Quantitative Integration of Multiple Near-Surface Geophysical Techniques for Improved Subsurface Imaging and Reducing Uncertainty in Discrete Anomaly Detection

A Dissertation Presented for the
Doctor of Philosophy
Degree
The University of Tennessee, Knoxville

Megan Estelle Carr
May 2013

DEDICATION

This dissertation is dedicated to Mrs. Phyllis Kroll, whose enthusiasm during my 8th grade Earth Science class and constant support throughout all the years afterwards has inspired and encouraged me to be the best student, teacher, and scientist possible. It was because of her that I took my first geology class when an accounting major; I have never looked back.

This dissertation is also dedicated to Ms. Marion Norton Carr Knox, my grandmother. Her devotion to continuous education and lifelong commitment to never settling on the status quo has driven me to do what makes me happy and achieve those things that most people only talk about.

ACKNOWLEDGEMENTS

I would like to thank my entire committee for providing the support, knowledge, open doors, and humor necessary to complete this dissertation. I would like to especially thank Greg Baker for bringing me over to the "dark side" of geophysics and providing me with many amazing opportunities to grow as a scientist, professor, and professional. His patience during my time at Tennessee has spoken volumes and I will be forever grateful for the life experiences along the way.

I would like to thank Michael Newman with UT's Statistical Consulting Center for his many hours of assistance and availability to discuss SAS coding. I also thank Gerald Ragghianti with UT's Newton HPC Program for his help in running my models, cutting down processing time by years.

I would like to thank the many undergraduate field assistants who gave up their time and energy to collect data used in this research. I thank my students who encouraged me to keep going with their kind words and positive attitudes.

I would like to thank my parents for their unquestioning support and unconditional love throughout all my years. Their lessons of sacrifice, honor, excellence, and perseverance in the face of adversity have been the cornerstones of this research reaching completion. And most importantly, I thank God for His countless blessings and the gifts bestowed upon me that have guided me throughout my educational journey.

ABSTRACT

Currently there is no systematic quantitative methodology in place for the integration of two or more coincident data sets collected using near-surface geophysical techniques. As the need for this type of methodology increases—particularly in the fields of archaeological prospecting, UXO detection, landmine detection, environmental site characterization/remediation monitoring, and forensics—a detailed and refined approach is necessary. The objective of this dissertation is to investigate quantitative techniques for integrating multi-tool near-surface geophysical data to improve subsurface imaging and reduce uncertainty in discrete anomaly detection. This objective is fulfilled by: (1) correlating multi-tool geophysical data with existing well-characterized "targets"; (2) developing methods for quantitatively merging different geophysical data sets; (3) implementing statistical tools within Statistical Analysis System (SAS) to evaluate the multiple integration methodologies; and (4) testing these new methods at several well-characterized sites with varied targets (i.e., case studies). Three geophysical techniques utilized in this research are: ground penetrating radar (GPR), electromagnetic (ground conductivity) methods (EM), and magnetic gradiometry. Computer simulations are developed to generate synthetic data with expected parameters such as heterogeneity of the subsurface, type of target, and spatial sampling. The synthetic data sets are integrated using the same methodologies employed on the case-study sites to (a) further develop the necessary quantitative assessment scheme, and (b) determine if these merged data sets do in fact yield improved results. A controlled setting within The University of Tennessee Geophysical Research Station permits the data (and associated anomalous bodies) to be spatially correlated with the locations of known targets. Error analysis is then conducted to guide any modifications to the data integration methodologies before transitioning to study sites of unknown subsurface features. Statistical analysis utilizing SAS is conducted to quantitatively evaluate the effectiveness of the data integration methodologies and determine if there are significant improvements in subsurface imaging, thus resulting in a reduction in the uncertainty of discrete anomaly detection.

TABLE OF CONTENTS

1. Introduction 1
    1.1 Motivation 2
    1.2 Hypothesis 3
    1.3 Objective and Goals 4
        1.3.1 Hypothesis 1 4
        1.3.2 Hypothesis 2 5
    1.4 Expected Outcomes 5
        1.4.1 Geoscience Curriculum Article 5
        1.4.2 Methodology and Statistics Focused Article 6
        1.4.3 Case Study Focused Article 7
    1.5 Broader Impacts 8

2. Background 10
    2.1 Previous Work 11
        2.1.1 Data Integration Issues 11
        2.1.2 Possible Data Integration Methods 12
        2.1.3 Approaches to Data Integration in the Past 14
            2.1.3.1 Interpretive Approach 14
            2.1.3.2 Computer Graphic 16
            2.1.3.3 Mathematical Transformation 19
            2.1.3.4 Statistical 22
    2.2 Geology of Site Location 25
        2.2.1 Geologic Overview 25
        2.2.2 The University of Tennessee Geophysical Research Station 26
    2.3 Geophysical Techniques 28
        2.3.1 Introduction 28
        2.3.2 Ground Penetrating Radar 28
        2.3.3 Ground Conductivity (EMI) 30
        2.3.4 Magnetic Gradiometry 34
    2.4 Data Acquisition 35
        2.4.1 Ground Penetrating Radar Data Acquisition 35
            2.4.1.1 Ground Penetrating Radar Equipment 35
            2.4.1.2 Acquisition Parameters 37
        2.4.2 EM-31 (Ground Conductivity) Data Acquisition 37
            2.4.2.1 EM-31 (Ground Conductivity) Equipment 37
            2.4.2.2 Acquisition Parameters 38
        2.4.3 Magnetic Gradiometry Data Acquisition 39
            2.4.3.1 Magnetic Gradiometry Equipment 39
            2.4.3.2 Acquisition Parameters 40
        2.4.4 Spatial Location 40

2.5 Data Processing 41
2.5.1 Ground Penetrating Radar 41
2.5.1.1 Introduction 41
2.5.1.2 Dewow 41
2.5.1.3 Migration 42
2.5.1.4 Velocity Conversion 43
2.5.1.5 Automatic Gain Control 43
2.5.2 Ground Conductivity (EMI) 44
2.5.2.1 Introduction 44
2.5.2.2 Smoothing 44
2.5.2.3 Destagger 45
2.5.3 Magnetic Gradiometry 46
2.5.3.1 Introduction 46
2.5.3.2 Clipping 46
2.5.3.3 Interpolate 47
2.5.3.4 Despike 48
2.5.3.5 Dedrift 48
2.5.3.6 Destagger 49
2.5.3.7 Deslope 49

3. Quantitative Integration of Geophysical Data for Enhancing Subsurface Features (Part 1): Methodology 50
Abstract 51
3.1 Introduction 52
3.2 Methods 55
3.2.1 Modeling of Data 55
3.2.1.1 GprMax 56
3.2.1.2 Geophysica 58
3.2.2 Initial Method for Testing 60
3.2.3 Revised Method for Testing 61
3.2.3.1 Creating a Model and Data Manipulation Scheme 61
3.2.3.2 Applying the Data Manipulation Scheme 68
3.2.4 Statistical Analysis 85
3.2.4.1 Logistical Regression 87
3.2.4.2 Proc GLM 88
3.2.4.3 Proc Frequency 89
3.2.4.4 Univariate 90
3.3 Results 90
3.4 Discussion 98
3.4.1 Significance of Variables 98
3.4.2 Quantitative Integration 99
3.4.3 Uncertainty Level Improvements 103
3.5 Conclusion 103

4. Quantitative Integration of Geophysical Data for Enhancing Subsurface Features (Part 2): Application of Methodology 106
Abstract 107
4.1 Introduction 109
4.1.1 Hypthesis 110
4.1.2 Field Site Description 111
4.2 Methods 116
4.2.1 Data Collection and Processing 116
4.2.1.1 Ground Penetrating Radar 116
4.2.1.2 EM-31 (Ground Conductivity) 117
4.2.1.3 Magnetic Gradiometry 119
4.2.2 Preparing Data for Analysis 120
4.2.3 Processing of Grids 126
4.2.4 Statistical Analysis 140
4.2.5 Mosaic of Data Grids 141
4.3 Results 142
4.4 Discussion 159
4.4.1 Data Resolution 159
4.4.2 Statistical Analysis 160
4.4.3 Data Visualization 162
4.4.4 Error Analysis 163
4.5 Conclusion 164

5. Summary and Conclusions 167
5.1 Summary 168
5.2 Objectives Met 168
5.3 Overall Impact 170
5.4 Future Work 171

References 173

Appendix 179

Vita 220

LIST OF TABLES

Table 1. Summation of Variables Tested for Synthetic Data Sets…….63

Table 2. Calculated optimal exponent for each geophysical technique and material type. This table was generated in conjunction with curves (*as shown in Figure 9*)……..…..70

Table 3. Representative example table to display the ranking of data manipulation schemes for the modeled stainless steel data. Ranking is based on highest accuracy for the cells [across the grid] that contain the target and are calculated to contain the target (1:1)……86

Table 4. Summary table for the minimum percentage accuracy that results from all data manipulation schemes across all material types and thresholds. The rankings are from all 50, 674 schemes being divided into groups, where 100 includes all schemes; 50 includes the top 25,337 schemes; and so forth…….91

Table 5. Summary table to display the statistical significance for the accuracy percentages (*as seen in Table 3*) that result from all data manipulation schemes across all material types and thresholds. The rankings are from all 50, 674 schemes being divided into groups, where 100 includes all schemes; 50 includes the top 25,337 schemes; and so forth……92

Table 6. Equations chosen to be applied to authentic data in the next phase of the research (i.e. control site). Each equation results in an accuracy of at least 94.21% with the "1:1" comparisons and 100% accuracy for the "0:0" comparisons…….93

Table 7. Summary table to show location and dimension of modeled iron sphere once data points are manipulated with various schemes……96

Table 8. Improvement in Uncertainty Level by Applying Data Integration Methodology …….97

Table 9. Target Descriptions for Field Site……..……115

Table 10. Dataset name as determined by geophysical technique, resolution, and depth of investigation……..…122

Table 11. Summary table showing what tables are integrated together to form the new *geoData* table to be used in evaluation of anomaly detection over the survey site…..…136

Table 12. Summary for percentage accuracy in detecting targets across the control site survey area defined by thresholds for manipulating the data sets by raising each normalized value by its optimal exponent prior to integration. Resolution type contains three depths: (a) 0.33m, (b) 0.66m, and (c) 1.00m. Grouping are for (1) No Shift in grid positioning, (2) Shift along the X axis (North/South), (3) Shift along the Y axis (East/West), (4) Difference in accuracy between groupings 2 and 3, (5) Difference in accuracy between groupings 1 and 3, (6) Difference in accuracy between groupings 1 and 2……143

Table 13. Summary for percentage accuracy in detecting targets across the control site survey area defined by thresholds for manipulating the data sets adding each normalized value together. Resolution type contains three depths: (a) 0.33m, (b) 0.66m, and (c) 1.00m. Grouping are for (1) No Shift in grid positioning, (2) Shift along the X axis (North/South), (3) Shift along the Y axis (East/West), (4) Difference in accuracy between groupings 2 and 3, (5) Difference in accuracy between groupings 1 and 3, (6) Difference in accuracy between groupings 1 and 2……...…144

Table 14. Summary for differences between data manipulation schemes [addition and optimal exponent] in percentage accuracy in detecting targets across the control site survey area defined by thresholds. Resolution type contains three depths: (a) 0.33m, (b) 0.66m, and (c) 1.00m. Grouping are for (1) No Shift in grid positioning, (2) Shift along the X axis (North/South), (3) Shift along the Y axis (East/West), (4) Difference in accuracy between groupings 2 and 3, (5) Difference in accuracy between groupings 1 and 3, (6) Difference in accuracy between groupings 1 and 2……...145

Table 15. Summary of accuracy in detecting targets across the control site survey area defined by data manipulation schemes (*see Tables 10 and 11*) when combining all depth slices of GPR together. Accuracies are also defined by type of shifting applied to grids. Example is given for Resolution Type 1 at a 95% threshold……...146

Table 16. Target Descriptions for TINGS Field Site……..…..…191

Table 17. Examples of Assignments Given to Students for TINGS Course……...…195

Table 18. Outline of Course Topics and Industry Partners (Instructors)…….....198

LIST OF FIGURES

Figure 1. Location of the University of Tennessee Agricultural Extension Center, which contains the study site, The University of Tennessee Geophysical Research Station. The University of Tennessee, as seen in the Northeast corner of the expanded map in located in Knoxville, Tennessee. The red square located in the zoomed in map shows the location of control site…….……...…….27

Figure 2. Basic GPR schematic. The x-axis is oriented parallel to the long axes of the antennas. ε_0, ε_1, ε_2 are the relative permittivities and σ_1, σ_2 the conductivities of the respective media. R^{TE}_{ab} is the reflection coefficient of the broadside (TE) mode GPR waves incident at the boundary between the a and b media, where a and b can be 0 and 1 or 1 and 2 (van der Kruk et al. 2009)…….31

Figure 3. Schematic diagram of basic EM-31 principles (Pettersson and Nobes 2003)…33

Figure 4. A synthetic example with strong remnant magnetization. (a) Dipping body viewed from the southwest direction. It has a total magnetization direction which deviates significantly from the inducing field direction. (b) The total-field anomaly. (Dannemiller and Li 2006)……...…….…….……..36

Figure 5. Output from GprMax program. Modeled object is a concrete sphere with a diameter of 50cm and buried at a depth of 2m……..57

Figure 6. Geophysica program GUI with parameters selected, as well as the resulting data……59

Figure 7. Representative outputs of modeled data from initial testing. Grouping 1 shows the modeled (A) GPR data, (B) Magnetic data, and (C) EM-31 Ground Conductivity data with a 10cm sample spacing. Grouping 2 shows the modeled data [from Grouping 1] with a 10cm sample spacing with (A) Magnetic and GPR data combined, (B) EM-31 and Ground Conductivity data combined, and (C) Magnetic and EM-31 Ground Conductivity data combined. Grouping 3 shows all three modeled data sets combined with a (A) 10cm sample spacing, (B) 25cm sample spacing, and (C) 50cm sample spacing……...62

Figure 8. Representative sample of curves extracted from model outputs to determine the representative model used for each material type. Curves displayed are for all concrete curves used to find median slope (top) and an isolated curve (bottom)……...66

Figure 9. Representative curve indicating what the optimal exponent is for stainless steel magnetic data. The normalized data values are raised to some exponent; as the slope [of the newly calculated curves] increases, the rate at which the slope increases is decreasing. Once that rate reaches some asymptote, it is assumed that value is the optimal exponent to apply to the data for future manipulation schemes……69

Figure 10. Representative example of the *geoData* table, displaying the EM, GPR, and MAG data associated with each data point within the grid……...71

Figure 11. Representative example of the *geoData TG* table. This table incorporates the different data manipulation schemes applied to the EM, GPR, and MAG data for each data point within the grids……..73

Figure 12. Representative example of the *geoExpression* table. Querries within this table creates 50,653 expressions for the multiplication data manipulation schemes…….74

Figure 13. Representative example of the *geoData Max* query table. This table isolates the maximum value calculated across each grid and data manipulation scheme, shown as the EMGMx value……..76

Figure 14a. Representative example of *geoData Target 01* table with the EMGx Field completed. The "1" on the top line shows that this is the point within the grid that is expected to be directly over the buried target……78

Figure 14b. Representative example of the *geoData Target 01* query – EMGx percentage level will be transformed into a "0" (not over target) or "1" (over the target) according to various calculated thresholds…….……80

Figure 14c. Representative example of *geoData Target 01* query where the transformed "0" and "1" cells (*seen in Figure 14b*) are compared for accuracy against the actual target locations of the modeled object…….82

Figure 15. Representative example of the *geoData Accuracy Summary* query. This table displays the accuracy (by frequency of occurrence and percentage) of each data manipulation scheme according to threshold…….83

Figure 16. Representative example of raw modeled data (iron sphere, buried at 175 cm and a diameter of 100cm). Groupings show the data as a (1A) Magnetic contour map; (1B) Magnetic surface map; (2A) EM-31 Ground Conductivity contour map; (2B) EM-31 Ground Conductivity surface map; (3A) GPR contour map; and (3B) GPR surface map. Data was mapped using Surfer……..94

Figure 17. Representative example of modeled iron datasets combined utilizing various data manipulation schemes. Groupings show (1A) normalized values for Mag, EM, and GPR added together; (2A) Normalized values for all datasets combined with a scheme ranked at 50% accuracy; (3A) Normalized values for all datasets combined with a scheme ranked as number 1 of all 50,674 schemes; (1B, 2B, and 3B) give the specific location of data once cells are transformed to binary format at a 85% threshold; (1C, 2C, and 3C) give the specific location of data once cells are transformed to binary format at a 95% threshold……95

Figure 18. (a) Map showing location of field site (star indicates Knoxville, TN); (b) Close-up view of Knoxville, TN with yellow box designating The University of Tennessee Geophysical Research Station and star showing Knoxville, TN; (c) Close up view of yellow box from (b), with smaller yellow box indicating general location of data collection……..……112

Figure 19. Map displaying locations of buried targets. Grid is measured in meters. GPR and EM-31 surveys were conducted over the entire 40mx50m grid. The red box is the areal extent of the Magnetic Gradiometry survey. The shaded blue section is a metal fenced area where the data was removed from all processing. Target descriptions for each number on the map is given in Table 8……....114

Figure 20. Control site at The University of Tennessee Geophysical Research Station. The survey area is divided into 5m x 5m grids for data manipulation processes and statistical analysis…….123

Figure 21. Control site at The University of Tennessee Geophysical Research Station. The survey area is divided into 5m x 5m grids for data manipulation processes and statistical analysis and shifted East/West 2.5m to minimize the edge effect……...…124

Figure 22. Control site at The University of Tennessee Geophysical Research Station. The survey area is divided into 5m x 5m grids for data manipulation processes and statistical analysis and shifted North/South 2.5m to minimize the edge effect……...125

Figure 23. Representative example of datasets that are integrated together with similar resolution (i.e. sample spacing). This query incorporates a common GPS coordinate, designated as an XY location on each survey area……..127

Figure 24. The *geoData Expression* query; these equations were chosen as representative from the statistical results from the synthetic data in Part 1 of this article……..128

Figure 25. Representative example of the *geoData EMG* database. This query example is created to process the *geoData Expressions* and create the EMG data for each table…129

Figure 26. Representative example of a *geoData Resolution* table, identifying the cells that contain buried targets in the survey area. Example is of Resolution Type 1……131

Figure 27. Representative example of the *geoData Grid* table. Example table represents the Grid (No Shift) table……..132

Figure 28. The *geoData Max Values* query, displaying the 27 Max Value tables for each expression of each grid for the three different Grid Shift tables……..134

Figure 29. Representative example of the geoData EMG query. This example is one of 27 listed in Table 10 (i.e., Type 1 Resolution at 0.33m)……...……137

Figure 30. Representative example of the *geoData EMG Accuracy Summary*. Example query represents the combination of the EMG tables containing the three types of data shifts and three levels of resolutions……139

Figure 31 A. Original (raw) magnetic gradiometry data collected over the study site at the University of Tennessee Geophysical Research Station. Data were collected over a portion of the greater 50 meter by 40 meter area. Data in the top left quartile is not valid. Peaks are considered to be probable locations of buried targets…….147

Figure 31 B. Interpretation of original (raw) magnetic gradiometry data collected over the study site at the University of Tennessee Geophysical Research Station. Yellow circles indicate locations of buried targets that also show a peak in the data. This interpretation allows for the calculation of Type I and Type II error presence……..148

Figure 32 A. Original (raw) EM-21 ground conductivity data collected over the study site at the University of Tennessee Geophysical Research Station. Data were collected over the entire 50 meter by 40 meter area. Depressions in the surface are considered to be probable locations of buried targets…….149

Figure 32 B. Interpretation of original (raw) EM-31 ground conductivity data collected over the study site at the University of Tennessee Geophysical Research Station. Yellow circles indicate locations of buried targets that also show a peak in the data. This interpretation allows for the calculation of Type I and Type II error presence……...…150

Figure 33 A. Original (raw) ground penetrating radar data collected over the study site at the University of Tennessee Geophysical Research Station. Data were collected over the entire 50 meter by 40 meter area. Peaks are considered to be probable locations of buried targets……...151

Figure 33 B. Interpretation of original (raw) ground penetrating radar data collected over the study site at the University of Tennessee Geophysical Research Station. Yellow circles indicate locations of buried targets that also show a peak in the data. This interpretation allows for the calculation of Type I and Type II error presence……...152

Figure 34 A. Representative example of all datasets integrated together. Data have been manipulated by raising each normalized value to the optimal exponent. This example is at a 1.00 meter depth slice, Resolution Type 1, and has not been shifted. Peaks are considered to be probable locations of buried targets……..153

Figure 34 B. Interpretation the integrated dataset from all three techniques collected over the study site at the University of Tennessee Geophysical Research Station. Yellow circles indicate locations of buried targets that also show a peak in the data. Red circles indicate the locations of targets that were not detected by any of the techniques. This interpretation allows for the calculation of Type I and Type II error presence……...154

Figure 35. Representative example of all datasets integrated together. Data have been manipulated by raising each normalized value to the optimal exponent. This example is at a 1.00 meter depth slice, Resolution Type 1, and has shifted along the X axis (North/South) by 2.5 meters. Peaks are considered to be probable locations of buried targets……..……155

Figure 36. Representative example of all datasets integrated together. Data have been manipulated by raising each normalized value to the optimal exponent. This example is at a 1.00 meter depth slice, Resolution Type 1, and has shifted along the Y axis (East/West) by 2.5 meters. Peaks are considered to be probable locations of buried targets……..156

Figure 37. Representative example of all datasets integrated together. Data have been manipulated by raising each normalized value to the optimal exponent. This example combines all depth slices (0.33m, 0.66m, and 1.00m), Resolution Type 1, and has shifted along the X axis (North/South) by 2.5 meters. Peaks are considered to be probable locations of buried targets……157

Figure 38. Representative example of all datasets integrated together and plotted as a binary response of presence (1) or absence (0) of target for each data point in the grid. Data have been manipulated by raising each normalized value to the optimal exponent. This example combines all depth slices (0.33m, 0.66m, and 1.00m), Resolution Type 1, and has shifted along the Y axis (East/West) by 2.5 meters. Peaks are considered to be probable locations of buried targets. Blue peaks are targets detected by two or three techniques. Block shapes give a general sense of shape, size, and orientation of the buried objects……..158

Figure 39. (a) Map showing location of where TINGS course is offered (star indicates Knoxville, TN); (b) Close-up view of Knoxville, TN with yellow box designating The University of Tennessee Geophysical Research Station and star showing Knoxville, TN; (c) Close up view of yellow box from (b), with smaller yellow box indicating general location of where field work for TINGS course is conducted…….188

Figure 40. Map displaying locations of buried targets. Grid is measured in meters. Green box is the areal extent of the ground penetrating radar, magnetic, EM-31, and EM-61 surveys. Red box is the areal extent of the 3-D electrical resistivity survey. The blue line is the location of the seismic survey, taken outside the field area for teaching purposes only (data is not correlated with other techniques). Target descriptions for each number on the map is given in Table 15; stars along boundary of survey grid denote markers for the survey and are not targets……...190

Figure 41. Example of a student's data map for conductivity. North is to the right, parallel to the x-axis. Maps on the left are for the inphase component, with maps on the right for the quadrature component. Units are in mS/m, with each tick mark denoting 5 m of distance across the grid. North is towards the right of the images. Target locations in are interpreted as the depressions in the lower maps. The anomaly in the southwest corner of the lower maps is related to a metal fence and not the location of a target…….. ……..204

Figure 42. Example of student ground penetrating radar data interpretation. White circles are non-point source targets. Yellow circles are point source targets. Pink circles are monitoring wells. The "highlighted" area in the northeast corner of the grid suggests a high soil contrast, perhaps due to saturated soils (standing water) compared to the other sections of the grid. The blanked data region in the southwest corner of the lower maps is related to a metal fence and not the location of a target. ……205

Figure 43. Example of student interpretation of magnetic data. North is up, parallel to the y-axis. The top map is from a cesium vapor gradiometer and displays data from the top sensor only. The bottom map is from the same instrumentation, but gives a gradient reading. Red circles are where targets are thought to have remnant magnetism. Blue circles indicate induced magnetic bodies. The anomaly in the southwest corner of the lower maps is related to a metal fence and not the location of a target……...206

Figure 44. Step-by-step display of student's final project: (A) The location of the field site using Google Earth; (B) All data sets incorporated together, each as its own layer that can be selected separately or in groups to correlate the locations of anomalies in the data to better discriminate the targets; (C) Magnetic data as a layer, complete with GPS coordinates, a scale bar, north arrow, and scale for units measured; and (D) The interpreted location of the buried targets (red-metallic objects, green-nonmetallic objects, blue-well heads). …….207

Figure 45. Grade trends for students for each year TINGS course was offered. Data points are the average grade among all students for each assignment, with assignment 9 being the final project; assignments increased in quantitative complexity throughout the course. The overall increase in grades throughout the term despite the difficulty level increasing give evidence that the pedigogy behind the TINGS course is effective……..208

Figure 46. Students in “action” during the TINGS course. From the top-right image and moving clockwise: processing ground penetrating radar data; receiving instruction on how to set up an electrical resistivity survey; learning how to operate the EM-31; Surveying with the magnetometer. Photographs courtesy of Noah McDougall…….214

1. INTRODUCTION

1.1 Motivation

Currently there are no systematic quantitative methodologies developed for the integration of two or more geophysical data sets collected using near-surface geophysical techniques for discrete anomaly detection. Recently, Urs Böniger and Jens Tronicke (2010) proposed results of an integrative analysis and interpretation of different data sets that combines geophysical instruments with modern topographic data using a tracking total station (TTS). However, their results are limited to composite images of the combined data sets and illustrated with various color schemes, which, when interpreting the data, still consists of a qualitative assessment and does not quantify the data to the extent presented in this dissertation. The geophysical techniques utilized in the research presented here include ground penetrating radar (GPR), magnetic gradiometry, and other magnetic/electromagnetic (EM) methods.

For an individual technique being employed in a survey to be appropriate (i.e., useful for target detection/discrimination), there is a dependence on a difference in physical properties between the target and the surrounding material (a.k.a. "background") in terms of the subsurface characteristic to which the technique responds (e.g., dielectric permittivity, magnetic susceptibility, and electrical conductivity, respectively); if there is no difference in the physical property of the target vis–a-vis the surrounding material the target will not be detected. With this in mind, when a geophysical investigation is conducted over a region containing targets that are unknown—and only one technique is used—it is possible to miss certain types of targets completely. Therefore, integrating

multiple techniques into one unified data set can be used to more accurately identify and discriminate characteristics of targets with a greater degree of certainty.

Interpreting multiple geophysical data sets commonly involves a qualitative correlation of different geophysical data sets. This research seeks a more refined, quantitative approach to combining data sets. The underlying concepts of this project are that single geophysical methods are typically not able to detect all discrete target types, and that utilizing multiple techniques—and the integration of multiple technique data—should produce significant improvements in data quality and target detection.

1.2 General Hypotheses

The presented statistical and quantitative approach will aid in accomplishing the ultimate goal of the research, which is to have a quantitative assessment of data integration methodologies utilizing multiple near-surface geophysical techniques for discrete anomaly detection. The subsequent methodology is designed to test the following hypotheses:

1) Certain targets, given multiple variables and parameters, will be detected with a greater degree of certainty than others when a specified combination of processing and merging of disparate but coincident data is implemented.
2) Integration of two or more geophysical techniques will result in an improved subsurface image and reduce uncertainty in discrete anomaly detection.

1.3 Objective and Goals

The primary objective of this research is to improve success rates as defined through data quality and visualization techniques within geophysical surveys for discrete anomaly detection (e.g. locating clandestine underground tunnels, locating buried objects, mapping historical features). Each hypothesis tested will incorporate different goals to meet this objective.

1.3.1 Hypothesis 1

This hypothesis will be tested via developed and analyzed computer simulations in order to satisfy the following goals:

- Develop a comprehensive model for creating typical signal responses for various materials of objects found at The University of Tennessee's Geophysical Research Station
- Create synthetic data for each geophysical techniques utilized in this study
- Determine how resolution of data and data sampling heterogeneities affects integration of data sets and the resulting ability to discriminate targets with a higher degree of certainty
- Determine which variables involved with a geophysical survey are most significant in the discrimination of targets (e.g., target depth, target size, composition of target)

- Complete a statistical analysis to evaluate the effectiveness of the methodologies to quantitatively merge different geophysical data sets (e.g., addition, multiplication, exponential)

1.3.2 Hypothesis 2

This hypothesis will be tested via case studies to satisfy the following goals:

- Correlate geophysical data with known discrete "targets" by utilizing an integration of multiple geophysical techniques
- Complete a statistical analysis to evaluate the effectiveness of the data integration methodology and applications using authentic data
- Test the methods developed using synthetic data by application to various case studies of completely unknown discrete targets

1.4 Expected Outcomes

1.4.1 Geoscience Curriculum Article

Although not focused specifically on the scientific advances associated with data integration methodology, the first planned manuscript will be used to discusses the need for pedagogical developments that emphasize enhancing student quantitative skills and knowledge of how to carry out field work effectively—needs which are increasingly apparent in today's job market. The Tennessee Intensive Near-surface Geophysics Study (TINGS) program introduces multiple near-surface geophysical techniques and allows the students to (1) become familiar with the theory behind each technique, (2) gain

experience operating the geophysical equipment, and (3) be trained in the software packages specific to each technique by processing their own data. This manuscript-ready section presents the framework of The University of Tennessee's contribution towards meeting the aforementioned industry needs. Individual students, as well as the overall effectiveness of the program, are assessed by means of a comprehensive final project where all associated data sets are correlated together to discriminate types of subsurface features and targets that are present at the experimental field site. Emphasis is placed on proper survey design and working in a team environment to implement the plans successfully. Additionally, the types of errors associated with geophysical surveys are discussed, leading to an understanding of the importance in the development of a quantitative data integration methodology for improving subsurface imaging and reducing uncertainty in discrete anomaly detection. Some of the data collected during the course will be used in the later portions of this research. The manuscript section is found in the Appendix of this dissertation.

1.4.2 Methodology and Statistics Focused Article

The second manuscript-ready section addresses the hypothesis that certain targets, given multiple variables and parameters, can be detected with greater degree of certainty than others when a specified combination of processing and merging of data is implemented. Essentially, this section will serve as an explanation to the dissertation's methodology in how data sets are integrated and the statistical tests associated with each merged data set to satisfy the objective of improving data quality and visualization techniques within geophysical surveys for discrete anomaly detection.

Initial statistical analysis methods will be developed on the synthetic data. Simulated geophysical data will include magnetics and electromagnetics (using the *Geophysica* program in MATLAB, developed by Alan Witten) and GPR (from Sensors of Software's *GPR Max* GUIs, originally developed by Antonis Gainnopoulos). Data integration techniques are to be given in full detail, as well as statistics (in both data preparation and in data integration), that is the quantitative assessment of the integration methodologies. Statistics utilizing SAS demonstrate the merging techniques that are best suited for various scenarios (type of target, geologic setting, etc.), indicate how significant the integration of data sets is in discrete anomaly detection, and elucidate how the uncertainty level in anomaly detection changes through implementing the proposed methodology.

1.4.3 Case Study Focused Article

The third manuscript-ready section of this dissertation is used to test the hypothesis that the integration of two or more geophysical techniques will result in an improved subsurface image and reduce uncertainty in discrete anomaly detection. Essentially, the section will use the TINGS data (described in Section 1.4.1 and the Appendix) and review the results (using statistical techniques described in Section 1.4.2 and covered in Section 3) to assess anomaly locations. The chapter will summarize basic statistics of utilizing different techniques to detect various targets. Additionally, there will be a discussion concerning the number of each target type that is detected with the use of all three techniques, why certain targets might not have been detected, and the occurrence of Type I/II errors in general. Once a qualitative assessment is performed, the data are

subject to the methodology set up in the second manuscript section (covered in Section 3) as a case study using authentic data in a controlled setting. An introduction to Cherokee Farm and utilizing the developed data integration methodology in an area of completely unknown targets will also be given to direct future works.

1.5 Broader Impacts

While this research was originally designed to enhance archaeological geophysical surveys by incorporating improved multi-tool geophysics, the resulting methodology for quantitatively merging different types of geophysical data together shows potential for utilization in several additional areas of interest. These may include—but are not limited to—environmental site characterization/monitoring (i.e. contaminant transport, groundwater studies), UXO detection, clandestine underground tunnel detection (national security), locating discrete stratigraphic features of the subsurface (e.g. localized geologic variability such as faulting), and mining/exploration (e.g. minerals and natural resources location). The development of methods for quantitatively merging different types of geophysical data allow for the enhancement of structures and features in the data through signal-to-noise (S/N) enhancement of features detectible through more than one technique. Additionally, by improving visualization methods of the data the interpretations are seen more clearly and in application to other studies may be more convincing to the scientist conducting the investigation, helping to quickly and accurately meet the objectives of the individual project. Of particular interest for this dissertation, the data integration methodologies give archaeologists additional tools in their planning

and choosing of locations and methods of excavations, saving project managers valuable time, money, and/or other resources. The methodologies developed through this dissertation satisfy the ever growing need within the private sector and scientific community for a powerful time- and cost-effective approach for integrative analysis of multiple geophysical data sets.

2. BACKGROUND

2.1 Previous Work

2.1.1 Data Integration Issues

The remote sensing community—in the traditionally-used connotation of satellite and airborne imagery—has been integrating multi-technique geophysical data extensively and successfully for nearly two decades to better discriminate targets, such as, mineral deposits and specific types of vegetative land cover (Ansmann et. al., 2010; Metternicht et. al., 2003; and Mishchenko et. al., 2007). Currently, however, there is no purely quantitative methodology in place for the integration of two or more geophysical data sets collected using near-surface geophysical techniques such as ground penetrating radar (GPR), Magnetic Gradiometry, and EM methods. Most traditional near-surface land-based data integration studies involve one geophysical technique with additional data such as boreholes (e.g., Hornby 2007, Ferré 2003) or general geological data such as soil surveys and other maps (e.g., Galicia 2001, Rahman et a. 2008, Allen et al. 2008).

