the near surface rocks and soils have been shown to be the important factors controlling the electrical conductivity (the ability of a media to conduct electric current) of geologic media (Zhdanov, 2009). The relationship between the conductivity of the porous medium and the water filled pore spaces is well known and described using the Archie's law (Archie, 1942, 1947, 1950) as,

$$\sigma_t = a\sigma_w W^m \quad (2.1)$$

where, σ_t is the bulk electrical conductivity of the geologic medium with units of Siemens per meter, σ_w is conductivity of the water in the pore spaces with units of Siemens per meter, W is the water content as a volume fraction of the medium, and a and m are empirical parameters accounting for the physical characteristics of the medium. Because the conduction of electrical current through geologic media takes place almost entirely through the fluid within the pore spaces, it is influenced by the same constraining factors that influence fluid flow such as porosity, connectivity, and tortuosity (Lesmes and Friedman, 2005).

In the case of solid electrolytes which account for the majority of rock-forming minerals, the electrical conductivity is largely dependent on quantity and mobility of free charged particles. The number and mobility of these charged particles is significantly influenced by multiple factors including water content, chemical composition, weathering process, and temperature. Typical conductivities of unsaturated rock and sediment are under 100 milliSiemens per meter (mS/m) but can be amplified to multiple thousand

mS/m depending on the water content and salinity (Zhdanov, 2009). In geophysical surveys the measurement of electrical resistivity (ρ) measured in units of ohm-meters is often used in place of electrical conductivity. This is done with the understanding that for isotropic materials electrical resistivity is simply defined as the reciprocal of electrical conductivity shown below:

$$\rho = \sigma^{-1} \quad (2.2)$$

The key step in quantifying the hydro-geophysical relationship is the transformation of the measured geophysical data to their corresponding hydraulic properties. Studies done in various geologic settings have shown multiple types of relationships including direct, inverse, and log relationships with widely varying degrees of accuracy ranging from an R^2 of 0.5 to 0.96 depending on the medium and wave frequency employed (Slater and Lesmes, 2002). These studies indicate a complex relationship that is dependent both on the physical and chemical characteristics of the fluid and geologic media as discussed above. Because of this, we are inclined to believe that the relationship is too complex to employ universal functions to describe such relationship and that a site dependent approximation would be more applicable.

2.2.1 Electromagnetic Survey Measurements

The frequency domain electromagnetic (FDEM) method is commonly employed in geologic and hydrologic investigations due to its versatility, high density data coverage, and range of coverage (Paine and Minty, 2005). FDEM surveys typically map the primary measurement of apparent electrical resistivity (ρ^*) and the secondary

measurement magnetic field strength (H) measured in ohm-meters ($\Omega\text{-m}$) and nano Teslas (nT) respectively.

The primary FDEM measurement, apparent electrical resistivity (ρ^*), which measures how strongly a material opposes the flow of electrical current, is of special interest in geologic and hydrologic investigations due to the relationship between the conduction of electrical current and fluid flow (see Section 2.2). Apparent electrical resistivity (ρ^*) differs from electrical resistivity in that it is the volume average of a heterogeneous material (**Figure 2.3**) rather than the intrinsic property of microscopic volume of material.

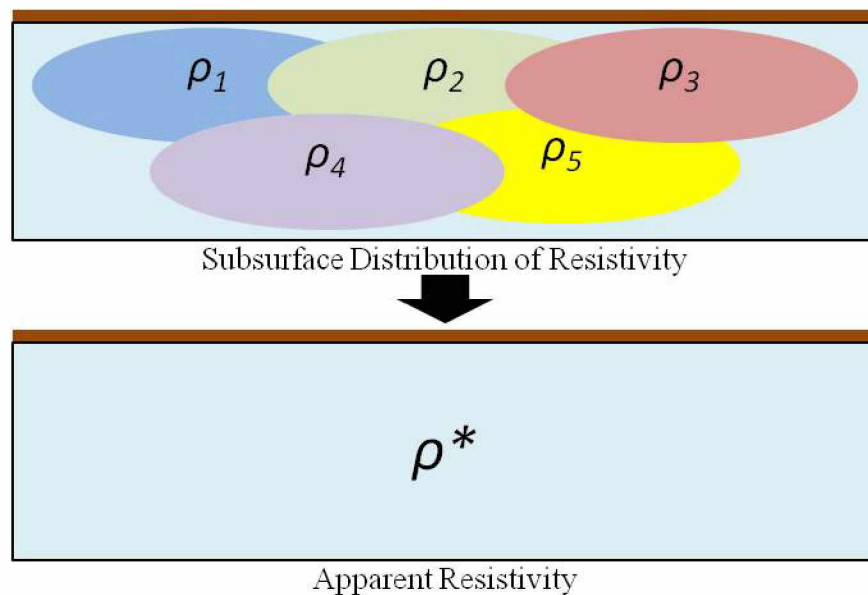


Figure 2.3: Illustration of apparent resistivity as a volume average.

The secondary FDEM measurement of magnetic field strength (H), measures changes in earth magnetic field. The magnetic field strength is not influenced by changes in the shallow subsurface hydraulic parameters but has been shown to be useful in hydrogeologic investigations through their ability to delineate basin geometry and certain types of faults, paleochannels, eolian deposits, and igneous intrusions that could have implications with respect to regional groundwater flow (Paine and Minty, 2005).

The magnetic survey generally employs the use of the International Geomagnetic Reference Field (IGRF), which is a mathematical description of the earth's main magnetic field to map deviations from the expected. By removing the IGRF, flight lines flown at different times can be merged and appropriately compared with previous surveys. Deviations from the expected field strength indicate changes in magnetic permeability of the media. Magnetic permeability is the ability of a material to be magnetized through the application of an external field. The magnetic field strength and magnetic permeability can be related through Maxwell's third constitutive equation relating the behavior of electromagnetic waves to the electrical properties of the earth (Huang and Fraser, 2001):

$$B = \mu H \quad (2.3)$$

where H is the magnetic field strength with units of teslas, μ is the magnetic permeability with units of henries per meter of free space, and B is the resulting magnetic induction with units of weber per square meter (Zhdanov, 2009).

2.2.2 Airborne Electromagnetic Surveys

The airborne FDEM survey equipment is typically composed of multiple transmitter and receiver coils housed together in a “bird”. The bird is towed behind an airborne platform on a predesigned survey grid. Throughout the operation, each transmitter coil emits an EM field at a fixed frequency. The electromagnetic fields propagate through the ground interacting with the geologic medium inducing eddy currents. These eddy currents in turn induce a secondary electromagnetic field. The receiver coils measure the resulting EM field, which is the super position of the primary and secondary fields as shown in **Figure 2.4**.

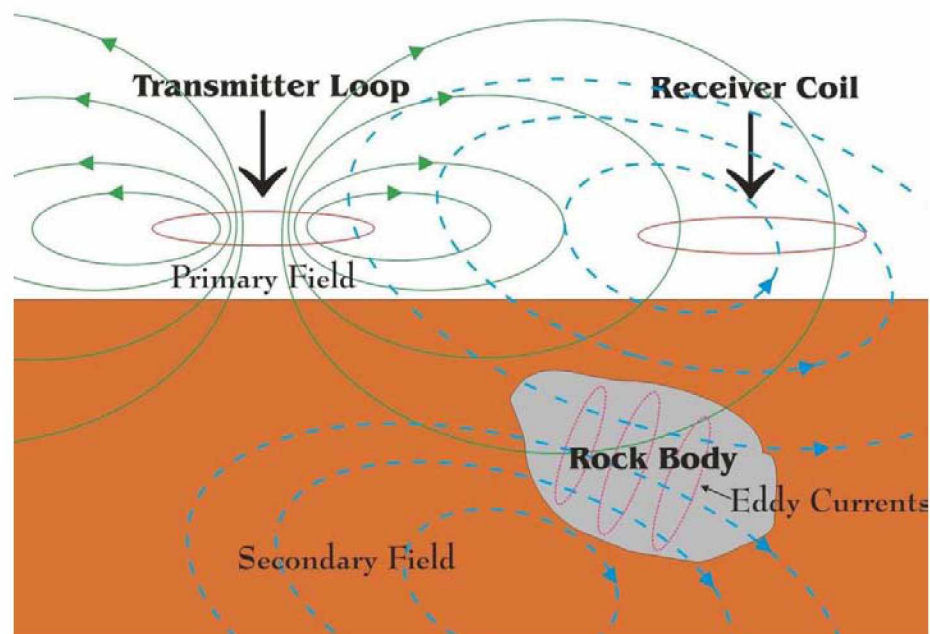


Figure 2.4: Basic electromagnetic system (Source: Scrivens (2005)).

The strength of the secondary EM field and depth of exploration is a complex relationship between the transmitted frequency and the electrical resistivity of the ground. Generally, as the conductivity of the ground increases, the strength of the secondary field increases, while the depth of exploration decreases. Due to inverse relationship between exploration depth and both EM frequency and ground conductivity, FDEM surveys employ multiple frequencies to provide a relative representation of how the conductivity of the ground changes with increasing depth. The skin depth can be approximated as the depth at which the transmitted field strength is reduced to thirty seven percent of its original strength. The exploration depth, also known as skin depth can be calculated using the equation

$$\delta = 503 \sqrt{\frac{\rho}{f}} \quad (2.4)$$

where δ is the skin depth (meters), ρ is the resistivity (ohm-m), f is the frequency (cycles/s) (Telford et al., 1990). While the exploration depth will vary depending on the frequency and conductivity of the subsurface, the maximum achievable penetration for low frequencies is around 500 m (Kearey et al. 2002).

The magnetic field strength is generally measured using magnetometer attached to the “bird” with the EM transmitter coils. The cesium magnetometer, which was employed in this study, measures the magnetic field strength by quantifying its influence on cesium – 133 atoms contained within the sensor. The instruments are extremely sensitive to alterations in magnetic field with changes as small as 0.001 nT being registered.

While AEM surveys provide an excellent tool in geological and hydrologic studies, there are significant limitations and potential sources of error accrued during both the collection and interpretation of the data. AEM surveys measure the local deviations of apparent resistivity and magnetic field strength from what is assumed to be a constant background fields (Kearey et al., 2002). In reality these fields are not constant and can be influenced by factors such as temperature and solar wind. While background measurements are taken at multiple times throughout the survey, the changes can be non-linear varying over time and difficult to account for. The changes in background fields have been shown to be the most significant source of poor data quality (Valleau, 2000). Other potential sources of error accrued during the survey include, non-geologic noise from electronics, vibration, and the helicopter or plane which can influence measurement of local EM waves.

In addition to the potential error incorporated through the AEM survey the interpretation of the data can be ambiguous. Because the AEM data is being used to deduce the physical properties of unknown geologic media, determining what represents an accurate measurement and what might be attributed to measurement errors can be difficult. AEM measurements do not respond uniquely to certain geologic media and physical characteristics and the interpretation of the data requires general knowledge of the subsurface hydrology and geology is necessary to interpret the data. Even with an extensive knowledge of the survey area various measurements are likely to be interpreted and attributed incorrectly to a certain physical characteristic of the medium. The AEM

data is used in this study with an understanding of the limitations and error likely introduced through its use.

2.3 ESTIMATION OF HYDRAULIC CONDUCTIVITY

Due to the lack of available hydraulic data within the study area alternate methods of estimating the subsurface hydraulic parameters were needed. The issue of insufficient hydraulic data is not unique to this study due to the costs associated with the collection of supplemental data. Due to the need for alternative methods in subsurface hydraulic characterization, functions have been developed that estimate relationships between soil hydraulic parameters and more easily acquired soil data, such as the porosity, bulk density, grain size distribution, organic carbon, along with multiple other physical properties depending on the method (Børgesen et al., 2008). Bouma (1989) introduced the term pedo-transfer function (PTF), which he described as the translation of available soil data into soil data needed. By employing the long established relationship between the grain size distribution of a porous medium and hydraulic conductivity (Freeze and Cherry, 1979), PTFs can be employed to augment limited hydraulic data sets using grain size distribution information collected during geotechnical studies such as the case in this research.

Due to the large degree of difficulty in including all the possible variables in a porous medium, PTF's are not universally applicable with their accuracy varying greatly depending on the geologic setting. Studies evaluating the applications of different PTF's to the same porous geologic medium have shown that their estimates can differ by factors of ten or twenty (Vokovic and Soro, 1992). The incorporation of PTFs in any study,

needs to come with an understanding of the limitations. The fact that no PTF can account for all of the physical variations in a porous medium influencing fluid flow is shown by the continual development of new methods and alterations of existing methods. The high degree of uncertainty expected in the estimated hydraulic conductivity (K^*) values of one-half to one order of magnitude (Schapp and Leij, 1998) demonstrates that they are not a adequate replacement for direct hydraulic conductivity measurements and should only be used as an additional tool or where limited data is available, such as the case in this research. PTF's are employed in this study with a full understanding of the limitations and error likely introduced through their use.

To minimize the error incurred through the use of PTF's, three relevant and widely used methods were chosen based on a survey of available literature. The use of the three different methods allows for the comparison of the estimated values and the selection of the most accurate method for the given soil. The first PTF selected was developed by Hazen (1911) for predicting the hydraulic conductivity (or permeability) of saturated sands:

$$K^* = C_H D_{10}^2 \quad (2.5)$$

where K^* is the estimated saturated hydraulic conductivity (cm/s), C_H is Hazen's empirical coefficient which ranges between 1 to 1000 but is often assumed to be 100, and D_{10} is the grain size that is 10% finer by weight (effective grain size) (cm). While this equation was developed for clean sands with a coefficient of uniformity (D_{60}/D_{10}) of less than two, it has been widely employed to a range of in-situ soils due to its simplicity, with varying degrees of success (Carrier, 2003).

The second PTF selected was developed by Kozeny (1927) and Carman (1938):

$$K^* = (\gamma/\mu) \left(1/C_{K-C}\right) \left(1/S_0^2\right) \left[e^3/(1+e)\right] \quad (2.6)$$

where γ is the unit weight of the fluid (g/cm^3), μ is the dynamic viscosity of the fluid (cp), C_{K-C} is the Kozeny-Carman empirical coefficient usually assumed equal to 5, S_0 is the specific surface area of the porous media per unit volume, and e is the void ratio. The Kozeny-Carmen PTF is one of the most widely accepted methods to estimate hydraulic conductivity using the physical characteristics of the soil medium (Odong, 2007). The fact that this function requires a greater depth of knowledge concerning the physical properties of both the fluid and soil has limited its application even though simple methods of estimation for these parameters have been shown to be quite accurate (Carrier, 2003).

The third PTF selected was developed by Cosby et al., (1984):

$$K^* = 2.54 * 10^{(-0.6+0.012(SA)-0.0064(CL))} \quad (2.7)$$

where SA is the sand content percentage, and CL is the clay content percentage. This method was developed using multi-linear regression with 1448 U.S. soil samples to estimate the model parameters of the Campbell equation (1974). Recent studies have shown that the hydraulic conductivity estimated by the PTF compares well with measured data (Budiman and McBratney, 2000; Mermoud and Xu, 2006). The three PTF's will be referred to hereinafter as Hazen, KC, and Cosby, respectively.

3.0 DATA AND METHODS

There is a need to quantify hydro-geophysical relationship of the Alaska Highway corridor between Delta Junction and Tok, Alaska and provide an alternate method mapping the spatial variability of hydraulic conductivity. To address this problem relevant data from the study area was compiled and alternate methods of analysis were investigated. This chapter provides a brief description of the types of data used as well as the pattern learning methods employed to characterize the hydro-geophysical relationship.

3.1 DATA

All data used in the research has been compiled from reports and surveys by state and federal government agencies as well as private contractors. The data employed in this study can be grouped into three types; geophysical, geologic, and hydraulic. The following details the source and content of the three types of data.

3.1.1 Geophysical Data

The electromagnetic and magnetic (geophysical) data employed in this study was collected by a helicopter based frequency domain electromagnetic survey. The survey was performed using a Resolve EM system by Fugro Airborne surveys for the Alaska Department of Geologic and Geophysical Surveys (ADGGS). The geophysical survey was conducted on a sixteen mile swath along the Alaska Highway corridor from Delta Junction, Alaska to the Canadian border that covered roughly 3,045 square miles (Burns et al., 2006). The survey measured the magnetic field and the inphase and quadrature

components of six frequencies; 391, 1800, 3245, 8184, 39500, and 132760 hertz (Hz).

The flight lines were flown with a one-quarter mile spacing and the magnetic and electromagnetic data collected from each frequency was interpolated onto an eighty meter grid using a modified Akima technique (Akima, 1970).

3.1.2 Geologic Data

The geologic data employed in this study consists of 546 borehole samples collected between 1964 and 1993 by the Alaska Department of Transportation (ADOT) (Livingston, 1964, 1969; Slater, 1976; Brazo, 1980; Grahek, 1981, 1983, 1984; Brazo, 1987, 1993; Butler, 1993) at depths ranging from zero to twenty three feet. For each sample the grain size distribution analysis was performed while the natural moisture content was only recorded for 180 of the samples.

3.1.3 Hydraulic Data

The hydraulic data consists of 211 soil permeability measurements collected during a United States Department of Agriculture (USDA) soil survey (Swanson, 2009). The measurements are not spatially referenced and were collected based on soil type. The samples are concentrated at the northern extent of the study area but encompass a wide range of soil types found throughout the study area.

3.2 METHODS

Due to the complex nature of the hydro-geophysical relationship, the characterization of hydraulic conductivity using both geophysical and geologic input parameters can be difficult to achieve mathematically and numerically. Recent studies have shown machine learning algorithms, which are adaptable pattern learning methods of analysis, to be a powerful and versatile tool, especially useful in pattern recognition, signal processing, and function approximation (e.g. Rizzo and Dougherty, 1994; Tamari et al., 1996). Their strength comes in their ability to approximate highly complex non-linear relationships, such as the case in this research.

In this study the hydro-geophysical relationship was analyzed using three types of machine learning algorithms. The machine learning algorithms employed are artificial neural network regression (ANNR), support vector regression (SVR), and artificial neural network classification (ANNC). Before employing the machine learning algorithms the input parameters were first individually analyzed to characterize their statistical, distributional, and spatial properties. The understanding of the statistical, distributional, and spatial characteristics of each of the hydraulic, geophysical, and geologic parameters is integral before their relationships can be used in a predictive manner. We structure the rest of the methodology into three subsections. These subsections provide a brief background of the machine learning algorithms and their associated parameters employed in this study.

3.2.1 Artificial Neural Networks

Recent studies have shown, Artificial Neural Networks (ANN), which is an adaptable pattern learning method of analysis, to be a powerful and versatile tool especially useful in pattern recognition, signal processing, and function approximation (e.g., Rizzo and Dougherty, 1994; Tamari et al., 1996; Ghanbarian-Alavijeh et al., 2010). Hence, ANN was used to provide a regression and classification of the data to simulate the hydraulic conductivity using input parameters of apparent resistivity, spatial coordinates, soil indicator, and natural moisture content.

The architecture of an ANN is based on that of biological neurons such that it is composed of a network of interconnected neurons each with multiple inputs and outputs. The neurons are generally structured into three layers (**Figure 3.1**): The input layer, which provides the input parameters that are believed to have a relationship with the target output to each node of the first hidden layer. The hidden layer/s, which multiply the inputs by the associated weights, sums the weighted inputs and biases, and then operates on the sum using an activation function, the output layer. The output layer, which applies a second set of weights, sums the weighted inputs and associated bias, and operates on the sum with an activation function, with the result being the output of the model. Activation functions are used within the network to account for the non-linearity and multidimensionality of complex data sets (Sarle, 2002) providing the network the ability to model nonlinear functions. The weights and bias of each neuron are generally optimized using a subset of the total data set with a training algorithm that iteratively

adjusts the weights and bias of each neuron using a subset order to minimize the error between the predicted output and target output.

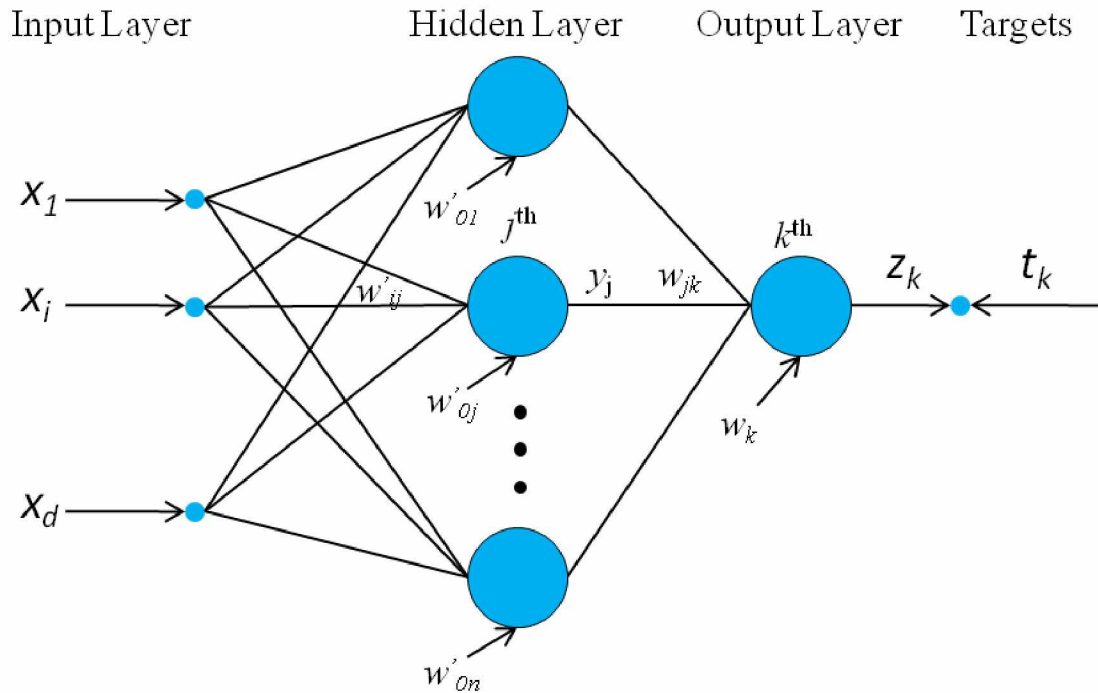


Figure 3.1: Feed forward multilayered artificial neural network.

The ANN analysis in this study was performed using the MATLAB[®] Neural Network Tool box which employs the structure of a two layer feed forward network. To account for the non-linearity of the data, sigmoid activation were employed due to their ability to fit multi dimensional mapping problems and have been shown to be most applicable for continuous value targets with a bounded range (Jordan, 1995) such as the case with this research. The network was trained using the Levenberg-Marquardt backpropagation training algorithm. Recent studies have shown that the Levenberg-

Marquardt backpropagation training algorithm which is a variation of Newton's method to be a computationally efficient, robust and accurate algorithm due to the absence of second derivatives and a learning rate that changes based on rate of change in the error function (Hagan et al., 2002).

The use of ANNC offers an alternative to regression techniques through the recognition of classifications rather than numerical values. In this study we employed the use of classification to represent ranges of hydraulic conductivity values. The hydraulic conductivity values were grouped into three clusters (high, medium, and low) using the k -means clustering algorithm (Serber, 1984). The algorithm employs the square Euclidean distance between the points to minimize the inter cluster difference in the specified k number of clusters. The ANNC analysis was performed using Neuroshell Classifier (Ward Systems, Frederick, MD, USA 2008), a neural network simulator, which employs a proprietary training algorithm (TurboProp2, 2003) to simultaneously minimize the number of misclassifications and optimize the number of neurons.

3.2.2 Support Vector Machines

Support Vector Machines (SVM) is a relatively new development in the field of machine learning algorithms. Studies have shown SVM to have significant advantages over ANN such as the absence of local minima, simple geometric interpretation, and sparseness of solution (Frag and Mohamed, 2004). These advantages are in part due to the utilization of the structural risk minimization principle, which has been shown to be superior to the empirical risk minimization principle employed by ANN (Gunn, 1998). The structural risk minimization principle helps address the issue of over fitting a model by balancing the complexity of the model with its ability to fit the training data.

SVM was first developed for classification problems and has been successfully employed in hydrological engineering investigations in several instances (Asefa et al., 2004; Twarakavi et al., 2006; Wohlberg et al., 2006). SVM was extended to regression (SVR) problems when Vapnik (1995) introduced the ϵ -insensitive loss function. The ϵ -insensitive loss function sets limits to the amount the target values that can deviate from the modeled regression line, where deviations larger than ϵ are not accepted.

The regression problem can be stated as:

Given a training data set $D = \{(x_i, t_i) | i = 1, 2, \dots, n\}$ of input vectors x_i , and target outputs t_i , the goal of SVR is to determine a function $f(x)$ that minimizes the error between the predicted output and target output (y_i), using the least complex model possible (degree of flatness). The key difference between SVR and ANN is the inclusion of the ϵ -insensitive loss function which helps reduce the risk of overtraining. This is done by only training the model using samples that lie outside specified margin of error.

These instances are known as support vectors. The support vectors define the model's hyper-plane (regression line) as shown in **Figure 3.2**. All data points that fall within the error margin ϵ , do not contribute to the optimization of the regression line. The number of support vectors and distance from the regression line correlates to the error and uncertainty within in the model. In this study we identify areas of high uncertainty and error by splitting each input parameter into ten equally distributed intervals and analyze the quantity of support vectors in each range. This allows for the identification of gaps in the data set as well as what ranges of variables represent the strongest relationship.

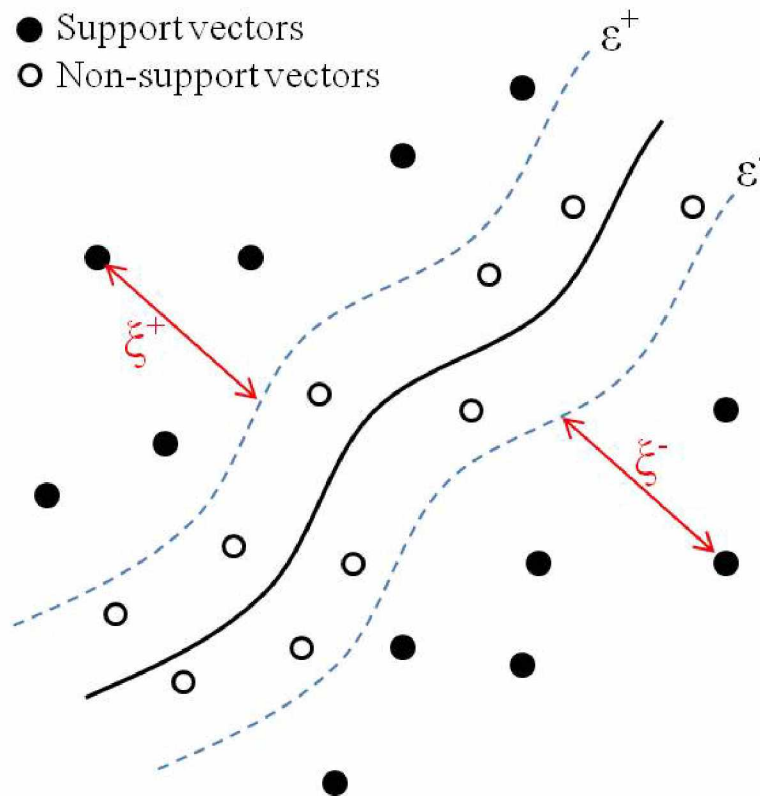


Figure 3.2: Nonlinear SVR using Vapnik's ϵ -insensitive loss function.

In complex and non-linear problems it may not be feasible to develop a function $f(x)$ where all errors are $< \varepsilon$. To accommodate such problems, the “soft margin” loss function (Bennett and Mangasarian, 1992) was incorporated into SVR. The “soft margin” loss function incorporates slack variables ξ_i^+ , ξ_i^- leading to the optimization problem as stated in Vapnik, (1995).

$$\begin{aligned} \text{minimize} \quad & \frac{1}{2} \|w\|^2 + C \sum_{i=1}^1 (\xi_i^- + \xi_i^+) & (3.1) \\ \text{subject to} \quad & \begin{cases} (\langle w, x_i \rangle + b) - t_i \leq \varepsilon + \xi_i^- \\ t_i - (\langle w, x_i \rangle + b) - t_i \leq \varepsilon + \xi_i^+ \\ \xi_i^-, \xi_i^+ \geq 0 \end{cases} \end{aligned}$$

Where w is the weight vector associated with the input parameters, b is the bias determined by its location relative to the origin of the input space, ξ_i^+ , ξ_i^- are the slack variables accounting for error outside of the margin ε , and C is a constant used to control the magnitude of the penalty that is associated with error that fall outside of the margin. By controlling the magnitude of penalties associated with the errors greater than ε , the constant $C > 0$ determines the tradeoff between the flatness of the function $f(x)$ and the amount to which error deviations are tolerated (Frag and Mohamed, 2004; Twarakavi et al., 2005).

In complex problems, issues of dimensionality and nonlinearity may arise and need to be overcome to characterize the relationship between the input space (predictor variables) and target space (output). In order to accomplish this characterization, the input space is mapped into a higher dimensional feature space using a nonlinear mapping

function called a kernel function. In this study, we employ the Gaussian radial basis kernel function (3.2) due to its wide degree of use.

$$K(x, x') = \exp\left(\frac{\|x - x'\|^2}{2\gamma^2}\right) \quad (3.2)$$

With the inclusion of the Gaussian radial basis kernel, three SVR model parameters need to be optimized during training: the Gaussian radial basis function parameter γ , magnitude of error penalty constant C , and the error margin ε . For this study, we implemented a grid search algorithm within the training of the model that optimized the parameters based on coefficient of determination between the target and predicted output. The SVR algorithm employed for this research is available in the *e1071* package (Dimitriadou et al., 2007) with code adapted from Oommen and Baise (in-press) was employed for the analysis.

4.0 EXPLORATORY DATA ANALYSIS

A statistical, spatial, and correlational analysis was done to derive inferences on the distribution and spatial characteristics of the samples containing apparent resistivity, magnetic field strength, grain size distribution, and natural moisture content data within the study area. Additionally the grain size distributions obtained from the ADOT borehole samples provided the necessary information to estimate supplemental hydraulic conductivity data to augment the insufficient hydraulic data set. The three major analysis techniques used were the histogram analysis, semi-variogram analysis, and correlation analysis. The histogram analysis was carried out to understand the statistical distribution of the data, while the semi-variogram analysis was performed to provide valuable information on the spatial continuity and roughness of a data set that descriptive statistics and histograms fail to quantify. The correlation analysis was performed to measure the strength and direction of a linear relationship between parameters.

4.1 GEOPHYSICAL DATA ANALYSIS

4.1.1 Electromagnetic Data Analysis

The EM data collected during the ADGGS survey was analyzed to determine the statistical and spatial characteristics of the data along with how it corresponds to the surficial geologic units mapped by Reger et al. (2008) and known surface and subsurface water corridors such as the Tanana River and its tributaries. The units of measurement for the EM data are ohm-meters. Due to the relatively shallow depths of the borehole samples (zero to twenty three feet) only the high range frequencies of 40,000 (40K) and

140,000 (140K) Hz collected during the geophysical survey were used in this study. The resulting apparent resistivity measured at these frequencies has a higher sensitivity and pertain to the geologic media near the surface as previously discussed (Section 2.2.2).

From the summary statistics of the apparent resistivity at 140K and 40K Hz given in **Table 4.1** it is observed that the means are relatively low with values of 1294 and 985 Ω -m, respectively in comparison to maximums with values 4622 and 2770 Ω -m, respectively. The low means coupled with the positively skewed distributions shown in their respective histograms (**Figure 4.1**) indicate that the electromagnetic waves are responding to the less resistive surface water and moisture rich soil near the surface. While one would expect that resistivity surveyed at 140K Hz would be lower than that of 40K Hz due to its sensitivity to surface water, we can see from the histograms and variance that mean was skewed by the outlying data points. A more representative value for the apparent resistivity surveyed at 140K Hz would be the median with a value of 921.78 Ω -m. This value aligns more closely with the distribution shown in the histogram (**Figure 4.1a**).

Table 4.1: Summary statistics of the apparent resistivity data.

	ρ^* at 140K Hz	ρ^* at 40K Hz
COUNT	556	556
MEAN	1294.33 Ω -m	986.59 Ω -m
MEDIAN	921.78 Ω -m	930.81 Ω -m
MINIMUM	73.81 Ω -m	125.11 Ω -m
MAXIMUM	4622.56 Ω -m	2770.08 Ω -m
VARIANCE	808988.17 (Ω -m) ²	233204.78 (Ω -m) ²
STANDARD DEVIATION	898.61 Ω -m	482.47 Ω -m
SKEWNESS	1.21	0.64
KURTOSIS	1.07	0.22

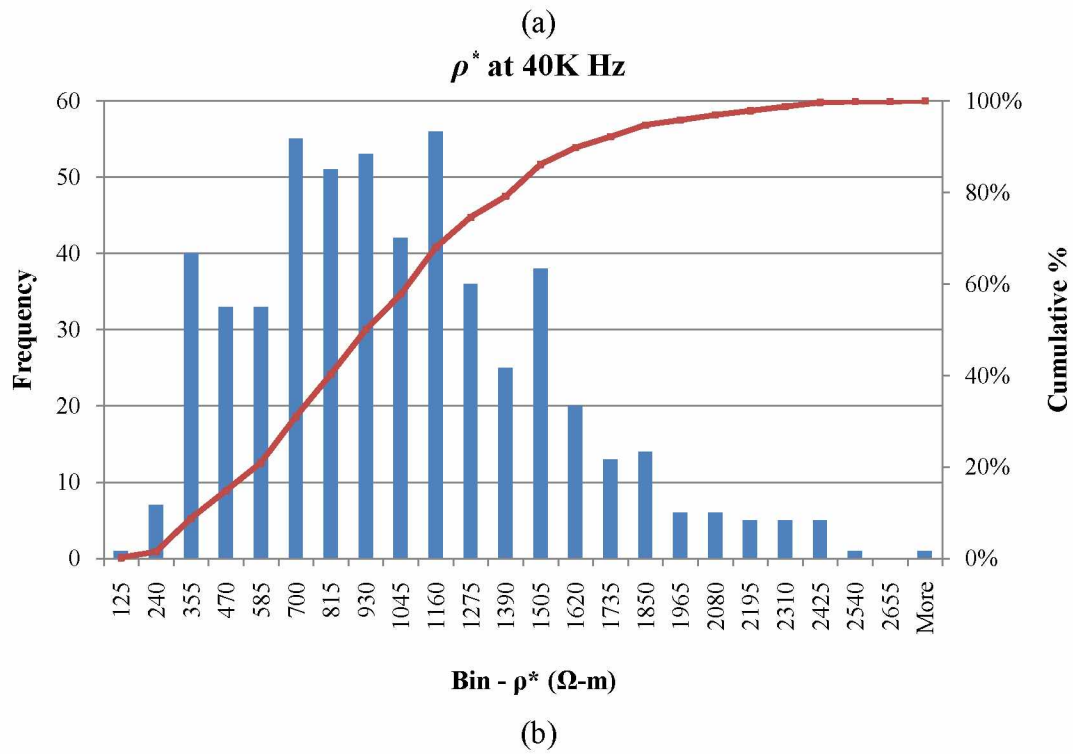
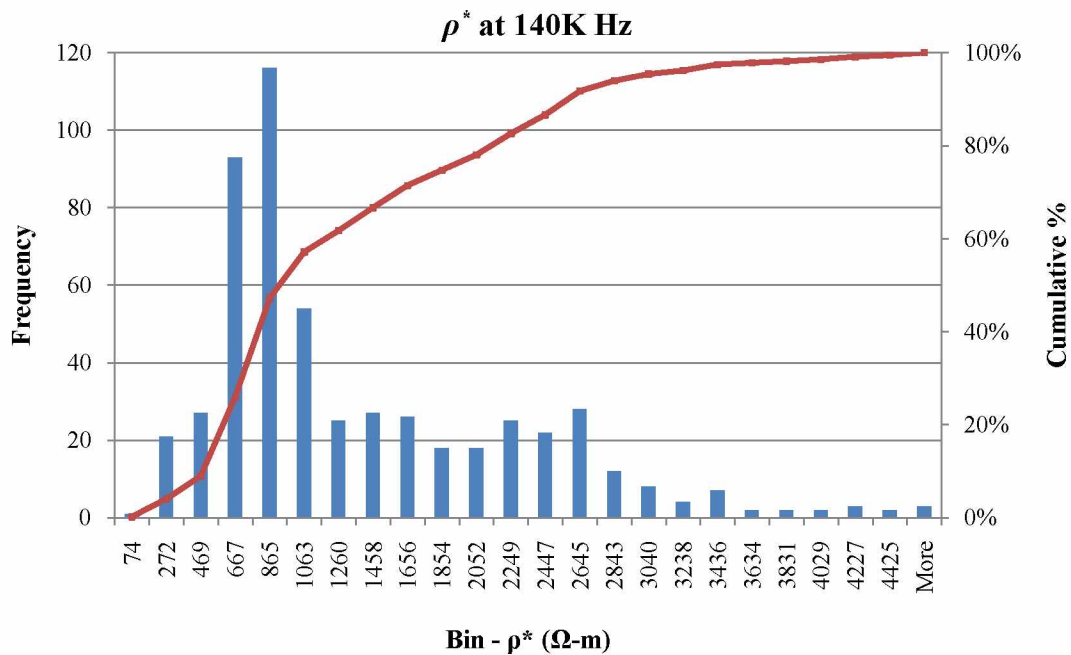


Figure 4.1: Histogram of apparent resistivity at 140K and 40K Hz.



From the contour plots of the resistivity at 140K and 40K Hz shown in **Figure 4.2**, it is observed that the areas of lowest resistivity correspond to the Tanana River, tributaries, and low lying areas . This strengthens our hypothesis that the EM fields are responding to the surface and subsurface volumetric water content. It is also observed that the areas of highest resistivity are concentrated on the southern edge of the area likely corresponding to the granitic outcroppings and the piedmont fans of the foot hills of the Alaska Range described by Reger et al. (2008).