Colombo and De Stefano (2007) state that when modeling an integration of techniques, conversions of parameters from one geophysical domain to another have traditionally been performed rigidly by means of empirical functions. This occurs by modeling or inverting the separate geophysical domains followed by parameter transformations. While their research mainly deals in techniques not utilized in this project (seismic, magnetotellurics, gravity), Colombo and De Stefano (2007) do show evidence of the importance in setting up appropriate parameters in both data acquisition and processing to ensure the highest degree of certainty in the integrated data models.

Data sampling heterogeneities inherent in geophysical databases have been identified as the main source of data integration problems (e.g., Stock and Pullar 1999*).* These include (1) syntactic heterogeneity, that stems from the use of different data models to represent database elements (Bishr 1998); (2) schematic heterogeneity, that results from different classification schemes employed in the component databases or structuring of database elements in component databases (Kim and Seo 1991); schematic heterogeneities also result from different definitions of semantically similar entities, missing attributes, and different representations for equivalent data; and (3) semantic heterogeneity, that occurs when there is a disagreement about the meaning, interpretation or intended use of the same or related data. Semantic heterogeneity results from the different categorizations employed by individuals when conceptualizing real world objects. Such categorizations differ among individuals depending on education, experience, and theoretical assumptions (Stock and Pullar 1999). Such semantic heterogeneities have been identified as the main cause of data sharing problems and are the most difficult to reconcile (e.g., Allen et al. 2008).

2.1.2 Possible Data Integration Methods

The more quantitative integration of two or more geophysical data sets has been explored by K.L. Kvamme (2007). His research outlines some of the advances in the management, portrayal, and interpretation of subsurface data through the use of geophysical instruments and the computer methods utilized in the display of those data sets. Furthermore, Kvamme (2007) explains the versatility that geophysics brings in the

realm of archaeological surveys and incorporates the idea of "data fusion" under the pretense that combined information—or use of more than one data set—will lead to more insight than simply using only one type of data to interpret subsurface features. This chapter is used to emphasize how crucial it is to survey an area with multiple techniques to more confidently interpret the features in the subsurface and the means in which the data is merged will prove important in that interpretation.

In order to establish patterns in anomalous behavior over the study site and avoid possible misinterpretations due to incomplete surveying, the importance of large-area survey coverage is stressed. Using multiple techniques and surveying larger areas are crucial to the accuracy of the data collected. Kvamme (2007) does not offer any innovative methodology or extension of what could be found in the literature, but does an excellent job of summarizing four possible data fusion methods and formulating a general plan that a geoscientist could incorporate into their project designs, bringing up points and rules of thumb that should not be overlooked. However, the methods employed throughout his work are qualitative in nature, which leaves a lot of room for misinterpretation of the data. There is some ambiguity in Kvamme's approach and until a more quantitative method is developed, there will always be a certain amount of error. Kvamme does not adequately discuss this issue, despite pointing out (in regard to data integration) "…because of limited prior work, there has been little effort to examine systematically and organize its legitimate domains and lines of inquiry." It would be expected to find at least one example in the literature of this being done to indicate any advancement in this field of study; however, it is encouraging to see Kvamme mention

this point, as this is the problem that my research will be solving. Understanding the area in which one is completing a geophysical survey (in both geology and anthropologic extent) is crucial in setting up an appropriate survey design and will help guide interpretation of the data correctly.

2.1.3 Approaches to Data Integration in the Past

Successful interpretation of geophysical data depends upon the careful and experienced eye of the interpreter. Knowledge of the surrounding area in a geological context is essential, as well as knowing how anomalies will appear in the data for any given type of geophysical technique. When combining (i.e. overlaying and/or comparing) several data sets, that knowledge becomes even more important as variables are added to the total data available because each type of technique responds uniquely to characteristics of the subsurface—magnetometry to soil magnetic susceptibility changes, electrical resistivity and electromagnetic (EM) induction to conductivity changes, and ground penetrating radar (GPR) to changes in dielectric permittivity contrasts.

2.1.3.1 Interpretive Approach

The aforementioned approach at integrating data is perhaps the simplest and most widely utilized method, as it does not require much technological effort beyond the data pre-processing stage. Geophysical findings within this phase of interpretation are traditionally subjective and incorporate deductive reasoning in order to draw conclusions about the subsurface features and locations of desired targets. Another advantage of the

method described in Section 2.1.3 includes the ability of the interpreter to incorporate an isolation technique to visualize and interpret data.

The isolation technique is defined as the determination of a set of parameters that will separate an event from its surroundings (Sheffield et al 2000). Unlike recognition of possible anomalies in a data map, isolation involves the use of viewing techniques on an existing data set. Using isolation techniques separately and in combination allows important events to become visible, particularly when these data maps from individual data utilize the isolation technique and highlight the significant anomalies; if the anomalies are matched together in the same local, researchers can then assume that location holds an important feature, worth investigating further. Additionally, it is advantageous to use the interpretive approach, as it is easy to incorporate with a GIS software package for visualization of the anomaly locations and surrounding environment. Each geophysical technique can be integrated separately as a different layer within the GIS program, as well as layers depicting geology and topography (Hill 2008) to complement the geophysical data and interpretation of the subsurface.

There are, however, disadvantages to this method, the first being that it is a qualitative approach to interpreting the data that may lead to various levels of error depending on the individual conducting the interpretation. As this method does not typically invoke the aid of computers, there is a manual tracing of features of interest (anomalies) that can be time consuming and tedious work. Should more than one interpreter complete this step, as the maps are compared to each other (side by side or as an overlay) some inconsistencies may emerge. For instance, if Geophysicist A makes her

interpretation of a GPR data set, it is possible that some less pronounced anomalies are overlooked, or perhaps interpreted as more significant to Geophysicist B. This discrepancy could become a greater problem if the region being studied is too large for one person to interpret and must be divided among different people, each with her own region of the area to complete processing and interpretation of all the techniques (EM, GPR, resistivity, etc.). A more systematic approach would be to have one individual evaluate the entire region for just one technique to avoid the aforementioned problem of anomaly classification differences. However, the interpretive approach is still extremely subjective and dependent on the experience of the interpreter, which is its biggest drawback.

2.1.3.2 Computer Graphic

The computer graphic (CG) method of integrating geophysical data has been used with increasing frequency in the last three decades, particularly with computer-aided interpretation following single-channel (temporal filtering, gain control, etc.) and multi-channel (migration, velocity analysis, etc.) processing steps (e.g., Kreisberg et al 1991). This is the first step towards reaching a quantitative means of analyzing geophysical data, although at this point it is still qualitative in nature. Two-dimensional overlays are created in a more sophisticated manner than previously described with the use of CG.

In cases where several geophysical techniques are utilized in the same area, it is relatively easy to have one data set displayed as a gray-scale image, another with isoline contours, another with shaded relief mapping, etc.; this will enable each type of data to be

displayed separately while at the same time exhibiting the relationships between each different data set when they are overlain. The negative aspect of the method, however, is that there can be a huge volume of data from different geophysical techniques for a project, each with its own set of interpreted horizons and faults,, and all of which can become hopelessly entangled unless an interpreter comes prepared with a well-conceived data management scheme for the project (e.g., Herron 2001).

Translucent overlays of varying opacity are another version of the CG method, but again, there is a need to be careful in the format of display to minimize misinterpretation of the primary features within the subsurface. The volume of data (and number of data sets) and the variety of formats often result in an overwhelming sea of details, through which the researcher must sift. Additionally, the advancement in the field of CG programs and the ability to visualize data in many different formats typically leads the processor to create images that are too "busy" and ultimately too difficult to differentiate. A CG needs to be simple and straightforward to prevent possible misinterpretations. The utilization of 3D modeling techniques as visualization tools offers an approach that allows the geologist to "see the forest beyond the trees," and develops ideas to ensure the full extraction of resources without adding to the existing environmental footprint of where his project is located (e.g., Kirkham et al 2003).

Red-green-blue (RGB) color CG composites—a sstandard for displaying satellite imagery for decades (e.g., Schowengerdt 1997)—have been adopted by the geophysical community to aid in delineating various features within the subsurface based on parameters set in place to classify anomalies. The power of the RGB CG method blends

as an intuitive visualization tool whereby the richness of structure and relative ease by which objects can be visually separated is very compelling. The next natural step in an interpretation workflow is to extract these structures as discrete anomalies.

Extracting accurate geological information from such a multi-attribute data space means addressing the issues of color-noise suppression and color volume based object delineation (ala Henderson et al 2008). For this reason, the use of RGB CG color composites can be a double edge sword when it comes to visualization of the geophysical data. A benefit of this tool, however, is that color-based co-rendering techniques work well when the individual data sources are naturally correlated to some degree, as is the case with multispectral satellite imagery or MRI data in medical imaging (Henderson et al 2008). However, the technique defines a linear mapping function and the user must choose whether to scale all three inputs equally or whether to utilize the maximum available dynamic range of each individual component. This linear mapping can lead to substantial changes in the appearance of the generated RGB CG image, as the former will preserve the relationship between absolute values in each component, while the latter masks this relationship. Clearly this must be taken into account when "interpreting" RGB blended volumes.

In the case of Kvamme's (2007) Army City data, he chooses to have each of the colors in the RGB CG scheme represent a different geophysical technique (EM resistivity—red; EM conductivity—green; and magnetic susceptibility—blue), resulting in different features being presented in the composite data map (image). The use of the RGB appears to be appropriate in this case, but the additional information from the

remaining data sets are put in place as overlays of other colors, which creates more confusion for the interpreter of that particular set of data. The use of computer graphics to aid in data integration can be extremely useful when done correctly. Simplicity assures that the important pieces of data are not overshadowed by a "busy" graphic displaying too much unnecessary data.

2.1.3.3 Mathematical Transformation

Mathematical transformations involve a less qualitative approach compared to the interpretive and computer graphic methods for integrating geophysical data sets. The two divisions of this method are those that utilize binary data and those that employ continuous measurements, both of which require pre-processing of the data to perform optimally (Kvamme 2007). One advantage in the use of binary data is that the presence of an anomaly is classified as a "1" and the absence of an anomaly is classified as a "0." This is helpful when mapping a large data set and trying to determine trends in the area; it also allows the interpreter to see clusters of anomalies and determine (with the use of geologic maps and knowledge of archaeological history of the area) what features might be present in the area. However, because this is a black and white case of anomaly detection, characteristics of the anomaly (intensity, shape, etc.) are not differentiated. Another disadvantage of the binary system is that relatively "quiet" datasets where the signal-to-noise ratio is low can be overlooked and the area in question will be assigned a "0" when it should have a "1"; these false negatives can have significant implications on interpretation of the study area.

Broad regional trends arising from soil changes or underlying geology frequently can mask discrete anomalies, meaning that the application of various filters are necessary to remove these trends prior to binary classification (Kvamme 2007). When integrating data, particularly in the binary realm, the occurrence of Type I and Type II errors are an important factor to consider. As a side note of reminder, Type I errors are when an target is thought to be detected but one doesn't exist, while Type II error are when no anomaly is detected but one exists.

The Boolean Union, used when data is integrated, shows a "1" when an anomaly is detected from any of the data sets being merged; this may be a problem if the detected anomaly isn't a true anomaly (Type I error) resulting in the study area as a whole now becoming misrepresented, leading to misinterpretation. The Boolean Intersection (BI) is a better approach, as it only registers a "1" if all the data sets detect an anomaly; this prevents Type I errors in the final interpretation. However, BI may lead to more Type II errors if only one of the data sets of the total six sets (as in the case of Army City, Kvamme 2007) has not been pre-processed appropriately.

The Binary Sum (BS) method combines the advantages of the RGB and BI CG methods, resulting in a ranking of anomaly systems that creates a map of varying confidence levels in the detection of anomalies. The BS method will yield a larger value for areas that have an increased number of the different geophysical techniques detecting presences of anomalies (e.g., EM+GPR+Res+Conductivity=4), allowing an interpreter to decide which areas are of more or less interest for excavation purposes (in the case of Army City or other archaeological sites). Another advantage of the BS method is that

when compared to graphical images, the user can determine the combination of techniques that might be useful in detecting various features within the subsurface. Additionally, the BS method can be taken one step further to set up a threshold whereby the interpreter may only be interested in areas of the survey in which 2-4 geophysical techniques were able to be used to detected anomalies, and only display those areas.

The second type of mathematical transformation for data integration involves operations on Continuous Data (CD) sets. This type of data transformation in the geosciences has typically been motivated by three objectives: (1) creating statistical normally-distributed data; (2) creating data that are additive; and (3) making errors constant across the range of the data (Stanley 2006). These transformations effectively convert data into a form that can be statistically manipulated, thus facilitating subsequent data analysis.

Operations on CD have an advantage over the BS method in that they can be combined more effectively into a color graphic image. Enhancement of the low-amplitude features is usually achieved through some type of image transform (e.g., Morris et al 2001). By far the simplest approach to this problem is modification of the input limits of the data—i.e., the operator selects a subset of the full dynamic range. These CD operations must also take into account the various measurement scales, different data ranges, and the even distributional forms that exist with each geophysical method (Kvamme 2007). Normalization of the data sets must take place prior to any integration method, or else resulting values will present a false representation of the combined data set and lead to possible misidentification of anomaly locations (Type I and

II errors). Kvamme (2007) gives a short explanation of the different procedures in integrating continuous data (Data Sum, Data Product, and Data Maximum), of which the standardized data illustrates anomalies effectively and to varying degrees of magnitude, This, however, is dependent on the operator and what parameters are placed on the data prior to integration.

The most serious disadvantage of the CD method is the possibility of a non-unique result. Additionally, it is the operator of the integration software that determines the mathematical transformation lends the clearest picture of the subsurface features, which ends up being a qualitative assessment of a quantitative method. With this method, therefore, no parameters are set in place with a range of values (i.e., S/N enhanced) that will definitively illuminate the superiority or inferiority of the transformation performed to another combination of approaches. Although BI approaches are able to display clear cut maps of the presence or absence of anomalies, the distinguishing advantage of the CD output—as Kvamme (2007) states—is that both robust and subtle anomalies can be simultaneously expressed, producing composite imagery with high information content.

2.1.3.4 Statistical

Mechanisms for data integration utilizing statistical methods can be divided into two classes: those that reduce dimensionality (principal component and factor analysis) and classification methods (K-means cluster analysis and binary logistic regression). In cases where the data has a significant number of anomalies that do not occur in all data sets, this type of analysis may be necessary because it will indicate if there is or is not a

correlation between the data correlated from one technique to another. For example, one would expect certain data sets to have more significant correlations (e.g., electrical resistivity with conductivity).

Statistical methods such as the principal component analysis (PCA) are multivariate analyses that can be used for data that are spatially distributed and have common geographical locations (Honarmand et al., 2002). For cases where multiple geophysical techniques are utilized within the same area, statistical methods serve as a major advantage. However, the survey design must result in individual data points for each technique falling in the same coordinate system, in order for the integration to be simplified compared to other techniques previously mentioned. The PCA method can also account for determining amounts of variance between data sets (Snyder et al 2001); when data is standardized, each variable within the statistical method contributes a variance of unity (Davis 2002). Another advantage of PCA is that as correlations of variables are identified and removed from the equation; thus, noise from the combined data sets can become isolated that may help in readjusting pre-processing procedures to further enhance the signal-to-noise ratio.

One result of the PCA is a factor analysis that rotates two-component axes of interest within the six-dimensional (or four or three-dimensional depending on how many different types of geophysical data you have) measurement space to equally distribute the variance and improve interpretation of the data due to clearer results. Singh (2007) argues that PCA maximizes the expected value of the between-sample distance in attribute space. Additionally, data scattering between two attributes can be maximized with PCA,

indicating there are no significant disadvantages in this method, other than if the data quality was poor or the pre-processing steps were not appropriate the results could be skewed.

Classification methods such as the K-means Cluster Analysis (KCA) and Binary Logistic Regression (BLR) attempt to define groups or classes in bodies of data through two different methods. The former is known as an unsupervised classification, which covers all classification techniques relying only on input data and not biased by the desired output (Coleou et al 2003). Its simplicity of implementation often makes it selected for multivariate statistical analysis. The object of KCA is to identify subclouds within the N-dimensional crossplot (Cormack 1971). Its purpose is not data reduction but data partitioning into disjointed subsets. Separation of the clusters is often based on the standardized Euclidean distance, the weights coming from normalization of the samples, or the more general Minkowski metric, among which the basic Manhattan or city block distance is found (Coleou et al 2003). For KCA, a prior knowledge of the number of clusters is required. Overall, it makes for an excellent filing system but does not describe the topological properties of geophysical data. It is known to perform well if data (i.e., N-dimensional crossplot) are organized into separated compact subclouds—also described as hyper ellipsoidal clusters with internal cohesion and external isolation (Cormack 1971).

The BLR method is considered a supervised classification, which is based on a multivariate normal model and produces continuous probability surfaces for anomalies of a single class. This is an advantage over other methods in that it simplifies the resulting

data map, highlighting only those anomalies that are of significance to the operator (low probabilities will be seen as an anomaly absence). Anomalies present within the data may also be divided into classes, similar to the BS method, but in a more quantitatively controlled nature. Another advantage of BLR is that differences between the data sets—with respect to the signal strength output relative to each geophysical technique—is maximized and aids in rescaling the data to standardize the output, offering a more cohesive interpretation. The predictive aspects of BLR data integration are excellent in that data patterns become more recognizable while less visible anomalous conditions are more noticeable. Unfortunately, there is a need for a large number of data points in order for the confidence level of this method to be within the desired range (Lado et al 2008). The selection of class (division of anomalies within the whole data set) incorporated in the algorithm is vital to achieve good results because the classification function that ultimately results from this method is optimized to patterns within the samples (anomalies). Should the classification scheme be too broad, the end result and ultimate interpretation might not be sophisticated enough; conversely, the data map may become too complex to be of much use if the classification scheme is too narrow.

2.2 Geology of Site Location

2.2.1 Geologic Overview

The whole project area is located within the Great Valley section of the Valley and Ridge Physiographic Province. The Great Smoky Mountains of the Blue Ridge to the east, and the Walden Ridge Division of the Appalachian Plateau to the west border this

province (Cagle 1948). The Tennessee River is the major drainage of the Great Valley section and is considered a general dividing line for topographic distinction. To the north of the river, relief is characterized by a series of parallel elongate ridges and intervening valleys trending northeast to southwest. South of the Tennessee River, the Great Valley section is expressed as a succession of rounded or conical hills and knobs that do not display orientation (e.g., Cagle 1948). The bedrock of the Valley and Ridge represents the westward thinning of a thick wedge of sediment that accumulated on the eastern shelf margin of North America throughout the Paleozoic (Byerly 1997). This region is part of a geosyncline where sediments were deposited and eroded. Late in the Paleozoic, approaching the Pennsylvanian (~310 MA), the region was lifted above the level of the sea and the strata were folded and faulted in the Appalachian orogeny (Cattermole 1958). The bedrock units within the project area all date to the Ordovician (~510 Mya) Knox Group.

2.2.2 The University of Tennessee Geophysical Research Station

The University of Tennessee Geophysical Research Station allows for the work described herein to be conducted in a controlled setting, located between Alcoa Highway 129 and the Tennessee River approximately 2 miles south of the University of Tennessee campus in Knoxville, Tennessee, as shown in **Figure 1**. Soil conditions in the site vary from residual soils developed directly on sedimentary bedrock near the highway, to loamy soils developed on a series of alluvial terraces at different elevations above the river. The University of Tennessee Geophysical Research Station contains known targets buried and surveyed in 1999, with the locations given by latitude, longitude, and depth



Figure 1. Location of the University of Tennessee Agricultural Extension Center, which contains the study site, The University of Tennessee Geophysical Research Station. The University of Tennessee, as seen in the Northeast corner of the expanded map in located in Knoxville, Tennessee. The red square located in the zoomed in map shows the location of control site.

beneath the surface. Additionally, information including size, shape, composition material, and orientation are given. This study assumes that there has been enough time for the ground to settle, soils to mature; thus, the majority of disturbance to the subsurface (and resulting signal in the data) has been minimized. It should be noted that there is no current surface expression of any of the buried objects visible to the naked eye.

2.3 Geophysical Techniques

2.3.1 Introduction

There are three geophysical techniques utilized in this project. The techniques were selected because of their traditional use in detecting small, discrete targets with a relatively quick pace in acquisition of the data. Understanding of the science of each technique used is vital for the integration process and are summarized in Sections 2.3.2-2.3.4).

2.3.2 Ground Penetrating Radar

The ground penetrating radar (GPR) technique utilizes propagating electromagnetic (EM) waves to detect changes in the dielectric properties of the shallow subsurface. The material properties that control electromagnetic energy transfer through media are conductivity, dielectric permittivity, and magnetic permeability (Powers 1997). Conductivity can be generally defined as a measure of charge transport through a material as a result of an applied EM field, whereby the charge transport associated with

charge displacements will only occur over the time duration of the particular polarization process. Dielectric permittivity is a measure of electric field energy stored and lost through induced charge displacements, just as magnetic permeability is the measure of magnetic field energy stored and lost through induced magnetization. According to Powers (1997), this parameter is often ignored in GPR studies because geoscientists assume many natural, near-surface materials have weak magnetic responses. However, sands and soils that exhibit significant magnetic responses and more measurements are needed to determine the extent of the magnetic effects in GPR surveys (i.e. clay-rich soils are high in magnetic response and cause scattering of the GPR waves).

The GPR technique is similar to seismic reflection methods; wave propagation velocity changes as a wave travels through the subsurface and generates reflected energy detectable at the surface (Baker et al., 2007). The dielectric permittivity contrast between the background material and the target determines the propagation velocity of EM waves (i.e., the controlling factor on the generation of reflections). Baker et al. (2007) defines the dielectric permittivity as the ability of a material to store and then permit the passage of EM energy when a field is imposed on the material, and this can be measured in the lab or *in situ*.

A GPR unit consists of transmitting and receiving antenna, where the transmitting antenna generates an EM pulse in the subsurface that travels into the subsurface, reflects off an interface or scatters off point sources (both caused by a contrast in dielectric permittivity). This reflected/scattered energy then travels back to the surface where it is recorded by the receiving antenna. The time it takes for the wave to travel down to an

interface and back up to the surface is called the travel time, and is used to determine the in situ propagation velocity of the subsurface material (Baker et al., 2007) and subsequently the estimated depth of the feature.

Any detected differences between the modeled GPR data and the real data may be due to scattering from inhomogeneities and the presence of multiple reflections (Baradello et al. 2004) in the real radargram. A limitation to the GPR technique is inadequate prior knowledge of either the electromagnetic properties of the subsurface or the geometrical and EM properties of the target in question. Through the use of the GprMax program (Giannoploulos 2005), data processors expect that this limitation will be minimized. **Figure 2** shows a representation of one configuration for antenna and the general pathway of the EM wave.

2.3.3 Ground Conductivity (EMI)

With the ground-conductivity EM inductive method (EMI), surveys can be carried out under most geologic conditions including those of high surface resistivity such as sand, gravel, and asphalt (McNeill 1980). Ground conductivity refers to the electrical conductivity of the shallow subsurface of the earth. The EMI detection of a buried target is accomplished by illuminating the subsurface with a time-varying primary field (e.g., Pasion et al. 2008). Furthermore, if a buried target is conductive, EM eddy currents and a subsequent chargeability decay be induced in the target, with those currents producing a secondary magnetic field which is detected by a receiver coil located at the surface a fixed distance away. According to Pasion et al. (2008), the rate of decay and the spatial

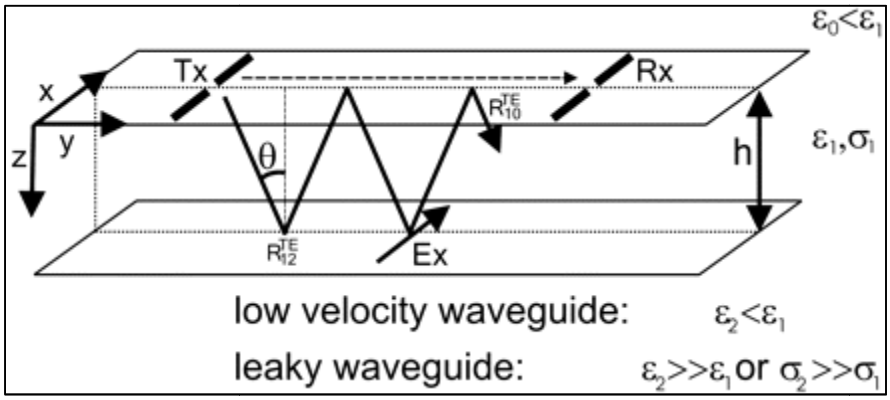


Figure 2. Basic GPR schematic. The x-axis is oriented parallel to the long axes of the antennas. ε_0, ε_1, ε_2 are the relative permittivities and σ_1, σ_2 the conductivities of the respective media. R^{TE}_{ab} is the reflection coefficient of the broadside (TE) mode GPR waves incident at the boundary between the a and b media, where a and b can be 0 and 1 or 1 and 2 (modified from van der Kruk et al. 2009).

behavior of the secondary field are determined by the target's conductivity, magnetic permeability, shape, and size. The electromagnetic response of the target will be primarily dipolar (Casey and Baertlein 1999) for the target/sensor geometries of metallic objects. Measurements are recorded in units of conductivity (typically milli-Siemens per meter, mS/m).

In addition to conductivity readings, the in-phase component of the electromagnetic field is recorded. The in-phase measurement is sensitive to the presence of metallic conductors and this measurement is used for metal detection. Abrupt spikes in the in-phase and conductivity measurements are indicative of locations of the desired targets within this study. It is important to remember that real field data have errors unaccounted for in the forward modeling operator (for example, inaccurate sensor positioning, noise spikes in the data, and sensor drift) that can lead to non-Gaussian error distributions (Pasion et al. 2008). Incorrect characterization of data statistics can bias the values of the recovered parameters and also invalidate the parameter variance analysis (e.g., Billings et al. 2003). Current methodologies for inverting EMI data and using recovered parameters to make classifications have been successful (Song et al. 2008); however, the technique has difficulty when anomalies arise from multiple targets. It is the goal of this research to analyze the data to better discriminate the types and number of targets present, and more closely approximate the spatial location of each target. **Figure 3** shows a schematic of how the EMI technique works.

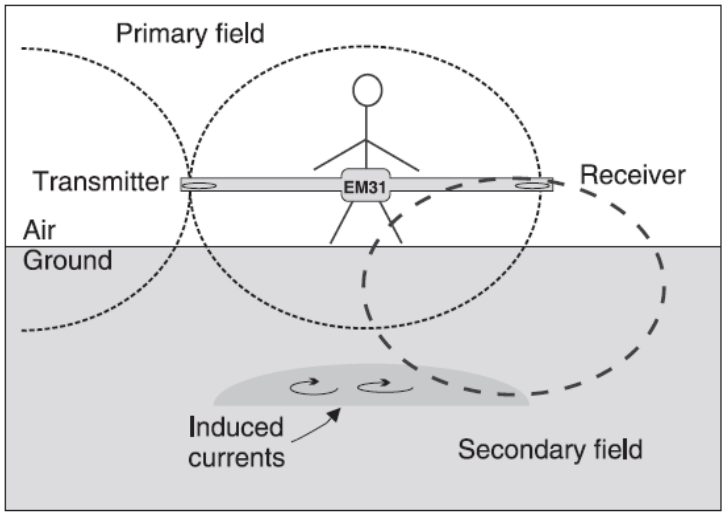


Figure 3. Schematic diagram of basic EM-31 principles (modified from Pettersson and Nobes 2003)

2.3.4 Magnetic Gradiometry

A magnetometer measures magnetic field strength at a specific measuring point. For purposes of this study, we are interested in measuring how much the strength of a magnetic field changes between two specific points, or the "gradient" of the field. Many of the subsurface features we are hoping to detect have magnetic characteristics that cause a disturbance in the earth's magnetic field in an area around the object. This disturbance can be detected as a magnetic field gradient by the gradiometer. Based on the use of gradients (derivatives) of the magnetic field anomalies, geometric parameters such as locations of boundaries and depths of the causative sources have been determined (Salem et al. 2002).

One technique becoming more common today is the approach of the analytic signal of magnetic anomalies, which was initially used in its complex function form and makes use of the properties of the Hilbert transform (Blakely 1995). The amplitude of the analytic signal (AAS) is defined as the square root of the squared sum of the vertical and two orthogonal horizontal derivatives of the magnetic field, where the horizontal and vertical derivatives of the magnetic field are Hilbert transform pairs (Debeglia and Corpel 1997) over 2D sources. The AAS of magnetic anomalies can be easily computed with both vertical and horizontal gradients being calculated in the frequency domain using conventional Fast Fourier Transform (FFT) techniques (Salem et al 2002).

The appeal of this geophysical technique is that the locations and depths of the sources are estimated with only a few assumptions about the nature of the source bodies, which are usually assumed as 2D magnetic sources. Salem (2002) explains that for these

geological models, the shape of the amplitude of the analytic signal is a bell-shaped symmetric function located directly above the source body. Depths estimates can be obtained from the lateral width (extant) of the anomalous AAS signal. When examining the data, remnant magnetization is determined to exist when the target anomaly is negative compared to the background value. A normally magnetized body would produce a positive anomaly relative to background (Dannemiller and Li 2006). We presume that the source body is metallic in nature if the data expresses a dipole magnetic anomaly. **Figure 4** illustrates a synthetic example of a target with a strong remnant magnetization and the resulting total-field dipole anomaly.

2.4 Data Acquisition

2.4.1 Ground Penetrating Radar Data Acquisition

2.4.1.1 Ground Penetrating Radar Equipment

The GPR units used for this research are a Sensors & Software, Inc., PulseEKKO Pro™ system and a Sensors & Software, Inc., Noggin™ unit. These systems both utilizes a "smart cart" configuration whereby the transmitting and receiving antennas are fixed in a common offset (constant antenna separation) type of survey design, resulting in a cross-sectional profile of the subsurface along the transect line. The resulting data yield information on depth to interfaces (when incorporating velocity information) as well as locations of anomalous discrete features. During this stage of data acquisition, 100 MHz frequency antenna pairs were used, depending on the nature of the site. Along-profile

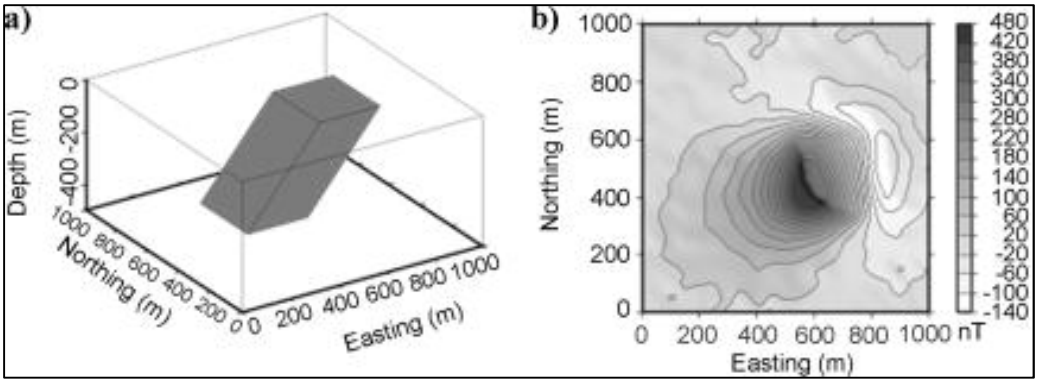


Figure 4. A synthetic example with strong remanent magnetization. (a) Dipping body viewed from the southwest direction. It has a total magnetization direction which deviates significantly from the inducing field direction. (b) The total-field anomaly. (Dannemiller and Li 2006)

data were collected at 10cm intervals, with each sample being controlled by an odometer on one of the wheels of the Smartcart™.

2.4.1.2 Acquisition Parameters

Data collection for this site utilized the PulseEKKO Pro™ system (100 MHz antennas). Profile lines were collected with a 0.5-meter spacing, with the acquisition alternating in a northward and southward direction (typically called a "zigzag"). A 0.5-meter buffer spacing surrounds the 50-meter by 40-meter grid, to account for the overall physical size of the GPR system. In-line data (in the direction of the profile) were collected every 10 cm. The resulting data volume was rectangular with one exception: a portion of the grid did not have data collected where a small 10-m by 10-m plot is fenced in, and a second buffer was created due to possible error readings along the edge of the voided area from the metal fencing.

2.4.2 Ground Conductivity Data Acquisition

2.4.2.1 Ground Conductivity (EM-31) Equipment

A Geonics™ EM-31™ terrain-conductivity meter was used for acquiring conductivity values for the survey area. The EM-31 is a single-operator device, composed of transmitter and receiver coils on either end of a 3.66-meter frame. The instrument utilizes an electromagnetic-inductive technique that allows measurements without invasive electrodes or ground contact. Effective exploration depth for this instrument is at most 6 meters, depending on the subsurface conductivity (more conductive ground yields

less depth of exploration). The system is a non-intrusive conductivity-measuring device, and data can be acquired at the speed in which the operator can walk—though the faster the horizontal speed the greater the "smear" of the data. Small changes in conductivity are measured with fairly good precision, and there is a continuous readout of data collection while traversing the survey area. Additionally, the recorded in-phase component is particularly useful for the detection of buried metallic structure and waste material. According to the Geonics™ website, the operating frequency is 9.8 kHz and has a measurement accuracy of +/- 5% at 20 mS/m. These have not been independently confirmed other than anecdotally, but the instrument is an industry standard and has been used around the world with success for the past 20 years. For all surveys, one measurements was acquired per second.

2.4.2.2 Acquisition Parameters

Survey design for the EM-31 utilized the same grid as described in Section 2.4.1.2 for the GPR plot: a line spacing of 1 meter was used. Data profiles were collected alternating using the zigzag acquisition scheme in eastward and westward directions. Data were collected continually with eight points per meter traversed in the profile direction, where the user set a pace such that 1-m marks in the field were passed every eight seconds. Because the sensor is oriented horizontally, it is important to note that the distance from the ground to the sensor is different for each operator (from having a different carrying height) and must be adjusted during data processing. As with the GPR grid, the fenced-in

portion of the 50-by-40-m survey area was skipped during acquision and a 0.5-m buffer was left surrounding the area..