From the contour map of the difference between the apparent resistivity measured at 140K and 40 K Hz shown in **Figure 4.3** it is observed that the majority of values lie within $\pm 500 \Omega\text{-m}$ of each other. The overall similarity correlates to the observations made from the summary statistics (**Table 4.1**) and individual contour plots (**Figure 4.2**). It can also be observed that the areas with a positive difference between approximately 1000 and 4000 $\Omega\text{-m}$ are at the base of many of the streams and rivers initiating from the Alaska Mountain Range. This indicates that lower frequency (40K Hz) waves penetrating to a deeper depth are likely encountering the subsurface groundwater flowing through the alluvial outwash whereas the higher frequency waves are not penetrating deep enough to interact with the less resistive material. The largest dissimilarities between the two frequencies lie on the southern most extent of the study area. This indicates that the lower frequency (40K Hz) responds to an increased amount of bed rock with depth. Overall it is observed that for the majority of the area, the two frequencies are responding to roughly the same subsurface material.

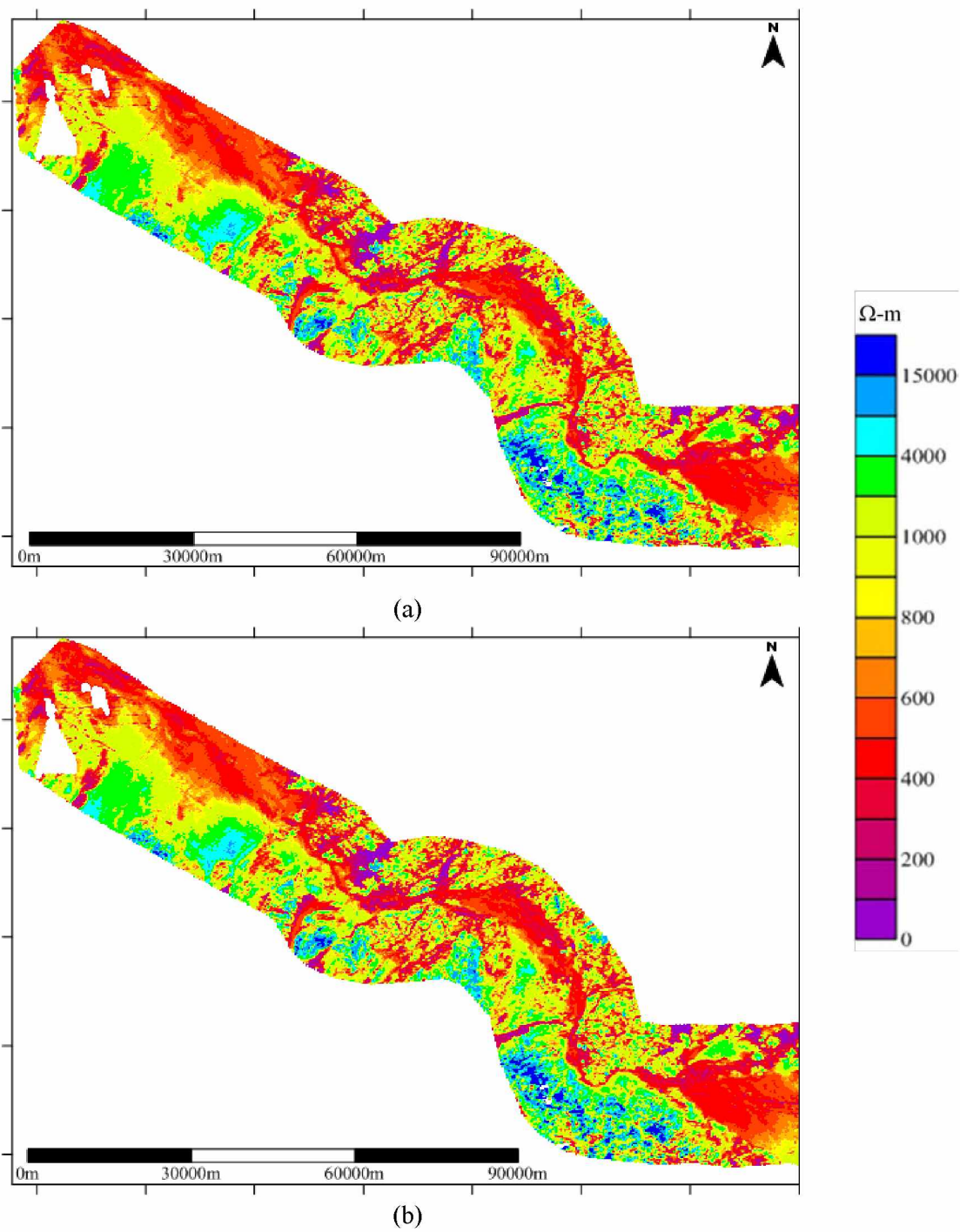


Figure 4.2: Contour plots of apparent resistivity

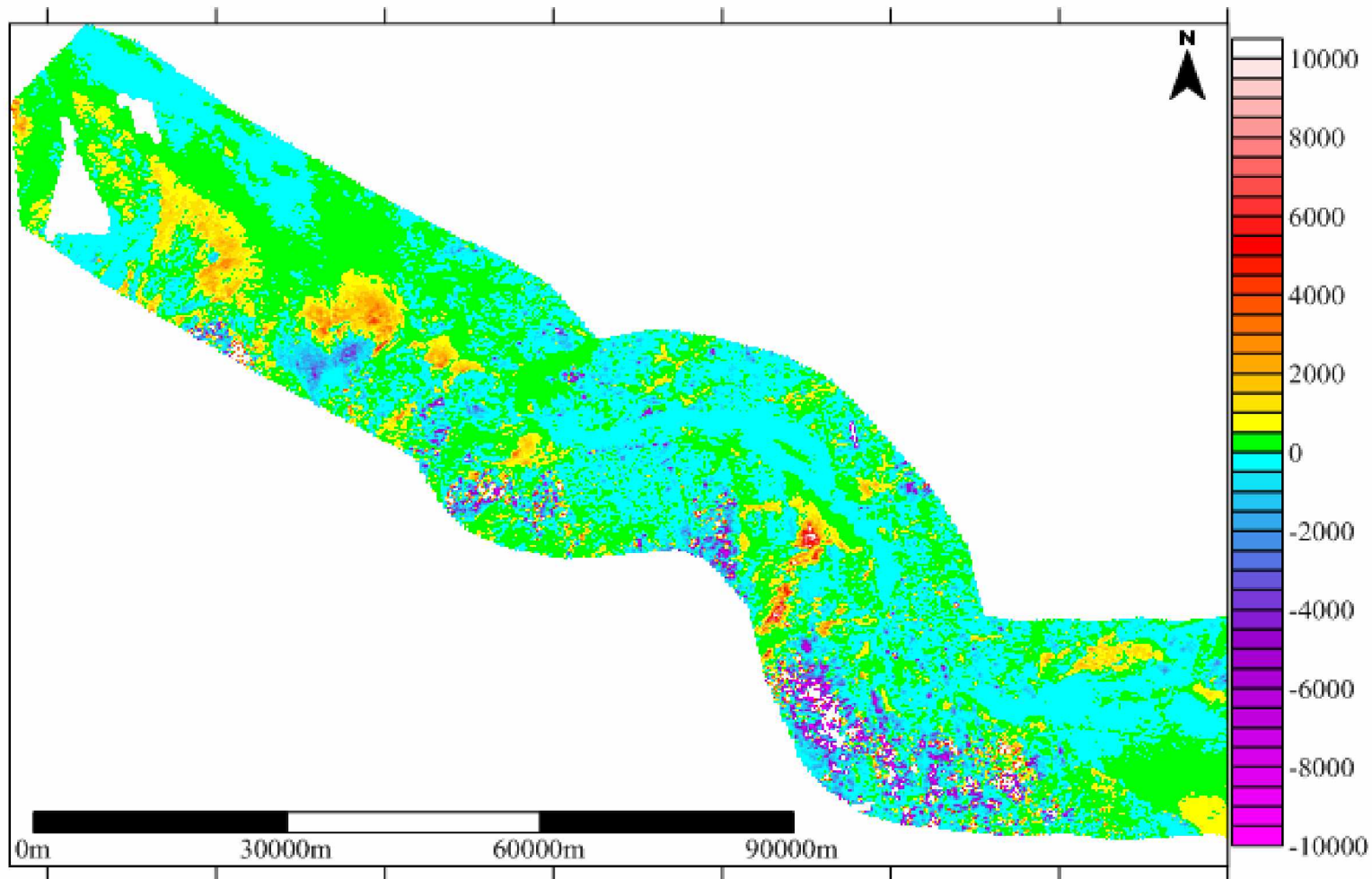


Figure 4.3: Contour plot of difference between ρ^* measured at 140K and 40K Hz.

The apparent resistivity at 140K and 40K Hz were tested for spatial variance using semi-variogram analysis. Semi-variogram analysis provides valuable information on the spatial continuity and roughness of a data set that descriptive statistics and histograms fail to quantify. **Figure 4.4** shows the experimental semi-variogram of apparent resistivity at 140K Hz evaluated perpendicular to the flight lines with a forty five degree tolerance. The estimated sill shown at approximately 800,000 (Ω -meter)² corresponds to the calculated variance in the summary statistics (**Table 4.1**). From the semi-variogram it is observed that the data points lose their spatial continuity at a lag distance of 6,500 meters. This roughly corresponds to the large surficial geologic units mapped by Reger et al. (2008). The experimental semi-variogram of apparent resistivity at 40K Hz shown in **Figure 4.5** roughly correlates to the variance shown in the summary statistics (**Table 4.1**) with a sill of 235,000 (Ω -meter)². The range of data correlation is shown to be at roughly 5,500 meters indicating that the geologic media increases in heterogeneity with an increase in depth.

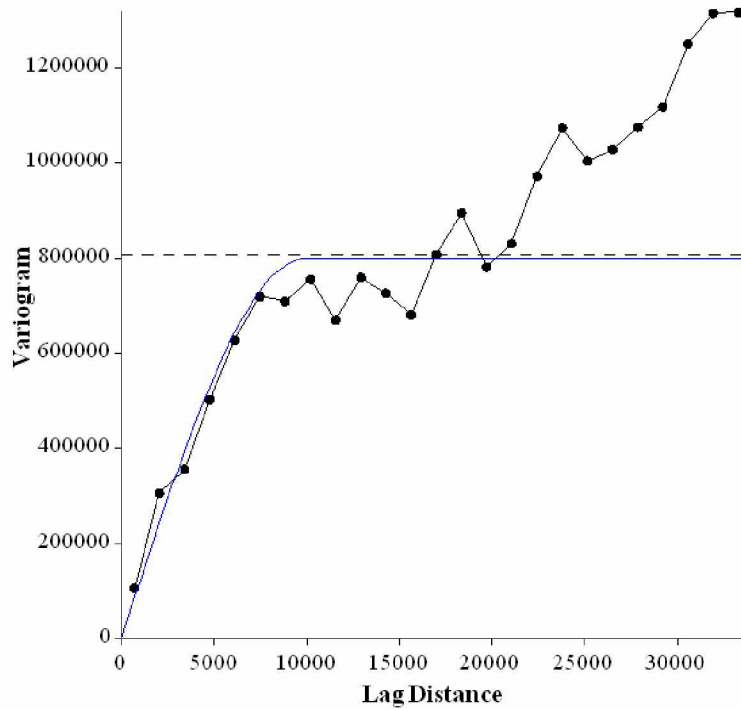


Figure 4.4: Semi-variogram ρ^* at 140K Hz.

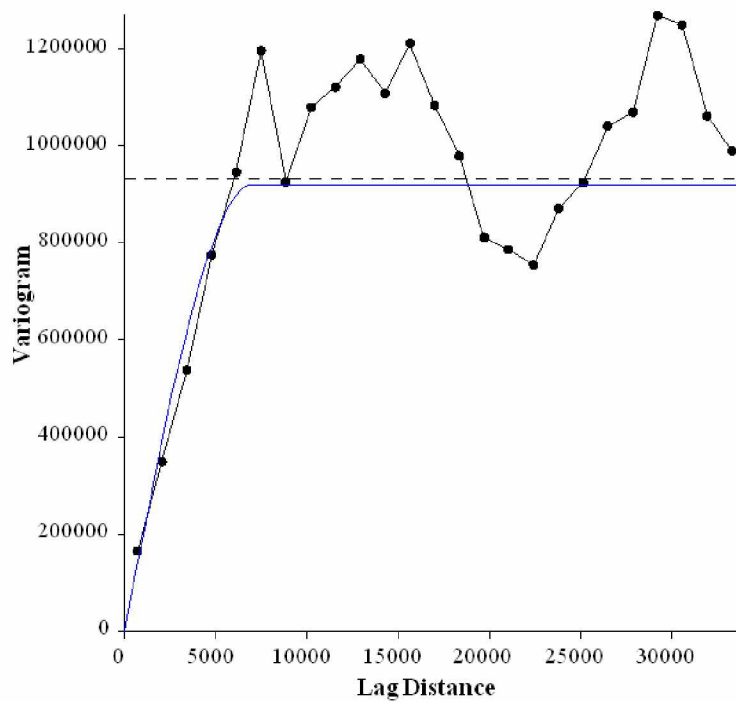


Figure 4.5: Semi-variogram of ρ^* at 40K Hz.

4.1.2 Magnetic Data Analysis

In addition to the apparent resistivity data, magnetic data was collected through the AEM survey. While the earth's magnetic field is not influenced by moisture content and geologic pore structure in the shallow surface, it can be useful in the identification of hydrologic basin geometry, igneous intrusions, faults, paleochannels, and eolian deposits (Paine and Minty, 2005). The magnetic data was analyzed to understand the characteristics of the data set and provide a supplemental input through its relationship to the subsurface geologic composition.

From the summary statistics of the magnetic data given in **Table 4.2**, and histogram in **Figure 4.6**, it is observed that the data is concentrated around the mean of 57093 nT with a range of 794 nT. Previous studies interpreting magnetic data have used magnetic values exceeding ± 500 nT from the mean as symbolizing anomalous values (Oommen, 2006). The narrow range of the data set (56803 to 57597 nT) combined with the low number of anomalous values shown in contour map (**Figure 4.7**) indicate that the magnetic susceptibility of the geologic material is relatively homogenous throughout the study area.

Table 4.2: Statistical summary of magnetic data.

COUNT	556	VARIANCE	16774.92
MEAN	57093.86 nT	STANDARD DEVIATION	129.40
MEDIAN	57046.85 nT	SKEWNESS	1.32
MINIMUM	56803.81 nT	KURTOSIS	2.15
MAXIMUM	57597.21 nT		

While the summary statistics (**Table 4.2**) and histogram (**Figure 4.6**) indicate a homogenous smooth data set, the experimental semi-variogram of the magnetic data evaluated perpendicular to the flight lines with a forty five degree tolerance shown in **Figure 4.8** indicates a low degree of spatial continuity accompanied by an estimated lag distance of less than 4,000 meters. This observation indicates that the magnetic field is likely responding to the changes in the subsurface geology. While the variations in magnetic susceptibility have not been shown to directly correlate to changes in the hydraulic structure of the geologic media they might provide an indicative parameter and therefore are used for further analysis.

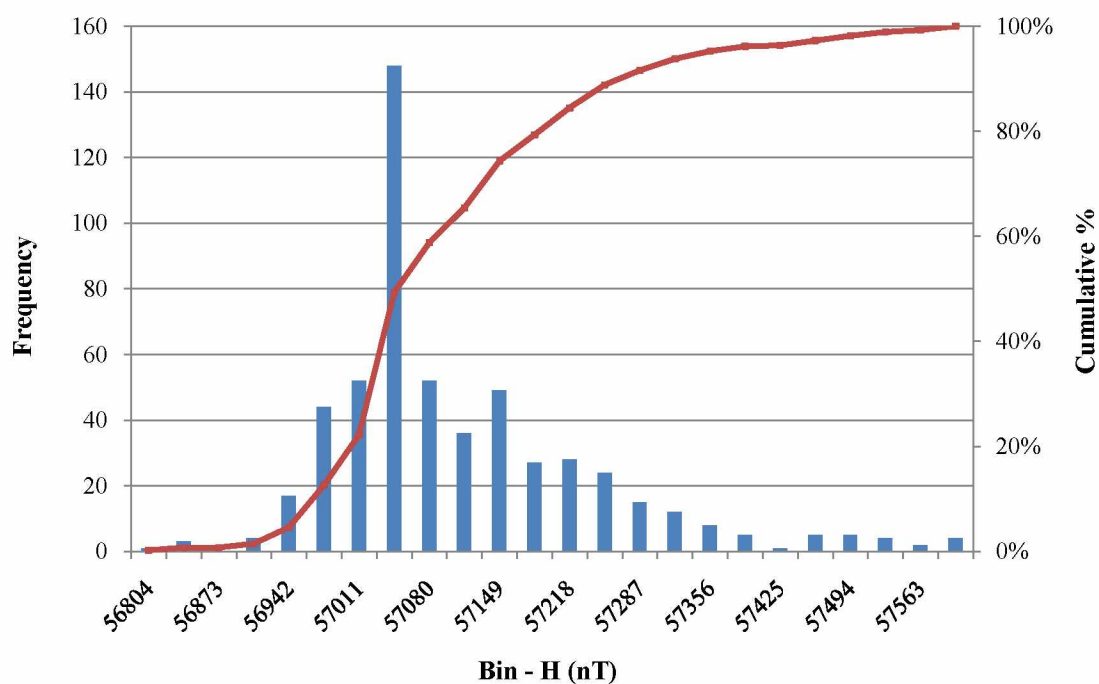


Figure 4.6: Histogram plot of magnetic data.

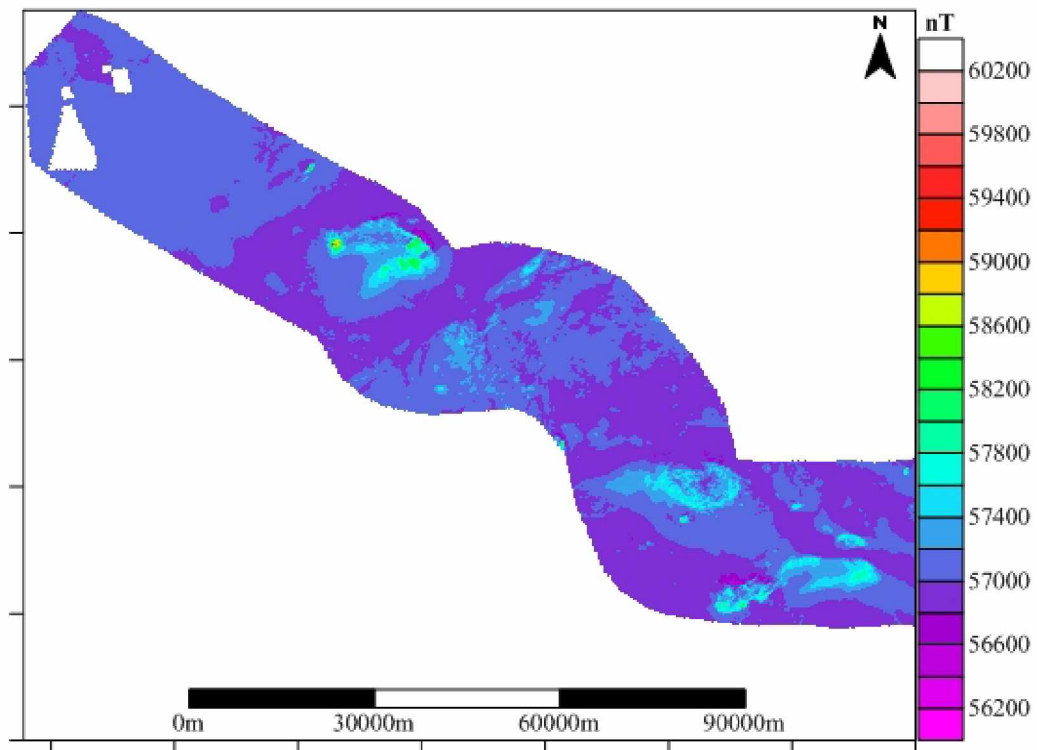


Figure 4.7: Contour plot of magnetic data.

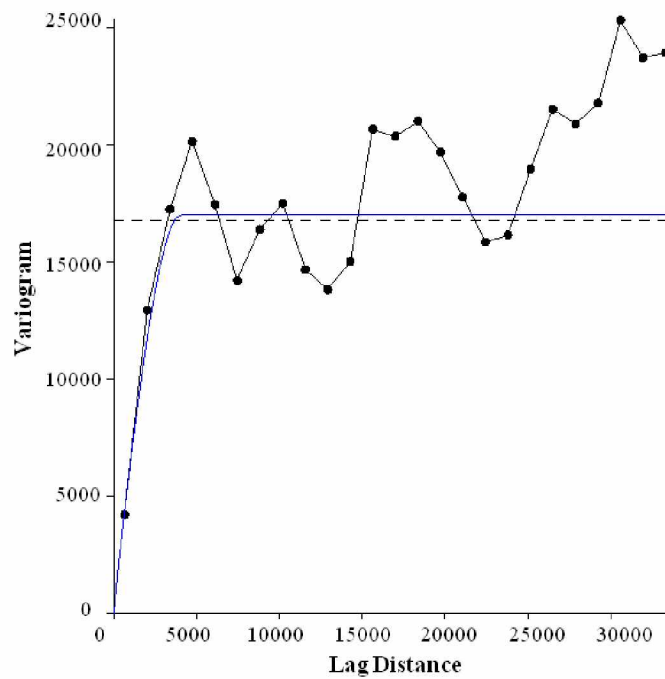


Figure 4.8: Semi-variogram magnetic data.

4.2 GEOLOGIC DATA ANALYSIS

4.2.1 Grain Size Distribution Analysis

The borehole grain size distributions were condensed into three major constituents; gravel (retained on a #4 sieve), sand (retained on #200 sieve), and fines (passing #200 sieve) based on their percentage of the total by weight. Along with the major soil constituents effective grain size (D_{eff}) was calculated for each grain size distribution using the equation developed by (Carrier, 2003):

$$D_{eff} = 100\% / \left[\sum (f_i / D_{ave i}) \right] \quad (4.1)$$

where f_i is the fraction of particle between two sieve sizes and D_{ave} is the average particle size. The effective grain size is representative of the overall distribution of gravel, sand, and fines as well as being a primary input for the estimation of hydraulic conductivity using the PTF developed by Carmen (1938) and Kozeny (1927). The distribution of the three constituents (gravel, sand, and fines) and the effective grain size are significant due to their physical implications controlling subsurface groundwater flow.

From the summary statistics shown in **Table 4.3** and histogram plots shown in **Figure 4.9**, it is observed that sand has the highest mean percentage of 42.89% with the distribution equally distributed around the mean, while gravel has the second highest mean with 34.89% and appears to have a negatively skewed bimodal distribution with high concentrations of samples falling around zero and fifty percent. The fine grained soil had the lowest mean with 22% and a negatively skewed distribution. The more

equally distributed percentages of sand and gravel compared to the fines indicate a porous media, which correlates to the previous hydrogeologic observations noted in Section 2.1.1.

Table 4.3: Statistical summary of borehole sample grain size distributions

	%Gravel	%Sand	%Fines	D_{eff}
COUNT	546	546	546	546
MEAN	34.87	42.89	22.18	1.80E-02
MEDIAN	40	42	12	1.36E-02
MINIMUM	0	0	1	1.89E-03
MAXIMUM	98	98	100	0.277
VARIANCE	523.40	252.55	514.36	3.30E-04
STANDARD DEVIATION	22.88	15.89	22.68	1.82E-02
SKEWNESS	-0.21	0.66	1.60	6.20
KURTOSIS	-1.09	1.38	1.70	76.64

From **Figure 4.9d** it is observed that the effective grain size distribution is negatively skewed with values concentrated at the low end of its range between 0.014 and 0.026 with a mean of 0.02 cm. This indicates that hydraulic conductivity of the samples will likely be negatively skewed and concentrated at the low end of the spectrum as well. The few outlying samples that extend the range to 0.28 cm are composed of samples with high percentages of gravel.

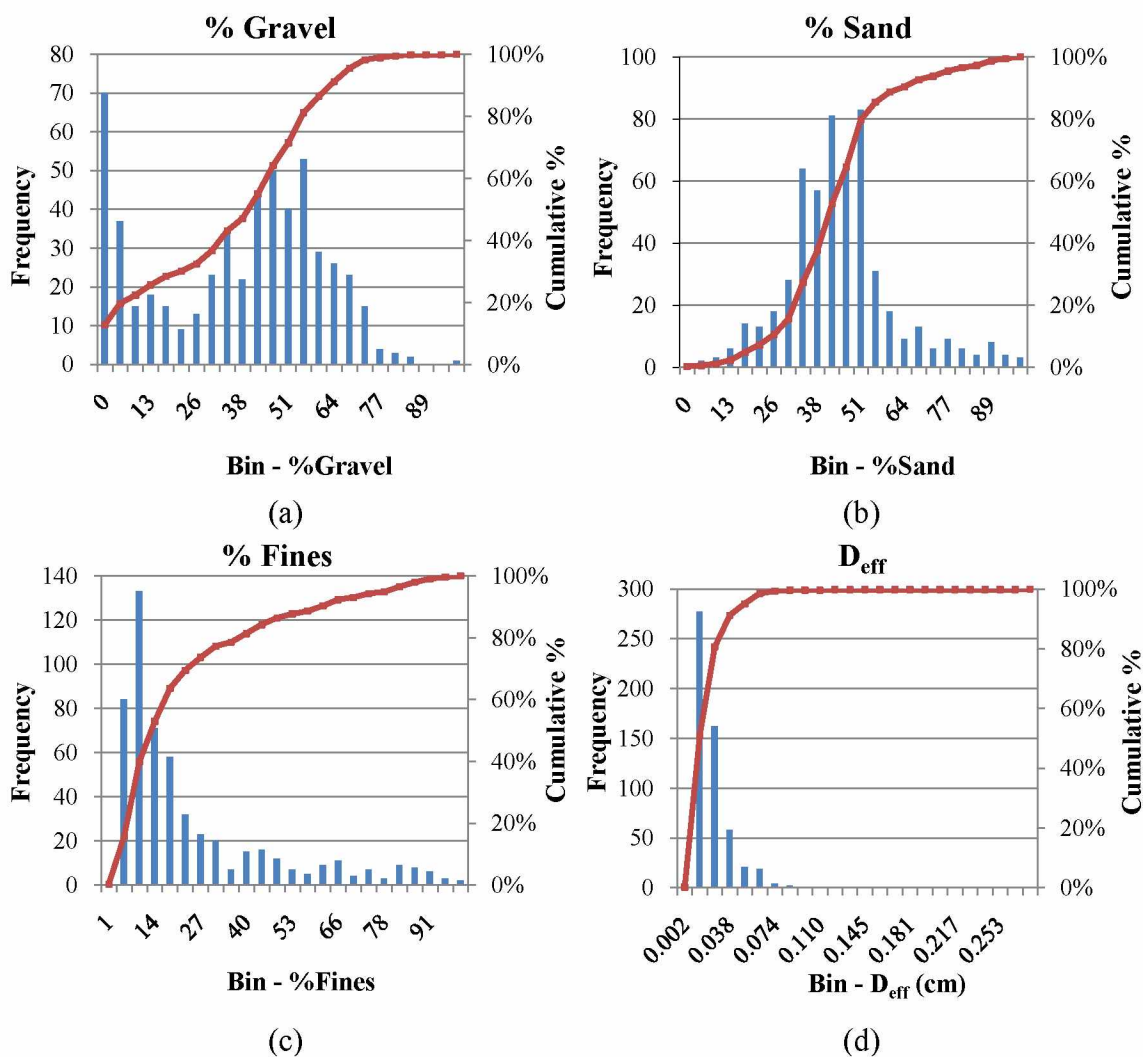


Figure 4.9: Histograms of primary grain size distribution parameters.

To further group the grain size distributions and provide a supplemental input, soil indicators (SI) were employed. Indicator variables have long been used in geostatistical modeling techniques as a means of characterizing heterogeneity in terms of hydrogeologic lithofacies (Fogg et al., 1998). To provide this classification input parameter, the soils were grouped into nine subsets. The subsets are based on the degree

of sorting and percentages of fines, using the unified soil classification system (Howard, 1988).

Due to the absence of information concerning organic contents and plasticity characteristics in the grain size analysis data, all soils with greater than 50% of the material passing a No. 200 (0.074 mm) sieve were grouped into the single classification of fine grained soils, while sand, gravel, clay, and silt mixtures were grouped based on their degree of sorting and percentage of fines. From **Figure 4.10** it is observed that the soil SI grouping encompassing the largest number of samples was the poorly graded gravelly-sand-clay-silt (GMC) classification with 180. GMC includes soils with more than 50% of the material by weight passing the No. 4 sieve (5 mm) and greater than 12% passing the No. 200 sieve. The remaining soil indicators all classified based on their percentage passing No. 4 and No. 200 sieve sizes range from four to ninety five samples.

From the semi-variogram of the soil indicator shown in **Figure 4.11** it is observed that there is no spatial continuity within the data set. The complete lack of spatial dependence is counterintuitive of both the common geologic deposition characteristics and observations made in geologic investigations in the area (Reger et al., 2008). The most probable cause for this lack of spatial dependence may be attributed to the irregular sampling pattern and depth of the borehole samples. If the number of borehole samples was increased and the samples were taken at smaller intervals, SI would likely demonstrate a spatial dependence correlating to the depositional characteristics of the area.

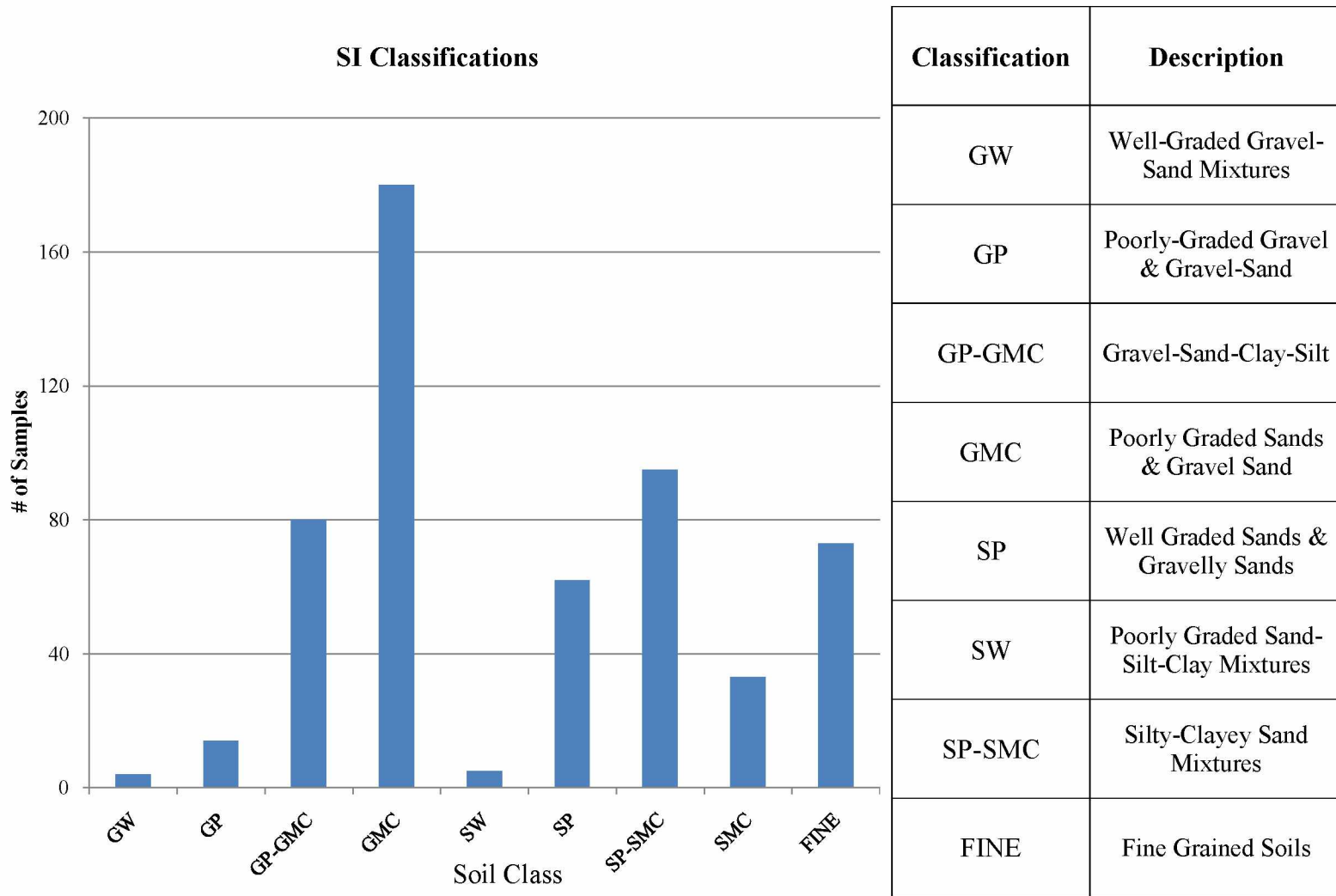


Figure 4.10: Soil classifications description and frequency

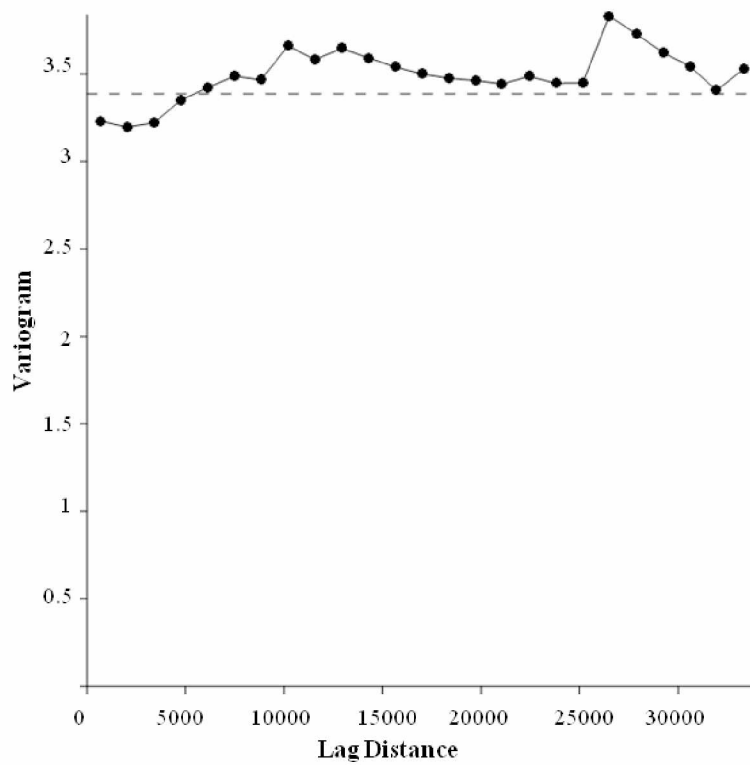


Figure 4.11: Semi-variogram of soil indicator

4.2.2 Natural Moisture Content Analysis

Of the 546 borehole samples, 180 included natural moisture (NM) content values. From the summary statistics shown in **Table 4.4** it is observed that the mean natural moisture content is roughly 16% with the data ranging from one to over a hundred percent. The high values, specifically the values exceeding one hundred percent likely correlate to ice lenses and ice rich soils, while the low values likely correlate to well drained soils located high above the water table. We can see from the histogram plot shown in **Figure 4.12** that the data is negatively skewed with the highest concentration of values lying between six and twelve percent. The concentration of values at the low end of the range confirms deep water table and relatively well drained soils noted in previous hydrologic investigations in the area.

Table 4.4: Statistical summary of natural moisture content from borehole samples.

	%NM
COUNT	180
MEAN	16.40
MEDIAN	10.65
MINIMUM	1.3
MAXIMUM	122.5
VARIANCE	318.75
STANDARD DEVIATION	17.85
SKEWNESS	2.59
KURTOSIS	9.47

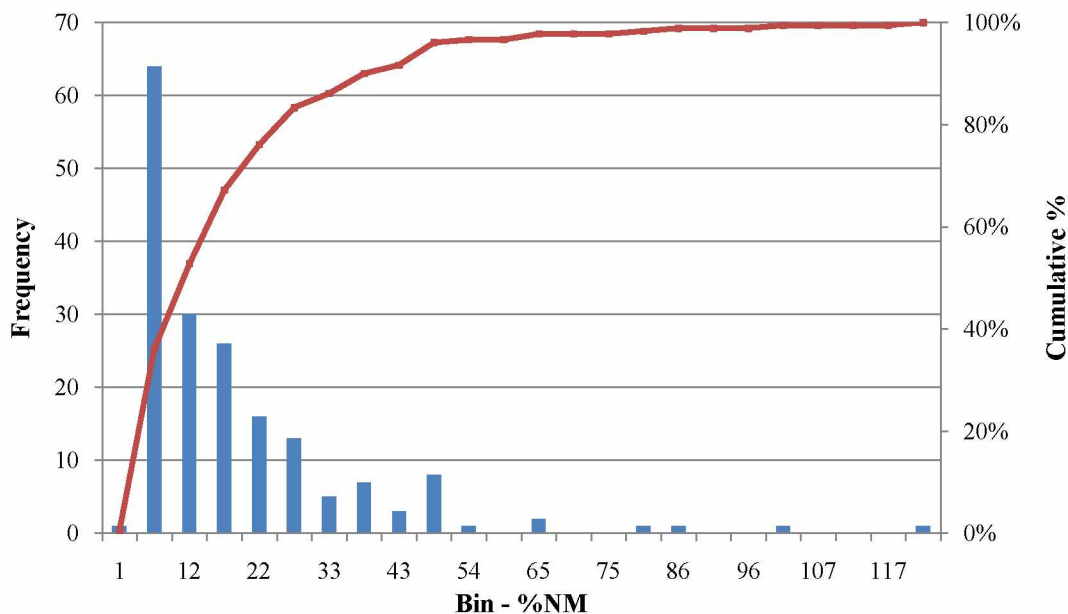


Figure 4.12: Histogram natural moisture content percentage.