2.4.3 Magnetic Gradiometry Data Acquisition

2.4.3.1 Magnetic Gradiometry Equipment

The instrument used during this portions of the project was the Bartington™ 601-2 single-axis magnetic gradiometer (magnetometer). The instrument is designed to detect disturbances in the geomagnetic field caused by the contrasts in magnetic susceptibility; for example, it is robust in detecting thermo-remanance in kilns and bricks for archaeological projects. Single-axis gradiometers, such as the one being used in this study, measure magnetic gradient in a single vector-direction (such as vertical, in-line horizontal, or cross-line horizontal). The Grad 601 is a single-axis, vertical-component fluxgate gradiometer with incorporated data logger and two cylindrical sensor assemblies.. The Bartington™ website indicates that each sensor assembly contains two fluxgate magnetometers with a one meter vertical separation, together with electronics and non-volatile memory for calibration data. The gradiometer has a linear range of 100nT, with a resolution of 0.1nT and a total range of 1000nT with a resolution of 1nT. Under ideal conditions, the depth of investigation is typically two meters with a surface spatial resolution of 0.25 meters, as fixed by the survey design. Data acquisition spacing as well as the speed with which the survey is conducted is set by the operator.

2.4.3.2 Acquisition Parameters

Survey design for this geophysical technique utilized the same grid as the GPR and EM-31 plots as described in Sections 2.4.1.2 and 2.4.2.2, respectively, with a profile spacing of 0.5 meter. Data profiles were collected in a zigzag alternating eastward and westward. In-line data points were collected continually with eight points per meter traversed, set by the pace of the user. A similar portion of the grid did not have data collected, with possible error readings along the edge of the voided area due to a metal fence.

2.4.4 Spatial Location

Differential real-time GPS (dGPS) measurements were integrated with each of the geophysical techniques in this study. The dGPS receiver is a Trimble™ Pathfinder ProXRT, incorporated with a Trimble™ Ranger handheld computer for data logging, and a subscription with the Omnistar™ service for real-time corrections. By integrating dGPS with each geophysical data point, the location of each anomalous feature is more easily identifiable due to the sub-meter horizontal special locating, making plotting of the data more robust. In order to effectively reduce uncertainty in their discrimination, it is particularly important for this study that the locations of the anomalies are as precise as possible when attempting to merge datasets that vary greatly in resolution.

2.5 Data Processing

2.5.1 Ground Penetrating Radar

2.5.1.1 Introduction

All GPR data were processed using standard data processing methods for a common offset GPR configuration via Sensor and Software's™ *Ekko_View*, *Ekko_Deluxe*, and *Ekko_Mapper* programs. The dGPS coordinates for each data point were integrated into the data with the *Ekko_Deluxe* program, allowing for a more robust 3D representations of the data. Processed data was exported in a grid file to the program Surfer (Golden Software™). The resulting grid file wass displayed in various ways (contour, image, shaded relief, and surface maps) to allow for user-specific 3D visualization of the data.

2.5.1.2 Dewow

The first processing step for the GPR data was to run a dewow filter, which is a type of frequency filter that is used to reduce or remove low frequency components of GPR traces and/or also remove DC shift or DC bias (Baker et al., 2007). When collecting GPR data, the component of the transmitted signal below 1 MHz may have induced a slowly decaying low-frequency DC shift (or "wow") on the recorded time-varying data traces; this "wow" then was superimposed on the high frequency reflections within the data. Frequency filtering worked to enhance or remove specific frequencies or frequency ranges in the data (Baker et al., 2007). A dewow filter is a high-pass frequency filter, which means that the filter passes through the high-frequency component of the data and attenuates the low frequency component (including the AC shift). The *EKKO_Mapper*

program used a running average filter on each trace to remove the low frequencies. A running average filter works by taking the average amplitude of one pulse at a particular point, and then subtracting the average amplitude within the pulse from the amplitude value at that point. This process continued by moving to the next point along the trace and running the same filter iteratively along the entire length of the trace. The artifacts generated by this filter were minimal, as the process has been optimized over the years through many experiments (*EKKO_Mapper* manual) as tested by Baker and others (2007).

2.5.1.3 Migration

Migration of GPR data is a process that focuses scattered signals by collapsing hyperbolic diffractions to their apex. This process is also called synthetic-aperture processing. Migration was used for common-offset data (*see* Section 4.2.1.2) to reassign the signal form undulating reflecting interfaces to their more true geometric positions and thus increase the horizontal resolution of the data. Migration was important both when dealing with dipping layers in the sub-surface and to focus diffracted energy from small-scale subsurface features. Dipping beds in unmigrated data were shown at their apparent dip - to see the true dip, the data needed to be migrated using the calculated subsurface velocity (Baker et al., 2007).

2.5.1.4 Velocity Conversion

In order to convert GPR data from time sections (as it is recorded in the field) to more usable depth sections, the EM-propagation velocity of the subsurface materials was calculated using the data. Velocity was measured from curve-fitting on the common midpoint data by using the "direct wave" method: by examining the slope of the direct EM wave (which is the raypath going directly from the transmitting antenna to the receiving antenna through the ground), velocity (in m/ns) can be calculated by taking the inverse slope (ns/m) of the first arrival of that energy. This calculated velocity was then used to convert sections from time-domain to depth-domain, and was also used in the migration calculations described in the previous Section. A critical assumption in this method of velocity estimation was that the surface layer velocity is representative of the velocity distribution throughout the volume that is imaged (Ambrose 2005). Although this assumption is often violated, for all the sites described here the shallow geology was relatively homogenous, at least down to the tops of the archaeological features being imaged; hence, the "direct wave" method of depth conversion was valid.

2.5.1.5 Automatic Gain Control

Autogain is a system to control the gain, or increase in the amplitude of an electrical signal, from the original input to the amplified output. Automatic gain control (AGC) is commonly used in seismic and GPR processing to improve the visibility of late-arriving events in which attenuation or wavefront divergence has caused amplitude decay (Baker et al., 2007). When autogain was implemented, the signal-to-noise ratio was reduced (as

noise is increased in amplitude late in time within the data). An advantage of using AGC was that it highlighted weak reflections; however, it also "created" some artifacts. It was important to keep in mind that the amount of gain applied to the geophysical data is different for each scenario and was adjusted accordingly.

2.5.2 Ground Conductivity (EMI)

2.5.2.1 Introduction

The Geonic™ program DAT31W is designed to aid in processing data that is acquired by the EM-31 instrument, allowing data to be transferred easily from the dignital data logger to a personal computer in the lab. The data files were used as input for the Geosoft™ *Oasis Montaj* and *Surfer* contouring packages where three-column (xyz) format was suitable. Overall processing of EM-31 data with this Microsoft™ Windows-based software resulted in much greater productivity, with readings arranged in profile lines consisting of unlimited segments. The DAT31W program used ASCII format for the data files.

2.5.2.2 Smoothing

The smoothing procedure can be applied to any number of selected survey lines and to any data type. Several methods of smoothing are available—such as a 3-point linear smooth, 5-point linear smooth, etc.—and may be applied to a selected set of data several times. Ultimately, the method of smoothing (or generating residual curves) and the degree of smoothing depended on the particular data setand desired method of

presentation;thus, the optimum parameters were data dependent. Smoothing the data enhanced S/N by capturing the patterns within the data that were significant while at the same time filtering out noise from the data. The most common way of smoothing data within DAT31W is called "curve fitting." While smoothing gives a general idea of relative changes of value with little attention paid to the close matching of data values, curve fitting concentrated on achieving as close a match as possible (Hastie 1990).

2.5.2.3 Destagger

Destagger describes a process used to compensate for data collection errors associated with the alternating direction taken by the operator during the zigzag acquisition scheme, if the midpoint of the instrument is precisely centered on the location of the body of the operator. This step shifted individual profiles (with the same direction of acquisition horizontally in space forward (and/or backwards) by a specified number of intervals. The correction offset (or destagger) was either applied to just the outbound or in both directions, depending on the collection issue at the time. The new values for each data point were then repositioned using a cubic-spline fitting algorithm to produce a smooth curve from the available data. Applying this method allowed for multiple sensor arrays. Additionally, if the data extended the full width of the grid, missing data points at the start of each traverse were filled in by extrapolating the existing values in the traverse, and points at the end of the traverse were discarded as they extended into adjacent grids or beyond the extent of the composite data map. Because this process was dependent on sensor configuration and data collection patterns, it was carried out prior to any

interpolation. In this project, the destagger processing method was utilized with the magnetic gradiometry data in addition to the ground conductivity data.

2.5.3 Magnetic Gradiometry

2.5.3.1 Introduction

For processing of the magnetic gradiometry data was acquired by the Bartington ™ 601 and the program *ArcheoSurveyor* (DW Consulting, Inc.) was utilized. This program is specifically designed to input, assemble the geometry of, process, and visualize the 2D geophysical data gathered with geophysical instruments such as ground-conductivity meters and magnetometers. *ArcheoSurveyor* recognizes two main categories of data: grids and composites. Grids are relatively small blocks of data typically collected by hand-held instruments in a structured and non-automatic manner. All data points in the grid are an exact interval apart in both X & Y axes (or north and south, if the grids are aligned to the compass directions). The grids are then assembled into a larger composite (consisting of multiple grids), still using the same X & Y intervals. In the case of dGPS based systems, it may not use specific X & Y axes at regular intervals, depending on the geometry mapped by the dGPS data. *ArcheoSurveyor* is therefore able to import dGPS-acquired datasets directly to composites.

2.5.3.2 Clipping

Clipping replaces all values in the current data grid (or composite) outside a specified minimum and maximum range with the minimum (or maximum) values. These values

can be specified or can be automated through statistically calculations of the standard deviation of the data. This clipping process was used to remove extreme data point values. Extreme values force the display to show all values in the center of the histogram in the same color and thus mask finer details. Excluding these extreme values via clipping allowed the details to show to be more visible.

2.5.3.3 Interpolate

This process describes both increases (via interpolation in the traditional sense) or decreases (via down sampling) to the resolution of the selected data volume. When increasing, interpolation generated an extra data point between every existing data point in either the X or Y direction. The values for the extra points were calculated using a cubic-spline algorithm. This produced a smooth curve to fit the available data points. Decreasing simply removed (or decimated) every other point/line in the data. Though it appeared to improve the data resolution, any improvement was artificial and excessive interpolation eventually created artifacts that have no basis in the source data. Due to this side effect of interpolation, it was important to have a strong understanding of local geology and target descriptions to avoid misinterpretation. Additionally, every doubling in any horizontal direction also doubled the processing time for each subsequent process and prevented some comparisons across the interpolated layer.

2.5.3.4 Despike

Despike (similar to a de-noise routine) was applied by a process of scanning the data with a uniformly-weighted window and tagging data points that exceed the mean (or median) of the window by a specified threshold amount. When found, the point was replaced by either the mean, median, or threshold (user specified). The despike filter is typically used with magnetometer data to remove spikes caused by small surface metallic anomalies. These anomalies are generally the result of modern metal 'rubbish' in the topmost layers, and typically cause very strong but highly localized signals.

2.5.3.5 Dedrift

The dedrift process was used to correct for long-wavelength drift in the readings taken by an instrument. The process applied a progressive correction to every data point within a range of points in a grid. Because of the source of the problem corrected by this process is grid dependent, it wasonly applied to individual grids. The actual correction value was the difference between the averaged beginning and ending values of the grid divided by the number of data points between those start and end points. Each data point between the start and end point was then reduced by the correction value multiplied by its distance from the start point. This correctly allowed for zigzag either or parallel data collection methods. All data points after the end point were reduced by the full difference.

2.5.3.6 Destagger

See the description in Section 2.5.2.3.

2.5.3.7 Deslope

This technique is primarily intended to correct the 'waterfall' errors seen in magnetometer data caused by large metal objects near a survey area, and similar to a detrending that is non-linear. The process calculated a curve for each row or column of data based on specified parameters; the curve was then subtracted from the actual data. Deslope is a selection based process sof no selection was made prior to starting the process, this step was applied to the whole survey. In the case of the synthetic data sets, no selection was made. However, for the authentic data, selections were made.

3. METHODOLOGY

This chapter is based on a manuscript to be submitted by Megan E. Carr and Gregory S. Baker to the journal *Geophysics as Part 1 of a two part series*:

Carr, Megan E. and Gregory S. Baker, Quantitative Integration of Geophysical Data for Enhancing Subsurface Features: Part 1, Geophysics, v. pp.

My contributions to this paper include (i) development of the modeled data, (ii) formulation of the programs utilized for data manipulation, (iii) statistical analysis of data sets, (iv) visualization of the data, (v) significant portion of the writing.

Abstract

This article will serve as an description to the methodology developed for a quantitative integration of multiple near-surface geophysical data sets to investigate the hypothesis that certain targets, given multiple variables and parameters, can be detected with a greater degree of certainty than others if a specified statistically-derived combination of data processing and data merging is implemented. The main objective is to improve data quality and visualization techniques within near-surface geophysical surveys for discrete anomaly detection. Initial statistical analysis methods are developed on synthetic data and include simulations for magnetic gradiometry, electromagnetics and GPR Data integration techniques are given in full detail, as well as the statistical treatment for data preparation as well as data integration). Statistics will (1) illuminate the specific merging protocols that are best suited under various scenarios—type of target, geologic setting, target geometry, etc., (2) indicate how statistically significant the integration of data is in discrete anomaly detection, and (3) be used to indicate how the uncertainty level in anomaly detection changes through implementation of the proposed methodology.

3.1 Introduction

Currently the literature contains no systematic, quantitative methodology for the integration of two or more geophysical data sets collected using near-surface geophysical techniques for discrete anomaly detection. Recently, Urs Böniger and Jens Tronicke (2010) propose successful results from an integrative analysis and interpretation of different data sets that combines geophysical instruments with modern topographic data using a tracking total station (TTS). However, their results are limited to composite images of the combined data sets and illustrated with various color schemes, which, when interpreting the data, still consists of a qualitative assessment and does not quantify the data interpretation to the extent presented here.

In order for an individual technique being employed in a survey to be appropriate (i.e., be useful for target detection/discrimination), there is a dependence on a difference in physical properties between the target and the surrounding material in terms of the subsurface characteristic that the technique responds to, such as dielectric permittivity, magnetic susceptibility, and electrical conductivity, respectively. With this in mind, when a geophysical investigation is conducted over a region containing targets that are unknown—and only one technique is used—it is possible to miss certain types of targets completely (e.g., Type I errors). Therefore, integrating multiple techniques into one unified data set may be used to more accurately identify and discriminate characteristics of targets with a greater degree of certainty.

Most “data integration” studies have involved a single near-surface geophysical technique with additional data such as boreholes (e.g., Hornby 2007, Ferré 2003) or

general geological data such as soil surveys and other maps (e.g., Galicia 2001, Rahman et a. 2008, Allen et al. 2008). Colombo and De Stefano (2007) state that when modeling an integration of techniques, conversions of parameters from one geophysical domain to another have traditionally been performed rigidly by means of empirical functions. While their research mainly deals in techniques not utilized in this project, Colombo and De Stefano (2007) do show evidence of the importance in setting up appropriate parameters in both data acquisition and data processing in order to ensure the highest degree of certainty in the integrated data models.

Data sampling heterogeneities inherent in geophysical databases have been identified as the main source of data integration problems (e.g., Stock and Pullar 1999, Bishr 1998, and Kim and Seo 1991). Semantic heterogeneity, which occurs when there is a disagreement about the meaning, interpretation, or intended use of the same or related data has been identified as the main cause of data sharing problems and are the most difficult to reconcile (Allen et al. 2008). The research presented here attempts to minimize these data integration problems by (1) setting up consistent parameters within geophysical techniques utilized; (2) develop a comprehensive model for integration of data; and (3) utilize consistent visualization techniques to represent the processed data for a higher degree of confidence when interpreting the results.

This manuscript is the first of a two part series, which will investigate two underlying hypotheses: (Part 1) Certain targets, given multiple variables and parameters, can be detected with a greater degree of certainty than others when a specified combination of processing and merging of data is implemented; and (Part 2) Integration of two or more

geophysical techniques will result in an improved subsurface image and reduce uncertainty in discrete anomaly detection. The primary objective is to improve success rates as defined through data quality and visualization techniques within geophysical surveys for discrete anomaly detection (e.g. locating clandestine underground tunnels, locating buried objects, mapping historical features). Each hypothesis tested will incorporate different goals to meet this objective.

Hypothesis (Part 1) will be tested via developed and analyzed computer models in order to satisfy the following goals:

- Develop a comprehensive model for creating typical signal responses for various materials of objects found at The University of Tennessee's Geophysical Research Station
- Create synthetic data sets for each geophysical techniques utilized in study
- Determine how resolution of data and data sampling heterogeneities affects integration of data sets and the resulting ability to discriminate targets with a higher degree of certainty
- Determine which variables involved with a geophysical survey are most significant in the discrimination of targets (e.g., target depth, target size, composition of target)
- Complete a statistical analysis to evaluate the effectiveness of the methodologies to quantitatively merge different geophysical data sets

This first hypothesis and the results of its testing is the focus of Part 1.

Initial statistical analysis methods are developed for the synthetic data. Simulated geophysical data include magnetic gradiometry, ground conductivity, and ground penetrating radar (GPR). Data integration techniques are to be given in full detail, as well as statistics (in both data preparation and in data integration), which is the quantitative assessment of the integration methodologies. Statistics demonstrate the merging techniques that are best suited for various scenarios (type of target, geologic setting, etc.), indicate how significant the integration of data sets is in discrete anomaly detection, and describe how the uncertainty level in anomaly detection changes through implementing the proposed methodology.

3.2 Methods

3.2.1 Modeling of Data

Modeling potential targets and subsurface parameters had a potentially significant impact on being able to meet the objectives of this research. Primarily, these models will aid in determining probable locations of various targets and reducing the uncertainty in discrete anomaly detection. Two different graphical user interfaces (GUI) were utilized to simulate the resulting anomalies for multiple scenarios (i.e. target characteristics, subsurface characteristics) for the geophysical techniques of GPR, EM-31 (ground conductivity), and Magnetic Gradiometry. The synthetic data created with these simulations are correlated with the data collected in the field for each of the geophysical techniques, that when combined with ground-truth data, result in a more robust interpretation of the geophysical data to discriminate (not simply detect) various features as demonstrated in the case study of Part 2. Utilizing the GUIs prior to identifying areas

of potential interest for any geophysical investigation in areas of unknown target types is particularly important, as geophysical data commonly result in non-unique signatures, increasing the uncertainty factor. It is expected that the incorporation of these models will reduce the uncertainty in anomaly characterization.

3.2.1.1 *GprMax2D*

The *GprMax2D* software, created by Antonis Giannopoulos, is a computer program that implements a finite-difference time-domain method to generate simulations. This version of the tool was chosen for it's GPR signature simulation and ease in exporting the processed synthetic data to other software programs for visualization. A simple ASCII (text) file to define the model's parameters with special commands to instruct the software to perform specific functions is available, depending on the type of model the user wants to create (*see* Giannopoulos, 2005). The *GprMax2D* program can do an excellent job of overcoming the issue of modeling open boundary problems like GPR, of which one issue is the truncation of the computational domain at a finite distance from the sources and targets. By adding approximate conditions like absorbing boundary conditions (ABC), waves impinging on the targets are absorbed, simulating an unbound space, yet also limiting the computational space within the model. An example of simulated GPR scans from the program over a simple discrete target is given in **Figure 5**.

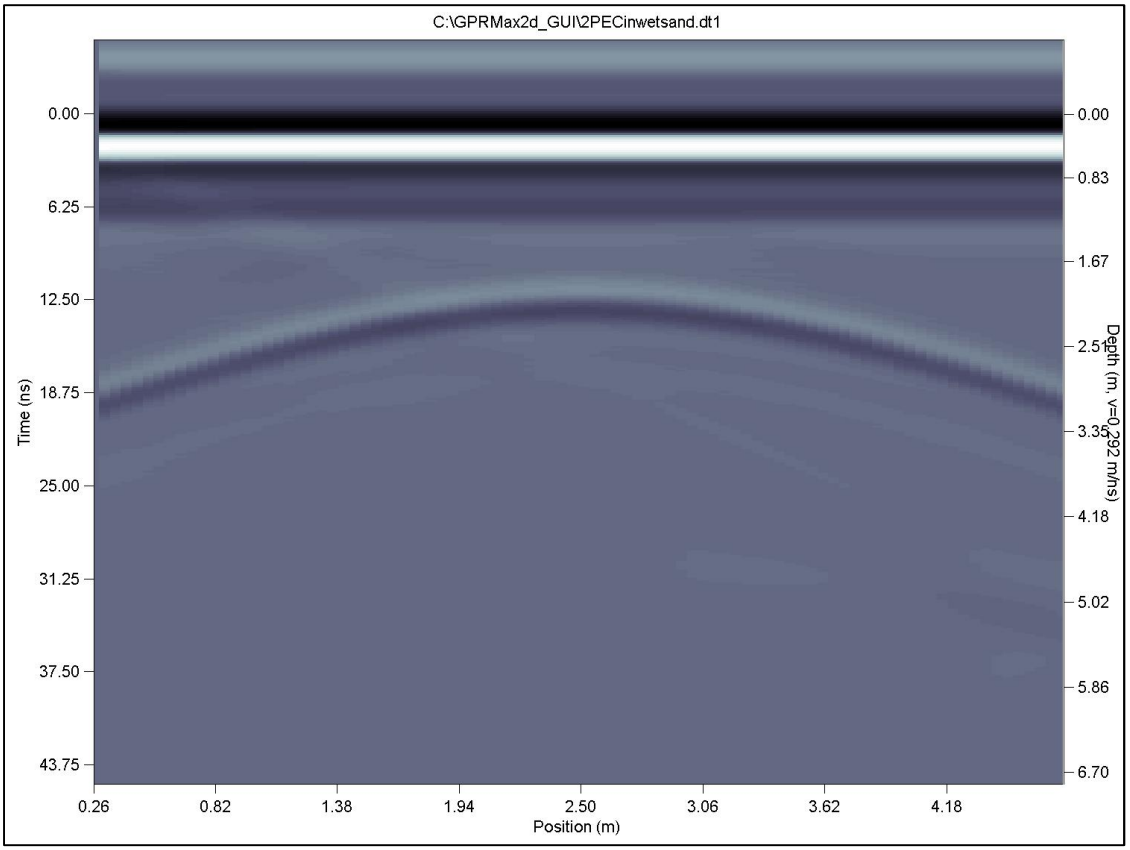


Figure 5. Output from *GprMax2D* program. Modeled object is a concrete sphere with a diameter of 50cm and buried at a depth of 2m. Velocity (v) adjusted to 0.292 m/ns.

3.2.1.2 *Geophysica*

Geophysica is a MATLAB-based software tool for the simulation, display, and processing of near-surface geophysical data; for this project it was found to be an appropriate approach to model magnetometry (specifically magnetic gradiometry) and electromagnetic induction (EM ground conductivity) targets. The software program was initially developed to provide practical experience in the design of field studies and data interpretation through the use of numerical simulations. Created by Alan Witten, the program is essentially a series of MATLAB m-files (code files) with target and sampling parameters specified throughout the GUIs and output data is graphically displayed. The EM ground conductivity simulations are performed in the frequency domain for an assumed co-located transmitter and receiver (Witten 2002), with the target simulated being a sphere and the dipole moment of both the transmitter and receiver being vertical.

For a user-selected target location (x_0, y_0, z_0) and radius, the in-phase and quadrature components of the vertical component of the secondary field is computed over a user-specified horizontal grid at the particular user-specified frequency. In terms of modeling magnetics, both field and gradient measurements can be modeled where the anomalous induction is assumed to be a scalar component of the vector field along the direction of the Earth's magnetic induction. For gradient measurements, Witten assumes that measurements are made at two positions separated by the vector distance and set the code to compute the difference between the anomalous inductions at these two points. Additionally, three types of gradient measurements can be simulated: vertical gradient, north-south horizontal gradient, and east-west horizontal gradient. **Figure 6** gives an

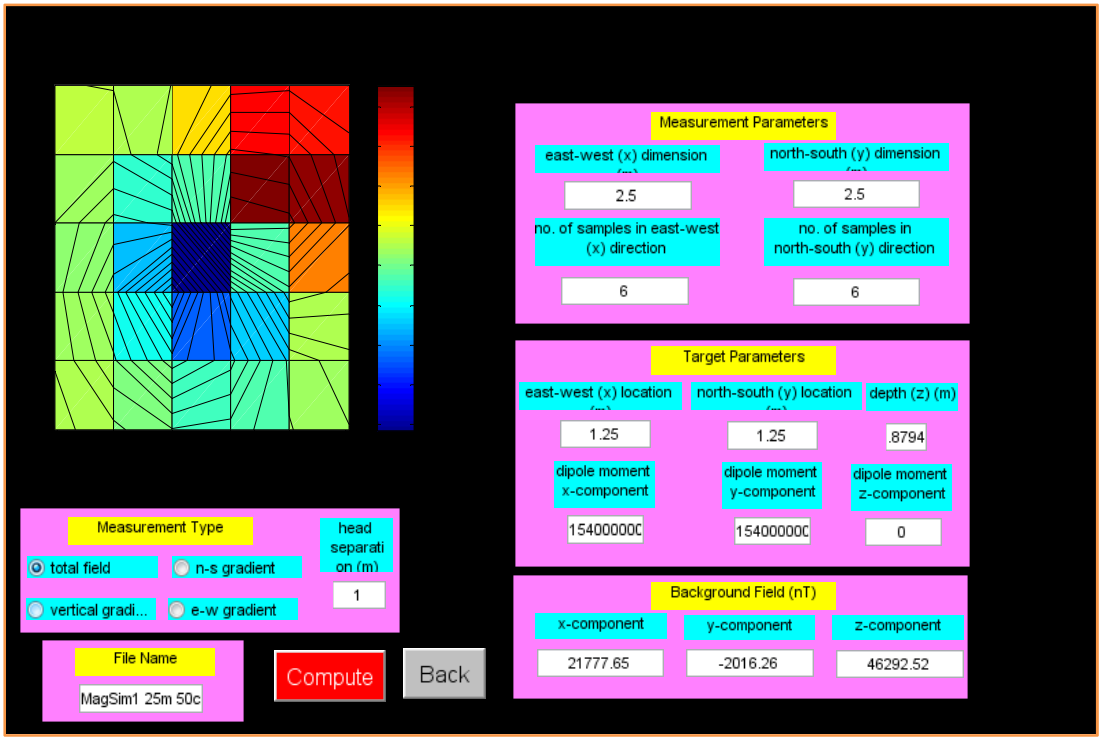


Figure 6. Geophysica program GUI with parameters selected, as well as the resulting data.

example of the *Geophysica* program GUI with parameters selected, as well as the resulting data.

3.2.2 Initial Method for Testing

Synthetic geophysical data for all techniques utilized in this study (GPR, Magnetic, EM-31) are generated over grids of varying size (2.5m x 2.5m, 5m x 5m, 10m x 10m, and 20m x 20m) using the pre-existing graphical user interfaces described in Section 3.2.1. Sample spacing of each grid also is varied (10cm, 25cm, and 50cm). Additionally, modeled targets will be a solid stainless steel sphere and vary in diameter (25cm, 50cm, 1m) buried at 1-m depth.

Once generated, the synthetic data sets are normalized and merged by means of summation (i.e. simply adding each corresponding data point—matching XY locations—across the grid) to get an initial understanding of the variables needed for further investigation. Furthermore, the magnetic data undergoes a reduction-to-pole process. Varying grid size and sample spacing determines how much of an influence the "rarity" of the target is in the integration process (i.e. number of cells with target vs. cells without target is significantly lower). Data sets are statistically analyzed to determine the variable in the integration process that has the highest significance and also the combination of variables that produces the "best" representation of the true target. Statistical analysis included spatial analysis to compare the modeled location of the target to the data represented location of the target; logistic regression; chi-square test; and frequency tables to show the accuracy of the model in displaying the true location of the target. Data

sets are visually presented via maps created in *Surfer*. **Figure 7** gives a representation of the initial findings.

3.2.3 Revised Method for Testing

3.2.3.1 Creating a Model and Data Manipulation Scheme

From the initial results described in Section 3.2.2, it is determined that the optimal grid size to model is 5m by 5m. This is due to the "rarity" issue, therefore rendering the logistic regression (transformation of normalized values according to some threshold) to not function properly. The optimal sample spacing for modeling is determined to be 10cm. Additionally, data sampling heterogeneities among techniques have been confirmed as a problem in data integration methods.

This stage of testing involves a new set of synthetic data with the grid size and sample spacing fixed, but with other variables expanded to evaluate their significance using a more refined, quantitative approach. Simulated targets were a solid spheres of various material types (cement, iron, stainless steel, and plastic), that are representative of those targets found at the control site case study discussed in Part 2. Targets vary in diameter (12.5cm, 25cm, 50cm, 1m, 2m), depth buried in "wet soil" background conditions (25cm, 50cm, 75cm, 1.25m, 1.75m, 2.5m, 3.5m) and are located for the simulation at the center of each grid (X=2.5m, Y=2.5m); this is summarized in **Table 1**.

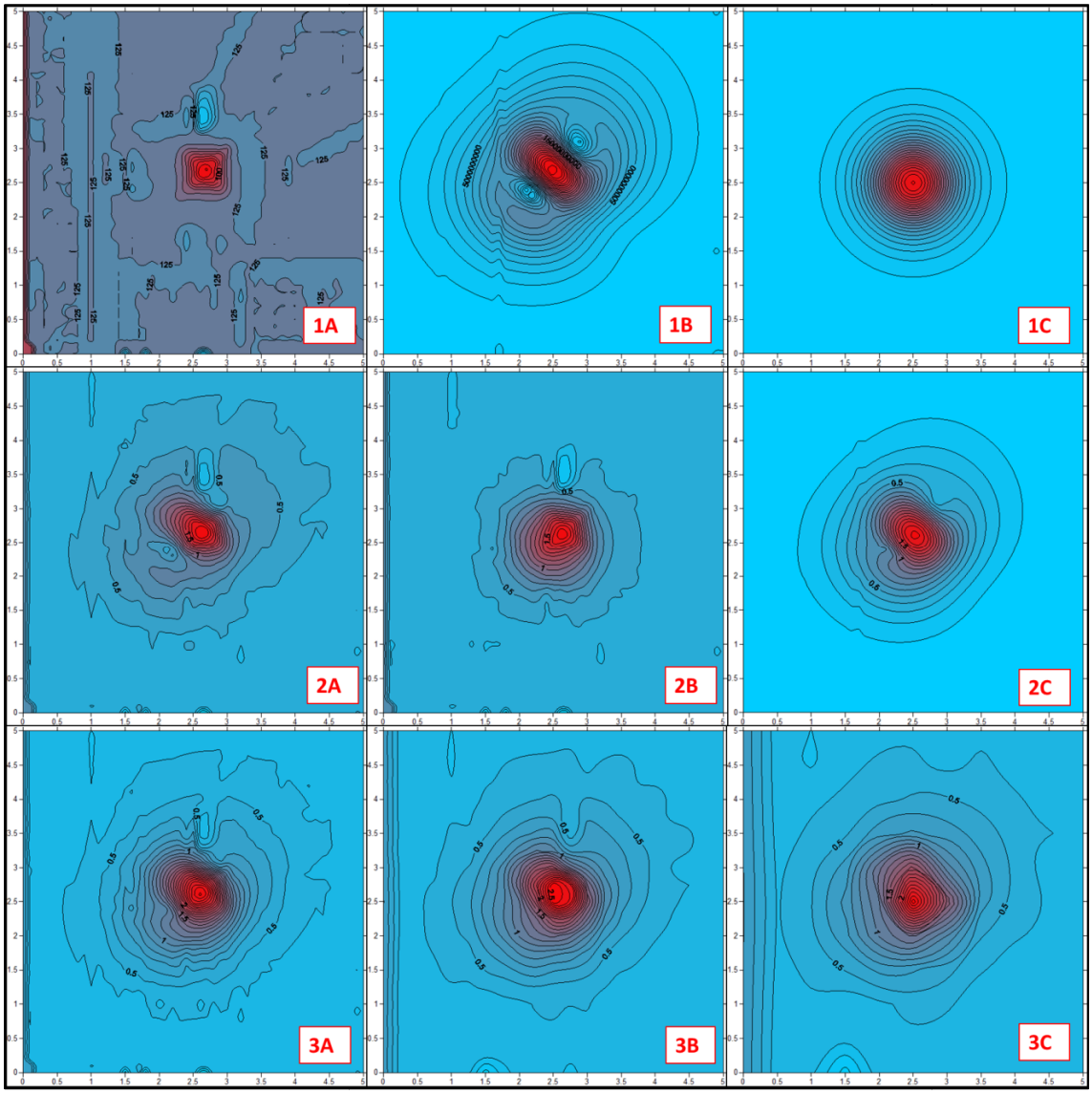


Figure 7. Representative outputs of modeled data from initial testing. Grouping 1 shows the modeled (A) GPR data, (B) Magnetic data, and (C) EM-31 Ground Conductivity data with a 10cm sample spacing. Grouping 2 shows the modeled data [from Grouping 1] with a 10cm sample spacing with (A) Magnetic and GPR data combined, (B) EM-31 and Ground Conductivity data combined, and (C) Magnetic and EM-31 Ground Conductivity data combined. Grouping 3 shows all three modeled data sets combined with a (A) 10cm sample spacing, (B) 25cm sample spacing, and (C) 50cm sample spacing.

Table 1. Summation of Variables Tested for Synthetic Data Sets

Variable Type	Variable Tested
Geophysical Technique	Magnetic Gradiometry, Ground Conductivity, Ground Penetrating Radar
Material Type	Concrete, Iron, Plastic, Stainless Steel
Depth of Burial	50cm, 75cm, 125cm, 175cm, 250cm, 350cm
Diameter of Target	12.5cm, 25cm, 50cm, 100cm, 200cm

The *Geophysica* program outputs the modeled data in a map-view grid, and 140 EM-31 (ground conductivity) and 140 magnetic models are produced. The GprMax2D program only produces 2D data, of which all data lines are manually inserted into a grid when importing into *Microsoft Access*, *Excel*, and *Surfer* for further analysis. Taking into account a sample spacing of 10cm, a total of 7,140 GPR simulations are run to generate the equivalent of 140 grids.