The semi-variogram of NM shown in **Figure 4.13** indicates a complete lack of spatial dependence in NM with a large nugget. While NM likely has a high degree of natural spatial variability throughout the area, the most probable reason for the complete lack of any spatial dependence indicated in the semi-variogram is the inconsistency in sample spacing, depth, and temporal variability of the measurements. The NM measurements were collected between early spring and late fall over the course of forty years. Due to the large seasonal fluctuations in surface and subsurface runoff, the NM content is likely to vary widely throughout the year. Even with the time differences and reduced data set size, NM has been included in this analysis with the knowledge of its limitations due to its significance in the hydro-geophysical relationship.

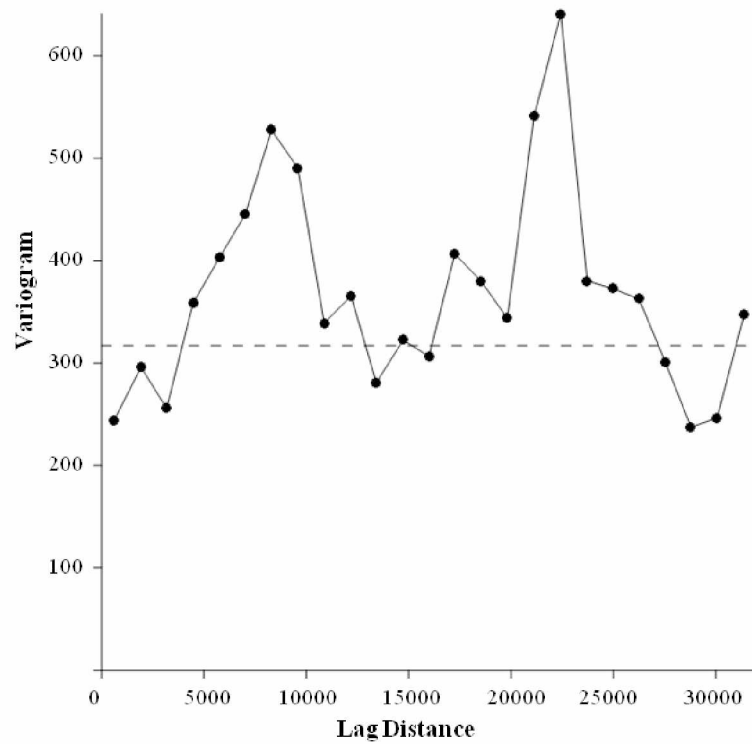


Figure 4.13: Semi-variogram of natural moisture content percentage.

4.3 HYDRAULIC DATA ANALYSIS

4.3.1 Estimation of Hydraulic Conductivity

By using the 546 grain size distributions obtained during the ADOT geotechnical investigations (Livingston, 1964, 1969; Slater, 1976; Brazo, 1980; Grahek, 1981, 1983, 1984; Brazo, 1987, 1993; Butler, 1993), the previously discussed PTF's (Section 2.3); Hazen (1911), Kozeny (1927), Carmen (1938), and Cosby et al. (1984) were employed to estimate the hydraulic conductivity using the available grain size distribution data. The accuracy of each PTF was judged by comparing their statistical distribution in comparison to the data set collected during the USDA soil survey (Swanson, 2009).

4.3.2 Comparison of Hydraulic Conductivity Estimates

From the basic statistics shown in **Table 4.5**, it is observed that the mean hydraulic conductivities estimated using Hazen's, KC, and Cosby's PTF are skewed towards the lower end of the range with means of .0225, .0098, and .00053 cm/s, respectively. While the mean of the estimated values using Cosby's PTF resembles the mean of the USDA values most closely with an absolute difference of 3.25E-04 cm/s, to accurately compare the values estimated using the PTFs to the USDA measured values the distributions must be normalized to test if they come from the same population. Past studies have shown that the log transformed data best represents the distribution of hydraulic conductivity values (Freeze, 1975). The histograms of the lognormally transformed data shown in **Figure 4.14** confirm that such transformation holds true for our data set. From the lognormal distribution parameters shown in **Table 4.6**, it is

observed that the means of the lognormal populations shifted to larger values for the Hazen, Karmen-Cozeny, and Cosby PTFs with values of 0.068, 0.0103, and 0.00055 cm/s respectively. The largest distributional change was observed with the values estimated using the Hazen PTF where the lognormal mean was more than triple the raw mean. The large distributional shift of the values estimated using Hazen's PTF is likely due to the large disparity between the mean and median of the estimated values of 1.69E-02 cm/s as well as the distribution being composed of discrete values rather than continuously distributed values (Figure 4.13a) due to its sole reliance on the D_{10} particle size.

The differences between the lognormal population mean, variance, and standard deviations of the estimated and USDA measured hydraulic conductivity values shown in **Table 4.7** confirm that Cosby's PTF provides the most accurate method of estimation with the smallest discrepancy between the lognormal means of 0.00204 cm/s. Even though the difference of 0.00204 cm/s between the mean of USDA measured values and Cosby PTF estimated values mean is substantial, it provides the best fit relative to the other methods and is therefore used in further analysis.

A spatial analysis of the hydraulic conductivity values estimated using Cosby's PTF indicates a lack of spatial dependence, shown by the large nugget effect observed in the semi-variogram (**Figure 4.15**). The lack of spatial dependence corresponds to both the high degree of natural variability in hydraulic conductivity as well as the irregular sample spacing and depth. Measurements taken at smaller intervals such as ten feet would likely show some degree of spatial continuity.

Table 4.5 Statistical properties of estimated K^* .

PTF	Hazen	K-C	Cosby	USDA
COUNT	546	546	546	211
MEAN	2.25E-02	9.81E-03	5.38E-04	2.58E-03
MEDIAN	5.63E-3	3.09E-03	4.81E-04	1.41E-03
MINIMUM	2.50E-05	9.77E-05	4.06E-05	1.41E-05
MAXIMUM	1.81E-01	2.47E-01	2.95E-03	1.41E-02
VARARIANCE	2.04E-03	5.29E-04	1.38E-07	1.42E-05
STANDARD DEVIATION	4.51E-02	2.30E-02	3.71E-04	3.75E-03
SKEWNESS	2.23	6.33	2.82	2.40
KURTOSIS	4.19	50.52	11.15	4.76

Table 4.6: Lognormal distribution parameters of K^* .

PTF	Hazen	K-C	Cosby	USDA
τ^*	6.82E-02	1.03E-02	5.49E-04	2.59E-03
ω^2	6.6824	1.14E-03	1.61E-07	2.39E-05
ω	2.5850	3.38E-02	4.01E-04	4.89E-03

τ^* , ω^2 , ω - estimated mean, variance, and standard deviation respectively of the lognormal population

Table 4.7: Comparison of population parameters between estimated and USDA measured K .

	USDA (K) vs.		
	Hazen (K*)	K-C (K*)	Cosby (K*)
μ^*_{diff}	0.38997	0.91094	1.00849
s^2_{diff}	5.75732	0.94287	1.08762
s_{diff}	1.46589	0.33691	0.57714
τ^*_{diff}	6.56E-02	7.74E-03	2.04E-03
ω^2_{diff}	6.68241	1.12E-03	2.00E-05
ω_{diff}	2.58015	2.89E-02	4.49E-03

μ^* , s^2 , s are the estimated mean, variance, and standard deviation

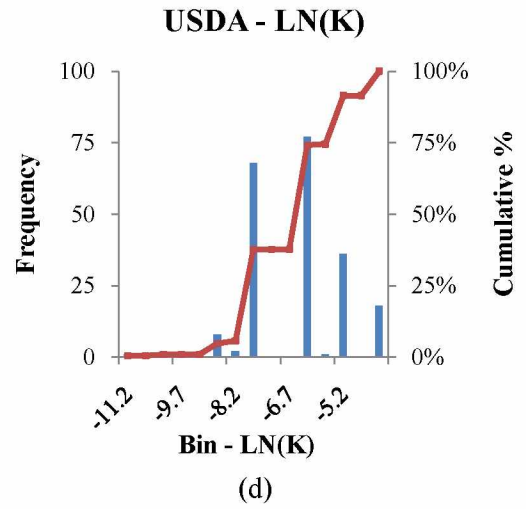
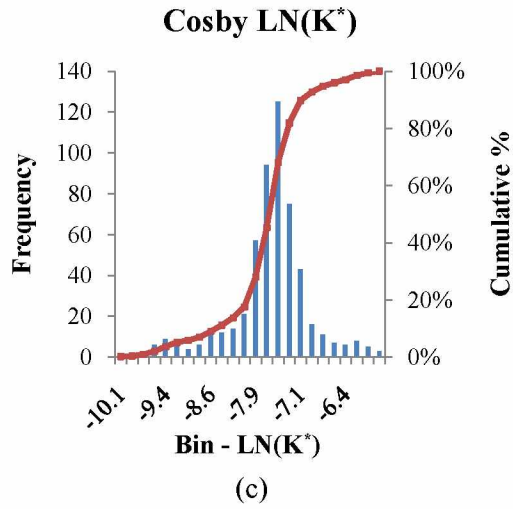
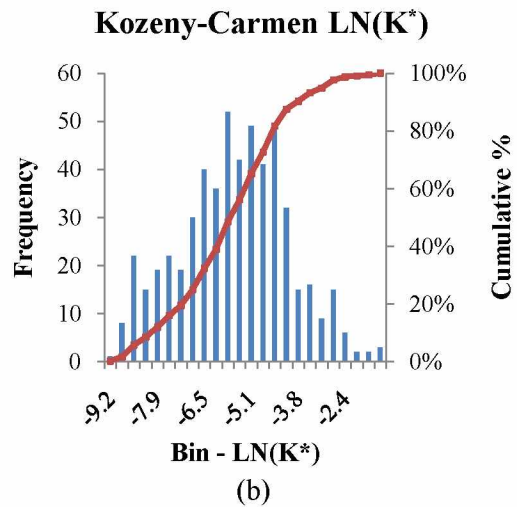
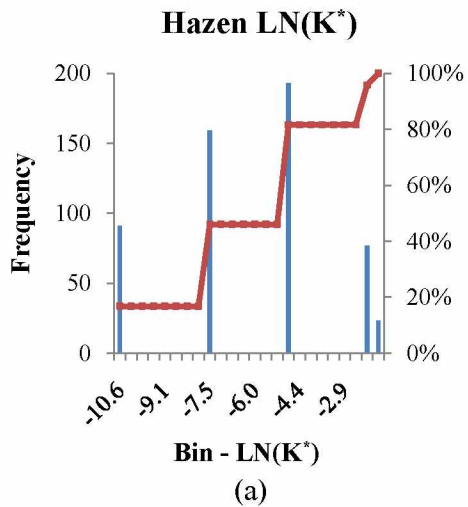


Figure 4.14: Histogram plots of natural-log transformed K^* distributions.



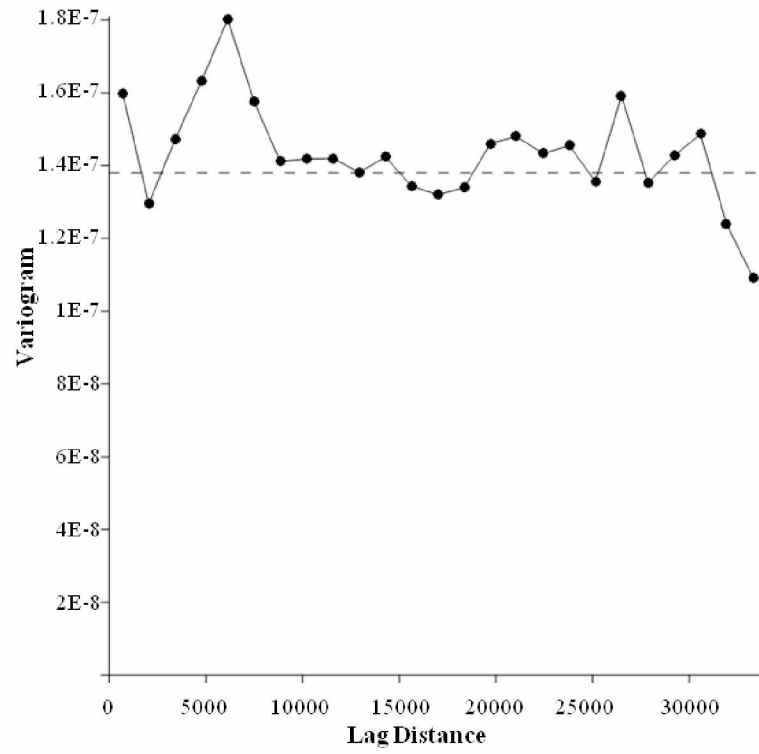


Figure 4.15: Semi-variogram of K^* obtained using (Cosby et al., 1984) PTF

4.4 CORRELATION ANALYSIS

Prior to using the compiled dataset in the development of predictive models, the relationships between hydraulic, geophysical, and geologic parameters need to be understood. From the basic correlation coefficient analysis shown in **Table 4.8** it can be observed that the strongest correlation between parameters is that between the apparent resistivity at 140K and 40K Hz. with a coefficient of correlation of 0.81. This corresponds to both the similarities shown in the contour plots (**Figure 4.2**) and our assumption that the two frequencies are responding to roughly the same geologic media. From **Table 4.8** it is also observed that the estimated hydraulic conductivity is shown to have a correlation of 0.54 with the soil indicator. This indicates that while there is a relationship between the estimated hydraulic conductivity and the soil indicator it is likely due to both parameters being derived from the same grain size distribution. When the hydraulic conductivity is plotted against the soil indicator classes (**Figure 4.16**) it can be observed that while a positive correlation is visible, the ranges of hydraulic conductivity values in each soil indicator class overlap and vary widely from 0.0022 to 0.0002 cm/s. This allows for the inclusion of the soil indicator in predictive model development without the parameters influence overshadowing the potential contribution of the other predictive variables.

From **Table 4.8**, it can be observed that the relationship between the estimated hydraulic conductivity and the geophysical parameters of apparent resistivity at 140K and 40K Hz, and magnetic field strength is observed to be weak with correlation coefficients of -0.11,-0.13, and -0.15, respectively. It can also be observed that the relationship

between the relative difference between the apparent resistivity measured at 140K and 40K Hz. (ρ_{diff}^*) and the estimated hydraulic conductivity is lower than that of the individual parameters with correlation coefficient of -0.06. Because of this the apparent resistivity measured at the two frequencies (140K and 40K Hz.) will be used as individual parameters in further analysis rather than the relative difference.

Table 4.8: Coefficient of correlation between parameters

	X	Y	ρ^{*140K}	ρ^{*40K}	ρ_{diff}^*	H	SI	NM
X	-							
Y	-0.97	-						
ρ^{*140K}	-0.35	0.36	-					
ρ^{*40K}	-0.13	0.10	0.81	-				
ρ_{diff}^*	-0.44	-0.47	0.87	0.42	-			
H	0.12	-0.15	-0.03	0.10	-0.13	-		
SI	-0.04	0.04	-0.10	-0.10	-0.07	-0.15	-	
NM	0.03	-0.01	0.27	0.19	0.25	0.14	-0.59	-
K^*	-0.03	0.04	-0.11	-0.13	-0.06	-0.15	0.54	-0.47

The scatter plot of hydraulic conductivity against the apparent resistivity at 140K Hz (**Figure 4.17a**) and magnetic field strength (**Figure 4.17b**) shows a homogenous distribution with no apparent relationship or trend. This strengthens the hypothesis that this relationship is too complex to employ universal functions or basic regression techniques to quantify the relationship and that a site dependent approximation using machine learning algorithms would be more suitable.