Considering that in some cases this systematic methodology results in unrealistic scenarios, such as a 3.5m diameter target buried at 25cm, those combinations are selectively removed from all evaluation schemes, leaving a total of 116 simulation results to consider for each geophysical technique; the scenario was unrealistic because the object would be partially exposed. We recommend utilizing a high processing computer (HPC) system to run the GPR models, as each will take a considerable amount of time when executed on a standard laptop. Setting up all GPR models to run through the University of Tennessee's Newton HPC Program cut processing time down to just over 4 weeks for all 7,140 models, compared to an estimated 2.5 years if run on a standard laptop computer.

Once all models were completed and the data points were exported into *Excel* and given their respective X-Y assignments within the grid, the data are normalized with the following formula:

(1) $T^* = (d_i - d_{min})/(d_{max} - d_{min}) = 0\text{-}1$

Whereas, T = Normalized value

di = Original data value at the xy location being calculated

dmax = Maximum data value across entire grid

dmin = Minimum data value across entire grid

In the interest of efficiency, a representative model for each material type is selected for the quantitative aspects of integrating data sets together. For any subsurface feature to be enhanced within the data (i.e. a map showing anomalies indicates possible locations of targets), the boundary between the area containing the target and not containing the target should be as sharp as possible. This, in theory, will reduce the uncertainty there is concerning the shape of the target, depth and orientation (i.e. horizontal or vertical and at what azimuth) it is buried, and the size of the target.

To determine which of the simulations is most representative to illustrate the sharpest boundary, a line graph is created plotting the maximum amplitude values for the data line Y=2.5 along the entire X range (0.0-5.0) of the simulated data grid. The slope of the curve for this data plot is calculated, with the median slope (and associated model) chosen as representative (as shown in **Figure 8**); a number of models on each side of the median were also noted. Considering that the curve is non-linear, the portion of the curve chosen for each slope calculation is where the slope is highest. This was conducted for all

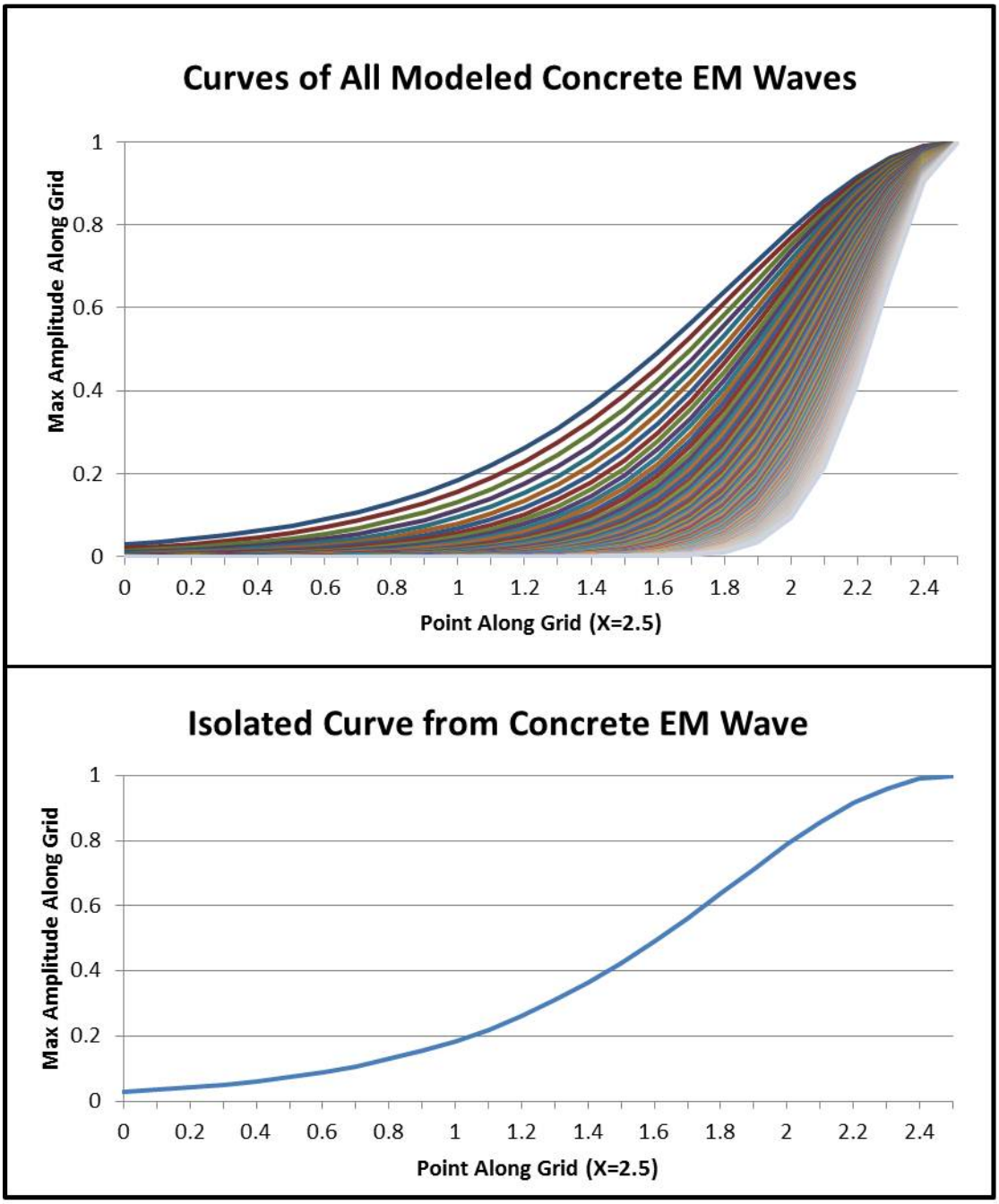


Figure 8. Representative sample of curves extracted from model outputs to determine the representative model used for each material type. Curves displayed are for all concrete curves used to find median slope (top) and an isolated curve (bottom).

material types and all geophysical techniques simulated for this study. Subsequently, the models selected for each technique are evaluated across material types, and the closest fit is selected to be the designated parameter for that material type. For example, if the model that was 50 cm in diameter and buried 1 m was within the range of median values for all three techniques (GPR, EM ground conductivity, and magnetics) then that specific simulation would be the designated simulation.

Synthetic data were merged in different ways using a number of mathematical functions applied to them in the process. These functions, referred to hereafter as "schemes," include simple addition, multiplying by some constant, and raising each data point to some exponent. With the addition scheme, the normalized values for each geophysical technique are added together. This is the simplest of all schemes being evaluated, and represents what typically is done (currently) when combining datasets. For example, Z= GPR+ MAG + EM.

In regard to the multiplication schemes, each of the three techniques will be multiplied by a constant ranging from 1.0 to 10.0 at an increment of 0.25, and when added together, will result in 50,653 total combinations. For example, a possible multiplication scheme could be Z= GPR(5.25) + MAG(9.75) + EM(2.50). To determine the exponent that is most appropriate for the data, the slope of the curve from each designated model is visited again. As the normalized values are raised to some selected power in the sequence (ranging from 1.0 to 50.0 with a 0.1 increment), the higher the exponent applied to the data, the greater the slope, as expected. Consequently, as the slope increases (thus sharpening the boundary around the target) when applying the

increasing exponents, the rate at which the slope increases is decreasing. Once that rate reaches some asymptote, it is assumed that value is the optimal exponent to apply to the data (as shown in **Figure 9**). **Table 2** displays the calculated optimal exponent for each of the material types and geophysical technique. An example of one exponentiation scheme is Z= GPR^(4.525) + MAG^(4.3) + EM^(3.125). When evaluating each of the eleven exponential scheme, both the optimal exponent and averages of the optimal exponents of each technique are utilized.

3.2.3.2 Applying the Data Manipulation Schemes

To effectively organize the synthetic data, carry out calculations, and analyze all 50, 674 different data manipulation schemes, *Microsoft Access* was utilized. Once all processing steps were completed, an extensive amount of data was created; the resulting master database exceeds two billion records and also incorporates simulated global positioning system (GPS) coordinates. Due to the size and quantity of data created by this system, sample database systems were populated to simplify explanations and give clarification on how data is being managed, manipulated and created.

The *Materials Data System* (MDS) is a database system I created to process 4 different types of material data sets (concrete, iron, plastic, and stainless steel) containing EM-31 Ground Conductivity (EM), GPR, and Magnetic (MAG) data that are combined using the simulated GPS coordinates to establish a common data link between the different data sets. The materials database system is comprised of 50,674 individual data tables (referred to as *geoData*, see **Figure 10**) that identify data points with a record ID,

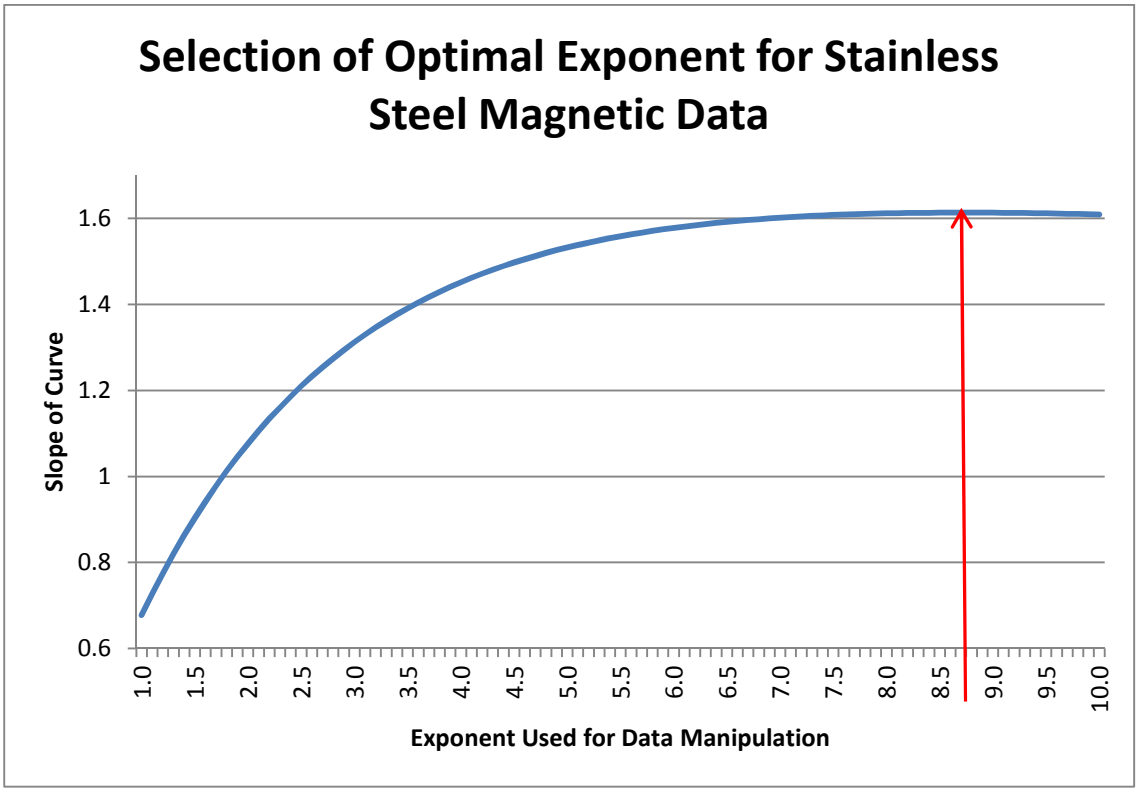


Figure 9. Representative curve indicating what the optimal exponent is for stainless steel magnetic data. The normalized data values are raised to some exponent; as the slope of the newly calculated curves increases, the rate at which the slope increases is subsequently decreasing. Once that rate reaches some asymptote, it is assumed that value is the optimal exponent to apply to the data for future manipulation schemes.

Table 2. Calculated optimal exponent for each geophysical technique and material type. This table was generated in conjunction with curves as shown in Figure 9.

	EM	**Magnetic**	**GPR**	*Average*
Concrete	2.6	3.3	3.3	*3.067*
Iron	2.6	3.5	3.5	*3.200*
Plastic	2.6	1.7	6.6	*3.633*
Stainless	4.7	8.7	4.7	*6.033*
Average	*3.125*	*4.3*	*4.525*	

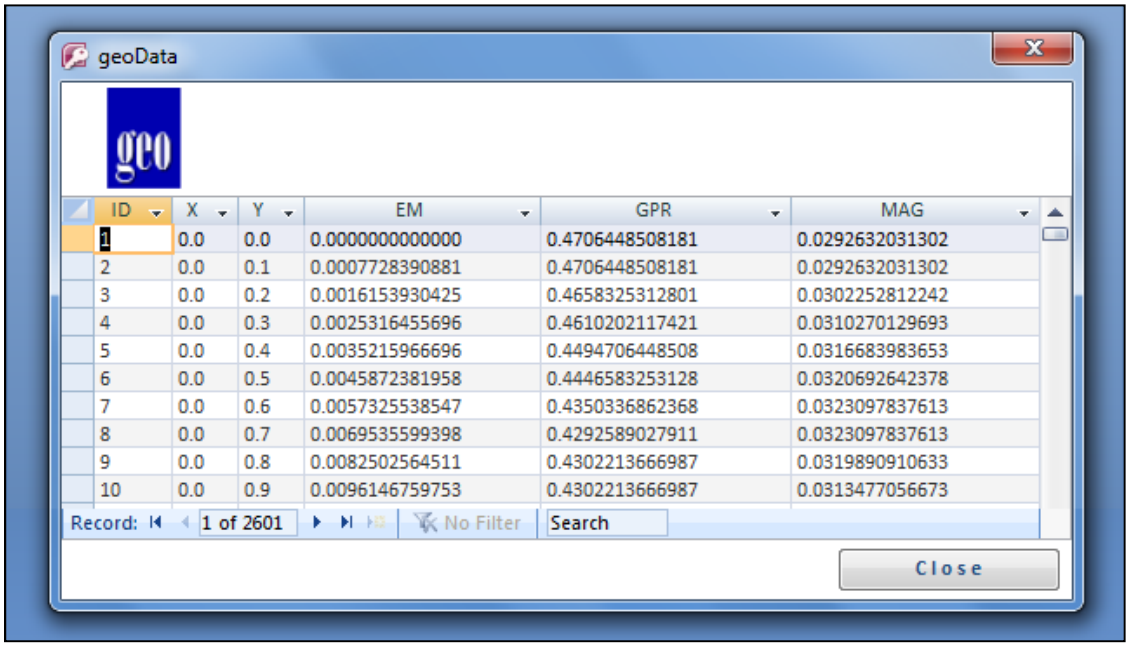


Figure 10. Representative example of the *geoData* table, displaying the EM, GPR, and MAG data associated with each data point within the grid.

X and Y coordinate and the EM, GPR, and MAG amplitude values for all 2,601 data points for each grid (i.e. the individual output from each data manipulation scheme for the designated representative model of that specified material type).

The MDS is also used to maintain data tables that contain the EM, GPR, and MAG data, along with a series of queries and tables created to establish where the location of the modeled target is located over the grid. Once each data manipulation scheme is applied to the normalized simulated data, the newly calculated data points can be compared to the original target location to determine if the modeled target has been enhanced with sharper boundaries to identify the target with a greater degree of certainty.

To determine certainty, a number of tables are integrated together. First, the *geoData TG* table is created and for the 2,601 data points of each grid when the data point is over the modeled target,it is designated a "1" value., and if it is not over the target it is designated a "0" value. **Figure 11** is a screen capture of a portion of the *geoData TG* table.

Once the target has been established for every simulated GPS coordinates on the *geoData* table, the next step is to process each coordinate's EM, GPR, and MAG data amplitude with a series of expressions defined in the *geoExpression* query (see **Figure 12**). The *geoExpressions* table creates a mathematical expression and also generates every possible unique combination for that expression totaling 50,674 expressions, utilizing the following equations:

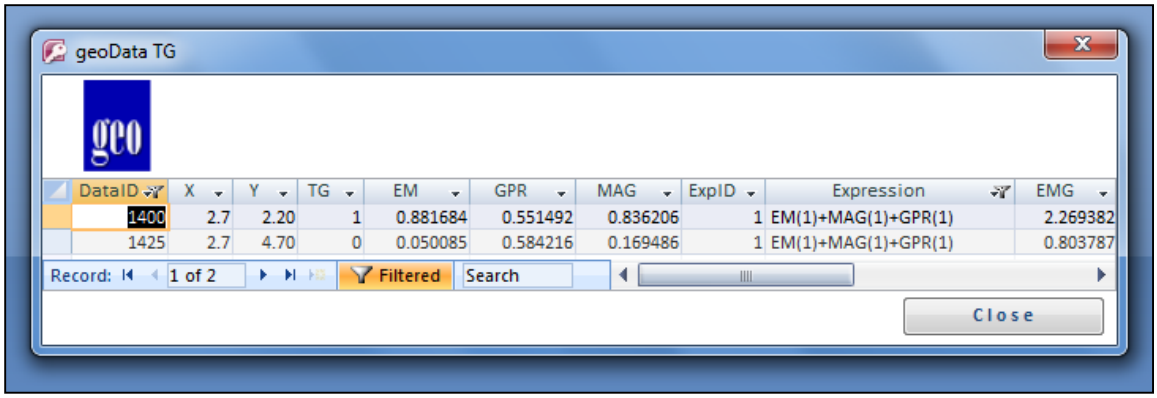


Figure 11. Representative example of the *geoData TG* table. This table incorporates the different data manipulation schemes applied to the EM, GPR, and MAG data for each data point within the grids.

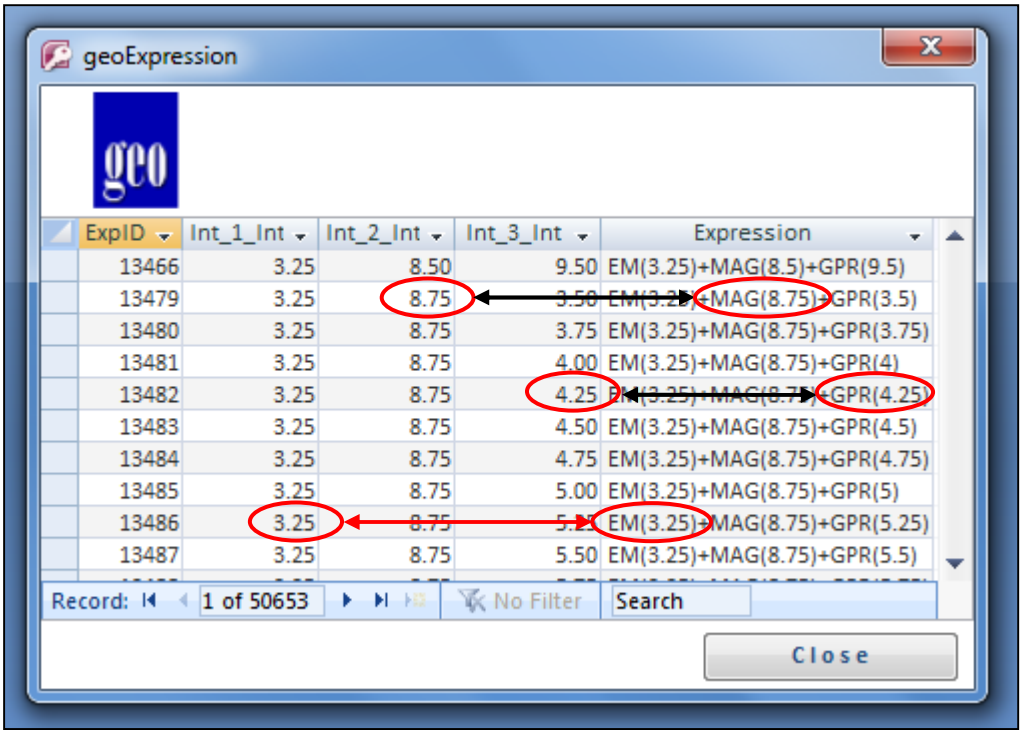


Figure 12. Representative example of the *geoExpression* table. Queries within this table creates 50,653 expressions for the multiplication data manipulation schemes.

(1) EMG = (EM * x) + (GPR * y) + (MAG * z)

Where: x= Int_1_INT, y= Int_2_INT and z= Int_3_INT

(2) EMG = (EM * Int_1_INT) + (MAG * Int_2_INT) + (GPR * Int_3_INT)

The intervals for x(Int_1_INT), y(Int_2_INT), and z(Int_3_INT) are 1.0 to 10.0 with a 0.25 interval for a total of 37 intervals for each x, y, and z. By combining x, y and z (37 x 37 x 37), the *geoExpression* queries created 50,653 possible expressions under the multiplication data manipulation schemes; the exponentiation and addition manipulation schemes are added to this database via another table set up in a similar format. The *geoExpression* query is created to be integrated with *geoData* tables, and both are combined in the *geoData TG* query. The MDS uses the *geoData* tables and are processed with *geoExpression* query for a total record output of 131,803,074 (2,601 records per grid x 50,674 expressions) for each material. When considering each material has its' own designated *geoExpressions* query system, a total of 527,212,296 records are produced.

Once each expression has been applied and the EMGMx (EM+MAG+GPR) field has been created, the materials database system uses the *geoData Max* query (see **Figure 13**) to create a database for what each maximum calculated value is within each individual grid (of the 2,601 data points). Another table is created to assign thresholds to this maximum value, identified as *geoData Accuracy Percentage*. This additional table aids in transforming the calculated data points into a "1" or "0," comparing these new data points to the actual location of the modeled target, and determining which of the

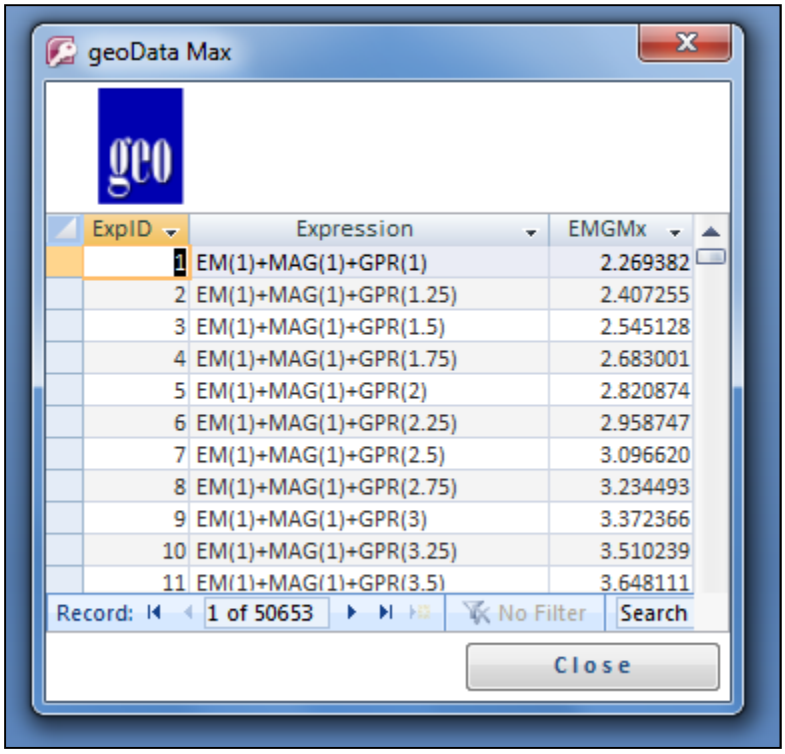


Figure 13. Representative example of the *geoData Max* query table. This table isolates the maximum value calculated across each grid and data manipulation scheme, shown as the EMGMx value.

50,674 data manipulation schemes result in the greatest accuracy to identify target locations. The databases *geoData Max* and *geoData Acccuracy Percentage* are used to integrate with the *geoExpressions* query system and generate a new database referred to as *geoData Target 01*.

The *geoData Target 01* query processes every record on the *geoData TG* query (131,803,074 records for each material database) and using a combination of equations and statements, identifies the percentage accuracy for each EMG value generated by the *geoData TG 01* query. To establish the percentage of the EMG, the *geoData TG 01* query uses the formula:

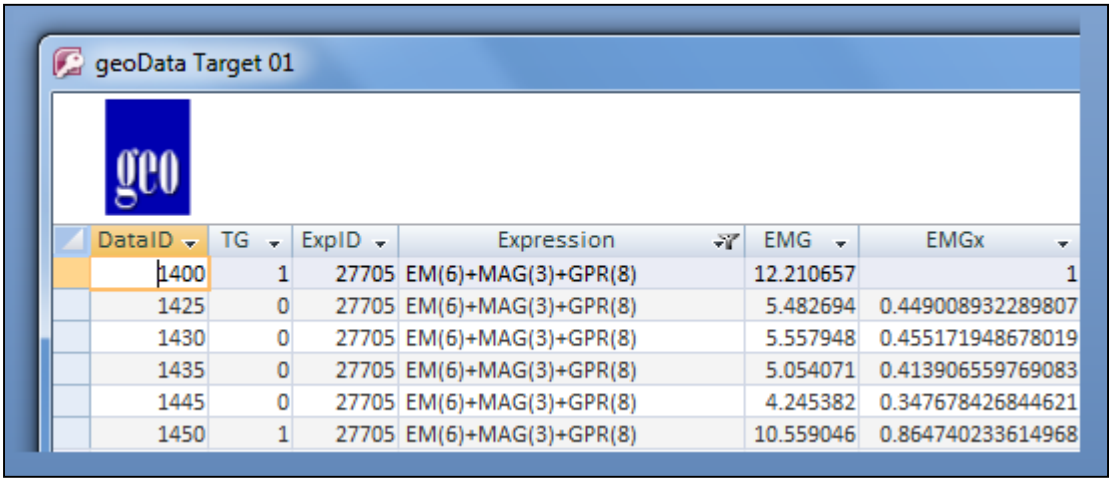
(1) EMGx = EMG/EMGMx

where,

EMG = Calculated value of associated data point once manipulation scheme is applied

EMGMx = Maximum value calculated within the associated grid (i.e. value found in the *geoData Max* table)

When the EMG and EMGMx are the same it will result in a 1.00, representing 100% of the grid and is expected to be the location of the modeled target (see **Figure 14a**). All other values for the EMGx field represent what percentage of the grid's maximum value that data point consists of; the closer to 0.00 the EMGx value is, the further away from the modeled target it is.



DataID	TG	ExpID	Expression	EMG	EMGx
1400	1	27705	EM(6)+MAG(3)+GPR(8)	12.210657	1
1425	0	27705	EM(6)+MAG(3)+GPR(8)	5.482694	0.449008932289807
1430	0	27705	EM(6)+MAG(3)+GPR(8)	5.557948	0.455171948678019
1435	0	27705	EM(6)+MAG(3)+GPR(8)	5.054071	0.413906559769083
1445	0	27705	EM(6)+MAG(3)+GPR(8)	4.245382	0.347678426844621
1450	1	27705	EM(6)+MAG(3)+GPR(8)	10.559046	0.864740233614968

Figure 14a. Representative example of *geoData Target 01* table with the EMGx Field completed. The "1" on the top line shows that this is the point within the grid that is expected to be directly over the buried target.

Once the EMGx has been establish, the *geoData Target 01* query incorporates the *geoData Accuracy Percentage* table to transform the EMGx value 0.00 into a "1" or "0" (illustrated in **Figure 14b**) using the following equations:

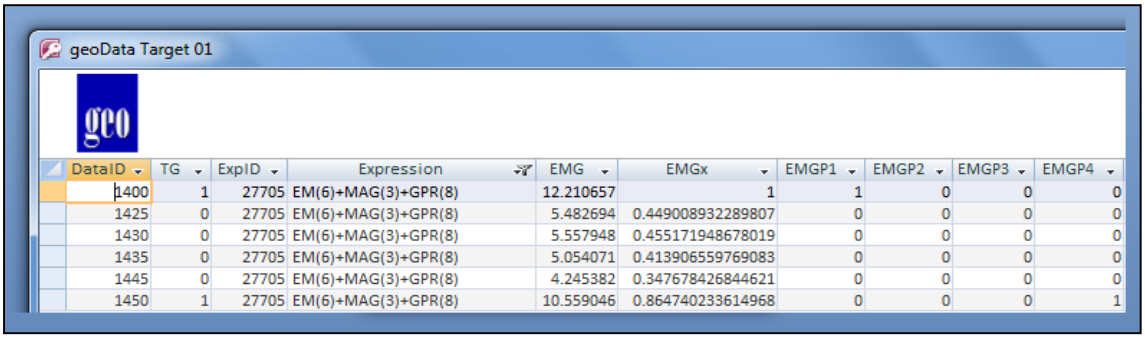
(2) EMGP1: If (EMG/Max of EMGMx) > 99%(PP1), then EMGP1 =1 if false 0

(3) EMGP2: If (EMG/Max of EMGMx) between 99%(PP1) and 95%(PP2), then EMGP2 =1 if false 0

(4) EMGP3: If (EMG/Max of EMGMx) between 95%(PP2) and 90%(PP3), then EMGP3 =1 if false 0

(5) EMGP4: If (EMG/Max of EMGMx) between 90%(PP3) and 85%(PP4), then EMGP4 =1 if false 0

After the data points have been transformed according to the thresholds applied, the *geoData Target 01* uses the preceeding formulae to yield an accuracy value for each threshold for every record on the *geoData* by comparing the newly transformed "0" and "1" values (i.e. calculated to be over the modeled target or not) and the actual location of the modeled target, also designated with "0" and "1" values. If these two data sets match, it is given a new value of "1" and if they do not match, it is given a new value of "0."



DataID	TG	ExpID	Expression	EMG	EMGx	EMGP1	EMGP2	EMGP3	EMGP4
1400	1	27705	EM(6)+MAG(3)+GPR(8)	12.210657	1	1	0	0	0
1425	0	27705	EM(6)+MAG(3)+GPR(8)	5.482694	0.449008932289807	0	0	0	0
1430	0	27705	EM(6)+MAG(3)+GPR(8)	5.557948	0.455171948678019	0	0	0	0
1435	0	27705	EM(6)+MAG(3)+GPR(8)	5.054071	0.413906559769083	0	0	0	0
1445	0	27705	EM(6)+MAG(3)+GPR(8)	4.245382	0.347678426844621	0	0	0	0
1450	1	27705	EM(6)+MAG(3)+GPR(8)	10.559046	0.864740233614968	0	0	0	1

Figure 14b. Representative example of the *geoData Target 01* query – EMGx percentage level will be transformed into a "0" (not over target) or "1" (over the target) according to various calculated thresholds.

Ideally, we want to have 100% accuracy for those data points calculated as a "1" and are actually a "1" (avoiding Type I errors); we also want 100% accuracy for the "0" to "0" comparisons (avoiding Type II errors). **Figure 14c** depicts the breakdown of these accuracies, with the cells populated using the following equations:

(6) P1_1_1: If Target(TG)=1 And ([emgp1] =1 then P1_1_1=1 if false 0

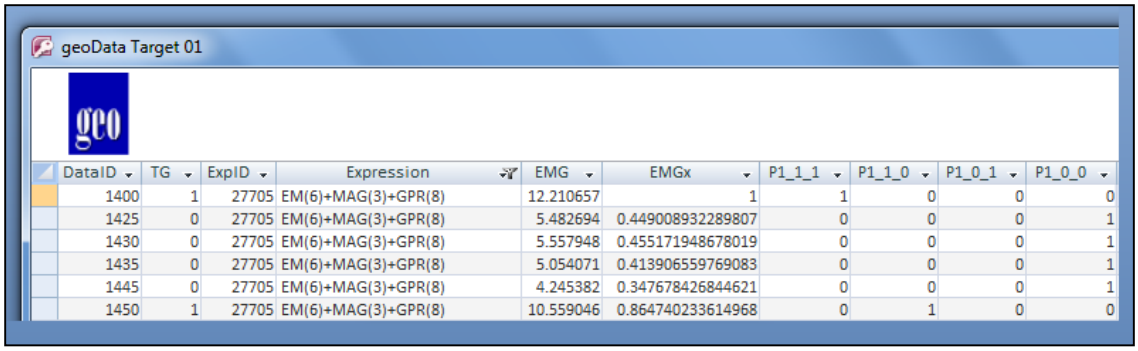
(7) P1_0_1: If Target(TG)=0 And ([emgp1] =1 then P1_1_1=1 if false 0

(8) P1_1_0: If Target(TG)=1 And ([emgp1] =0 then P1_1_1=1 if false 0

(9) P1_0_0: If Target(TG)=0 And ([emgp1] =0 then P1_1_1=1 if false 0

These equations are duplicated for each of the thresholds, with distinction given to each P1, P2, P3 and P4 correlations within the master database.