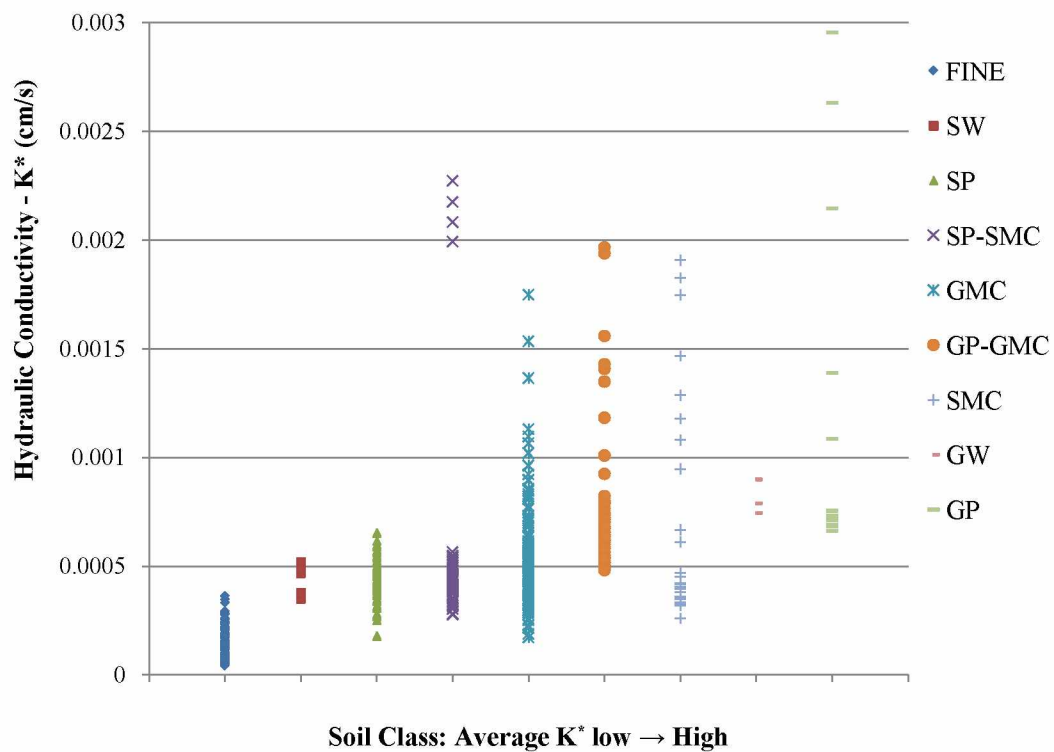


Figure 4.16: Distribution of K^* with respect to soil class

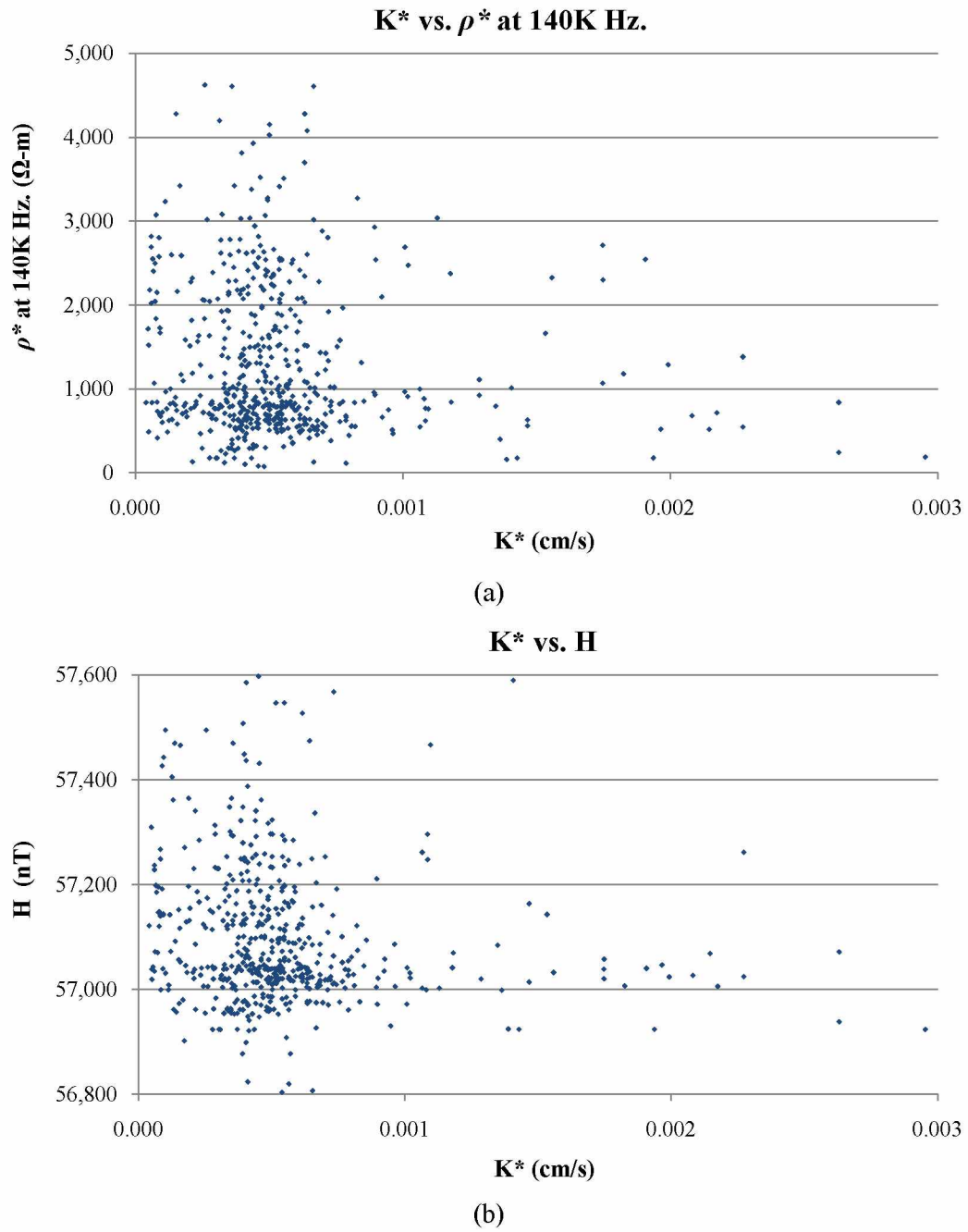


Figure 4.17: Scatter plot of geophysical parameters ρ^* at 140K and H vs. K^*

5.0 RESULTS AND DISCUSSION

The key objective of this study is to characterize the hydro-geophysical relationship in a manner that may be employed in a predictive model thus limiting the amount of borehole samples needed (or possibly eliminating the need of borehole sampling) to accurately characterize subsurface hydraulic conductivity. The correlation analysis in the previous chapter (Section 4.4) demonstrated that if there is a relationship between hydraulic conductivity and the geologic and geophysical parameters it is likely complex and non-linear. In order to quantify this complex relationship we employ machine learning algorithms thus capitalizing on their ability to capture complex patterns and physical relationships.

5.1 MEASURES OF PERFORMANCE

One of the main issues in the development of predictive models is the verification of accuracy and resilience. Ideally this would be done by testing the models on a supplemental data set. Due to the additional costs and time requirements involved in the collection of supplemental data it is not always feasible. One of the main concerns when using a single data set to both train and test a model is over-fitting. Over-fitting can occur when a model is accurately fit to a specified data set but when provided with an independent data set for verification the performance is significantly reduced. An alternative method that both minimizes the risk of over training and avoids the costs of collecting supplemental data is to divide the data set into two statistically similar subsets. One subset is used to develop the model and the second is used to validate the accuracy

of the model. This has shown to be the most reliable method to verify the accuracy of a model when additional data is not available (Dutta, 2006).

Due to the limitations in the number of natural moisture content measurements with respect to the other parameters, two data sets were created. This was done to test the significant of natural moisture content as an input parameter while not limiting the number of samples in the data set.

Two distinctly different data sets were used in this research for the characterization of hydraulic conductivity. One set was the complete data set excluding the natural moisture content data with a total of 546 samples (Total Dataset). The second data set was a reduced set of data that included the samples with natural moisture content as a parameter with a total of 180 samples (NM Dataset). Each data set was split into two statistically similar subsets of 80% and 20% using a genetic algorithm developed by Ganguli et al., (2003) with their statistical properties shown in **Table 5.1** and **Table 5.2** respectively. The subset composed of 80% of the data was used to develop the models while the smaller subset (20%) was used to test the model. The use of 80% of the data for development and 20% to test the model is a widely employed technique (Bowden et al., 2002; Oommen, 2006) that ensures that a sufficient number of points are supplied to train the model.

While the above methods help ensure the accuracy and reliability of the models developed, measures of performance are needed to quantitatively compare the results. In this study the accuracy is based on difference between the actual values (target data) and predicted values. This was calculated using performance measures such as:

Coefficient of Determination (R^2):

$$R^2 = \left\{ \frac{\sum_{i=1}^N \{(P_i - \bar{P})(T_i - \bar{T})\}}{[\sum_{i=1}^N (P_i - \bar{P})^{0.5}] [\sum_{i=1}^N (T_i - \bar{T})^{0.5}]} \right\}^2 \quad (5.1)$$

Coefficient of Correlation (r):

$$r = \frac{\sum_{i=1}^N \{(P_i - \bar{P})(T_i - \bar{T})\}}{\{[\sum_{i=1}^N (P_i - \bar{P})^2] [\sum_{i=1}^N (T_i - \bar{T})^2]\}^{0.5}} \quad (5.2)$$

Root Mean Square Error (RMSE):

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (P_i - T_i)^2}{N}} \quad (5.3)$$

Mean Absolute Error (MAE):

$$MAE = \frac{(\sum_{i=1}^N |P_i - T_i|)}{N} \quad (5.4)$$

where P_i is the predicted value, \bar{P} is the mean of the predicted values, T_i is target value, \bar{T} is the mean of the predicted values, and N is the number of samples.

The use of the multiple performance measures listed above provides a more comprehensive understanding of the model's accuracy. The coefficient of determination (R^2) serves as a primary measure by representing the proportion of variability in the target data set that is accounted for by the predictive model. The second key performance measure is the coefficient of correlation (r) which measures the strength and direction of a linear relationship between the target and predicted data set. While these two

correlation based measures are the most widely used methods to evaluate the performance of predictive models, their limitations with respect to sensitivity to extreme values and insensitivity to additive and relative differences between the target and predicted data (Oommen, 2006) necessitate the inclusion of supplemental measures.

By employing RMSE and MAE as additional measures of performance analysis, the predictive characteristics can be better clarified. RMSE gives a relatively high weight to large errors while MAE weights all averages equally. By comparing the RMSE and MAE, the degree of outlying errors can be assessed. For example, a model with a large RMSE but with a small MAE would indicate that while most data points are being predicted with a low degree of error, a few samples are not being accounted for and are producing large errors. When the RMSE is equal to the MAE the magnitudes of all the errors are equal.

Table 5.1: Statistical summary of training and testing subsets from Total Dataset

COUNT	TRAINING 437 (80%)		TESTING 109 (20%)	
	INPUT	MEAN	STDEV	MEAN
K^*	0.00053 cm/s	0.00037	0.00053 cm/s	0.00036
X	615031.4 m	22717.46	614951.7 m	22922.95
Y	7072783m	10862.17	7072791 m	11117.98
SI	4.42	1.85	4.43	1.81
ρ^*140K	1294.97 Ω -m	891.21	1291.75 Ω -m	935.91
ρ^*40K	986.29 Ω -m	479.43	987.79 Ω -m	498.88
H	57093.86 nT	128.87	57093.8 nT	132.68

Table 5.2: Statistical summary of the training and testing subsets from NM Dataset

COUNT	TRAINING 144 (80%)		TESTING 36 (20%)	
	INPUT	MEAN	STDEV	MEAN
K^*	0.00056 cm/s	0.00049	0.00051 cm/s	0.00045
X	612999.36 m	21995.42	610984.68 m	18262.49
Y	7073462.9 m	10595.85	7074292.90 m	8706.98
SI	4.09	2.48	3.43	2.10
ρ^*140K	1234.93 Ω -m	790.03	1329.48 Ω -m	983.50
ρ^*40K	954.48 Ω -m	406.55	1003.57 Ω -m	477.54
H	57103.28 nT	135.91	57140.96 nT	174.63
NM	15.59 %	15.66	19.39 %	24.79

5.2 ANALYSIS OF HYDRO-GEOPHYSICAL RELATIONSHIP

5.2.1 Neural Network Regression Analysis

Multiple networks were developed to test different combinations of input parameters and to determine the most accurate modeling method. **Table 5.3** shows the performance measures obtained for the validation of the ANNR models. From **Table 5.3** it is observed that using apparent resistivity as the only input parameter is insufficient to characterize the hydraulic conductivity (with our data) with coefficient of determination of .04. The regression plot of the target values against the predicted values (**Figure 5.1**) shows that values are being both under and over predicted with all of the values greater than $7.5E-04$ cm/s being under predicted. This lack of trend with respect to the target and predicted values is shown with a low correlation coefficient of 0.21. The inclusion of the magnetic field strength, spatial coordinates, soil indicator, and natural moisture content individually, improved the performance with coefficients of determination 0.11, 0.16, 0.37, and 0.36 respectively. While the results are improved the ANNR models failed to accurately characterize the relationship. The increased accuracy achieved through the addition of the soil indicator and natural moisture content indicate that these parameters each provide information in the characterization of the hydro-geophysical relationship but are insufficient without supplemental parameters.

The most accurate results using the Total Dataset was found using ρ^* , spatial coordinates, and SI with an R^2 of 0.44. While the accuracy is still poor, the correlation between the target and predicted values is significantly higher with a correlation

coefficient of 0.65. The improved correlation is shown in the scatter plot (**Figure 5.2**) with predicted values being more evenly distributed along the regression line. The best overall results were found using the NM Data set using inputs of ρ^* , spatial coordinates, soil indicator, and natural moisture content with an R^2 of 0.64. From the regression plot (**Figure 5.3**) it can be observed that the overall accuracy as well as prediction of the larger values is improved over the other models. The correlation of coefficient of 0.8 indicates that the strength of the relationship between the target and predicted values is high although the overall accuracy is still lacking.

Table 5.3: Performance measures for the ANNR model analysis.

Model Inputs					RMSE	MAE	r	R ²
ρ^*					0.00036	0.00023	0.21	0.04
ρ^*	H				0.00035	0.00022	0.34	0.11
ρ^*	XY				0.00034	0.00021	0.40	0.16
ρ^*	SI				0.00029	0.00017	0.61	0.37
ρ^*	NM				0.00039	0.00026	0.60	0.36
ρ^*	H	XY			0.00037	0.00025	0.23	0.05
ρ^*	H	SI			0.00032	0.00021	0.57	0.33
ρ^*	XY	SI			0.00028	0.00017	0.66	0.44
ρ^*	H	NM			0.00041	0.00025	0.56	0.31
ρ^*	XY	NM			0.00045	0.00030	0.40	0.16
ρ^*	SI	NM			0.00038	0.00024	0.64	0.40
ρ^*	H	XY	SI		0.00028	0.00018	0.65	0.43
ρ^*	H	XY	NM		0.00038	0.00025	0.62	0.38
ρ^*	H	SI	NM		0.00039	0.00024	0.62	0.38
ρ^*	XY	SI	NM		0.00029	0.00020	0.80	0.64
ρ^*	H	XY	SI	NM	0.00033	0.00022	0.73	0.53
AVERAGE					0.00035	0.00026	0.54	0.32

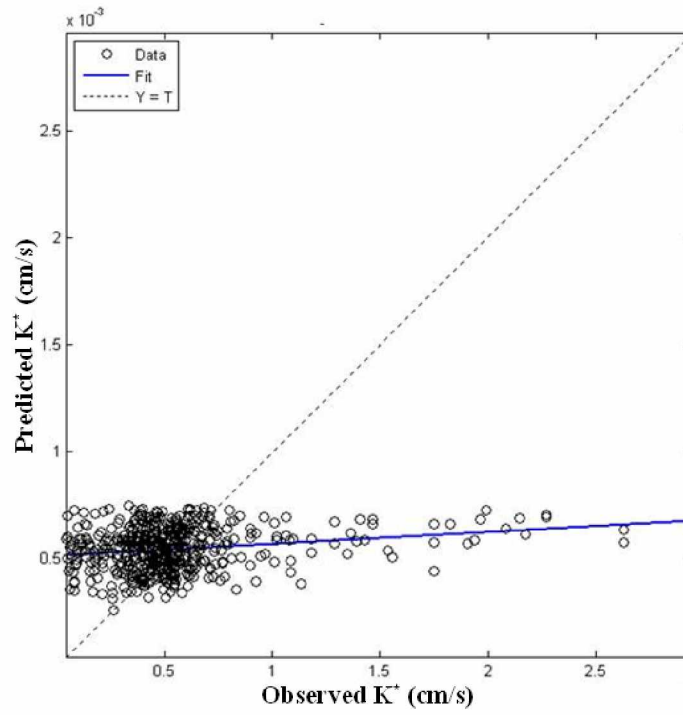


Figure 5.1: Regression plot ANNR model results with an input of ρ^* .

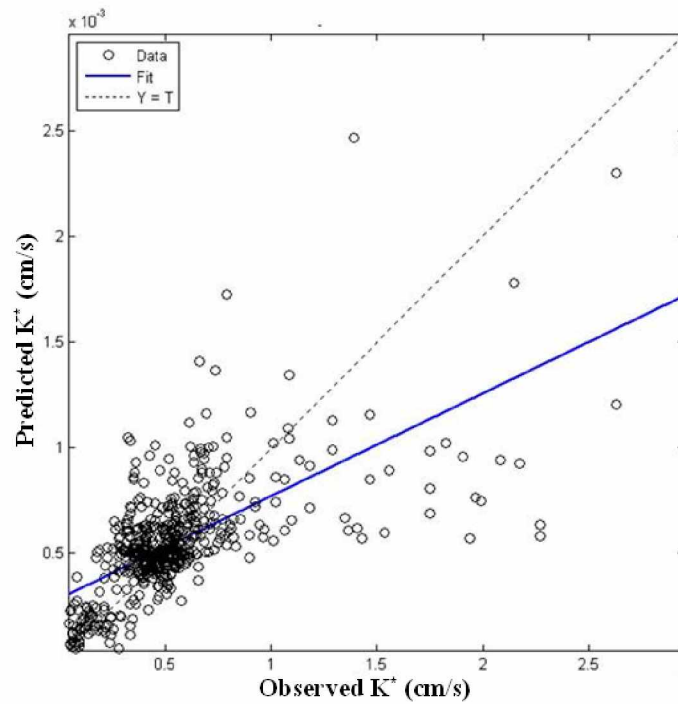


Figure 5.2: Regression plot of ANNR model results with inputs of ρ^* , SI , and XY .

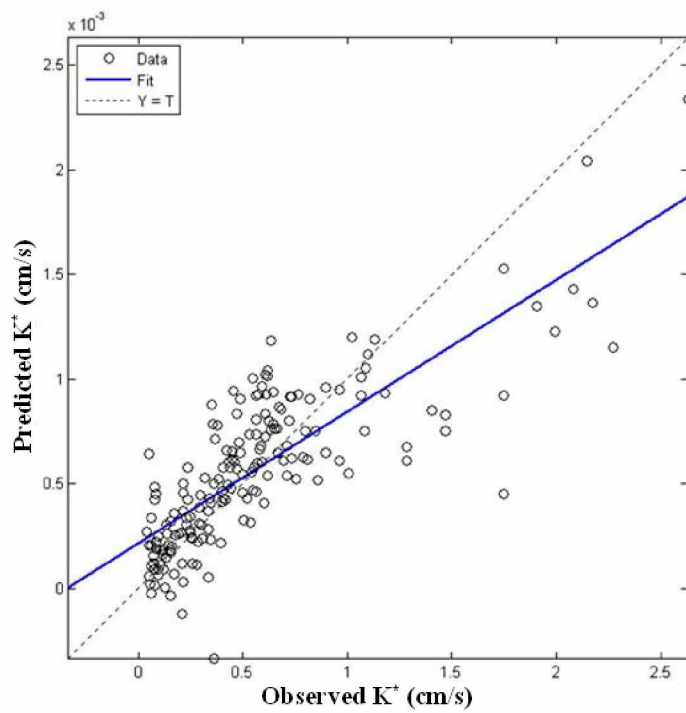


Figure 5.3: Regression plot of ANN model results with inputs of ρ^* , XY , SI , NM .

5.2.2 Support Vector Regression Analysis

Due to the inability of ANNR to successfully characterize the hydro-geophysical relationship with an average R^2 of 0.32, SVR was employed as an alternative method of analysis. **Table 5.4** shows the performance measures, optimum model parameters, and number of support vectors for each of the models developed. From **Table 5.4** it can be observed that similar to ANNR, using the apparent resistivity as the only input parameter is insufficient to characterize the hydraulic conductivity with R^2 of 0.03. The low degree of accuracy is correlated with the high degree of uncertainty found in the data set with all 437 samples in the training data set being employed as support vectors. The regression plots of the target values against the predicted values shown in **Figure 5.4** indicates that the model is roughly being fit to the mean of the hydraulic conductivity ($5.38E-4$ cm/s) of the data set without accounting for any deviation.

The incorporation of magnetic field (H), spatial coordinates (X and Y), and the soil indicator (SI) provides the best results with an R^2 of 0.57. The regression plot of the target values against the predicted values shown in **Figure 5.5** indicates that the model's accuracy decreases as the magnitude of the target value increases. While the uncertainty is decreased relative to the use of apparent resistivity as the sole input, it is still quite high with 271 (61.59%) of the training instances being classified as support vectors.

The best performance using the reduced data set with natural moisture content was an R^2 of 0.49 using input parameters as apparent resistivity, magnetic field, soil indicator, and natural moisture content. The high degree of uncertainty in the data set

and complexity of the model is shown with 100% of the training data being used as support vectors.