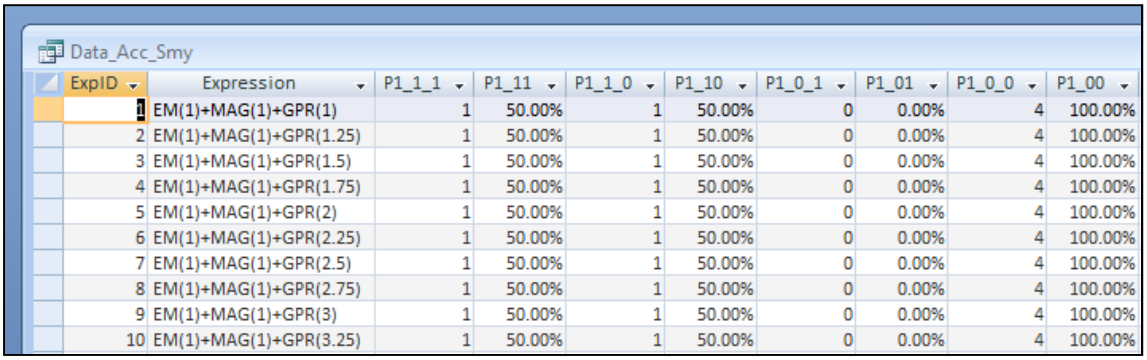
Finally, the output of the *geoData Target 01* query is used to create the *geoData Target 01 Accuracy* query. The *geoData Target 01 Accuracy* query is the summary of the *geoData Target 01* query and is also used to create a "percentage accuracy" score for all 50,674 expressions (i.e. data manipulation schemes). The accuracy is a simple calculation of the proportion of the frequency tables generated in the stage of the database. For example, if there are 121 data points that are correctly co-located the modeled target and given a distinction of "1" but only 97 of these data points are actually calculated as a "1" then there would be an 80.16% accuracy. Within the table illustrated in **Figure 15**, the column headings can be deciphered with the following explanations:



geoData Target 01

DataID	TG	ExpID	Expression	EMG	EMGx	P1_1_1	P1_1_0	P1_0_1	P1_0_0
1400	1	27705	EM(6)+MAG(3)+GPR(8)	12.210657	1	1	0	0	0
1425	0	27705	EM(6)+MAG(3)+GPR(8)	5.482694	0.449008932289807	0	0	0	1
1430	0	27705	EM(6)+MAG(3)+GPR(8)	5.557948	0.455171948678019	0	0	0	1
1435	0	27705	EM(6)+MAG(3)+GPR(8)	5.054071	0.413906559769083	0	0	0	1
1445	0	27705	EM(6)+MAG(3)+GPR(8)	4.245382	0.347678426844621	0	0	0	1
1450	1	27705	EM(6)+MAG(3)+GPR(8)	10.559046	0.864740233614968	0	1	0	0

Figure 14c. Representative example of *geoData Target 01* query where the transformed "0" and "1" cells (*seen in Figure 14b*) are compared for accuracy against the actual target locations of the modeled object.



Data_Acc_Smy

ExpID	Expression	P1_1_1	P1_11	P1_1_0	P1_10	P1_0_1	P1_01	P1_0_0	P1_00
1	EM(1)+MAG(1)+GPR(1)	1	50.00%	1	50.00%	0	0.00%	4	100.00%
2	EM(1)+MAG(1)+GPR(1.25)	1	50.00%	1	50.00%	0	0.00%	4	100.00%
3	EM(1)+MAG(1)+GPR(1.5)	1	50.00%	1	50.00%	0	0.00%	4	100.00%
4	EM(1)+MAG(1)+GPR(1.75)	1	50.00%	1	50.00%	0	0.00%	4	100.00%
5	EM(1)+MAG(1)+GPR(2)	1	50.00%	1	50.00%	0	0.00%	4	100.00%
6	EM(1)+MAG(1)+GPR(2.25)	1	50.00%	1	50.00%	0	0.00%	4	100.00%
7	EM(1)+MAG(1)+GPR(2.5)	1	50.00%	1	50.00%	0	0.00%	4	100.00%
8	EM(1)+MAG(1)+GPR(2.75)	1	50.00%	1	50.00%	0	0.00%	4	100.00%
9	EM(1)+MAG(1)+GPR(3)	1	50.00%	1	50.00%	0	0.00%	4	100.00%
10	EM(1)+MAG(1)+GPR(3.25)	1	50.00%	1	50.00%	0	0.00%	4	100.00%

Figure 15. Representative example of the *geoData Accuracy Summary* query. This table displays the accuracy (by frequency of occurrence and percentage) of each data manipulation scheme according to threshold.

P1_1_1: Threshold 1 (99%) total number (i.e. frequency) of data points calculated as a "1" that are in truth the modeled "1" location

P1_11: Percentage of P1_1_1 compared to the total number of modeled "1" locations

P1_1_0: Threshold 1 (99%) total number (i.e. frequency) of data points calculated as a "1" that are in truth the modeled "0" location; also referred to as false positives.

P1_10: Percentage of P1_1_0 compared to the total number of modeled "1" locations

P1_0_1: Threshold 1 (99%) total number (i.e. frequency) of data points calculated as a "0" that are in truth the modeled "1" location; also referred to as false negatives.

P1_01: Percentage of P1_1_1 compared to the total number of modeled "1" locations

P1_0_0: Threshold 1 (99%) total number (i.e. frequency) of data points calculated as a "0" that are in truth the modeled "0" location

P1_00: Percentage of P1_1_0 compared to the total number of modeled "0" locations

3.2.4 Statistical Analysis

Performing a complete and appropriate suite of statistical analyses of the synthetic geophysical data is crucial to meeting the objectives of this study, in terms of evaluating the effectiveness of the data integration methodologies and applications. The statistics have a qualitative and a quantitative component. While it is the ultimate intent to develop a quantitative-only method for integrating multiple geophysical data sets, one must first evaluate the ability of each technique to "detect" various targets in a qualitative sense. To accomplish this task, more general statistics are explored first by correlating the simulated geophysical data produced within the *GprMax2D* and *Geophysica* programs with the known "targets" that were modeled. Part II (Section 4) expands the statistical treatment to include a more sophisticated suite of approaches for evaluating the effectiveness of the merging methodologies on real-world case-study data. To effectively evaluate the manipulation schemes that result in the highest percent accuracy (i.e. minimizing both Type I and Type II errors) and to identify the optimal threshold, the program SAS is utilized. Due to the overwhelmingly large number of records that will need to be evaluated, *Statistical Analysis System* (SAS) software is the most appropriate tool for determining statistical significance among the results.

All of the 50,674 data manipulation scheme outputs in terms of percentage accuracy for the P1_1_1, P1_1_0, etc. are all ordered with percent accuracy for the 1:1 comparisons ranked from highest to lowest. This can be seen with a representative sample from the master database in **Table 3** for the Stainless Steel simulation scheme.

Table 3. Representative example table to display the ranking of data manipulation schemes for the modeled stainless steel data. Ranking is based on highest accuracy for the cells [across the grid] that contain the target and are calculated to contain the target (1:1).

Rank	Expression	Count 1:0	Accuracy	Count 1:1	Accuracy	Count 0:1	Accuracy	Count 0:0	Accuracy
1	Add Avg Exp EM_GPR	0	0.00%	25	100.00%	2	0.08%	2574	99.92%
1	Add Avg Exp GPR_MAG	0	0.00%	25	100.00%	0	0.00%	2576	100.00%
1	Add EM_GPR Opt Exp	0	0.00%	25	100.00%	0	0.00%	2576	100.00%
1	Add GPR_MAG Opt Exp	0	0.00%	25	100.00%	0	0.00%	2576	100.00%
1	ALL ADD Norm	0	0.00%	25	100.00%	3	0.12%	2573	99.88%
1	ALL ADD Opt Exp	0	0.00%	25	100.00%	0	0.00%	2576	100.00%
1	Exp Avg GPR	0	0.00%	25	100.00%	4	0.16%	2572	99.84%
1	Exp Opt GPR	0	0.00%	25	100.00%	4	0.16%	2572	99.84%
2	EM(1)+MAG(1)+GPR(7.75)	1	0.83%	120	99.17%	69	2.78%	2411	97.22%
2	EM(1)+MAG(1)+GPR(8)	1	0.83%	120	99.17%	95	3.83%	2385	96.17%
2	EM(1)+MAG(1)+GPR(8.25)	1	0.83%	120	99.17%	121	4.88%	2359	95.12%
2	EM(1)+MAG(1)+GPR(8.5)	1	0.83%	120	99.17%	156	6.29%	2324	93.71%
2	EM(1)+MAG(1)+GPR(8.75)	1	0.83%	120	99.17%	198	7.98%	2282	92.02%
2	EM(1)+MAG(1)+GPR(9)	1	0.83%	120	99.17%	252	10.16%	2228	89.84%
2	EM(1)+MAG(1)+GPR(9.25)	1	0.83%	120	99.17%	320	12.90%	2160	87.10%
2	EM(1)+MAG(1)+GPR(9.5)	1	0.83%	120	99.17%	385	15.52%	2095	84.48%
2	EM(1)+MAG(1)+GPR(9.75)	1	0.83%	120	99.17%	438	17.66%	2042	82.34%
2	EM(1)+MAG(1)+GPR(10)	1	0.83%	120	99.17%	496	20.00%	1984	80.00%
2	EM(1)+MAG(1.25)+GPR(9)	1	0.83%	120	99.17%	88	3.55%	2392	96.45%
2	EM(1)+MAG(1.25)+GPR(9.25)	1	0.83%	120	99.17%	114	4.60%	2366	95.40%
…	…	…	…	…	…	…	…	…	…
50669	EM(10)+MAG(3)+GPR(1)	30	25%	91	75%	0	0%	2480	100%
50670	Exp Avg EM	8	32.00%	17	68.00%	0	0.00%	2576	100.00%
50671	Exp Opt EM	12	48.00%	13	52.00%	0	0.00%	2576	100.00%
50672	Norm MAG	13	52.00%	12	48.00%	10	0.39%	2566	99.61%
50673	Add EM_MAG Norm	16	64.00%	9	36.00%	6	0.23%	2570	99.77%
50674	Norm EM	16	64.00%	9	36.00%	12	0.47%	2564	99.53%

All 50,674 sets of results are included in the statistical suite for the first pass, although separated out by threshold (i.e. only those results with a 99% threshold are compared to each other). When it is found that there is a statistical significance between the results, the grouping is divided and all statistics are run again. This is repeated until the grouping is small enough to where there is no statistical significance between all evaluated data manipulation schemes producing similar accuracies in identifying the location of the modeled target. Within this grouping no statistical significance exists and any of the variables for that range of data manipulation schemes can be selected to enhance the subsurface features of the simulations with similar results. Statistical methods employed during this stage of the research are *logistical regression*, *proc GLM*, *proc frequency*, and *univariate* as described in Sections 3.2.4.1-3.2.4.4, respectively

3.2.4.1 Logistical Regression

Binary responses (for example, the presence or absence of a target) set up logistic regression analysis to be the most appropriate means to investigate the relationship between these discrete responses and a set of explanatory variables. The logistic procedure fits linear logistic regression models for discrete response data by the method of maximum likelihood (Lado et al. 2008). The process can also be used to perform conditional logistic regression for binary response data and exact conditional logistic regression for binary response data. The logistic procedure enables the user to specify categorical variables—also known as classification or class variables—or continuous variables as explanatory variables. Any term specified in the model is referred to as an

"effect" whether it is a continuous variable, a class variable, an interaction, or a nested term. An effect in the model that is not an interaction or a nested term is referred to as a "main effect." When there are no interaction terms, a main effect can enter or leave a model in a single step based on the p-value of the score or Wald statistic (Lado et al. 2008). *Logistic regression* was used to validate the coding within *Microsoft Access* in transforming the normalized data points into a binary response format.

3.2.4.2 Proc GLM

The GLM procedure uses the method of least squares to fit general linear models. Among the statistical methods available in *proc GLM* are regression, analysis of variance, analysis of covariance, multivariate analysis of variance, and partial correlation. *Proc GLM* is used to analyze data within the framework of general linear models; it also handles models varying from one or several continuous dependent variables to one or several independent variables. The independent variables may be either classification variables,that divide the observations into discrete groups, or continuous variables. *Proc GLM* enables the user to specify any degree of interaction (crossed effects) and nested effects. It also provides for polynomial, continuous-by-class, and continuous-nesting-class effects. Through the concept of estimability, the GLM procedure can provide tests of hypotheses for the effects of a linear model regardless of the number of missing cells or the extent of confounding (Davis 2002). *Proc GLM* results in displays of the sum of squares (SS) associated with each hypothesis tested and, upon request within the

software, the form of the estimable functions employed in the test. *Proc GLM* can be used to produce the general form of all estimable functions.

3.2.4.3 Proc Frequency

The *proc Freq* procedure produces one-way to n-way frequency and contingency (crosstabulation) tables (Rahman et al. 2008). For two-way tables, this statistical method computes tests and measures of association. For n-way tables, a stratified analysis is provided by computing statistics across, as well as within, strata. For one-way frequency tables, a goodness-of-fit test is computed for equal proportions or specified null proportions, and confidence limits and tests for binomial proportions are provided, including tests for noninferiority and equivalence. For contingency tables, this method can be used to compute various statistics to examine the relationships between two classification variables. For some pairs of variables, the user may want to examine the existence or strength of any association between the variables. To determine if an association exists, chi-square values are computed. To estimate the strength of an association, *proc Freq* is used to compute measures of association that tend to be close to zero when there is no association and close to the maximum (or minimum) value when there is perfect association. *Proc Freq* is also used to compute asymptotic standard errors, confidence intervals, and tests for measures of association and measures of agreement. Exact p-values and confidence intervals are available for many test statistics and measures. This test also performs analyses that adjust for any stratification variables by computing statistics across, as well as within, strata for n-way tables.

3.2.4.4 Univariate

The *univariate* procedure provides several important pieces of information to compute summary statistics. The most appropriate applications of this statistical test is that it provides (1) descriptive statistics based on moments (including skewness and kurtosis), quantiles or percentiles (such as the median), frequency tables, and extreme values, (2) histograms that optionally can be fitted with probability density curves for various distributions and with kernel density estimates, (3) cumulative distribution function plots (cdf plots). Optionally, these can be superimposed with probability distribution curves for various distributions (4), such as quantile-quantile plots (Q-Q plots), probability plots, and probability-probability plots (P-P plots). These plots facilitate the comparison of a data distribution with various theoretical distributions (5), and goodness-of-fit tests for a variety of distributions including the normal. The *univariate* procedure is also used to produce graphical outputs to allow for ease in interpretation of the SAS results.

3.3 Results

The following tables and figures are summaries of the aforementioned methods and give additional representative samples of the results.

Table 4. Summary table for the minimum percentage accuracy that results from all data manipulation schemes across all material types and thresholds. The rankings are from all 50, 674 schemes being divided into groups, where 100 includes all schemes; 50 includes the top 25,337 schemes; and so forth.

		% Accuracy For 1:1 Comparisons				
Threshold	**Ranking (Percentage)**	*Concrete*	*Iron*	*Plastic*	*Stainless Steel*	*Average*
99	100	76.03	95.87	80.00	64.00	78.98
	50	89.25	98.35	100.00	97.52	96.28
	25	90.91	N/A	N/A	98.35	96.90
	10	91.74	N/A	N/A	99.17	97.32
	5	92.56	N/A	N/A	N/A	97.52
	<1	94.21	99.17	N/A	100.00	**98.35**
95	100	36.00	86.77	20.00	36.00	44.69
	50	90.08	97.52	100.00	95.04	95.66
	25	90.91	98.35	N/A	96.69	96.49
	10	91.73	N/A	N/A	N/A	96.69
	5	92.56	N/A	N/A	N/A	96.90
	<1	94.21	99.17	N/A	98.35	**97.93**
90	100	0.08	4.13	8.00	0.00	3.05
	50	88.43	95.04	100.00	76.86	90.08
	25	89.26	N/A	N/A	80.99	91.32
	10	90.08	95.86	N/A	89.26	93.80
	5	90.91	96.69	N/A	92.56	95.04
	<1	91.74	97.52	N/A	97.52	**96.70**
85	100	83.47	84.29	82.64	66.94	79.34
	50	89.26	92.56	92.56	73.55	86.98
	25	90.91	95.04	98.35	76.03	90.08
	10	91.74	95.86	100.00	77.68	91.32
	5	92.56	N/A	N/A	79.34	91.94
	<1	94.21	98.34	N/A	82.65	**93.80**

Table 5. Summary table to display the statistical significance ($p<0.05$) for the accuracy percentages (as seen in Table 4) that result from all data manipulation schemes across all material types and thresholds. The rankings are from all 50, 674 schemes being divided into groups, where 100 includes all schemes; 50 includes the top 25,337 schemes; and so forth.

		Statistical Significance ($p<0.05$) For 1:1 Comparisons			
Threshold	**Ranking (Percentage)**	*Concrete*	*Iron*	*Plastic*	*Stainless Steel*
99	100	Yes	Yes	Yes	Yes
	50	Yes	No	No	No
	25	Yes	No	No	No
	10	Yes	No	No	No
	5	Yes	No	No	No
	<1	No	No	No	No
95	100	Yes	Yes	Yes	Yes
	50	Yes	Yes	No	Yes
	25	Yes	No	No	No
	10	Yes	No	No	No
	5	Yes	No	No	No
	<1	No	No	No	No
90	100	Yes	Yes	Yes	Yes
	50	Yes	Yes	No	Yes
	25	Yes	Yes	No	Yes
	10	Yes	Yes	No	Yes
	5	Yes	No	No	Yes
	<1	No	No	No	No
85	100	Yes	Yes	Yes	Yes
	50	Yes	Yes	Yes	Yes
	25	Yes	Yes	Yes	Yes
	10	Yes	Yes	No	Yes
	5	Yes	Yes	No	Yes
	<1	No	No	No	No

Table 6. Equations chosen to be applied to authentic data in the next phase of the research (i.e. control site). Each equation results in an accuracy of at least 94.21% with the "1:1" comparisons and 100% accuracy for the "0:0" comparisons.

ID	Variable Explored	Expression
1	GPR multiplied, all others constant	EM(1)+MAG(1)+GPR(6.75)
2	MAG multiplied, all others constant	EM(1)+MAG(8.5)+GPR(1)
3	EM multiplied, all others constant	EM(9.25)+MAG(1)+GPR(1)
4	EM only held constant	EM(1)+MAG(7.5)+GPR(8.25)
5	GPR only held constant	EM(8.75)+MAG(9.5)+GPR(1)
6	MAG only held constant	EM(5)+MAG(1)+GPR(9.75)
7	All are multiplied	EM(9.5)+MAG(10)+GPR(8.5)
8	All are held constant	EM(1)+MAG(1)+GPR(1)
9	All are raised to average optimal exponent	EM^(3.125)+MAG^(4.3)+GPR^(4.525)

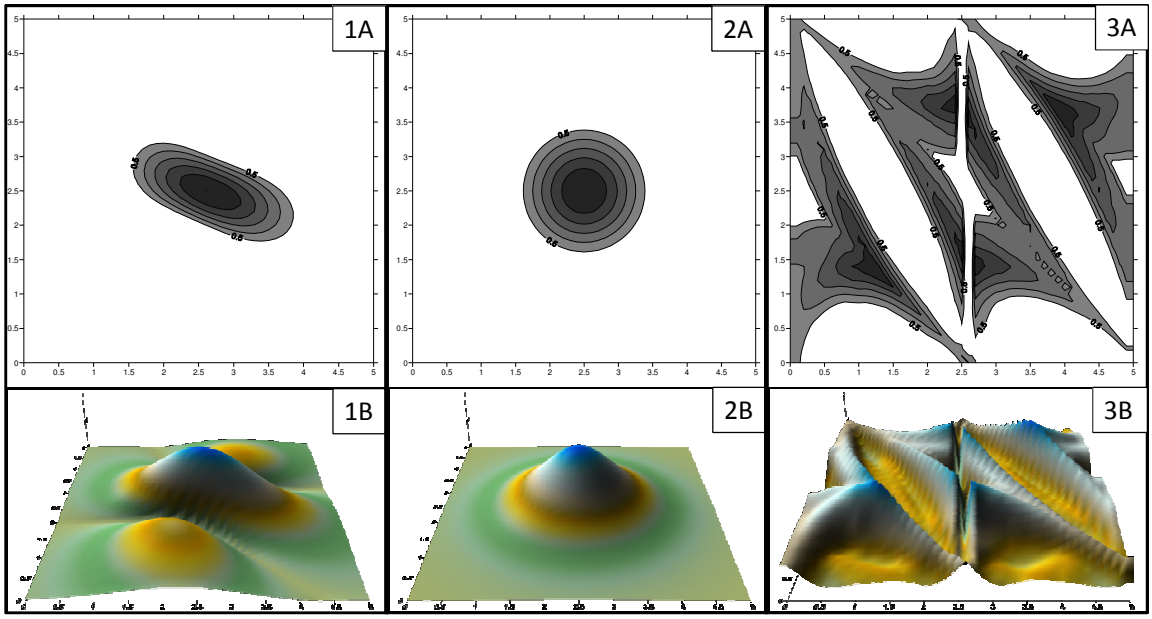


Figure 16. Representative example of raw modeled data (iron sphere, buried at 175 cm and a diameter of 100cm). Groupings show the data as a (1A) Magnetic contour map; (1B) Magnetic surface map; (2A) EM-31 Ground Conductivity contour map; (2B) EM-31 Ground Conductivity surface map; (3A) GPR contour map; and (3B) GPR surface map. Data was mapped using *Surfer*.

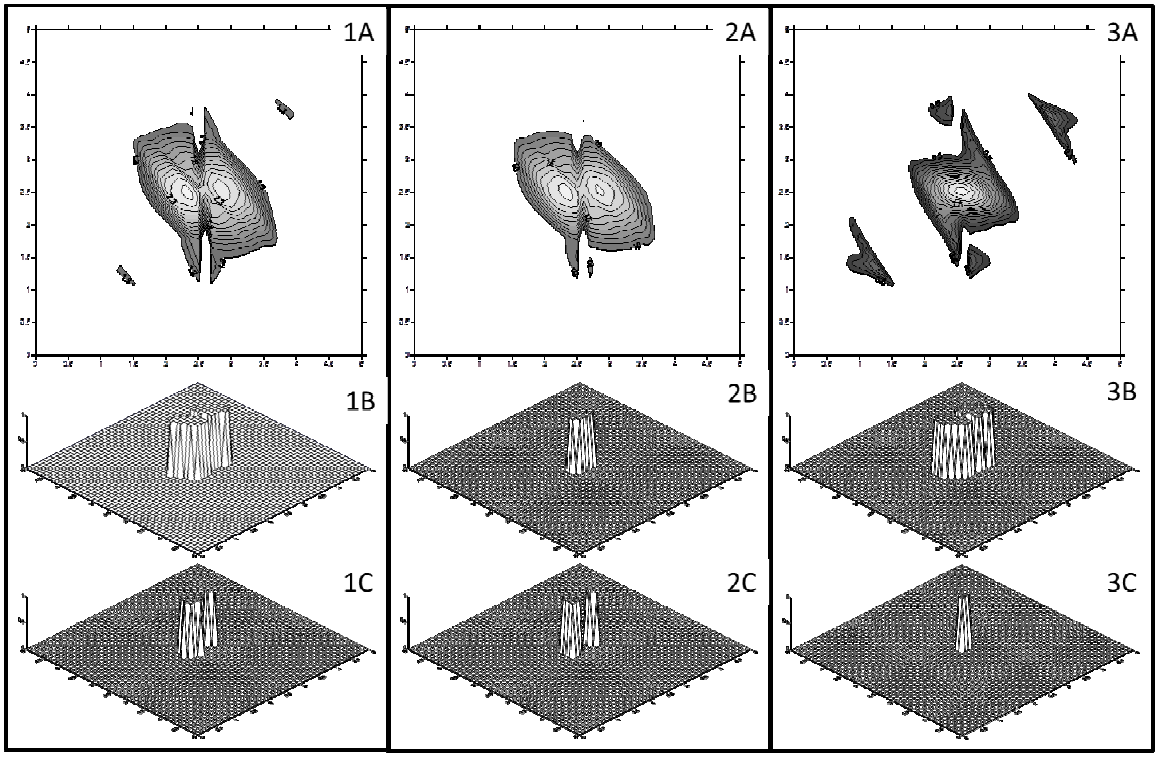


Figure 17. Representative example of modeled iron datasets combined utilizing various data manipulation schemes. Groupings show (1A) normalized values for Mag, EM, and GPR added together; (2A) Normalized values for all datasets combined with a scheme ranked at 50% accuracy; (3A) Normalized values for all datasets combined with a scheme ranked as number 1 of all 50,674 schemes; (1B, 2B, and 3B) give the specific location of data once cells are transformed to binary format at a 85% threshold; (1C, 2C, and 3C) give the specific location of data once cells are transformed to binary format at a 95% threshold.

Table 7. Summary table to show location and dimension of modeled iron sphere once data points are manipulated with various schemes.

	85% Threshold				95% Threshold			
	Node 1		***Node 2***		***Node 1***		***Node 2***	
	Area (m^2)	*Center Point*	*Area (m^2)*	*Center Point*	*Area (m^2)*	*Center Point*	*Area (m^2)*	*Center Point*
Lowest Rank (Simple Addition)	0.235	2.2, 2.5 trending NW	0.265	2.9, 2.5 trending NW	0.06	2.25, 2.5 trending NW	0.01	2.85, 2.5 trending NW
50% Rank (Constant Multiples)	0.21	2.25, 2.5 trending NW	0.275	2.9, 2.5 trending NW	0.04	2.3, 2.5 trending NW	0.015	2.85, 2.55 trending N/S
Top Rank (Optimal Exponent Add)	0.09	2.55, 2.5 trending E/W	n/a	n/a	0.015	2.55, 2.55	n/a	n/a

Table 8. Improvement in Uncertainty Level by Applying Data Integration Methodology

		Concrete		Iron		Plastic		Stainless Steel	
		Addition	***Exponential***	***Addition***	***Exponential***	***Addition***	***Exponential***	***Addition***	***Exponential***
99%	***Actual 1, Calculated 0***	12.00%	0.00%	1.65%	0.83%	4.00%	4.00%	8.00%	0.00%
	Actual 1, Calculated 1	88.00%	100.00%	98.35%	99.17%	96.00%	96.00%	92.00%	100.00%
	Actual 0, Calculated 1	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.04%	0.04%
	Actual 0, Calculated 0	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	99.96%	99.96%
95%	***Actual 1, Calculated 0***	32.00%	8.00%	2.48%	0.83%	36.00%	0.00%	0.00%	0.00%
	Actual 1, Calculated 1	68.00%	92.00%	97.52%	99.17%	64.00%	100.00%	100.00%	100.00%
	Actual 0, Calculated 1	1.40%	0.08%	0.00%	0.00%	0.08%	0.00%	0.12%	0.00%
	Actual 0, Calculated 0	98.60%	99.92%	100.00%	100.00%	99.92%	100.00%	99.88%	100.00%
90%	***Actual 1, Calculated 0***	56.00%	4.00%	33.06%	4.13%	60.00%	20.00%	8.00%	0.00%
	Actual 1, Calculated 1	44.00%	96.00%	66.94%	95.87%	40.00%	80.00%	92.00%	100.00%
	Actual 0, Calculated 1	16.03%	0.00%	30.40%	32.00%	15.49%	0.31%	19.25%	38.16%
	Actual 0, Calculated 0	83.97%	100.00%	69.60%	99.68%	84.51%	99.69%	80.75%	61.84%
85%	***Actual 1, Calculated 0***	4.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
	Actual 1, Calculated 1	96.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%
	Actual 0, Calculated 1	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
	Actual 0, Calculated 0	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%

3.4 Discussion

Discussion begins by noting the variables that are most significant (i.e. target size, depth, composition) in accurately identifying target location. After statistical analysis, the effectiveness of each data manipulation scheme to quantitatively integrate the geophysical datasets is covered, along with the significance of each result. Finally, an explanation on how the uncertainty level in anomaly detection changes through implementing the proposed methodology will be given.

3.4.1 Significance of Variables

During the initial testing of variables, it was found that a sample spacing of 10cm between data points along a 5m by 5m grid would provide the greatest degree of certainty in anomaly detection. Further testing of additional variables added to the complexity of the models and statistics:

Target Composition (i.e. Material Type): Cement, Iron, Plastic, Stainless Steel

Target Diameter: 12.5cm, 25cm, 50cm, 100cm, 200cm

Target Depth: 25cm, 50cm, 75cm, 125cm, 175cm, 250cm, 350cm

Traditionally, when designing the survey parameters, the geophysicist has to consider the target she is trying to detect with respect to the surrounding geology. Certain geophysical techniques are not appropriate depending upon the geology (e.g. GPR waves will attenuate rapidly in clay rich soils due to the interference from high iron concentrations). Additionally, if the geophysicist is trying to detect a target that is 20cm

in diameter, she will want her survey lines to be less than 10cm apart to avoid completely passing over the target unintentionally (using the Nyquist rule). For the purpose of this research, there was some built in advantages with the models (and subsequent synthetic data) because (1) the smallest object simulated was 12.5cm and the survey lines were spaced 10cm apart, and (2) the matrix containing all objects was simulated as wet sand, with very low iron content. These advantages are considered acceptable at this stage of the research because we are testing a methodology for data integration as a primary objective, rather than a universal test of all possible Earth combinations.

Throughout the process of determining which of the 140 model outputs for all of the EM-31 ground conductivity, GPR, and magnetic gradiometry simulations would be representative for each of the different material types (concrete, iron, plastic, stainless steel), specific variables such as burial depth and sphere diameter were consistently found to be significant ($p<0.05$) in the statistical analysis of the simulations. For the concrete and plastic representative models, both had a diameter of 25cm and a buried depth of 175cm. For the iron representative simulation, diameter was 100cm and buried depth was 175cm. For the stainless steel representative simulation, diameter was 25cm and buried depth was 250cm.

3.4.2 Quantitative Integration

Among the dozens of different data tables and queries that comprise the *Microsoft Access* based *Materials Database System (MDS)*, data thresholds are also be evaluated. For the results presented here, only four thresholds were systematically executed during

statistics (99%, 95%, 90% and 85%). However, one could extend this methodology and further refine the thresholds to decide what confidence level would be crucial for that particular project's objectives. For example, if it is unclear the threshold that is most appropriate between 95% and 99%, the same procedure can be repeated using a 1% increment (compared to the 5% increment used in this research).

Through calculations of the synthetic datasets, it was found that the 95% threshold should be applied to all future datasets (*see* ***Table 4***). Both the 99% and 95% thresholds returned the accuracies for each of the material types modeled of at least 89.25% with the top 50% of the data manipulation schemes. Even at the top 1% of the data manipulation schemes, the 99% and 95% thresholds were similar, with only a 0.42% accuracy difference averaged across all four material types. However, the 95% threshold is more appropriate to choose for application towards authentic data because the shape and size of the targets are not altered as severely when compared to the 99% threshold and plotted as binary responses, giving a better representation of the modeled object's true characteristics overall.

The statistical analysis for each of the different material types simulated returned slightly varying results dependening on the material type and threshold applied to the data. Once all of the *geoData Accuracy Summary* queries were constructed, and each data manipulation schemes accuracies were ranked, statistics were conducted to evaluate the significance of each grouping (*see* ***Table 5***). In most cases, the change in when a grouping no longer had statistical significance ($p<0.05$) is reflected in the results of **Table 4**. For example, in Table 4, for Iron at a 99% threshold, all data manipulation schemes at

the 50% and higher all return an accuracy of 98.35%; Table 5 shows that there is no statistical significance within this range of manipulation schemes. Understanding how to group the different data manipulation schemes together will allow future data geophysicists to choose from an assortment of schemes to apply to the data and have confidence that the results will be similar in detection of subsurface features. This also gives further evidence that the 95% threshold is appropriate to apply to the authentic data because there is a large number of schemes that can be applied across all material types.

In some cases, there will only be one or two geophysical techniques utilized during a survey. The suggested data manipulation schemes for those instances are provided in **Table 6.** There were a number of other schemes that could be applied, but these are representative of the range of schemes resulting from this methodology and chosen to be applied to the authentic data. For those cases that the geophysicist would only have one geophysical technique, data could be manipulated in cases like those provided with equation ID 1-3 *(see Table 6)*. For those cases that she would have two geophysical techniques, data could be manipulated in cases like those provided with equation ID 4-6. Equation ID 8 represents a simple addition of the normalized data sets. One must remember, particularly when looking at the results displayed in Table 6, that there is no single "best fit" for the data, but instead there are several schemes that work well in enhancing subsurface features dependent upon your target type and project objectives.

Once all data have been integrated together, targets become enhanced, as seen when comparing **Figures 16 and 17**. The raw, original data is displayed in **Figure 16**, although there are some deviations in the simulated target's actual physical location within the

model, centered on X=2.5, Y=2.5 and the location (and shape) of the anomaly produced after processing. Particularly in the case of the GPR data, the image is distorted due to boundary conditions of the computer model (i.e. programming); it is unclear why in the data the center point has the highest wave amplitude, which is expected, but is showing a depressed value in the image generated using *Surfer* (this is most likely a numerical artifact of the curve fitting). Once the datasets are integrated, however, that feature within the data drops out and the simulated anomalies take on the expected shapes and location within the model grid. One peculiar result can be seen in **Figure 17**, parts 1A-C and 2A-C.

For the addition-only data manipulation scheme, the simulated target is imaged close to the central point, but is found to have two dominating "nodes" in the signal. The dataset with a 50% ranked scheme applied has the same effect but as extreme in position and size as the former. The top ranked data manipulation scheme—raising each normalized value to the averaged optimal exponent—displays the anomaly as a whole feature (i.e. not divided into nodes) over the central point but slightly shifted along the X-axis (*see* ***Table 7***). Once the data are held to a 95% threshold, the general shape is maintained, although there is an increase in the sharpness of the anomaly shape and the central point of the simulated anomaly is closer to the central point of the grid. Additionally, the anomaly featured in 3C suggests that the target location may be identified, as well as very accurately and clearly indicate the center of the target.

3.4.3 Uncertainty Level Improvements

There were some inconsistencies between the threshold levels in the improvements in accuracy percentages when comparing the lowest ranking data manipulation scheme (tradition addition) with the top scheme (optimal exponent), as identified in **Table 8**. The 85% threshold appears to be the best choice in applying to the authentic data when looking at these two schemes isolated from the other 50,672 schemes. However, when evaluating all schemes, that is not the case. The purpose of **Table 8** is to illuminate how the uncertainty level for accurately locating the simulated targets changes when applying various data manipulation schemes. For every example of threshold and material type, there is an improvement in the percentage accuracy when comparing the addition scheme to the optimal exponent scheme. In some cases (e.g. concrete at 90% threshold), that improvement is by 52%.

3.5 Conclusions

The proposed methodology may be used to identify a purely quantitative statistical approach for integrating two or more geophysical data sets collected using near-surface geophysical techniques for discrete anomaly detection. It should be noted, that while these results are focused on only three techniques (GPR, magnetic gradiometry, and ground conductivity), the described methodology may be used for any number of different types of geophysical techniques; the three were chosen due to the accessibility of the equipment (that are applied in Part 2, Section 4) and processing software. Our results may be used to minimize common data integration problems by: (1) setting up

consistent parameters within the geophysical techniques utilized; (2) developing a comprehensive model for integration of data; and (3) utilizing consistent visualization techniques to represent the processed data for a higher degree of confidence when interpreting the results.

The *Materials Data Systems* (MDS) created in *Microsoft Access* proves to be a robust and beneficial tool for processing the dozens of data tables and queries that were interconnected during analysis of multiple variables. The program SAS is utilized for statistical analysis and found to be appropriate for determining statistical significance of the results due to the sheer volume of data simulations and the software's flexibility for integrating multiple statistical tests within one run of the program. Statistical tests included *logistical regression*, *proc GLM*, *proc frequency*, and *univariate*.

The variables found to be most significant in the accurate discrimination of targets are the diameter of the target and the depth to which it is buried. The composition (i.e. material) of the target was also significant. A 95% threshold was determined as most appropriate to be applied to the data to get at least a 94.21% accuracy in correctly identifying the location of targets (minimizing Type I errors) and 100% accuracy in knowing where targets are not located (eliminating Type II errors) and this is consistent across all material types. Additionally, our methodology can be used to determine ranges of data manipulation schemes that can be applied to data and maintain similar results in the accuracy of target identification.

Upon integration of synthetic datasets, modeled targets are precisely identified with the center of each target located at the center of the grid (as modeled) with only a 5-cm

shift in the location. This is well within the acceptable error range for standard dGPS systems. Should there be dGPS coordinates integrated with geophysical equipment during collection of authentic data, this methodology would prove beneficial. By integrating multiple geophysical datasets utilizing the systematic and quantitative methodology proposed in this research, the uncertainty level in discrete anomaly detection is significantly improved, in some cases by 52% when the threshold level, material type, and data manipulation scheme applied to the data are considered.

4. APPLICATION OF METHODOLOGY

This chapter is based on a manuscript to be submitted by Megan E. Carr and Gregory S. Baker to the journal Geophysics as Part 2 of a two-part submission:

Carr, Megan E. and Gregory S. Baker, Quantitative Integration of Geophysical Data for Enhancing Subsurface Features Part 2, Geophysics, v. pp.

My contributions to this paper include (i) collection of data, (ii) development of the methodology discussed, (iii) formulation of the programs utilized for data manipulation, (iv) statistical analysis of data sets, (v) visualization of the data, (vi) significant portion of the writing.

Abstract

This article will serve as a description of the application ofthe methodology developed for a quantitative integration of multiple near-surface geophysical data sets to investigate the second of two hypotheses for my dissertation research: Integration of two or more geophysical techniques will result in an improved subsurface image and reduce uncertainty in discrete anomaly detection. The main objective is to improve data quality and visualization techniques within geophysical surveys for discrete anomaly detection, with testing conducted in a controlled, relatively noise-free environment. Statistical analysis methods will be developed on the integration techniques for authentic data (magnetic gradiometry, electromagnetics, and ground penetrating radar).

Data integration techniques are given in full detail, as well as statistics (in both data preparation and in data integration), which is the quantitative assessment of the integration methodologies. Statistical analysis of which resolution type, depth of investigation, and shifting direction (or no direction) of the data points has the most significant impact on accurately locating the buried targets are also discussed.

Additionally, a brief commentary on why certain targets might not have been detected within the study site is given, along with an assessment of the general occurrence of Type I (false positives) and Type II (false negatives) errors are present in the data.

4.1 Introduction

In order for an individual geophysical technique being employed in a survey to be appropriate (i.e., by useful for target detection/discrimination), there is a dependence on a difference in physical properties between the target and the surrounding material in terms of the subsurface characteristic that the technique responds to (e.g., dielectric permittivity, magnetic field variations, and electrical conductivity, respectively. With this in mind, when a geophysical investigation is conducted over a region containing targets that are unknown - and only one technique is used - it is possible to miss certain types of targets completely. Therefore, integrating multiple techniques into one unified data set will more accurately identify and discriminate characteristics of targets with a greater degree of certainty.

Most subsurface data integration studies have involved one geophysical technique with additional data such as boreholes (e.g., Hornby 2007, Ferré 2003) or general geological data such as soil surveys and other maps (e.g., Galicia 2001, Rahman et a. 2008, Allen et al. 2008). Colombo and De Stefano (2007) state that when modeling an integration of techniques, conversions of parameters from one geophysical domain to another have traditionally been performed rigidly by means of empirical functions. While their research mainly deals in techniques not utilized in this project, Colombo and De Stefano show evidence of the importance in setting up appropriate parameters in both data acquisition and processing to ensure the highest degree of certainty in the integrated data models. Recently, Urs Böniger and Jens Tronicke (2010) have proposed results of an integrative analysis and interpretation of different data sets that combines geophysical

instruments with modern topographic data using a tracking total station (TTS). However, their results are limited to composite images of the combined data sets and illustrated with various color schemes, which, when interpreting the data, still consists of a qualitative assessment and does not quantify the data to the extent that this research presents.

Additionally, data sampling heterogeneities inherent in geophysical databases have been identified as the main source of data integration problems (Stock and Pullar 1999, Bishr 1998, and Kim and Seo 1991). Semantic heterogeneity, which occurs when there is a disagreement about the meaning, interpretation or intended use of the same or related data have been identified as the main cause of data sharing problems and are the most difficult to reconcile (Allen et al. 2008). The research presented here attempts to minimize these data integration problems by (1) setting up consistent parameters within geophysical techniques utilized; (2) developing a comprehensive model for integration of data; and (3) utilizing consistent visualization techniques to represent the processed data for a higher degree of confidence when interpreting the results.

4.1.1 Hypothesis

This article is part two of the previously submitted article from the preceding section of this dissertation. Essentially, this phase of the research addresses the second hypothesis, “Integration of two or more geophysical techniques will result in an improved subsurface image and reduce uncertainty in discrete anomaly detection” and will be tested via case studies to satisfy the following goals:

- Correlate geophysical data with known discrete "targets" by utilizing an integration of multiple geophysical techniques
- Complete a statistical analysis to evaluate the effectiveness of the data integration methodology and applications using authentic data
- Test the methods developed using synthetic data by application to various case studies of completely unknown discrete targets

The end result of this research, discussed within this article, is to supply supporting evidence that the integration of two or more geophysical techniques will result in an improved subsurface image and reduce uncertainty in discrete anomaly detection.

4.1.2 Field Site Description

Data collection was all conducted at the University of Tennessee Geophysical Research Station. Located between Alcoa Highway 129 and the Tennessee River (approximately two miles to the south of the University of Tennessee main campus in Knoxville, Tennessee), this site is also referred to as the Environmental Hydrology and Geophysics Teaching and Research Site (see **Figure 18**). Given the climate of East Tennessee during the Spring months, both wet and dry conditions are expected, which can affect results. Consequently, relative vadose zone saturation and water table elevations will likely vary among tests conducted on different days, possibly affecting the relative times of refracted first-arrivals among the seismic profiles (e.g., Gaines, 2010). Soil conditions across the site vary from residual soils developed directly on sedimentary

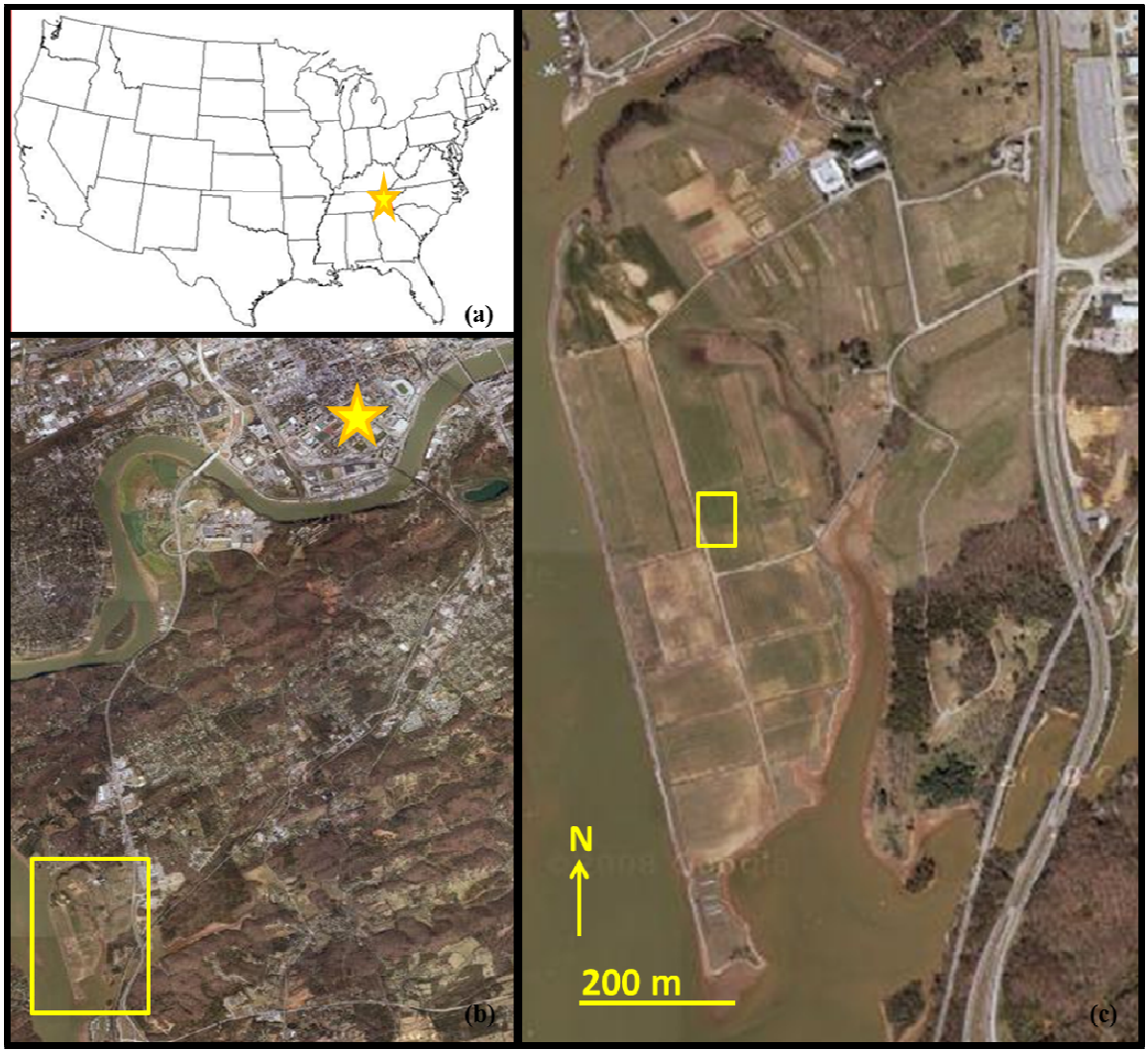


Figure 18. (a) Map showing location of field site (star indicates Knoxville, TN); (b) Close-up view of Knoxville, TN with yellow box designating The University of Tennessee Geophysical Research Station and star showing Knoxville, TN; (c) Close up view of yellow box from (b), with smaller yellow box indicating general location of data collection.

bedrock (near the highway) to loamy soils developed on alluvial terraces at elevations above the river. Silt or sandy silt dominates the top 6.1 m of strata, which overlies approximately 0.9 – 1.5 m of fine to medium sand and cemented sand. The remaining portion of the stratigraphic section is comprised of fractured shale till and limestone until reaching bedrock at a depth of approximately 11.6 m. Bedrock is Ottossee Shale, which is a Middle Ordovician member of the Chickamauga Group. As a whole, it is generally characterized by fine-grained calcareous shale with some interbedded limestone (Milici and Smith 1969).

The field site additionally contains known targets that were buried in the spring of 1999 having detailed positioning given by latitude, longitude, and depth within the sub-surface (**Figure 19**). Information including size, shape, composition material and orientation is given in **Table 9**. It is assumed that there has been sufficient time for the ground to settle and any disturbance to the subsurface (and resulting signal in the geophysical data) to be minimized. This has been assessed by noting that data of various types collected over back-filled holes where no object was buried yield no significant anomalies compared to the background.

Although fairly "quiet" from a geophysical noise perspective, the site is susceptible to some background noise from various sources. A relatively large water pump is used intermittently to supply a portion of the agricultural site is located about 200 meters ENE of the plot's NE corner. However, for the purposes of this research, this noise is not significant.

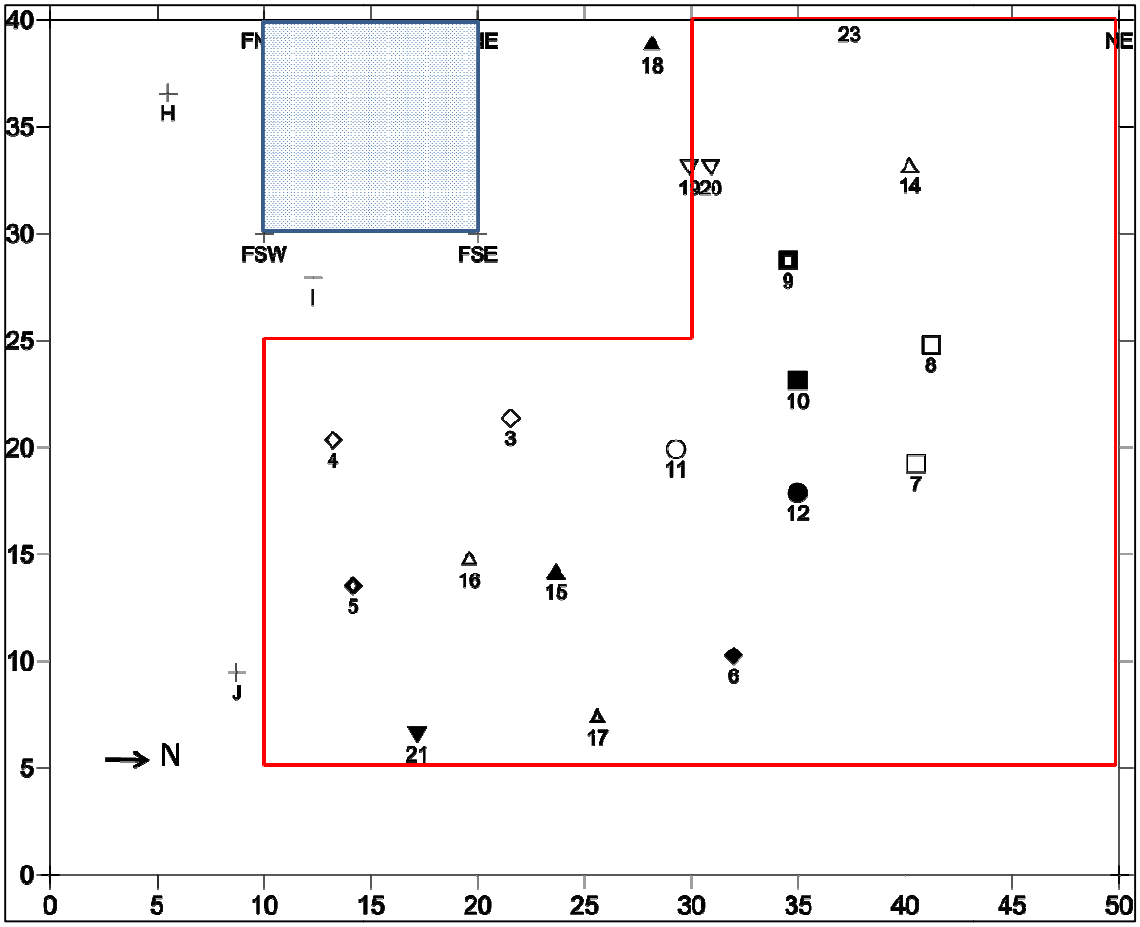


Figure 19. Map displaying locations of buried targets. Grid is measured in meters. GPR and EM-31 surveys were conducted over the entire 40mx50m grid. The red box is the areal extent of the Magnetic Gradiometry survey. The shaded blue section is a metal fenced area where the data were removed from all processing. Target descriptions for each number on the map are given in Table 9.

Table 9. Target Descriptions for Field Site

Map ID	Description*	Depth of Burial (m)
3	Vertical 55 gal drum	0.635
4	Horizontal 55 gal drum, N-S	0.686
5	Horizontal 55 gal drum, N-S	0.991
6	Horizontal 55 gal drum, E-W	0.660
7	Steel scrap, 3 pcs 3-4 feet long	0.635
8	Vertical 55 gal drum	0.775
9	Plastic 55 gal drum, freshwater and gravel filled	0.686
10	Vertical 55 gal drum	1.118
11	Plastic 55 gal drum, saltwater and gravel filled	0.635
12	Iron pipe, 3" diameter, 42" long	0.610
14	2 pcs styrofoam, 9'x2'x4", dipping N45E	0.915
15	Cement blocks, 1.5 cu feet pea gravel	0.686
16	Aluminum gutter; 5 pcs, 6.5-8 feet long	0.394
17	Coil of 12/3 copper wire	0.305
18	Solid iron rod, ~41" long, 1" diameter	0.331
19	Iron Pipe, 4" diameter, 80"? long	0.914
21	Two vertical drums, 33" center to center along N-S line	0.991
22	Iron Pipe, 4" diameter, 64" long	0.559
23	Two horizontal drums, 19" separation end to end, N-S	0.914
H, I, J	Well Locations, 8” metal cover	0.000

* For location of the buried targets in relationship to each other, refer to Figure 19; the surrounding agricultural plots are occasionally mowed or plowed, and the vehicle traffic can cause some issues for geophysical techniques.

4.2 Methods

4.2.1 Data Collection and Processing

Three geophysical techniques were utilized in this study: Ground Penetrating Radar (GPR), EM-31 Ground Conductivity, and Magnetic Gradiometry. Each technique's data were collected separately to avoid potential signal interference. Additionally, each data set was processed according to best practices for that respective technique prior to integration. Detailed explanations of each type of processes can be found in the software manufacturer's user manual. The issue of each technique having different data resolutions will be addressed throughout this paper.

4.2.1.1 Ground Penetrating Radar

The GPR unit used is a Sensors and Software PulseEKKO Pro system. This technique utilizes propagating electromagnetic (EM) waves to detect changes in the EM properties of the shallow subsurface. The propagation velocity of EM waves (i.e., the controlling factor on the generation of reflections) is determined by the dielectric permittivity contrast between the background material and the target. Dielectric permittivity is defined as the ability of a material to store and then permit the passage of EM energy when a field is imposed on the material (Baker et al., 2007) and can be measured in the lab or in situ. A GPR unit consists of transmitting and receiving antenna, where the transmitting antenna generates an EM pulse in the subsurface that travels into the subsurface, reflects off an interface or scatters off point sources (both caused by contrasts

in dielectric permittivity). This reflected/scattered energy then travels back to the surface where it is recorded by the receiving antenna.

Data lines were collected in a grid type pattern, with 0.5 meter spacing and lines collected in East/West directions. Data points along each line were collected every 0.1 meter. A differential GPS unit was integrated with the GPR data, allowing coordinates of any anomaly detected to be recorded, giving an exact XYZ location. The frequency utilized during this study was a 100 MHz antennae. The grid size for GPR is 40 meters by 50 meters.

Data was processed using EKKOView Deluxe and EKKOMapper 3 (Sensors and Software, Inc.). The EKKOView Deluxe software enables the GPR data to be manipulated using the processing steps of dewow, migration, and autogain. The GPS data collected in succession with the GPR data was incorporated with the data during processing. The EKKOMapper 3 software was then used to plot the GPR data into one composite image, with the grid being divided into depth slices, which allowed easy identification of anomalies within the grid. Processed data were exported in a grid file to the program Surfer. The grid file was able to be displayed as various maps (contour, image, shaded relief, and surface maps) to allow for a 3D visualization of the data.

4.2.1.2 EM-31 (Ground Conductivity)

A Geonics EM-31 terrain conductivity meter was used for this experiment. The EM-31 is a one-person device containing both transmitter and receiver coils on a 3.7-meter frame and uses an electromagnetic inductive technique that allows measurements without

electrodes or ground contact. With this inductive method, surveys can be carried out under most geologic conditions including those of high surface resistivity such as sand, gravel, and asphalt. Effective exploration depth for this instrument is about 6 meters. The system is a non-intrusive conductivity measuring device, and data can be collected at the speed in which the operator can walk. Terrain conductivity (EM-31) measurements are made by inducing an electromagnetic current into the ground from a transmitter coil, and recording the resulting secondary electromagnetic field at a receiver coil a fixed distance away. Measurements are recorded in units of conductivity called milli-Siemens per meter (mS/m). Abrupt negative spikes in the inphase and conductivity measurements are indicative of locations of the desired targets within this study.

Survey design for this geophysical technique is with the 40 meter by 50 meter grid having data lines with 1.0 meter spacing, with lines alternating in an eastward/westward direction and a sample spacing of 0.5 meter.

Grid data were downloaded directly from the instrument to the computer used for processing; the software used in processing was DAT31W (Geonics Limited, Inc.). Tools within DAT31W allowed data to be smoothed and lines to be corrected for linear drift. The terrain conductivity ASCII data were then converted into a Microsoft Excel spreadsheet and then processed and interpreted using Golden Software, Inc. Surfer software and used to construct plan views of EM data for the entire field survey area.

4.2.1.3 Magnetic Gradiometry

The instrument used during this phase of the project is the Bartington 601-2 single-axis magnetic gradiometer (magnetometer). Single-axis gradiometers, such as the one being used in this study, measure magnetic gradient in a single vector-direction. A magnetometer measures magnetic field strength at a specific measuring point. For purposes of our study, we are interested in measuring how much the strength of a magnetic field changes between two specific points, or the "gradient" of the field. The Grad 601 is a single-axis, vertical component, fluxgate gradiometer with data logger with two cylindrical sensor assemblies for use in geophysics and archaeology. Each sensor assembly contains two fluxgate magnetometers with a one meter vertical separation, together with electronics and non-volatile memory for calibration data. This gradiometer has a linear range of 100nT with a resolution of 0.1nT and a range of 1000nT with a resolution of 1nT.

Each grid was surveyed with 0.5 m spacing, with line data collected in a zig-zag pattern. The instrument's calibration and survey parameters will be set to collect 8 points of data for each meter that was traversed. For each grid, data collection began in the southwest corner and the first line of data was traversed moving towards the north; the second data line was in the south direction, alternating along the survey lines and generally moving towards the east within each grid.

Upon completion of data collection, Archeosurveyor version 2.3.0X (DW Consulting, Inc.) was used to assemble, process, and visualize the individual grids into one comprehensive data set. Grid files are generated by directly downloading data from the

instrument and importing it as individual files (plain text file). When the grids are assembled in the correct orientation based from the study site grid map, the data are digitally processed to produce the best possible interpretation. The following processes are performed to maximize the visualization of anomaly locations within the subsurface of the study area: Clip, Interpolate, Despike, Stretch, Destagger, and Deslope.

In addition to traditional processing of magnetic data, the data must undergo a reduction to poles procedure. This is easy to accomplish in a simple excel spreadsheet utilizing the following equations:

(10) $$\left|A(x,y)\right| = \left[\left(\frac{dT}{dx}\right)^2 + \left(\frac{dT}{dy}\right)^2 + \left(\frac{dT}{dz}\right)^2\right]^{1/2}$$

(11) $$\frac{dT(x,y)}{dx} = \frac{Ti+1,j - Ti-1,j}{2\Delta x}$$

(12) $$\frac{dT(x,y)}{dy} = \frac{Ti,j+1 - Ti,j-1}{2\Delta y}$$

(13) $$\frac{dT}{dz} = \frac{T-horiz(nTz)}{2\Delta z}$$

Where,

T	is magnetic field value at x,y location across grid
$\left\|A(x,y)\right\|$	is the amplitude of the analytic signal at (x,y)
Δx, y	is the distance between measured points

4.2.2 Preparing Data for Analysis

Each individual dataset is divided into 5 meter by 5 meter sections (i.e. grids), as it was determined during the initial testing phase (synthetic data analysis) that this is the

optimal grid size for enhancing subsurface features. Additionally, the GPR data were captured in three different depth slices (0.33 m, 0.66m, and 1.0m) to investigate if varying depth plays a significant role in anomaly detection. These arbitrary depths were chosen because (1) they represent the most frequent depths the targets are buried at the Agricultural Station and (2) are evenly distributed throughout the strata. The GPR data was evaluated in 2D as it is common practice to do so for near-surface applications of geophysics. This division resulted in 76 individual grids, with a 10 meter by 10 meter space removed from processing to accommodate the metal fenced off area from the greater survey area. Once all data points within each grid were manipulated and statistics carried out, the survey area was mosaicked back together, giving a complete image for comparative purposes (i.e. qualitative assessment).

To accommodate for the different resolutions that each technique collects data in, the data points across each grid, versions of each grid were created with points interpolated to match each of the respective geophysical techniques resolution. For example, the GPR data were collected with 0.5m (E/W) and 0.1m (N/S) sample spacing, but all data points would undergo kriging to transform it into a grid with sample spacing matching that of the EM-31 or Magnetic data grids. **Table 10** represents which sample spacing was used for the new data sets prior to integration.

A common problem with geophysical data is the "edge effect," in which should there be potential targets along the edge of the survey area the resulting anomaly will be distorted or possibly missed entirely. To account for this, the survey area's 76 grids were all shifted North/South and East/West by 2.5 meters and reprocessed. **Figures 20-22**

Table 10. Dataset name as determined by geophysical technique, resolution, and depth of investigation

Geophysical Technique	Resolution Type	Resolution	Original Resolution	Depth of Investigation	Dataset Name
EM-31 Ground Conductivity	Type 1	0.5m N/S, 0.1 E/W	GPR	n/a	EM Type 1
EM-31 Ground Conductivity	Type 2	1.0m N/S, 0.5 E/W	EM-31	n/a	EM Type 2
EM-31 Ground Conductivity	Type 3	0.5m N/S, 0.125m E/W	Mag	n/a	EM Type 3
Magnetic Gradiometry	Type 1	0.5m N/S, 0.1 E/W	GPR	n/a	Mag Type 1
Magnetic Gradiometry	Type 2	1.0m N/S, 0.5 E/W	EM-31	n/a	Mag Type 2
Magnetic Gradiometry	Type 3	0.5m N/S, 0.125m E/W	Mag	n/a	Mag Type 3
Ground Penetrating Radar	Type 1	0.5m N/S, 0.1 E/W	GPR	0.33 m	GPR Type 1a
Ground Penetrating Radar	Type 2	1.0m N/S, 0.5 E/W	EM-31	0.33 m	GPR Type 2a
Ground Penetrating Radar	Type 3	0.5m N/S, 0.125m E/W	Mag	0.33 m	GPR Type 3a
Ground Penetrating Radar	Type 1	0.5m N/S, 0.1 E/W	GPR	0.66 m	GPR Type 1b
Ground Penetrating Radar	Type 2	1.0m N/S, 0.5 E/W	EM-31	0.66 m	GPR Type 2b
Ground Penetrating Radar	Type 3	0.5m N/S, 0.125m E/W	Mag	0.66 m	GPR Type 2b
Ground Penetrating Radar	Type 1	0.5m N/S, 0.1 E/W	GPR	1.0 m	GPR Type 1c
Ground Penetrating Radar	Type 2	1.0m N/S, 0.5 E/W	EM-31	1.0 m	GPR Type 2c
Ground Penetrating Radar	Type 3	0.5m N/S, 0.125m E/W	Mag	1.0 m	GPR Type 3c

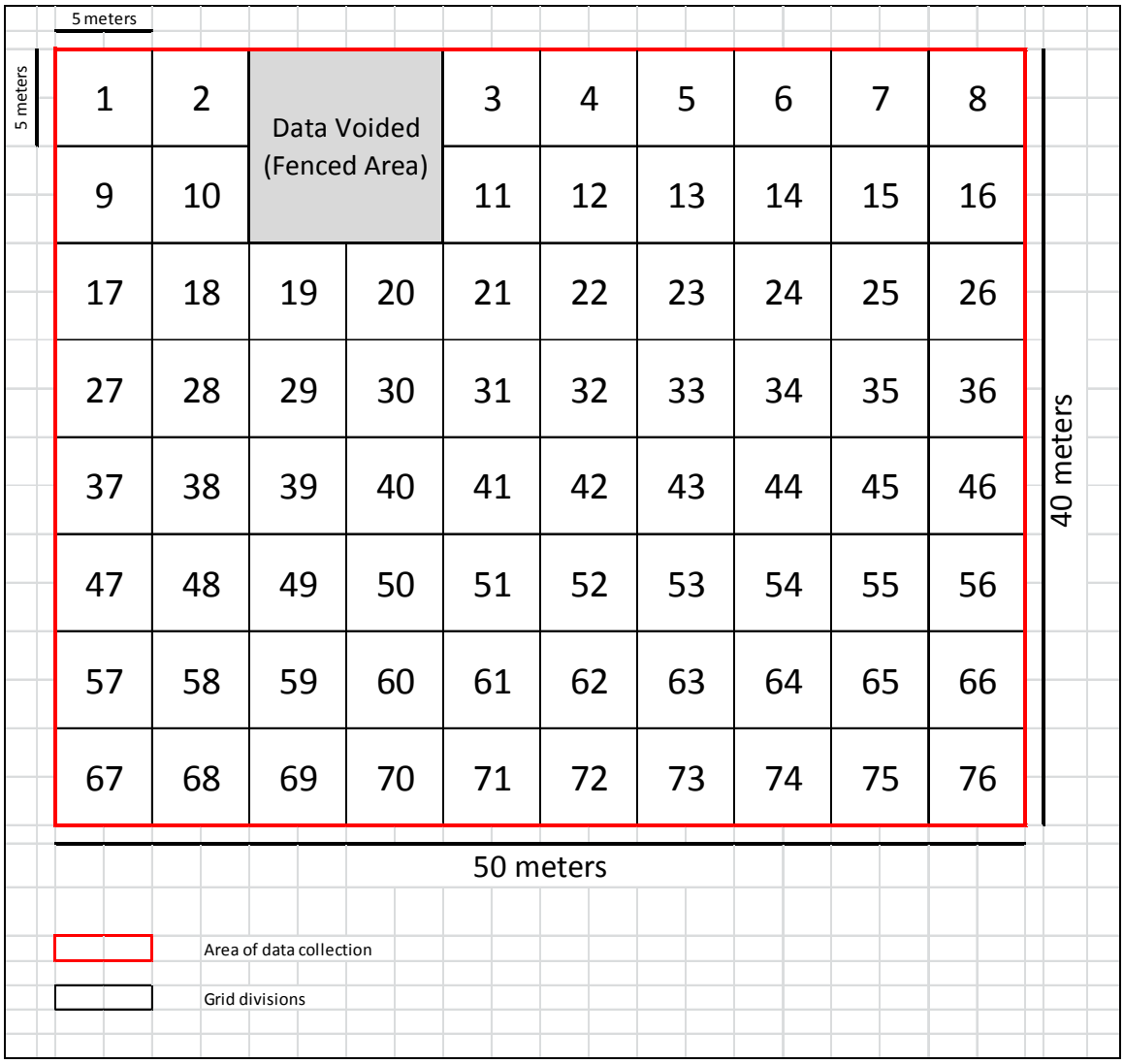


Figure 20. Control site at The University of Tennessee Geophysical Research Station. The survey area is divided into 5m x 5m grids for data manipulation processes and statistical analysis.

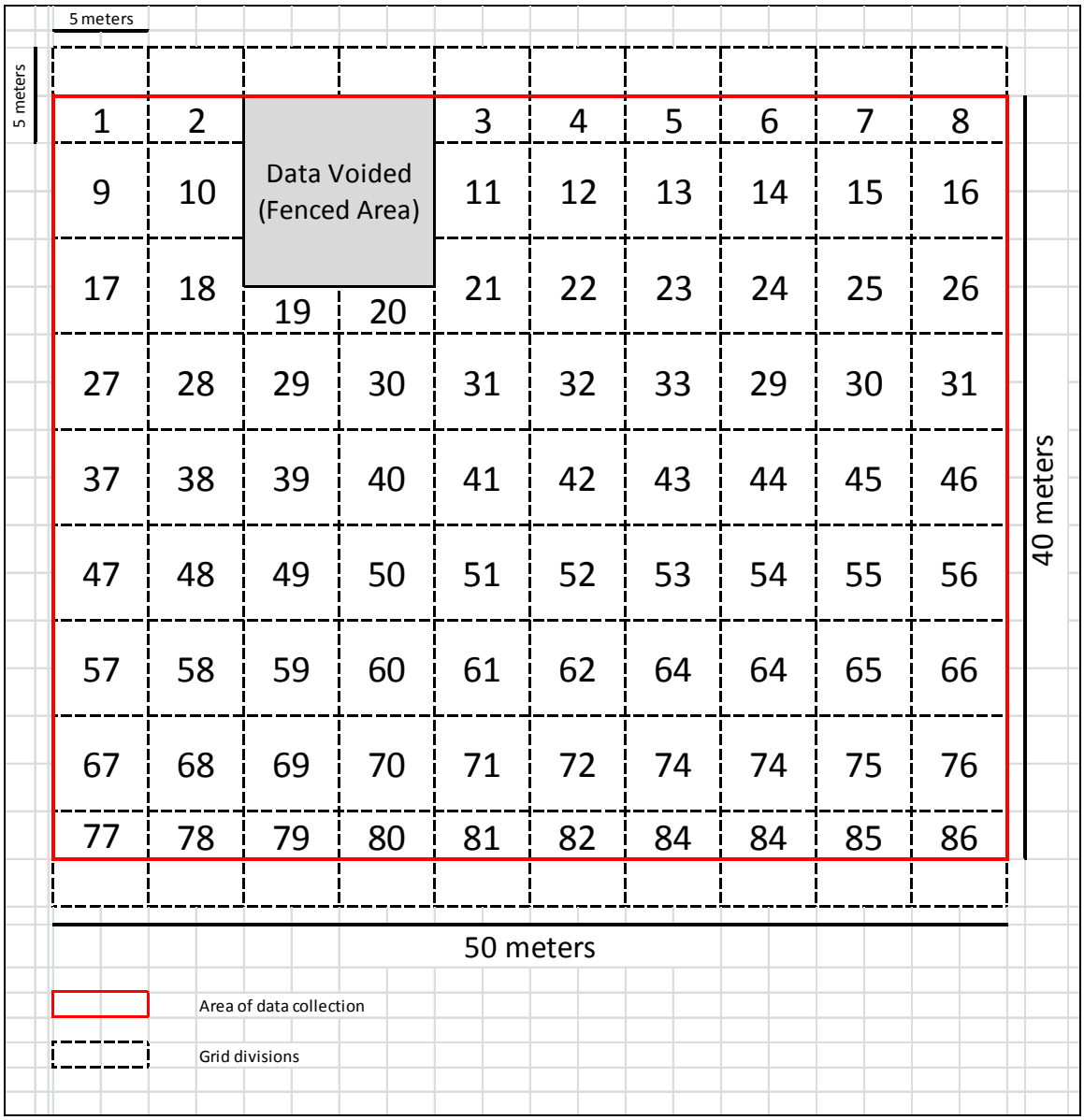


Figure 21. Control site at The University of Tennessee Geophysical Research Station. The survey area is divided into 5m x 5m grids for data manipulation processes and statistical analysis and shifted East/West 2.5m to minimize the edge effect.