The concept of support vector was then used to analyze the data set of the best performing combination of input parameters (ρ^* , XY, H, and SI) to better characterize the uncertainty within each parameter and to identify gaps in the data set that can be used to guide future data collection. It is important to state here that SVR method helps in determining data gaps in a model as opposed to the ANNR method, which is a significant strength of the SVR method (Oommen and Baise, in Press). **Figure 5.6** shows the distribution of the support vectors in the range of the input parameters. It is observed that of the input parameters, apparent resistivity at 140K Hz has the least amount of uncertainty and lowest percentage of support vectors, whereas, SI has the highest amount of uncertainty and highest percentage of support vectors.

It is observed from **Figure 5.6** that the distribution of support vectors within the ranges of spatial coordinates (X and Y) appears to be evenly distributed with roughly 60 to 80% of the training data as support vectors. The large amount of support vectors throughout the data ranges indicates that there is a relatively high degree of uncertainty in the spatial variability of K^* . This confirms what was observed in the variogram analysis of K^* shown in **Figure 4.14**. From the distribution of support vectors it is observed that of apparent resistivity data measured at 140K Hz which ranged from 74 to 4,723 Ω -m over 80% of values greater than 4,168 Ω -m employed as support vectors. This indicates the need for supplemental data from areas with higher apparent resistivity in order to obtain a more accurate predictive model for K^* characterization.

From the distribution of support vectors for magnetic field strength (**Figure 5.6**), it can be observed that values of magnetic field strength, which range from 56,804 to 57,597 nT are shown to be well sampled at the high (<57,597 nT) and low (>56,883 nT) end of the range with less than 50% being employed as support vectors. The magnetic strength values in the mid range between 57,518 and 56,883 nT were poorly sampled with between 70% and 80% being employed as support vectors. The high degree of uncertainty observed in the middle range indicates that the extreme values have a stronger relationship with changes in the hydraulic conductivity. This observation aligns with characteristics of magnetic field strength measurements in that it was not influenced by subtle changes in the shallow subsurface hydraulic parameter. The high and low ends of the range are likely recognizing more significant changes in the subsurface such as igneous intrusions or changes in deposition influencing the subsurface hydraulic conductivity.

From **Figure 5.6** it can be observed that the strongest relationship between the indicator parameters and hydraulic conductivity exists for apparent conductivity measured at 140K and 40K Hz, between values of 3,258 to 3,713 Ω -m and 2,241 to 2,506 Ω -m, respectively with less than 40% being employed as support vectors. The distribution of support vectors for both frequencies is similar in that the top end range is the most poorly sampled followed by well sampled ranges of data. The poor sampling at the top end of the data range for the apparent resistivity measured at both 140K and 40K could possibly be due to error incurred at these points during the survey or insufficient number of data points in this range to accurately characterize the changing relationship.

Table 5.4: Performance measures and parameters for the SVR model analysis.

Model Inputs					γ	ε	C	SV	RMSE	MAE	r	R ²
ρ^*					0.001	0.001	1	437	0.00042	0.00025	0.18	0.03
ρ^*	H				0.001	0.141	1	330	0.00042	0.00024	0.24	0.06
ρ^*	XY				0.281	0.481	1	151	0.00040	0.00023	0.31	0.1
ρ^*	SI				0.081	0.121	13	309	0.00034	0.00017	0.67	0.46
ρ^*	NM				0.011	0.001	5	138	0.00041	0.00026	0.47	0.49
ρ^*	H	XY			0.061	0.361	3	201	0.00041	0.00023	0.28	0.08
ρ^*	H	SI			0.161	0.181	19	255	0.00033	0.00018	0.70	0.50
ρ^*	XY	SI			0.101	0.381	19	141	0.00030	0.00017	0.75	0.56
ρ^*	H	NM			0.211	0.181	2	90	0.00037	0.00024	0.55	0.33
ρ^*	XY	NM			0.091	0.071	1	121	0.00036	0.00022	0.61	0.39
ρ^*	SI	NM			0.011	0.001	1	142	0.00040	0.00024	0.62	0.48
ρ^*	H	XY	SI		0.081	0.181	15	271	0.00030	0.00016	0.75	0.57
ρ^*	H	XY	NM		0.101	0.431	16	126	0.00041	0.00030	0.45	0.39
ρ^*	H	SI	NM		0.011	0.001	1	143	0.00040	0.00024	0.63	0.47
ρ^*	XY	SI	NM		0.011	0.001	5	140	0.00037	0.00023	0.65	0.49
ρ^*	H	XY	SI	NM	0.011	0.001	1	143	0.00039	0.00024	0.62	0.47
AVERAGE								196	0.00038	0.00023	0.53	0.37

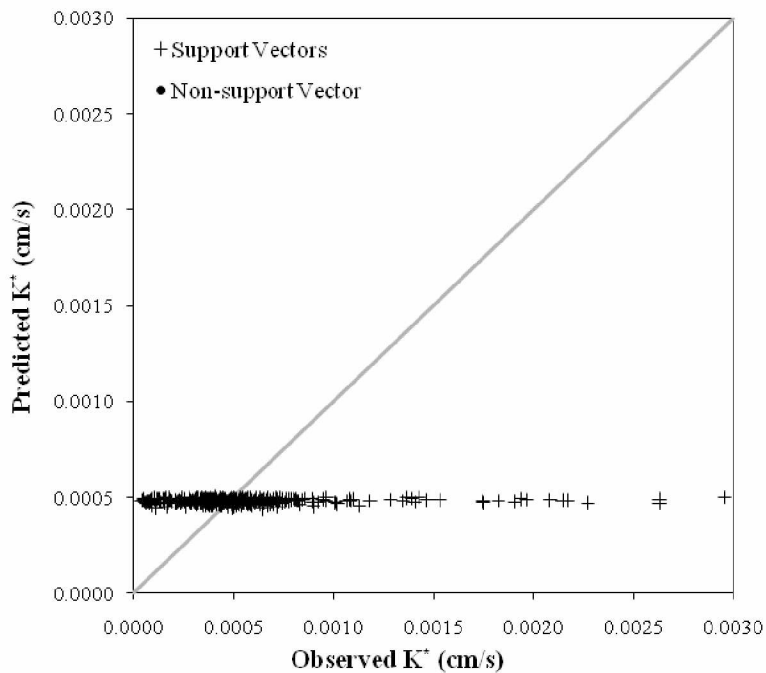


Figure 5.4: Regression plot of SVR model results with an input of ρ^*

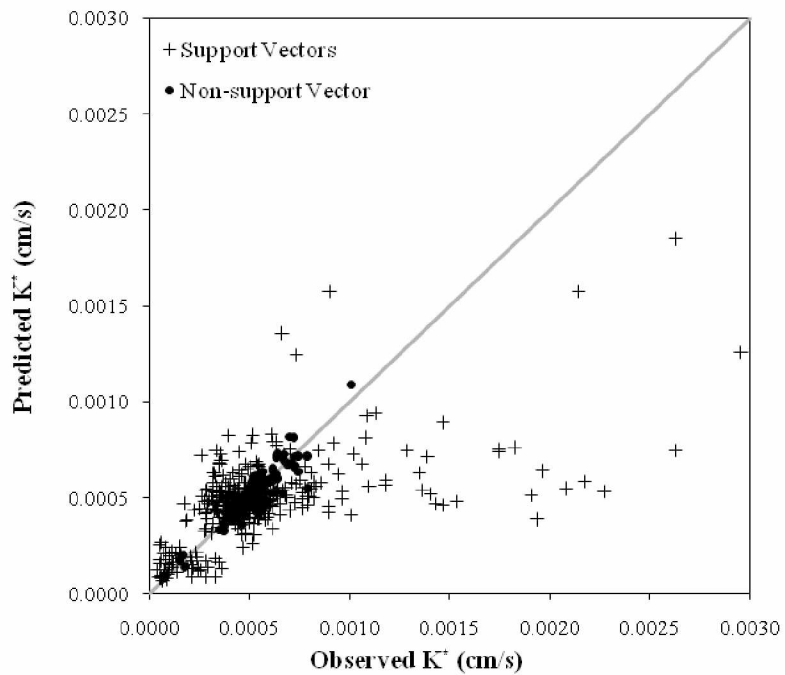


Figure 5.5: Regression plot of SVR model results with inputs of ρ^* , H , XY , and SI

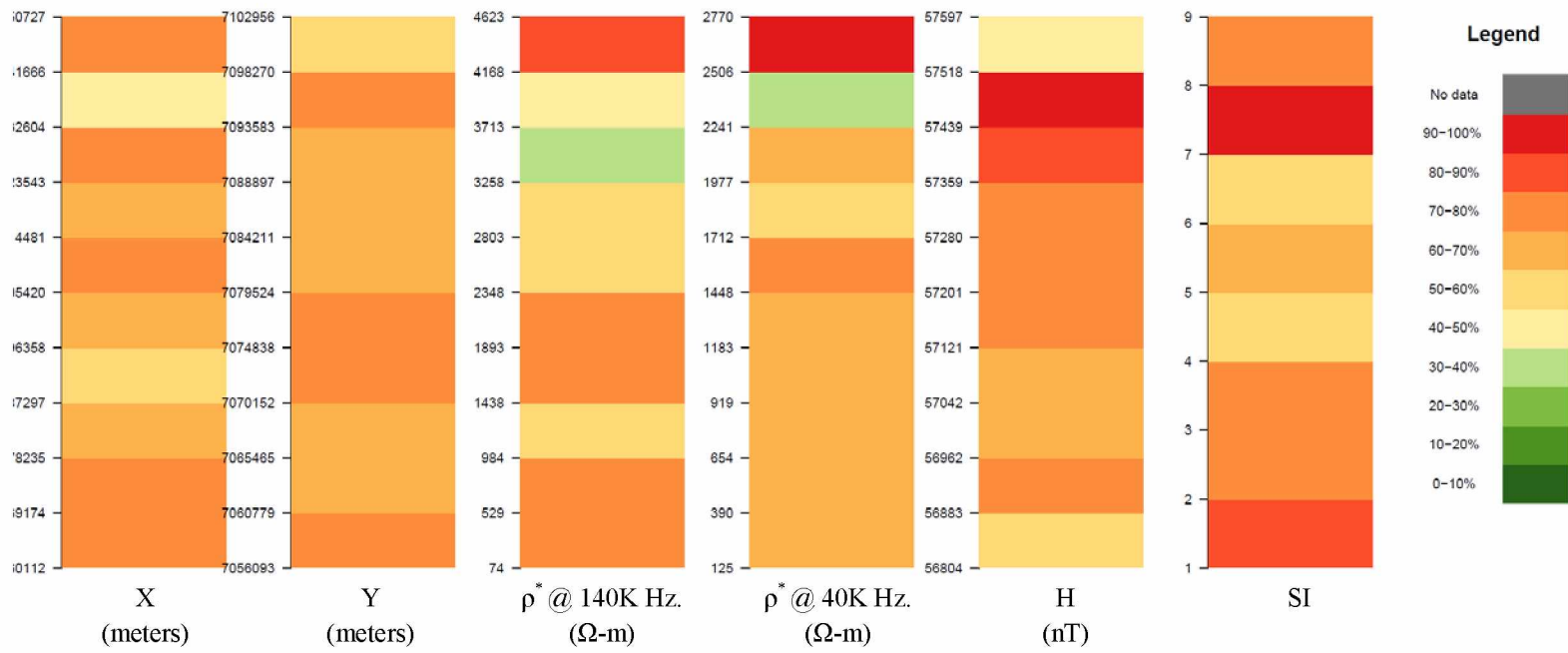


Figure 5.6: Percentage of support vectors in the range of predictor variables.

5.2.3 Neural Network Classification Analysis

Due to the relatively poor performance of the regression algorithms with the best results observed showing a R^2 of 0.64, alternative methods of analysis were explored.

Table 5.5 shows the statistical properties of the high, medium, and low clusters of hydraulic conductivity grouped using the *k-means* clustering algorithm. These classifications were used as the target output of the previously derived subsets shown in **Table 5.1** and **Table 5.2**.

Table 5.5: Statistical summary of k-means clusters

	CLASSIFICATION		
	HIGH	MEDIUM	LOW
COUNT	27	256	263
MEAN	1.84E-03	6.27E-04	3.18E-04
MEDIAN	1.75E-03	5.75E-04	3.54E-04
MINIMUM	1.29E-03	4.74E-04	4.06E-05
MAXIMUM	2.95E-03	1.18E-03	4.70E-04
VARIANCE	2.04E-07	2.26E-08	1.51E-08
STANDARD DEVIATION	4.52E-04	1.50E-04	1.23E-04
SKEWNESS	0.78	1.66	-0.80
KURTOSIS	4.38E-03	2.53	-0.58

From the analyses, the following observations were made:

- Similar to the regression analysis, multiple networks were developed to test the different combinations of parameters. From **Table 5.6** it is observed that the overall accuracy average of the classification models is relatively good with a value of 76%, however, the model's ability to correctly classify the high value cluster targets is poor with an average of 19%.

- Similar to the regression analysis the use of apparent resistivity as the sole input is shown to be insufficient with the worst overall accuracy of the classification models at 56% and 0% correct classification of the high value cluster targets. While the inclusion of magnetic field, spatial coordinates, and the soil indicator with the apparent resistivity provided the best results using the complete data set with an overall accuracy of 78% the prediction performance of the high value classification was still poor with only 15% correct classifications.
- The most accurate model was developed using the reduced data set with apparent resistivity, spatial coordinates, soil indicator and natural moisture content with an overall accuracy of 88%. Even with the most accurate model the high value clusters were only correctly classified 40% of the time.

ANNC's inability to accurately predict the high value cluster in comparison to the medium and low value clusters is likely in part due to the an insufficient number of sample points. From **Table 5.5** it can be observed that the medium and low classifications have 256 and 263 samples respectively while the high value classification has only twenty seven samples.

Table 5.6: Neural network classification results.

Model Inputs					# Neurons	% CORRECT			
						HIGH	MEDIUM	LOW	OVERALL
ρ^*					69	0%	43%	74%	56%
ρ^*	H				98	0%	58%	77%	64%
ρ^*	XY				57	11%	79%	69%	71%
ρ^*	SI				116	0%	58%	76%	64%
ρ^*	NM				138	0%	87%	86%	79%
ρ^*	H	XY			124	0%	67%	75%	67%
ρ^*	H	SI			149	15%	80%	74%	74%
ρ^*	XY	SI			145	15%	79%	83%	78%
ρ^*	H	NM			133	13%	88%	86%	81%
ρ^*	XY	NM			132	33%	91%	83%	82%
ρ^*	SI	NM			114	33%	90%	84%	82%
ρ^*	H	XY	SI		142	15%	79%	83%	78%
ρ^*	H	XY	NM		103	47%	88%	86%	84%
ρ^*	H	SI	NM		111	40%	93%	86%	85%
ρ^*	XY	SI	NM		104	40%	94%	91%	88%
ρ^*	H	XY	SI	NM	93	47%	94%	89%	87%
AVERAGE					114	19%	79%	81%	76%

5.2.4 Comparison of Model Predictions

The performance of the machine learning algorithm models developed in sections 5.2.1 and 5.2.2 have been objectively evaluated using the performance measures discussed in Section 5.1. **Table 5.7** summarizes the various performance measures of the regression models for the most accurate combination of input parameters from both the Total Dataset and the NM Dataset. From **Table 5.7** it is observed that SVR provided the best model (largest coefficient of determination) using the Total Dataset for mapping hydraulic conductivity along the Alaska Highway corridor between Delta Junction and Tok, Alaska. The best combination of input parameters was found to be the apparent resistivity at 140K and 40K Hz, spatial coordinates (X and Y), magnetic field strength, and soil indicator with a R^2 of 0.57. While the overall accuracy of the SVR model is poor, the r of 0.75 between the predicted and target outputs leads us to believe that there is a relationship between the input parameters and the hydraulic conductivity. The poor overall accuracy of the SVR model is likely due to a variety of reasons, such as the complexity of hydro-geophysical relationship shown by the high number of support vectors (61.59%) indicating a lack of data in certain aspects of the individual parameters (see **Figure 5.6**), errors incorporated through the estimation of hydraulic conductivity, large spatial distribution of data, and limitations of the method. If the sources of error can be minimized through the incorporation of supplemental hydraulic data with a higher spatial density, SVR might be able to capture the relationship indicated by the correlation coefficient and characterize the subsurface hydraulic parameters more accurately.

From **Table 5.7** it can also be observed that when using the NM Dataset, ANNR provided the best model for the prediction of hydraulic conductivity. The most accurate combination of input parameters using this regression technique was found to be apparent resistivity at 140K and 40K Hz, spatial coordinates (X and Y), soil indicator and natural moisture content with a R^2 of 0.64. While the overall accuracy of the model may be poor, the r of 0.79 indicates that the model is recognizing a relationship between the input parameters and target output of hydraulic conductivity. Although the natural moisture content should have contributed more significantly based on its importance with respect to electrical conductivity, the fact that the geophysical and natural moisture content measurements were not taken at the same time most likely contributed to the errors seen.

A comparison of average accuracy of the models developed using ANNR (**Table 5.3**) and SVR (**Table 5.4**) reveals that SVR provided a 25% and 10% improvement in the coefficient of determination for the Total Dataset and NM Dataset respectively. SVR's drop in improvement over ANNR between the Total and NM Datasets indicate that reduction in size of the training subset is more significantly impacting its accuracy. A comparison of the regression plots between the predicted and observed values (Section 5.2.1 and 5.2.2) indicates that the accuracy of both machine learning algorithm regression techniques decreases as the hydraulic conductivity increases in value. The most probable reason for the decrease in accuracy for the larger values is the insufficient number of sample points to train the models.

Due to the relatively poor performance of the regression models, ANNC was employed as an alternative method. Although the results cannot be directly compared

due to the difference in their target values, it provides an alternative method of classification when the delineation of ranges of values is sufficient or where the performance of the regression models proves insufficient.

Table 5.8 summarizes the various performance measures of the classification models for the most accurate combination of input parameters from both the Total and NM Datasets. From **Table 5.8**, it is observed that of the models developed using ANNC, the best results were achieved using the NM Dataset with input parameters of apparent resistivity, spatial coordinates (X and Y), soil indicator, and natural moisture content that provided an overall classification accuracy of 88%. As with both regression models, ANNC's ability to predict the high values is significantly lower than the overall average with only 40% accuracy in classifications. **Table 5.8** also shows that the best results achieved using the Total Dataset were with input parameters of apparent resistivity, spatial coordinates (X and Y), soil indicator, and magnetic field strength with an overall classification accuracy of 78%.

Table 5.7: Performance measures obtained using the best combination of input parameters from both the Total and the NM Dataset for the regression models discussed in section 5.3.

Method	Dataset	Model Inputs				RMSE	MAE	r	R ²
SVR	Total	ρ^*	XY	SI	H	0.00030	0.00016	0.75	0.57
ANNR	Total	ρ^*	XY	SI		0.00028	0.00017	0.66	0.44
SVR	NM	ρ^*	H	SI	NM	0.00039	0.00024	0.63	0.47
ANNR	NM	ρ^*	XY	SI	NM	0.00029	0.00020	0.79	0.64

Table 5.8: Accuracy of the best combination of input parameters from both the Total and NM Dataset for the classification model discussed in section 5.3.