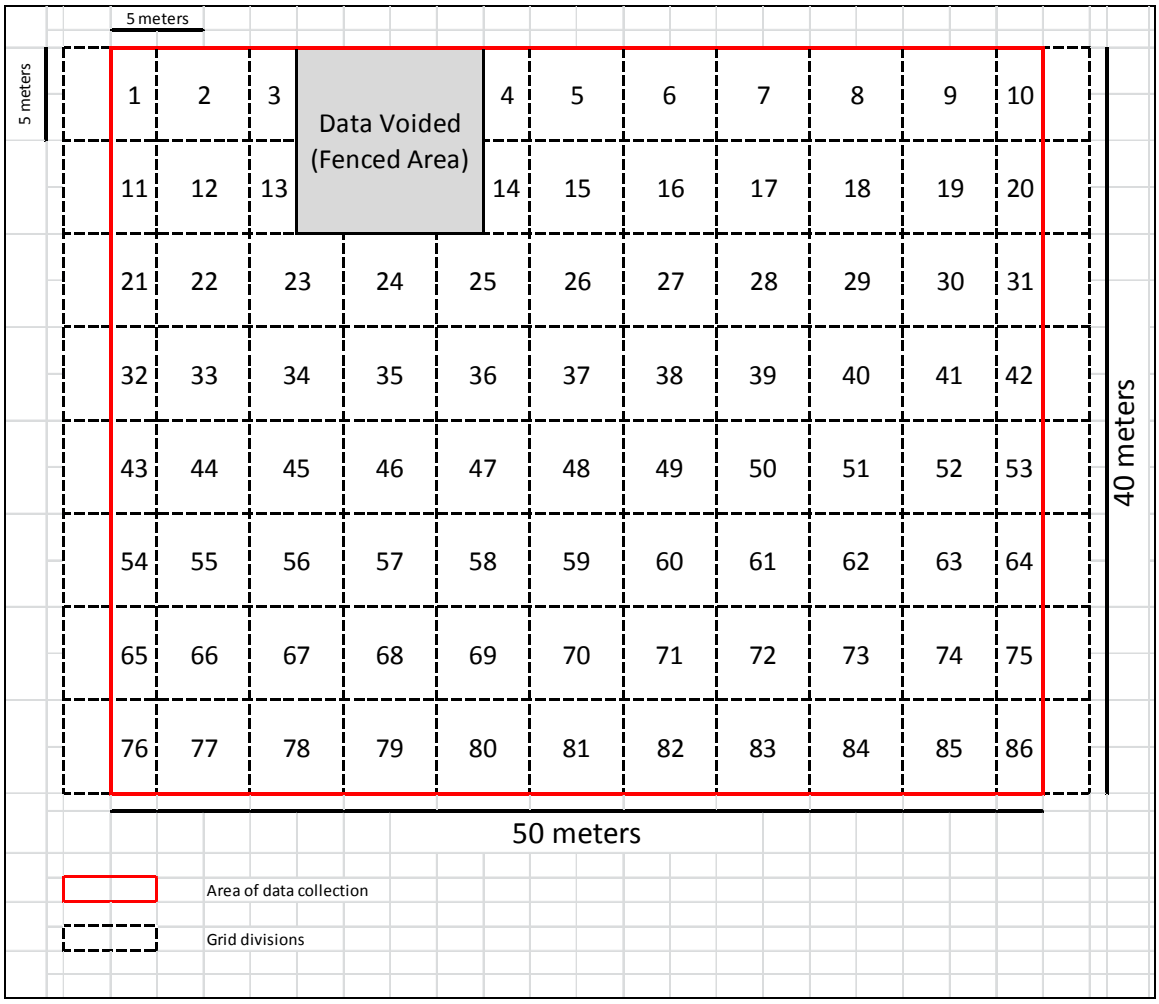


Figure 22. Control site at The University of Tennessee Geophysical Research Station. The survey area is divided into 5m x 5m grids for data manipulation processes and statistical analysis and shifted North/South 2.5m to minimize the edge effect.

illustrate how the survey area was divided and the resulting grids from the different shift directions.

4.2.3 Processing of Grids

Preparing the data tables within Microsoft Access for manipulation of the data points within each grid is similar to that discussed in the previous article, with an added complexity that the overall survey area is divided into sections and then later reassembled. However, there is some simplicity in that there are only nine data manipulation schemes to carry out, compared to the 50,674 for the synthetic datasets. An additional component of this Access table relationship setup is the incorporation of a GPS grid system to identify the coordinates of each target and compare those to the location of the anomalies within the authentic geophysical datasets.

The *geoTINGS EMG* database system has been created to analyze and process a series of 15 datasets (see **Table 10**) composed of EM-31 Ground Conductivity, Magnetic, GPR, and GPS data. These tables are combined using related GPS coordinates to establish a common data link between EM-31, Magnetic, GPR data with a series of internal queries and tables (see **Figure 23**). The *geoData Expression* query (**Figure 24**), in this stage, is a query containing nine specific expressions (i.e. data manipulation schemes). These nine expressions are applied to each new *geoData* table to create another set of nine tables that contain the EMG data (see **Figure 25**):

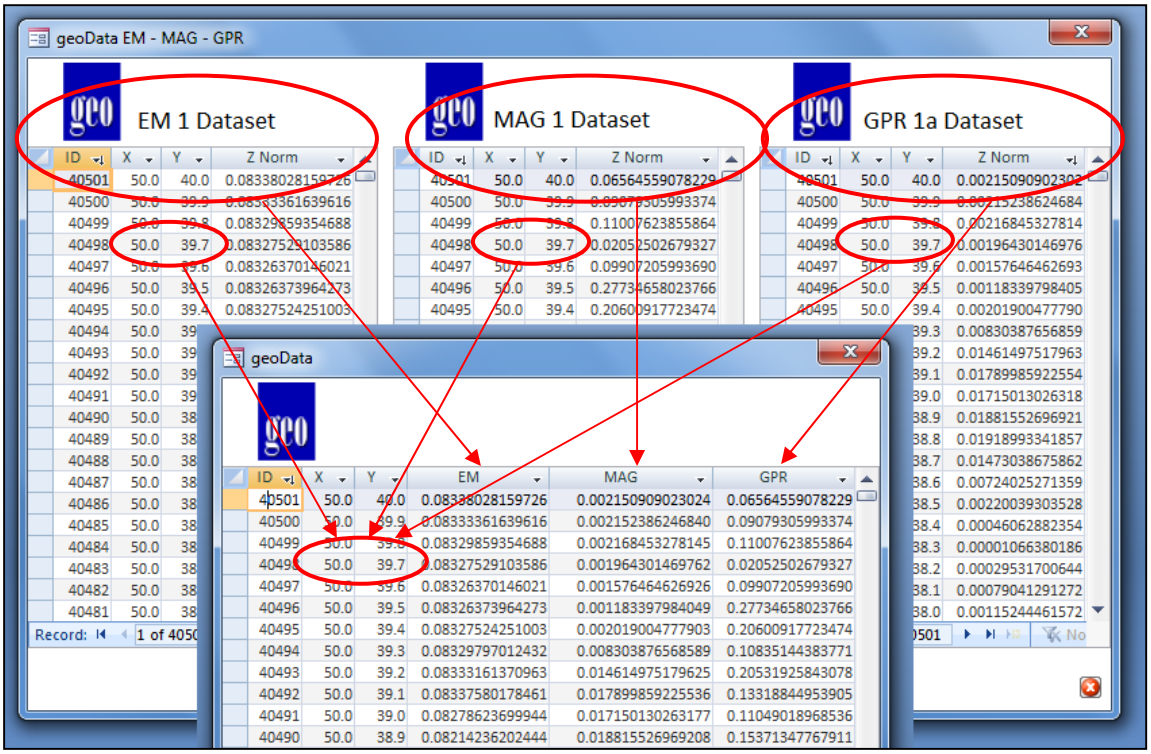


Figure 23. Representative example of datasets that are integrated together with similar resolution (i.e. sample spacing). This query incorporates a common GPS coordinate, designated as an XY location on each survey area.

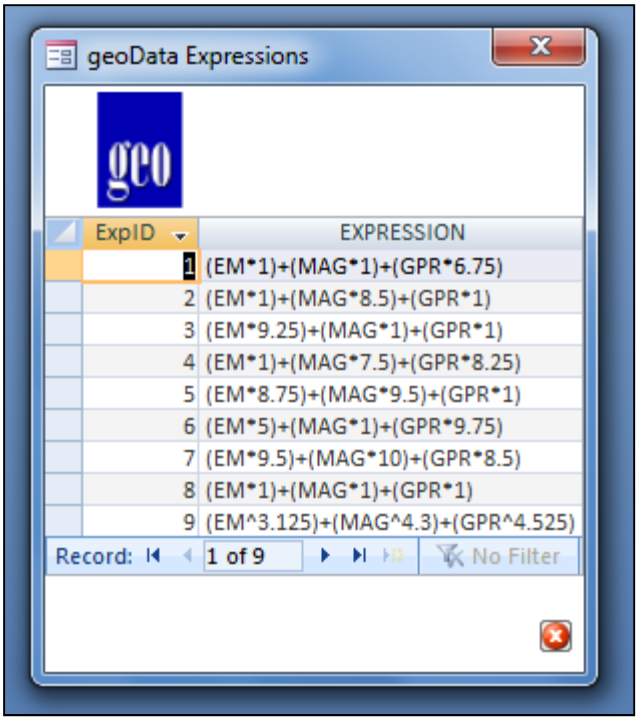


Figure 24. The *geoData Expression* query; these equations were chosen as representative from the statistical results from the synthetic data in Part 1 of this article.

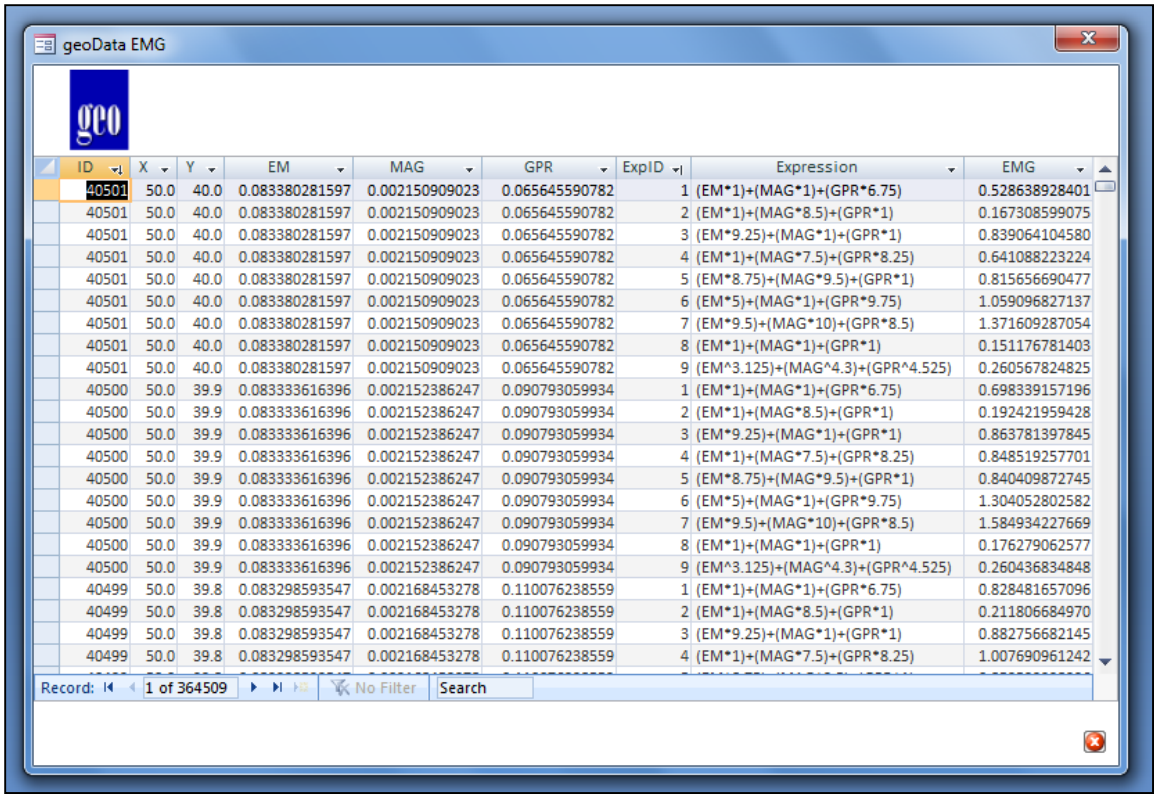


Figure 25. Representative example of the *geoData EMG* database. This query example is created to process the *geoData Expression* and create the EMG data for each table.

1. EM1 – MAG1 – GPR1a
2. EM1 – MAG1 – GPR1b
3. EM1 – MAG1 – GPR1c
4. EM2 – MAG2 – GPR2a
5. EM2 – MAG2 – GPR2b
6. EM2 – MAG2 – GPR2c
7. EM3 – MAG3 – GPR3a
8. EM3 – MAG3 – GPR3b
9. EM3 – MAG3 – GPR3c

Additionally, a series of formulas were applied to each coordinate to create and establish the Target (TG= 1 if true, 0 if False) map across the survey area using three different resolution tables (ResType 1, ResType 2, and ResType 3). The three different resolutions are defined in each new table as Type 1, Type 2, and Type 3 (see **Table 10**, **Figure 26**). Formulas used to identify the Target (TG) are as follows:

(14) If x is between x Start and x End then TGx=1 if false TGx=0

(15) If y is between y Start and y End then TGy=1 if false TGy=0

(16) If TGx=1 and TGy=1, then TG=1, if false TG=0

Once the target has been identified for all datasets, a series of grids is also applied to establish a grid number for all the GPS coordinates (**Figure 27**):

1. Grid – No Shift
2. Grid X Shift – Shifts GPS coordinates 2.5 meters North/South.
3. Grid Y Shift – Shifts GPS coordinates 2.5 meters East/West.

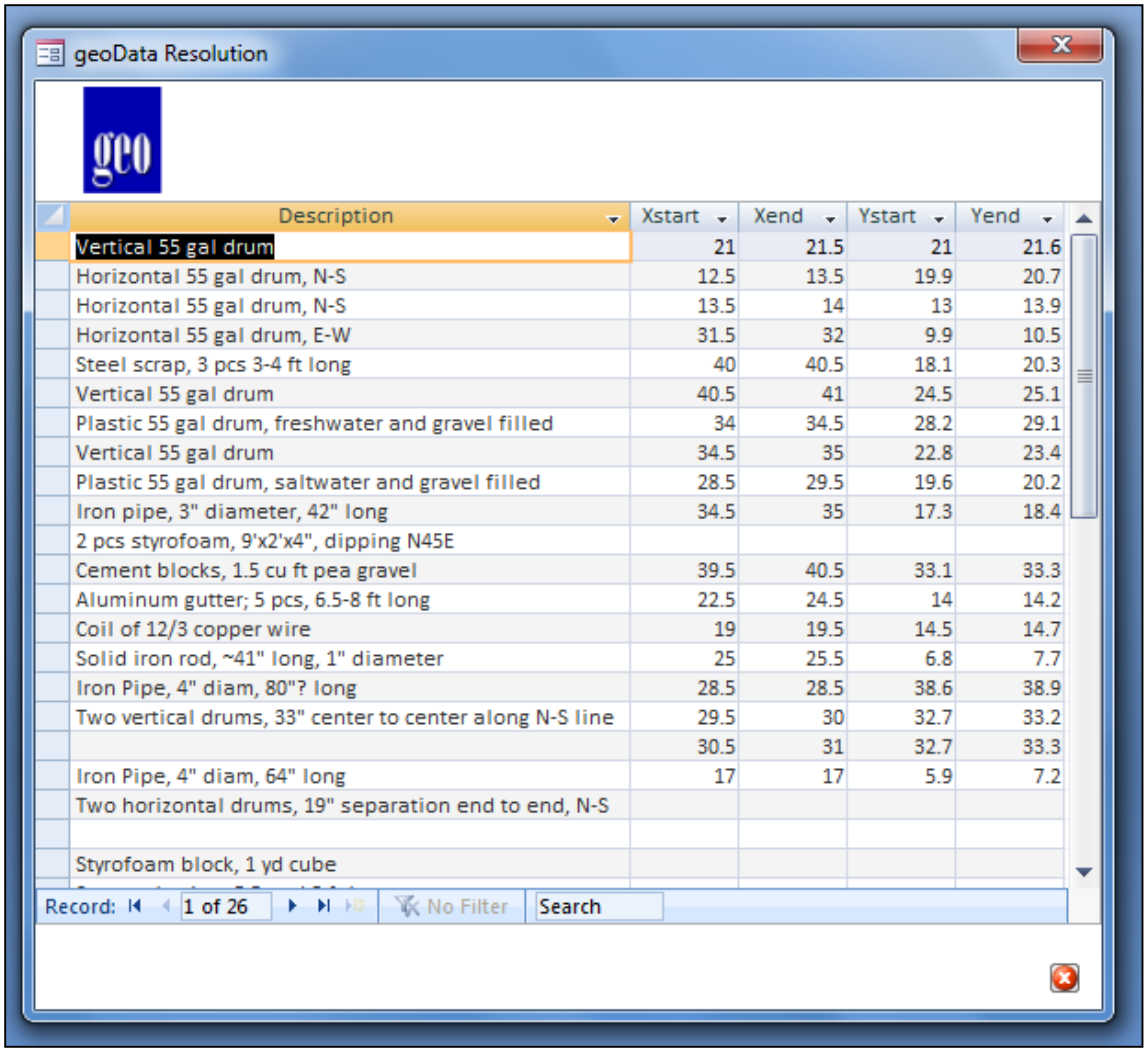


Figure 26. Representative example of a *geoData Resolution* table, identifying the cells that contain buried targets in the survey area. Example is of Resolution Type 1.

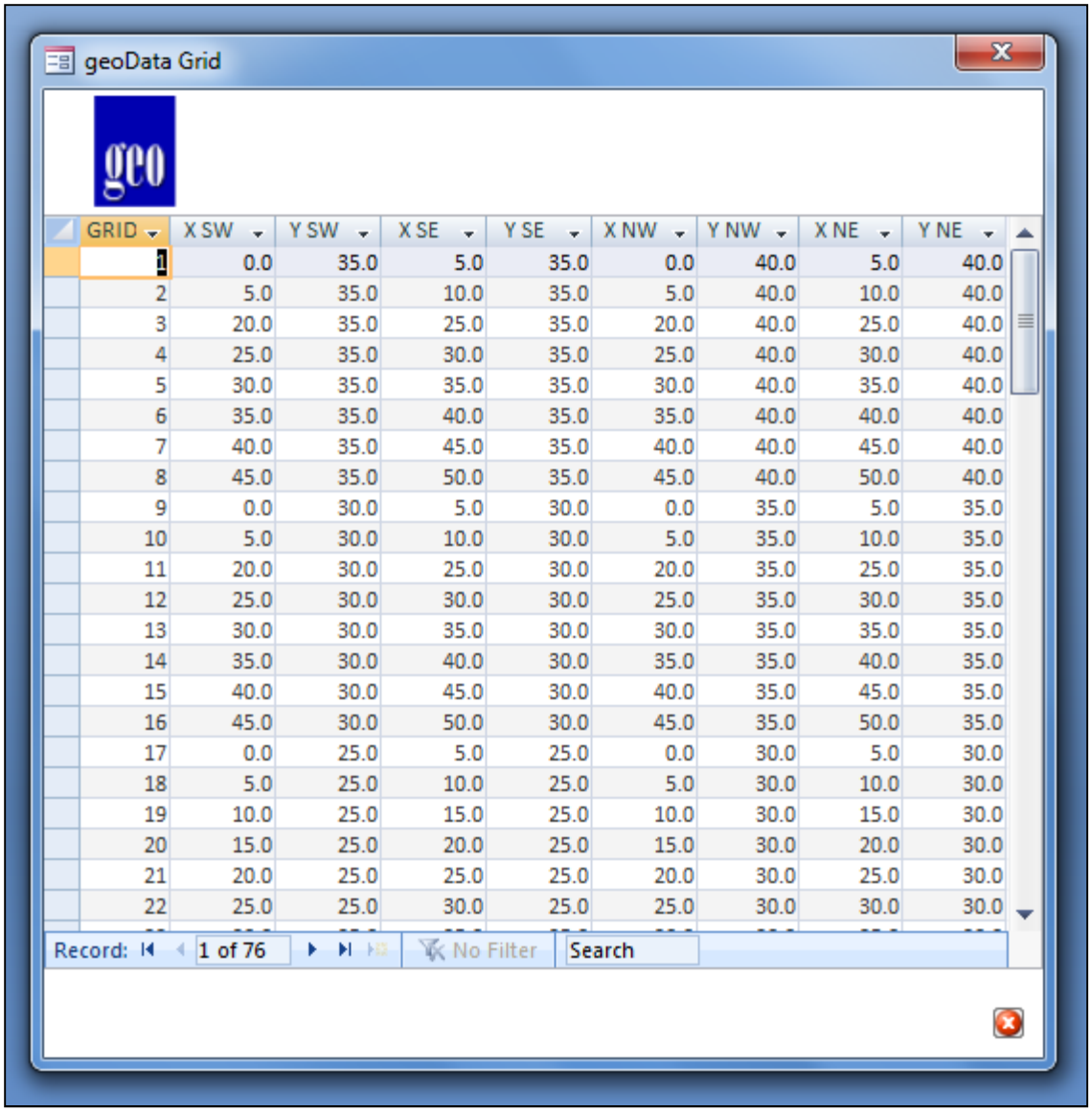


Figure 27. Representative example of the *geoData Grid* table. Example table represents the Grid (No Shift) table.

The Grid (1) table contains 76 grids and the x and y coordinates establishing the parameter of each grid. The Grid X Shift (2) table and Grid Y Shift (3) table each contain 86 grids. Both the Grid X Shift table and Grid Y Shift table use a formula to create a shift of 2.5 meters along the X or Y axes in either direction.

The three Grid tables are combined separately with the nine *geoData EMG* queries to create 27 different *Max Value* tables for the expressions and value of the EMG according to each grid. Each expression from the nine *geoData EMG* queries is assigned a max value for each grid on each grid table (**Figure 28**).

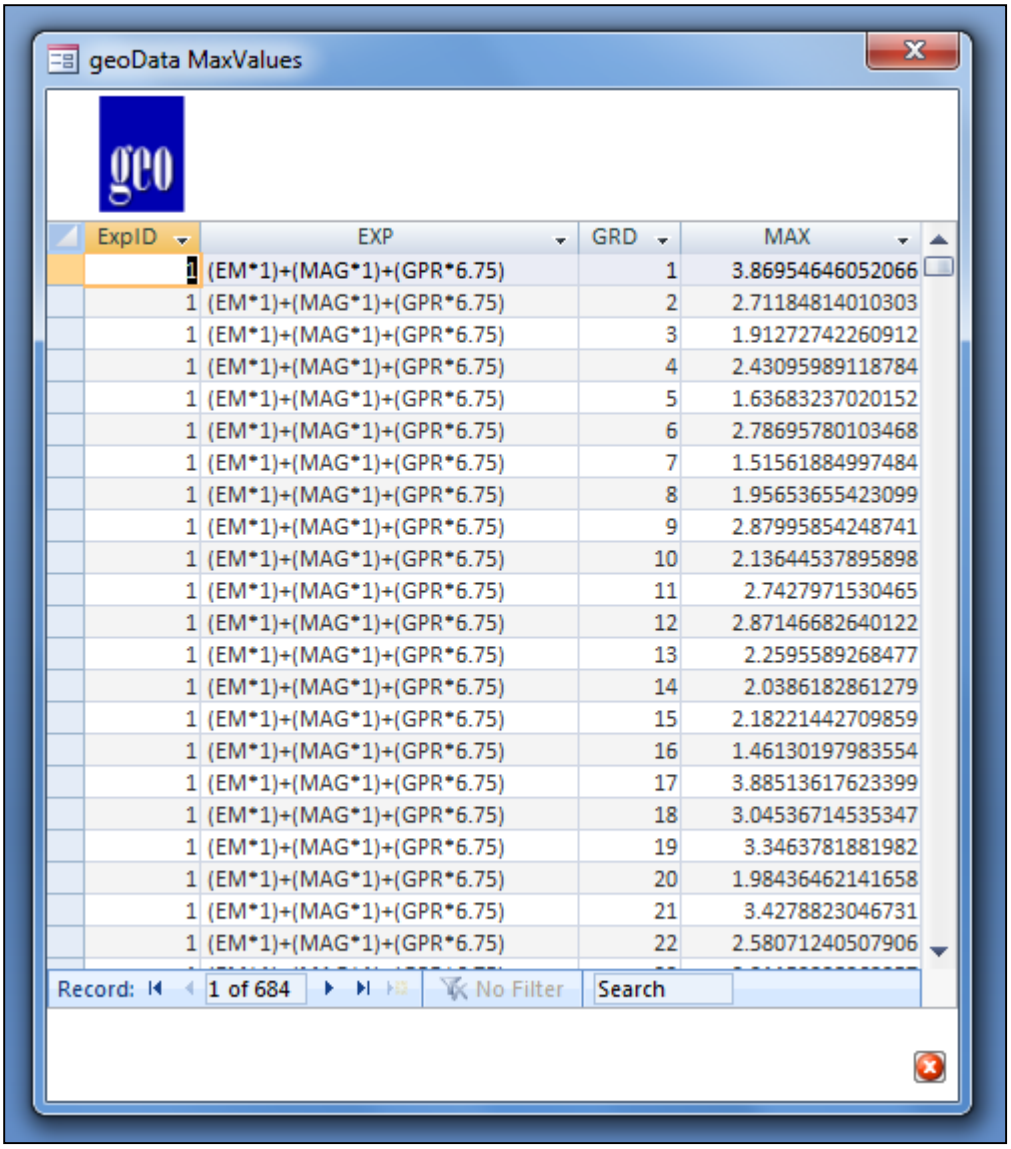


Figure 28. The *geoData Max Values* query, displaying the 27 Max Value tables for each expression of each grid for the three different Grid Shift tables.

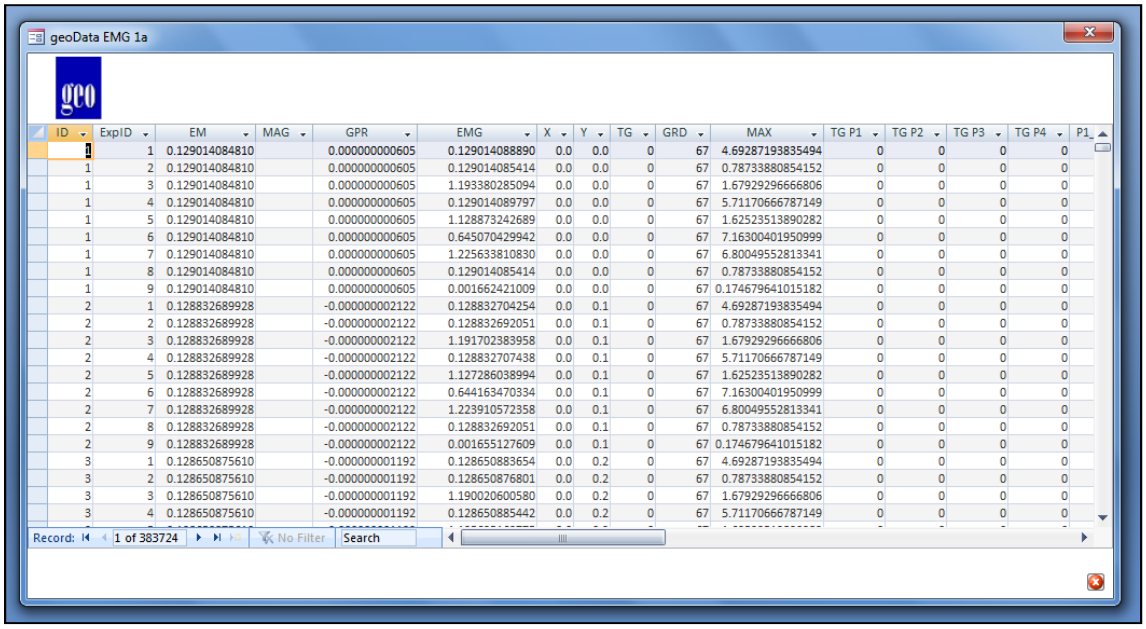
The *geoData Max* queries were combined along with the *geoData Accuracy Percent* table [similar to what was done with the synthetic data tables] and with the *geoData EMG* tables to create a query based on expression from the *geoData Expressions* query. The EMG data of each query is processed using the following formulas to create a query for each expression according to the grid the expression belongs to:

> If (EMG > (Max * 99%(PP1)), then TG P1(Target 99%) =1, if false= 0
>
> If (EMG between (Max * 99%(PP1)) and (Max* 95%(PP2)), then TG P2(Target 95%) =1, if false= 0
>
> If (EMG between (Max * 95%(PP2)) and (Max* 90%(PP3)), then TG P3(Target 90%) =1, if false= 0
>
> If (EMG between (Max * 90%(PP2)) and (Max* 85%(PP3)), then TG P4(Target 85%) =1, if false= 0

The nine *geoData EMG* tables were combined with the three different *geoData Resolution* tables, three different *geoData Grid* tables, *geoData Accuracy Percentage* table and 27 *geoData Max Value* tables to create 27 new *geoData* tables, shown in **Table 11**. **Figure 29** gives a representative view of one of these 27 new *geoData* tables.

Table 11. Summary table showing what tables are integrated together to form the new *geoData* Table to be used in evaluation of anomaly detection over the survey site.

Table ID	*geoData EMG* Table	Resolution Type	Grid Type	New *geoData* Table
1	EM - MAG - GPR1a	1	No Shift	Data Type 1a EMG
2	EM - MAG - GPR1a	1	X Shift	Data Type 1a X EMG
3	EM - MAG - GPR1a	1	Y Shift	Data Type 1a Y EMG
4	EM - MAG - GPR1b	1	No Shift	Data Type 1b EMG
5	EM - MAG - GPR1b	1	X Shift	Data Type 1b X EMG
6	EM - MAG - GPR1b	1	Y Shift	Data Type 1b Y EMG
7	EM - MAG - GPR1c	1	No Shift	Data Type 1c EMG
8	EM - MAG - GPR1c	1	X Shift	Data Type 1c X EMG
9	EM - MAG - GPR1c	1	Y Shift	Data Type 1c Y EMG
10	EM - MAG - GPR2a	2	No Shift	Data Type 2a EMG
11	EM - MAG - GPR2a	2	X Shift	Data Type 2a X EMG
12	EM - MAG - GPR2a	2	Y Shift	Data Type 2a Y EMG
13	EM - MAG - GPR2b	2	No Shift	Data Type 2b EMG
14	EM - MAG - GPR2b	2	X Shift	Data Type 2b X EMG
15	EM - MAG - GPR2b	2	Y Shift	Data Type 2b Y EMG
16	EM - MAG - GPR2c	2	No Shift	Data Type 2c EMG
17	EM - MAG - GPR2c	2	X Shift	Data Type 2c X EMG
18	EM - MAG - GPR2c	2	Y Shift	Data Type 2c Y EMG
19	EM - MAG - GPR3a	3	No Shift	Data Type 3a EMG
20	EM - MAG - GPR3a	3	X Shift	Data Type 3a X EMG
21	EM - MAG - GPR3a	3	Y Shift	Data Type 3a Y EMG
22	EM - MAG - GPR3b	3	No Shift	Data Type 3b EMG
23	EM - MAG - GPR3b	3	X Shift	Data Type 3b X EMG
24	EM - MAG - GPR3b	3	Y Shift	Data Type 3b Y EMG
25	EM - MAG - GPR3c	3	No Shift	Data Type 3c EMG
26	EM - MAG - GPR3c	3	X Shift	Data Type 3c X EMG
27	EM - MAG - GPR3c	3	Y Shift	Data Type 3c Y EMG



ID	ExpID	EM	MAG	GPR	EMG	X	Y	TG	GRD	MAX	TG P1	TG P2	TG P3	TG P4	P1_
1	1	0.129014084810		0.000000000605	0.129014088890	0.0	0.0	0	67	4.69287193835494	0	0	0	0	
1	2	0.129014084810		0.000000000605	0.129014085414	0.0	0.0	0	67	0.78733880854152	0	0	0	0	
1	3	0.129014084810		0.000000000605	1.193380285094	0.0	0.0	0	67	1.67929296666806	0	0	0	0	
1	4	0.129014084810		0.000000000605	0.129014089797	0.0	0.0	0	67	5.71170666787149	0	0	0	0	
1	5	0.129014084810		0.000000000605	1.128873242689	0.0	0.0	0	67	1.62523513890282	0	0	0	0	
1	6	0.129014084810		0.000000000605	0.645070429942	0.0	0.0	0	67	7.16300401950999	0	0	0	0	
1	7	0.129014084810		0.000000000605	1.225633810830	0.0	0.0	0	67	6.80049552813341	0	0	0	0	
1	8	0.129014084810		0.000000000605	0.129014085414	0.0	0.0	0	67	0.78733880854152	0	0	0	0	
1	9	0.129014084810		0.000000000605	0.001662421009	0.0	0.0	0	67	0.174679641015182	0	0	0	0	
2	1	0.128832689928		-0.000000002122	0.128832704254	0.0	0.1	0	67	4.69287193835494	0	0	0	0	
2	2	0.128832689928		-0.000000002122	0.128832692051	0.0	0.1	0	67	0.78733880854152	0	0	0	0	
2	3	0.128832689928		-0.000000002122	1.191702383958	0.0	0.1	0	67	1.67929296666806	0	0	0	0	
2	4	0.128832689928		-0.000000002122	0.128832707438	0.0	0.1	0	67	5.71170666787149	0	0	0	0	
2	5	0.128832689928		-0.000000002122	1.127286038994	0.0	0.1	0	67	1.62523513890282	0	0	0	0	
2	6	0.128832689928		-0.000000002122	0.644163470334	0.0	0.1	0	67	7.16300401950999	0	0	0	0	
2	7	0.128832689928		-0.000000002122	1.223910572358	0.0	0.1	0	67	6.80049552813341	0	0	0	0	
2	8	0.128832689928		-0.000000002122	0.128832692051	0.0	0.1	0	67	0.78733880854152	0	0	0	0	
2	9	0.128832689928		-0.000000002122	0.001655127609	0.0	0.1	0	67	0.174679641015182	0	0	0	0	
3	1	0.128650875610		-0.000000001192	0.128650883654	0.0	0.2	0	67	4.69287193835494	0	0	0	0	
3	2	0.128650875610		-0.000000001192	0.128650876801	0.0	0.2	0	67	0.78733880854152	0	0	0	0	
3	3	0.128650875610		-0.000000001192	1.190020600580	0.0	0.2	0	67	1.67929296666806	0	0	0	0	
3	4	0.128650875610		-0.000000001192	0.128650885442	0.0	0.2	0	67	5.71170666787149	0	0	0	0	

Figure 29. Representative example of the *geoData EMG* query. This example is one of 27 listed in Table 10 (i.e., Type 1 Resolution at depth of 0.33m).