Dataset	Model Inputs				% CORRECT CLASSIFICATION			
					HIGH	MEDIUM	LOW	OVERALL
Total	ρ^*	H	XY	SI	15%	79%	83%	78%
NM	ρ^*	XY	SI	NM	40%	94%	91%	88%

6.0 CONCLUSIONS AND FUTURE WORK

This study investigated the applicability of characterizing subsurface hydraulic parameters on a large scale using geophysical data from an AEM survey. Machine learning algorithms were employed to develop predictive models using both the AEM data and supplemental geologic parameters. Based on the study results the following major conclusions are derived:

- The relationship between geophysical properties and hydraulic properties of a porous media was found to be too complex or weak to use the apparent resistivity collected during the airborne electromagnetic survey (AEMS) of the Alaska Highway corridor as the sole parameter in the characterization of hydraulic conductivity. The inclusion of supplemental geophysical, geologic, and spatial parameters with apparent resistivity significantly aided in the characterization increasing the performance of the predictive models.
- SVR demonstrated its capability to model complex process by providing the best performing predictive model using the Total Dataset with input parameters of apparent resistivity, spatial coordinates (X and Y), soil indicator, and magnetic field strength. The models developed using SVR showed an overall average increase in R^2 of 25% and 10% for the Total and NM Datasets respectively over ANN.
- Artificial neural network regression (ANNR) provided the best performing predictive model using the smaller NM Dataset with input parameters of apparent resistivity, spatial coordinates (X and Y), soil indicator, and natural moisture content.

- Artificial neural network classification (ANNC) was shown to offer a viable alternative to regression techniques in situations where point measurements are not needed and broad spatial changes in hydraulic conductivity are desired. The best accuracy was found using input parameters of apparent resistivity, spatial coordinates, soil indicators, and natural moisture content with 76% accuracy in classifications.
- The accuracy of each of the machine learning algorithms was seen to decrease as the value of hydraulic conductivity increased. This is likely due to insufficient data in high range of values between 0.00129 and 0.00295 cm/s. Additional data in the high range from which to train the models would likely increase the accuracy significantly.
- While NM is theoretically the most important parameter, its inclusion was not as significant as expected. This is likely due to an insufficient number of samples and the temporal difference in data collected from the boreholes and when the geophysical survey was performed.

Recommendations for future work:

- The use of empirically derived hydraulic conductivity values rather than measured values likely contributed to the errors seen and over signified the importance of the soil indicator. Field derived hydraulic conductivity measurements sampled at regular intervals of spacing and depth would likely provide an improved characterization of the hydro-geophysical relationship.
- The inclusion of time relative natural moisture content data through additional sampling or satellite imagery would provide clarification with respect to its influence on hydro-geophysical relationship.

- An improved hydrologic analysis of the area including the modeling of subsurface flow would provide an improved reference from which to compare the predictive results.
- Additional analysis using methods such as Multiple Regressive Pattern Recognition Techniques (Oommen et al., accepted pending minor revisions) could provide improved results through the combination of geo-statistical techniques and machine learning algorithms.

REFERENCES

- Akima, H. 1970. A new method of interpolation and smooth curve fitting based on local procedures. *Journal of the Association of Computing Machinery* 17(4):589-602.
- Anderson, G. S. 1970. Hydrologic reconnaissance of the Tanana basin, central Alaska. U.S. Geological Survey.
- Archie, G. E. 1942. The electrical resistivity log as an aid in determining some reservoir characteristics. *Petroleum Transactions of AIME* 146:54-62.
- Archie, G. E. 1947. Electrical resistivity an aid in core-analysis interpretation. *American Association of Petroleum Geologists Bulletin* 31(2):350-366.
- Archie, G. E. 1950. Introduction to petrophysics of reservoir rocks. *American Association of Petroleum Geologists Bulletin* 34(5):943-961.
- Asefa, T., Kemblowski, M. W., Urroz, G., McKee, M., and Khalil, A. 2004. Support vectors-based groundwater head observation networks design. *Water Resour. Res.* 40(11):W11509.
- Bennett, K. P., and Mangasarian, O. L. 1992. Robust linear programming discrimination of two linearly inseparable sets. *Optimization Methods and Software* 1(1):23 - 34.
- Børgesen, C. D., Iversen, B. V., Jacobsen, O. H., and Schaap, M. G. 2008. Pedotransfer functions estimating soil hydraulic properties using different soil parameters. *Hydrological Processes* 22(11):1630-1639.
- Bouma, J. 1989. Using soil survey data for quantitative land evaluation. *Advances in Soil Science* 9:177-213.
- Bowden, G. J., Maier, H. R., and Dandy, G. C. 2002. Optimal division of data for neural network models in water resources applications. *Water Resour. Res.* 38(2):1010.
- Brazo, G. M. 1980. Engineering Geology and Hydrology Report, Delta Junction 10 Miles S.E., Project No. F-062-3(20). Alaska Department of Transportation and Public Facilities, Engineering Geology Section. Fairbanks, Alaska.
- Brazo, G. M. 1987. Engineering Geology & Soils Report, Alaska Highway, Johnson River to DOT Lake, Project No. A37672. Alaska Department of Transportation and Public Facilities, Materials Section. Fairbanks, Alaska.
- Brazo, G. M. 1993. Geotechnical Report, Alaska Highway, Mile 1386 North, Federal Project No. NH-IR-0A2-2(2). Alaska Department of Transportation and Public Facilities, Materials Section. Fairbanks, Alaska.

- Brooks, A. H., and Peters, W. J. 1900. A reconnaissance in the White and Tanana River Basin, Alaska, 1898: U.S. Geological Survey 20th Anniversary Report, 7:431-494
- Budiman, M., and McBratney, A. B. 2000. Evaluation and development of hydraulic conductivity pedotransfer functions for Australian soil. *Australian Journal of Soil Research* 38:905-926.
- Burns, L. E., Fugro Airborne Surveys Corp., and Stevens Exploration Management Corp. 2006. Line, grid, and vector data, and plot files for the airborne geophysical survey of the Alaska Highway corridor, east-central Alaska. A. D. o. G. G. Surveys.
- Butler, P. L. 1993. Geotechnical and Soils Report, Alaska Highway Upgrade, Milepost 1380 to 1386, Project No. 65408/1-0A2-2(1). Alaska Department of Transportation and Public Facilities. Fairbanks Alaska.
- Campbell, G. S. 1974. A Simple Method for Determining Unsaturated Conductivity From Moisture Retention Data. *Soil Science* 117(6):311-314.
- Carman, P. C. 1938. The determination of the specific surface of powders. *J. Soc. Chem. Ind. Trans.* 57:225.
- Carrier, W. D. 2003. Goodbye, Hazen; Hello, Kozeny-Carman. *Journal of Geotechnical and GeoEnvironmental Engineering* 129(11):1054-1056.
- Cosby, B. J., Hornberger, G. M., Clapp, R. B., and Ginn, T. R. 1984. A Statistical Exploration of the Relationships of Soil Moisture Characteristics to the Physical Properties of Soils. *Water Resour. Res.* 20(6):682-690.
- Dimitriadou, E., Hornik, K., Leisch, F., Meyer, D., and Weingessel, A. 2007. *e1071 Misc functions of the department of statistics*. TU Wein.
- Doussan, C., and Ruy, S. 2009. Prediction of unsaturated soil hydraulic conductivity with electrical conductivity. *Water Resour. Res.* 45(10):W10408.
- Dutta, S. 2006. Predictive performance of machine learning algorithms for ore reserve estimation in sparse and imprecise data: PhD Thesis. University of Alaska, Fairbanks
- Farg, A., and Mohamed, R. M. 2004. Regression Using Support Vector Machines: Basic Foundations. University of Louisville. Louisville, Kentucky.
- Fogg, G. E., Noyes, C. D., and Carle, S. F. 1998. Geologically based model of heterogeneous hydraulic conductivity in an alluvial setting. *Hydrogeology Journal* 6(1):131-143.
- Freeze, R. A. 1975. A stochastic-conceptual analysis of one-dimensional groundwater flow in non-uniform homogenous media. *Water Resour. Res.* 11(5):725-741.

Freeze, R. A., and Cherry, J. A. 1979. *Groundwater*. Prentice Hall Inc., Englewood Cliffs, New Jersey.

Friedman, S. P. 2005. Soil properties influencing apparent electrical conductivity: a review. *Computers and Electronics in Agriculture* 46(1-3):45-70.

Ganguli, R., Walsh, D. E., and Yu, S. 2003. Calibration of Online Analyzers Using Neural Networks. United States Department of Energy DE-FC26-01NT41058.

Ghanbarian-Alavijeh, B., Liaghat, A. M., and Sohrabi, S. 2010. Estimating saturated hydraulic conductivity from soil physical properties using neural networks model. *World Academy of Science, Engineering and Technology*(62):131-136.

Grahek, M. E. 1981. Engineering Geology and Soils Report, Sears Creek Realignment, Project No. BR-062-02(15). Alaska Department of Transportation and Public Facilities, Materials Section. Fairbanks, Alaska.

Grahek, M. E. 1983. Engineering Geology and Soils Report, Alaska Highway, Robertson River To Yerrick Creek, Project No. F-062-2(14). Alaska Department of Transportation and Public Facilities, Materials Section. Northern Region.

Grahek, M. E. 1984. Engineering Geology and Soils Report, Alaska Highway, Dot Lake to Robertson River, Project No. F-062-2(14). Alaska Department of Transportation and Public Facilities, Materials Section. Northern Region.

Gunn, S. 1998. Support vector machines for classification and regression. Technical Report, University of South Hampton.

Hagan, M. T., Demuth, H. B., and Beale, M. H. 2002. *Neural Network Design*. Campus Publishing Service, Colorado University Bookstore, Boulder, Colorado.

Hazen, A. 1911. Discussion of 'Dams on sand foundations' by A.C. Koenig. *Trans. Am. Soc. Civ. Eng.* 73:199-203.

Holmes, W. 1965. Geological Reconnaissance Along the Alaska Highway Delta River to Tok Junction, Alaska. U.S. Department of the Interior 1181-H.

Homes, G. W., and Benninghoff, W. S. 1957. Terrain study of the Army test area, Fort Greely, Alaska. U.S. Geological Survey.

Howard, A. K. 1988. *Unified soil classification system [microform] : test procedures*. GR (Series) ; 88-8. No. Accessed from <http://nla.gov.au/nla.cat-vn4012092>. Geotechnical Services Branch, Research and Laboratory Services Division, Denver Office, U.S. Dept. of the Interior, Bureau of Reclamation, Denver, Colo. .:

Huang, H., and Fraser, D. C. 2001. Mapping of the resistivity, susceptibility, and permittivity of the earth using a helicopter-borne electromagnetic system. *Geophysics* 66(1):148-157.

Jordan, M. I. 1995. Why the logistic function? A tutorial discussion on probabilities and neural networks. Massachusetts Institute of Technology Computational Cognitive Science Technical Report 9503. Boston, Massachusetts.

Kearey, P., Brooks, M., and Hill, I. 2002. *An Introduction to Geophysical Exploration*. 3 ed. John Wiley & Sons, Ltd., Chichester.

Kim, K., Zhou, W., and Huang, S. L. 2008. Frost heave predictions of buried chilled gas pipelines with the effect of permafrost. *Cold Regions Science and Technology* 53(3):382-396.

Kozeny, J. 1927. Ueber kapillare Leitung des Wassers im Boden. *Wien. Akad. Wiss.* 136(2a):271.

Lesmes, D. P., and Friedman, S. P. 2005. Relationships between the Electrical and Hydrogeological Properties of Rocks and Soils. In *Hydrogeophysics*, 87-128. Y. Rubin, and S. S. Hubbard, eds: Springer Netherlands.

Livingston, H. R. 1964. Materials Investigation, Johnson River to Mile Post 1386, Project No. F 062-3(12). Alaska Department of Highways, Materials Division. Fairbanks, Alaska.

Livingston, H. R. 1969. First Supplemental Materials Site Report, Johnson River to M.P. 1386, Project No. F-062-3(12). Alaska Department of Highways, Materials Division. Fairbanks, Alaska.

Mbonimpa, M., Aubertin, M., Chapuis, R. P., and Bussière, B. 2002. Practical pedotransfer functions for estimating the saturated hydraulic conductivity. *Geotechnical and Geological Engineering* 20(3):235-259.

Mermoud, A., and Xu, D. 2006. Comparative analysis of three methods to generate soil hydraulic functions. *Soil and Tillage Research* 87(1):89-100.

Mertie, J. B. 1937. The Yukon-Tanana Region, Alaska. U.S. Geological Survey Bulletin. 872:276.

Moffit, F. H. 1954. Geology of the eastern part of the Alaska Range and adjacent area. U.S. Geological Survey Bulletin. 989-D:63-218.

Odong, J. 2007. Evaluation of empirical formula for determination of hydraulic conductivity based on grain-size analysis. *Journal of American Science* 3(3):54-60.

- Oommen, T. 2006. Geodatabase development and GIS based analysis for resource assessment of placer platinum in the offshore region of Goodnews Bay, Alaska. University of Alaska, Mining and Geological Engineering, Fairbanks
- Oommen, T., and Baise, L. G. in-press. Model Development and Validation for Intelligent Data Collection for Lateral Spread Displacements. *Journal of Computing in Civil Engineering*.
- Oommen, T., Misra, D., Prakash, A., Bandopadhyay, S., Naidu, S.A., and Kelly, J.J. accepted pending minor revisions. Multiple Regressive Pattern Recognition Technique: An Adapted Approach for Improved Georesource Estimation, *Natural Resources Research*.
- Paine, J., and Minty, B. 2005. Airborne Hydrogeophysics. In *Hydrogeophysics*, 333-357. Y. Rubin, and S. S. Hubbard, eds: Springer Netherlands.
- Pewe, T. L. 1952. Preliminary report of multiple glaciations in the Big Delta area, Alaska. *Geological Society of America Bulletin*. 63:1289
- Pewe, T. L. 1955. Middle Tanana Valley. In *Permafrost and Groundwater in Alaska*, 126-130. D. M. Hopkins, ed: U.S. Geological Survey Professional Paper.
- Reger, R. D., and Solie, D. N. 2008. Reconnaissance interpretation of permafrost, Alaska Highway corridor, Delta Junction to Dot Lake, Alaska. Alaska Division of Geological & Geophysical Surveys PIR 2008-3C. Fairbanks, Alaska.
- Reger, R. D., Stevens, D. S. P., and Solie, D. N. 2008. Surficial Geology of the Alaska Highway Corridor, Delta Junction to Dot Lake, Alaska. Alaska Division of Geological and Geophysical Surveys PIR 2008-3a. Fairbanks, Alaska.
- Rizzo, D. M., and Dougherty, D. E. 1994. Characterization of aquifer properties using artificial neural networks: Neural kriging. *Water Resour. Res.* 30(2):483-497.
- Sarle, W. S. 2002. Neural Network FAQ, part 2 of 7: Learning. Available at: <ftp://ftp.sas.com/pub/neural/FAQ.html>.
- Saskal, R. 2008. Paths to the pipeline. San Francisco, CA: The Bond Buyer. Available at: <http://cdn.bondbuyer.com/media/newspics/20080717LMBSOEMP-1-0718trend.jpg>.
- Schapp, M. G., and Leij, F. J. 1998. Database-related accuracy and uncertainty of pedotransfer functions. *Soil Science* 163(10):765-779.
- Scrivens, S. 2005. A comparison between helicopter and fixed-wing time-domain electromagnetic systems. Carleton University, Earth Sciences, Ottawa

- Serber, G. A. F. 1984. *Multivariate Observations*. John Wiley & Sons, Inc., Hoboken, NJ.
- Slater, L., and Lesmes, D. P. 2002. Electrical-hydraulic relationships observed for unconsolidated sediments. *Water Resour. Res.* 38(10):1213.
- Slater, W. H. 1976. Engineering Geology & Soils Report, Alaska Highway Realignment, MP 1378 to 1376.5, Project No. HHS-062-2(11). Alaska Department of Highways, Materials Division. Fairbanks, Alaska.
- Solie, D. N., and Burns, L. E. 2007. Alaska Highway Corridor Geology and Geophysics. In *Alaska Geosurvey News*, 6. Alaska Division of Geological and Geophysical Surveys.
- Swanson, D. K. 2009. Soil Survey of Gerstle River Area, Alaska. U.S. Department of Agriculture, Natural Resources Conservation Service. N. R. C. S. U.S. Department of Agriculture.
- Tamari, S., Wosten, J. H. M., and Ruiz-Suarez, J. C. 1996. Testing an Artificial Neural Network for Predicting Soil Hydraulic Conductivity. *Soil Sci Soc Am J* 60(6):1732-1741.
- Telford, W. M., Geldart, L. P., and Sheriff, R. E. 1990. *Applied Geophysics*. 2nd ed. Cambridge University Press, New York.
- Topp, G. C., Davis, J. L., and Annan, A. P. 1980. Electromagnetic determination of soil water content: Measurements in coaxial transmission lines. *Water Resources Research* 16(3):574--582.
- Twarakavi, N., Misra, D., and Bandopadhyay, S. 2006. Prediction of Arsenic in Bedrock Derived Stream Sediments at a Gold Mine Site Under Conditions of Sparse Data. *Natural Resources Research* 15(1):15-26.
- Twarakavi, N. K., Misra, D., and Bandopadhyay, S. 2005. Optimization Based Upscaling of Hydrogeological Properties for Flow and Transport Simulation. EOS Trans.
- U.S. Army Corps of Engineers. 1959. Evaluation data for liquid radioactive waste disposal SM-1A--Army Package Power Reactor. U.S Army Corps of Engineers, . Anchorage, Alaska.
- Valleau, N. C. 2000. HEM data Processing - A Practical Overview. *Exploration Geophysics* 31:584-594.
- Vapnik, V. 1995. *The Nature of Statistical Learning Theory*. Springer-Verlag, Newyork.
- Vokovic, M., and Soro, A. 1992. *Determination of Hydraulic Conductivity of Porous Media from Grain-Size Composition*. Water Resources Publications, Littleton, Colorado.

Wallner, R. M., Feulner, A. J., and Tisdell, F. E. 1961. Ground-water movement in the Fort Greely area. Geological Society of America Special Paper. Alaska (abs.).

Wilcox, D. E. 1980. Geohydrology of the Delta-Clearwater Area, Alaska. U.S. Geological Survey AK 78-109. Anchorage, AK.

Wohlberg, B., Tartakovsky, D. M., and Guadagnini, A. 2006. Subsurface characterization with support vector machines. *Geosciences and Remote Sensing, IEEE Transactions on* 44(1):47-57.

Woo, M.-k. 1986. Permafrost Hydrology in North America. *Atmosphere-Ocean* 24(3):201-234.

Zhdanov, M. S. 2009. *Geophysical Electromagnetic Theory and Methods*. Methods in Geochemistry and Geophysics. Elsevier, Amsterdam, The Netherlands.