These 27 *geoData EMG* queries are individually summarized to sum the Target and sum each Percentage Target for every expression defined by the grid and data type and creates the *geoData EMG Accuracy* query (see **Figure 30**). This formula is used for the four results (1-1, 1-0, 0-1, 0-0) of each percentage 99%(P1), 95%(P2), 90%(P3), 85%(P4) stated on *geoData EMG Accuracy* query. Below is the detail format for P1(99%):

P1_11: [P1_1_1]/([P1_1_1]+[P1_1_0])

P1_10: [P1_1_0]/([P1_1_1]+[P1_1_0])

P1_01: [P1_0_1]/([P1_0_1]+[P1_0_0])

P1_00: [P1_0_0]/([P1_0_1]+[P1_0_0])

The *geoData EMG Accuracy* Summary is a sample query describing the process of all three grids and an individual summary has been created for each grid.

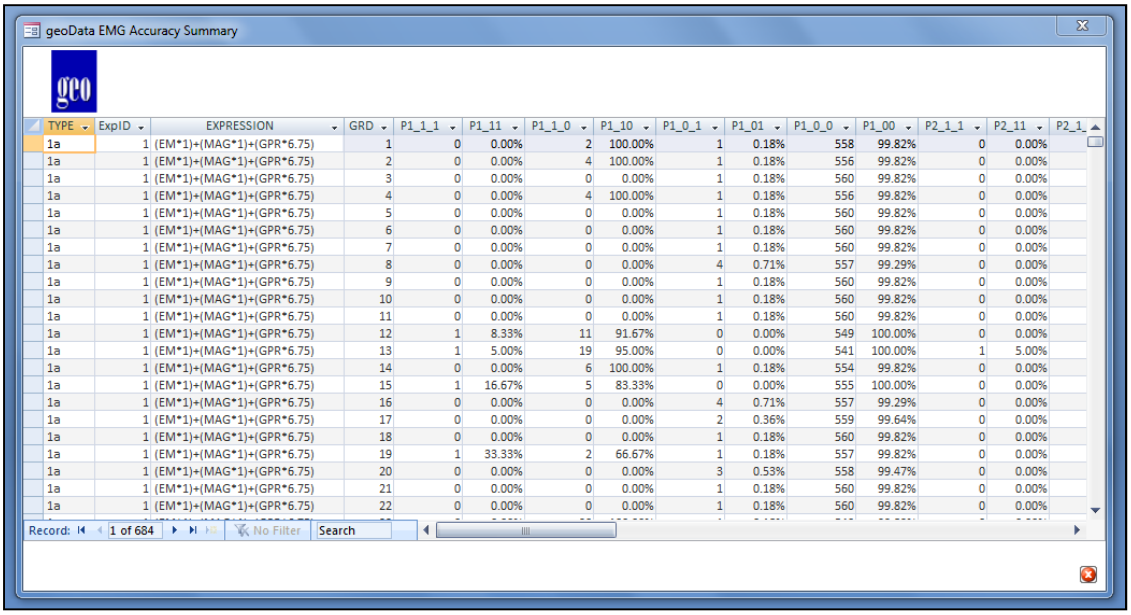


Figure 30. Representative example of the *geoData EMG Accuracy Summary*. Example query represents the combination of the EMG tables containing the three types of data shifts and three levels of resolutions.

4.2.4 Statistical Analysis

In order to run any statistical analysis in SAS, the grid file created that contains the location of the anomalies must be at the same resolution as the data points collected with the various geophysical techniques. As discussed previously, each geophysical technique, and respective data set, is at a different resolution. To remedy this problem, individual files were created to match the locations of the buried targets to the corresponding grid locations with each geophysical technique. For example, the magnetic data was collected with a grid size of 1.0 m by 0.125 m. The target location was then identified with an X-Y coordinate according to the "southwest" corner of the grid space within the magnetic data. It is also important to note that during this phase of the research the raw data values are used in preparing these grids and no interpolation has taken place.

The location of the targets within the Geophysical Research Station and associated descriptions can be found in **Figure 19** and **Table 10**. The same statistical methods of Part 1 were employed during this stage of the research: *logistical regression, proc GLM, proc frequency, and univariate*. These tools will be used to prepare the data for use in the testing of the developed data integration methodologies and aid in the quantitative assessment of target identification.

In addition to these tools, kriging was used to create the datasets of varying resolutions. Kriging is a group of geostatistical techniques to interpolate the value of a random field (e.g., the amplitude of the transmitting EM wave, z, of the subsurface as a function of the geographic location) at an unobserved location from observations of its value at nearby locations. Ordinary kriging was performed locally on each grid by using

only data points within a specified radius (10cm) of each grid point. Ordinary kriging is based on the following assumptions: (1) the observations are a partial realization of a random function Z(x), where x denotes spatial location; (2) the random function is second-order stationary, so the mean, spatial covariance, and semivariance do not depend upon x; (3) the mean is known and (4) the mean is constant of the regionalized variables are constant throughout the area of interest (Davis 2022).

Kriging of the data included trend data and the contouring of data was accomplished by defining the prediction grid point (node) locations. The prediction grid was rectangular, with the grid points population and spacing based on the available data in the GPR surveys. Based on the spatial distribution of the GPR data and the range of the linear covariance model, for each prediction location there were up to 16 neighboring data points that contributed to the prediction value. Nodes close to the boundaries of the prediction grid were sometimes not calculated within the kriging process, attributing to the edge effect of the survey. The size of these neighborhoods depended on the range of the specified covariance model that characterized the spatial continuity of the domain, and the prediction radius. The standard errors tended to increase toward the borders of the prediction area, beyond which no observations were available.

4.2.5 Mosaic of Data Grids

Spatial visualization software (Golden Software's Surfer 8) is used to interpolate, by means of kriging, irregularly spaced XYZ data into a regularly spaced grid and display the geophysical data in 2D and possibly 3D representations. Once all data are exported

from each individual instrument, grid files are created for quantitative merging of the data and to create grid-based maps including contour, image, shaded relief, and surface maps. The program Surfer works in conjunction with the Access databases and tables to take the individual grids from the survey field and mosaic them together for a complete, enhanced view of the subsurface. Advantages of using Surfer are the ease of use and the ability to manipulate large datasets.

In order to display and manipulate the data, they needed to be: (1) in the Surfer data format or the Surfer grid format, and (2) in a three column format, where the first column is the horizontal x location, the second column is the horizontal y location and the third column is the data amplitudes. Once the data is in the proper format, they could be converted to Surfer grids. In order to merge the two grids in the Surfer program, the grids needed to be the same size, meaning they have the same number of rows and columns (Ambrose 2005). It is very important to keep in mind throughout this process that there is no "best fit" for data integration. Depending on the subsurface features, expected target types, project objectives, etc., there will be a more appropriate combination of functions to merge the data sets. It is the ultimate goal of this research to determine what those variations are.

4.3 Results

The following tables and figures are summaries of the aforementioned methods and give additional representative samples of the results. These results clearly show that the hypothesis was proven correct.

Table 12. Summary for percentage accuracy in detecting targets across the control site survey area defined by thresholds for manipulating the data sets by raising each normalized value by its optimal exponent prior to integration. Resolution type contains three depths: (a) 0.33m, (b) 0.66m, and (c) 1.00m. Grouping are for (1) No Shift in grid positioning, (2) Shift along the X axis (North/South), (3) Shift along the Y axis (East/West), (4) Difference in accuracy between groupings 2 and 3, (5) Difference in accuracy between groupings 1 and 3, (6) Difference in accuracy between groupings 1 and 2.

		99% Threshold				95% Threshold				90% Threshold				85% Threshold			
	Resolution Type	*Actual 1, Calculated 0*	*Actual 1, Calculated 1*	*Actual 0, Calculated 1*	*Actual 0, Calculated 0*	*Actual 1, Calculated 0*	*Actual 1, Calculated 1*	*Actual 0, Calculated 1*	*Actual 0, Calculated 0*	*Actual 1, Calculated 0*	*Actual 1, Calculated 1*	*Actual 0, Calculated 1*	*Actual 0, Calculated 0*	*Actual 1, Calculated 0*	*Actual 1, Calculated 1*	*Actual 0, Calculated 1*	*Actual 0, Calculated 0*
1	1a	7.00%	93.00%	3.36%	96.64%	13.00%	87.00%	13.81%	86.19%	11.00%	89.00%	19.14%	80.86%	0.00%	100.00%	12.14%	87.86%
	1b	0.00%	100.00%	1.25%	98.75%	0.00%	100.00%	0.68%	99.32%	0.25%	99.75%	1.67%	98.33%	0.00%	100.00%	0.68%	99.32%
	1c	0.25%	99.75%	0.67%	99.33%	1.02%	98.98%	1.69%	98.31%	3.30%	96.70%	3.59%	98.28%	0.00%	100.00%	1.66%	98.34%
	2a	7.00%	93.00%	3.36%	96.64%	13.00%	87.00%	13.81%	86.19%	11.00%	89.00%	19.14%	80.86%	0.00%	100.00%	12.14%	87.86%
	2b	4.00%	96.00%	2.30%	97.70%	3.00%	97.00%	1.26%	98.74%	2.00%	98.00%	3.64%	96.36%	0.00%	100.00%	1.25%	98.75%
	2c	4.00%	96.00%	2.22%	97.78%	1.00%	99.00%	3.48%	96.52%	4.00%	96.00%	6.92%	93.08%	0.00%	100.00%	3.36%	96.64%
	3a	2.06%	97.94%	1.36%	98.64%	13.53%	86.47%	12.14%	87.86%	6.76%	93.24%	17.51%	82.49%	0.00%	100.00%	10.82%	89.18%
	3b	0.00%	100.00%	0.70%	99.30%	0.00%	100.00%	1.08%	98.92%	0.59%	99.41%	1.70%	98.30%	0.00%	100.00%	1.07%	98.93%
	3c	0.29%	99.71%	0.55%	99.45%	0.88%	99.12%	1.79%	98.21%	2.94%	97.06%	3.52%	96.48%	0.00%	100.00%	1.76%	98.24%
2	1aX	2.92%	97.08%	1.26%	98.74%	9.91%	90.09%	10.10%	89.90%	9.62%	90.38%	14.67%	85.33%	0.00%	100.00%	9.17%	90.83%
	1bX	0.00%	100.00%	1.32%	98.68%	0.52%	99.48%	1.27%	98.73%	1.29%	98.71%	3.81%	98.69%	0.00%	100.00%	1.26%	98.74%
	1cX	0.87%	99.13%	0.59%	99.41%	2.04%	97.96%	2.01%	97.99%	3.21%	96.79%	3.78%	97.95%	0.00%	100.00%	1.97%	98.03%
	2aX	8.86%	91.14%	3.49%	96.51%	10.13%	89.87%	12.35%	87.65%	12.66%	87.34%	16.26%	83.74%	0.00%	100.00%	10.99%	89.01%
	2bX	7.59%	92.41%	2.91%	97.09%	1.27%	98.73%	2.81%	97.19%	2.53%	97.47%	3.49%	96.51%	0.00%	100.00%	2.74%	97.26%
	2cX	7.59%	92.41%	2.77%	97.23%	3.80%	96.20%	4.56%	95.44%	3.80%	96.20%	8.14%	91.86%	0.00%	100.00%	4.36%	95.64%
	3aX	3.04%	96.96%	1.38%	98.62%	10.81%	89.19%	10.10%	89.90%	8.45%	91.55%	14.30%	85.70%	0.00%	100.00%	9.17%	90.83%
	3bX	0.68%	99.32%	0.71%	99.29%	1.69%	98.31%	1.65%	98.35%	0.34%	99.66%	2.18%	97.82%	0.00%	100.00%	1.63%	98.37%
	3cX	0.68%	99.32%	0.56%	99.44%	1.69%	98.31%	2.26%	97.74%	2.70%	97.30%	4.22%	95.78%	0.00%	100.00%	2.21%	97.79%
3	1aY	2.33%	97.67%	1.34%	98.66%	10.08%	89.92%	10.84%	89.16%	20.16%	79.84%	18.00%	82.00%	0.00%	100.00%	9.78%	90.22%
	1bY	0.00%	100.00%	1.32%	98.68%	0.52%	99.48%	1.27%	98.73%	1.29%	98.71%	3.81%	98.69%	0.00%	100.00%	1.26%	98.74%
	1cY	0.78%	99.22%	0.57%	99.43%	0.26%	99.74%	1.31%	98.69%	3.36%	96.64%	5.26%	98.64%	0.00%	100.00%	1.29%	98.71%
	2aY	7.84%	92.16%	3.22%	96.78%	10.78%	89.22%	13.25%	86.75%	19.61%	80.39%	20.78%	79.22%	0.00%	100.00%	11.70%	88.30%
	2bY	4.90%	95.10%	2.68%	97.32%	2.94%	97.06%	2.30%	97.70%	1.96%	98.04%	4.65%	95.35%	0.00%	100.00%	2.25%	97.75%
	2cY	5.88%	94.12%	2.32%	97.68%	3.92%	96.08%	4.51%	95.49%	8.82%	91.18%	10.29%	95.24%	0.00%	100.00%	4.31%	95.69%
	3aY	2.40%	97.60%	1.47%	98.53%	9.91%	90.09%	10.84%	89.16%	20.12%	79.88%	17.64%	82.36%	0.00%	100.00%	9.78%	90.22%
	3bY	0.00%	100.00%	0.79%	99.21%	0.60%	99.40%	1.89%	98.11%	1.50%	98.50%	3.67%	96.33%	0.00%	100.00%	1.86%	98.14%
	3cY	0.90%	99.10%	0.49%	99.51%	0.30%	99.70%	1.61%	98.39%	3.00%	97.00%	4.88%	95.12%	0.00%	100.00%	1.58%	98.42%
4	1a X-Y	0.59%	-0.59%	-0.08%	0.08%	-0.17%	0.17%	-0.74%	0.74%	-10.54%	10.54%	-3.33%	3.33%	0.00%	0.00%	-0.61%	0.61%
	1b X-Y	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
	1c X-Y	0.09%	-0.09%	0.02%	-0.02%	1.78%	-1.78%	0.70%	-0.70%	-0.15%	0.15%	-1.48%	-0.69%	0.00%	0.00%	0.68%	-0.68%
	2a X-Y	1.02%	-1.02%	0.27%	-0.27%	-0.65%	0.65%	-0.90%	0.90%	-6.95%	6.95%	-4.52%	4.52%	0.00%	0.00%	-0.71%	0.71%
	2b X-Y	2.69%	-2.69%	0.23%	-0.23%	-1.67%	1.67%	0.51%	-0.51%	0.57%	-0.57%	-1.16%	1.16%	0.00%	0.00%	0.49%	-0.49%
	2c X-Y	1.71%	-1.71%	0.45%	-0.45%	-0.12%	0.12%	0.05%	-0.05%	-5.02%	5.02%	-2.15%	-3.38%	0.00%	0.00%	0.05%	-0.05%
	3a X-Y	0.64%	-0.64%	-0.09%	0.09%	0.90%	-0.90%	-0.74%	0.74%	-11.67%	11.67%	-3.34%	3.34%	0.00%	0.00%	-0.61%	0.61%
	3b X-Y	0.68%	-0.68%	-0.08%	0.08%	1.09%	-1.09%	-0.24%	0.24%	-1.16%	1.16%	-1.49%	1.49%	0.00%	0.00%	-0.23%	0.23%
	3c X-Y	-0.22%	0.22%	0.07%	-0.07%	1.39%	-1.39%	0.65%	-0.65%	-0.30%	0.30%	-0.66%	0.66%	0.00%	0.00%	0.63%	-0.63%
5	1a N-Y	4.67%	-4.67%	2.02%	-2.02%	2.92%	-2.92%	2.97%	-2.97%	-9.16%	9.16%	1.14%	-1.14%	0.00%	0.00%	2.36%	-2.36%
	1b N-Y	0.00%	0.00%	-0.07%	0.07%	-0.52%	0.52%	-0.59%	0.59%	-1.04%	1.04%	-2.14%	-0.36%	0.00%	0.00%	-0.58%	0.58%
	1c N-Y	-0.53%	0.53%	0.10%	-0.10%	0.76%	-0.76%	0.38%	-0.38%	-0.06%	0.06%	-1.67%	-0.36%	0.00%	0.00%	0.37%	-0.37%
	2a N-Y	-0.84%	0.84%	0.14%	-0.14%	2.22%	-2.22%	0.56%	-0.56%	-8.61%	8.61%	-1.64%	1.64%	0.00%	0.00%	0.44%	-0.44%
	2b N-Y	-0.90%	0.90%	-0.38%	0.38%	0.06%	-0.06%	-1.04%	1.04%	0.04%	-0.04%	-1.01%	1.01%	0.00%	0.00%	-1.00%	1.00%
	2c N-Y	-1.88%	1.88%	-0.10%	0.10%	-2.92%	2.92%	-1.03%	1.03%	-4.82%	4.82%	-3.37%	-2.16%	0.00%	0.00%	-0.95%	0.95%
	3a N-Y	-0.34%	0.34%	-0.11%	0.11%	3.62%	-3.62%	1.30%	-1.30%	-13.36%	13.36%	-0.13%	0.13%	0.00%	0.00%	1.04%	-1.04%
	3b N-Y	0.00%	0.00%	-0.09%	0.09%	-0.60%	0.60%	-0.81%	0.81%	-0.91%	0.91%	-1.97%	1.97%	0.00%	0.00%	-0.79%	0.79%
	3c N-Y	-0.61%	0.61%	0.06%	-0.06%	0.58%	-0.58%	0.18%	-0.18%	-0.06%	0.06%	-1.36%	1.36%	0.00%	0.00%	0.18%	-0.18%
6	1a N-X	4.08%	-4.08%	2.10%	-2.10%	3.09%	-3.09%	3.71%	-3.71%	1.38%	-1.38%	4.47%	-4.47%	0.00%	0.00%	2.97%	-2.97%
	1b N-X	0.00%	0.00%	-0.07%	0.07%	-0.52%	0.52%	-0.59%	0.59%	-1.04%	1.04%	-2.14%	-0.36%	0.00%	0.00%	-0.58%	0.58%
	1c N-X	-0.62%	0.62%	0.08%	-0.08%	-1.02%	1.02%	-0.32%	0.32%	0.09%	-0.09%	-0.19%	0.33%	0.00%	0.00%	-0.31%	0.31%
	2a N-X	-1.86%	1.86%	-0.13%	0.13%	2.87%	-2.87%	1.46%	-1.46%	-1.66%	1.66%	2.88%	-2.88%	0.00%	0.00%	1.15%	-1.15%
	2b N-X	-3.59%	3.59%	-0.61%	0.61%	1.73%	-1.73%	-1.55%	1.55%	-0.53%	0.53%	0.15%	-0.15%	0.00%	0.00%	-1.49%	1.49%
	2c N-X	-3.59%	3.59%	-0.55%	0.55%	-2.80%	2.80%	-1.08%	1.08%	0.20%	-0.20%	-1.22%	1.22%	0.00%	0.00%	-1.00%	1.00%
	3a N-X	-0.98%	0.98%	-0.02%	0.02%	2.72%	-2.72%	2.04%	-2.04%	-1.69%	1.69%	3.21%	-3.21%	0.00%	0.00%	1.65%	-1.65%
	3b N-X	-0.68%	0.68%	-0.01%	0.01%	-1.69%	1.69%	-0.57%	0.57%	0.25%	-0.25%	-0.48%	0.48%	0.00%	0.00%	-0.56%	0.56%
	3c N-X	-0.39%	0.39%	-0.01%	0.01%	-0.81%	0.81%	-0.47%	0.47%	0.24%	-0.24%	-0.70%	0.70%	0.00%	0.00%	-0.45%	0.45%

Table 13. Summary for percentage accuracy in detecting targets across the control site survey area defined by thresholds for manipulating the data sets adding each normalized value together. Resolution type contains three depths: (a) 0.33m, (b) 0.66m, and (c) 1.00m. Grouping are for (1) No Shift in grid positioning, (2) Shift along the X axis (North/South), (3) Shift along the Y axis (East/West), (4) Difference in accuracy between groupings 2 and 3, (5) Difference in accuracy between groupings 1 and 3, (6) Difference in accuracy between groupings 1 and 2.

		99% Threshold				95% Threshold				90% Threshold				85% Threshold			
	Resolution Type	Actual 1, Calculated 0	Actual 1, Calculated 1	Actual 0, Calculated 1	Actual 0, Calculated 0	Actual 1, Calculated 0	Actual 1, Calculated 1	Actual 0, Calculated 1	Actual 0, Calculated 0	Actual 1, Calculated 0	Actual 1, Calculated 1	Actual 0, Calculated 1	Actual 0, Calculated 0	Actual 1, Calculated 0	Actual 1, Calculated 1	Actual 0, Calculated 1	Actual 0, Calculated 0
1	1a	10.00%	90.00%	1.46%	98.54%	0.00%	100.00%	0.47%	99.53%	4.00%	96.00%	0.71%	99.29%	0.00%	100.00%	0.47%	99.53%
	1b	0.51%	99.49%	1.17%	98.83%	0.25%	99.75%	1.38%	98.62%	1.02%	98.98%	1.55%	98.45%	0.00%	100.00%	1.37%	98.63%
	1c	0.25%	99.75%	0.45%	99.55%	0.00%	100.00%	0.32%	99.68%	2.03%	97.97%	0.35%	99.68%	0.00%	100.00%	0.32%	99.68%
	2a	10.00%	90.00%	1.46%	98.54%	0.00%	100.00%	0.47%	99.53%	4.00%	96.00%	0.71%	99.29%	0.00%	100.00%	0.47%	99.53%
	2b	8.00%	92.00%	2.24%	97.76%	0.00%	100.00%	1.26%	98.74%	1.00%	99.00%	2.07%	97.93%	0.00%	100.00%	1.25%	98.75%
	2c	3.00%	97.00%	1.63%	98.37%	1.00%	99.00%	0.39%	99.61%	0.00%	100.00%	0.71%	99.29%	0.00%	100.00%	0.39%	99.61%
	3a	1.76%	98.24%	0.27%	99.73%	0.29%	99.71%	0.32%	99.68%	1.18%	98.82%	0.57%	99.43%	0.00%	100.00%	0.31%	99.69%
	3b	0.29%	99.71%	1.03%	98.97%	0.29%	99.71%	1.41%	98.59%	0.88%	99.12%	1.55%	98.45%	0.00%	100.00%	1.39%	98.61%
	3c	0.29%	99.71%	0.45%	99.55%	0.00%	100.00%	0.30%	99.70%	1.76%	98.24%	0.39%	99.61%	0.00%	100.00%	0.30%	99.70%
2	1aX	1.17%	98.83%	0.24%	99.76%	1.75%	98.25%	0.33%	99.67%	1.17%	98.83%	0.58%	99.42%	0.00%	100.00%	0.33%	99.67%
	1bX	0.00%	100.00%	1.27%	98.73%	0.26%	99.74%	1.33%	98.67%	1.29%	98.71%	1.56%	98.67%	0.00%	100.00%	1.31%	98.69%
	1cX	0.87%	99.13%	0.43%	99.57%	0.29%	99.71%	0.37%	99.63%	2.33%	97.67%	0.45%	99.63%	0.00%	100.00%	0.37%	99.63%
	2aX	11.39%	88.61%	1.93%	98.07%	0.00%	100.00%	0.72%	99.28%	3.80%	96.20%	1.16%	98.84%	0.00%	100.00%	0.72%	99.28%
	2bX	10.13%	89.87%	2.70%	97.30%	1.27%	98.73%	1.63%	98.37%	1.27%	98.73%	2.40%	97.60%	0.00%	100.00%	1.60%	98.40%
	2cX	7.59%	92.41%	2.05%	97.95%	2.53%	97.47%	0.72%	99.28%	1.27%	98.73%	0.93%	99.07%	0.00%	100.00%	0.72%	99.28%
	3aX	1.69%	98.31%	0.30%	99.70%	0.68%	99.32%	0.33%	99.67%	1.69%	98.31%	0.63%	99.37%	0.00%	100.00%	0.33%	99.67%
	3bX	0.34%	99.66%	1.17%	98.83%	0.34%	99.66%	1.45%	98.55%	1.01%	98.99%	1.57%	98.43%	0.00%	100.00%	1.43%	98.57%
	3cX	0.34%	99.66%	0.44%	99.56%	0.34%	99.66%	0.35%	99.65%	2.36%	97.64%	0.55%	99.45%	0.00%	100.00%	0.35%	99.65%
3	1aY	1.81%	98.19%	0.24%	99.76%	1.03%	98.97%	0.38%	99.62%	0.78%	99.22%	0.67%	99.33%	0.00%	100.00%	0.38%	99.62%
	1bY	0.00%	100.00%	1.27%	98.73%	0.26%	99.74%	1.33%	98.67%	1.29%	98.71%	1.56%	98.67%	0.00%	100.00%	1.31%	98.69%
	1cY	0.26%	99.74%	0.48%	99.52%	0.00%	100.00%	0.34%	99.66%	2.07%	97.93%	0.45%	99.65%	0.00%	100.00%	0.34%	99.66%
	2aY	11.76%	88.24%	1.61%	98.39%	0.00%	100.00%	0.64%	99.36%	4.90%	95.10%	0.87%	99.13%	0.00%	100.00%	0.63%	99.37%
	2bY	5.88%	94.12%	2.40%	97.60%	0.00%	100.00%	1.35%	98.65%	0.98%	99.02%	2.07%	97.93%	0.00%	100.00%	1.33%	98.67%
	2cY	4.90%	95.10%	1.81%	98.19%	0.98%	99.02%	0.36%	99.64%	1.96%	98.04%	0.82%	99.64%	0.00%	100.00%	0.36%	99.64%
	3aY	2.40%	97.60%	0.29%	99.71%	0.60%	99.40%	0.38%	99.62%	1.50%	98.50%	0.70%	99.30%	0.00%	100.00%	0.38%	99.62%
	3bY	0.00%	100.00%	1.12%	98.88%	0.00%	100.00%	1.39%	98.61%	1.20%	98.80%	1.58%	98.42%	0.00%	100.00%	1.37%	98.63%
	3cY	0.30%	99.70%	0.50%	99.50%	0.00%	100.00%	0.36%	99.64%	1.80%	98.20%	0.45%	99.55%	0.00%	100.00%	0.36%	99.64%
4	1a X-Y	-0.64%	0.64%	-0.01%	0.01%	0.72%	-0.72%	-0.06%	0.06%	0.39%	-0.39%	-0.09%	0.09%	0.00%	0.00%	-0.06%	0.06%
	1b X-Y	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
	1c X-Y	0.62%	-0.62%	-0.05%	0.05%	0.29%	-0.29%	0.03%	-0.03%	0.27%	-0.27%	0.00%	-0.03%	0.00%	0.00%	0.03%	-0.03%
	2a X-Y	-0.37%	0.37%	0.32%	-0.32%	0.00%	0.00%	0.09%	-0.09%	-1.10%	1.10%	0.29%	-0.29%	0.00%	0.00%	0.08%	-0.08%
	2b X-Y	4.24%	-4.24%	0.29%	-0.29%	1.27%	-1.27%	0.28%	-0.28%	0.29%	-0.29%	0.33%	-0.33%	0.00%	0.00%	0.27%	-0.27%
	2c X-Y	2.69%	-2.69%	0.24%	-0.24%	1.55%	-1.55%	0.36%	-0.36%	-0.69%	0.69%	0.11%	-0.57%	0.00%	0.00%	0.36%	-0.36%
	3a X-Y	-0.71%	0.71%	0.01%	-0.01%	0.08%	-0.08%	-0.06%	0.06%	0.19%	-0.19%	-0.07%	0.07%	0.00%	0.00%	-0.06%	0.06%
	3b X-Y	0.34%	-0.34%	0.05%	-0.05%	0.34%	-0.34%	0.06%	-0.06%	-0.19%	0.19%	-0.02%	0.02%	0.00%	0.00%	0.06%	-0.06%
	3c X-Y	0.04%	-0.04%	-0.06%	0.06%	0.34%	-0.34%	-0.01%	0.01%	0.56%	-0.56%	0.10%	-0.10%	0.00%	0.00%	-0.01%	0.01%
5	1a N-Y	8.19%	-8.19%	1.22%	-1.22%	-1.03%	1.03%	0.08%	-0.08%	3.22%	-3.22%	0.04%	-0.04%	0.00%	0.00%	0.08%	-0.08%
	1b N-Y	0.51%	-0.51%	-0.10%	0.10%	0.00%	0.00%	0.06%	-0.06%	-0.28%	0.28%	-0.01%	-0.22%	0.00%	0.00%	0.06%	-0.06%
	1c N-Y	0.00%	0.00%	-0.04%	0.04%	0.00%	0.00%	-0.03%	0.03%	-0.04%	0.04%	-0.10%	0.03%	0.00%	0.00%	-0.03%	0.03%
	2a N-Y	-1.76%	1.76%	-0.14%	0.14%	0.00%	0.00%	-0.17%	0.17%	-0.90%	0.90%	-0.16%	0.16%	0.00%	0.00%	-0.17%	0.17%
	2b N-Y	2.12%	-2.12%	-0.17%	0.17%	0.00%	0.00%	-0.09%	0.09%	0.02%	-0.02%	0.01%	-0.01%	0.00%	0.00%	-0.09%	0.09%
	2c N-Y	-1.90%	1.90%	-0.18%	0.18%	0.02%	-0.02%	0.03%	-0.03%	-1.96%	1.96%	-0.11%	-0.35%	0.00%	0.00%	0.03%	-0.03%
	3a N-Y	-0.64%	0.64%	-0.02%	0.02%	-0.31%	0.31%	-0.07%	0.07%	-0.33%	0.33%	-0.13%	0.13%	0.00%	0.00%	-0.07%	0.07%
	3b N-Y	0.29%	-0.29%	-0.09%	0.09%	0.29%	-0.29%	0.02%	-0.02%	-0.32%	0.32%	-0.04%	0.04%	0.00%	0.00%	0.02%	-0.02%
	3c N-Y	-0.01%	0.01%	-0.04%	0.04%	0.00%	0.00%	-0.06%	0.06%	-0.04%	0.04%	-0.06%	0.06%	0.00%	0.00%	-0.06%	0.06%
6	1a N-X	8.83%	-8.83%	1.23%	-1.23%	-1.75%	1.75%	0.14%	-0.14%	2.83%	-2.83%	0.13%	-0.13%	0.00%	0.00%	0.14%	-0.14%
	1b N-X	0.51%	-0.51%	-0.10%	0.10%	0.00%	0.00%	0.06%	-0.06%	-0.28%	0.28%	-0.01%	-0.22%	0.00%	0.00%	0.06%	-0.06%
	1c N-X	-0.62%	0.62%	0.01%	-0.01%	-0.29%	0.29%	-0.05%	0.05%	-0.30%	0.30%	-0.10%	0.05%	0.00%	0.00%	-0.05%	0.05%
	2a N-X	-1.39%	1.39%	-0.47%	0.47%	0.00%	0.00%	-0.25%	0.25%	0.20%	-0.20%	-0.45%	0.45%	0.00%	0.00%	-0.25%	0.25%
	2b N-X	-2.13%	2.13%	-0.46%	0.46%	-1.27%	1.27%	-0.37%	0.37%	-0.27%	0.27%	-0.32%	0.32%	0.00%	0.00%	-0.36%	0.36%
	2c N-X	-4.59%	4.59%	-0.42%	0.42%	-1.53%	1.53%	-0.33%	0.33%	-1.27%	1.27%	-0.22%	0.22%	0.00%	0.00%	-0.33%	0.33%
	3a N-X	0.08%	-0.08%	-0.03%	0.03%	-0.38%	0.38%	-0.01%	0.01%	-0.51%	0.51%	-0.06%	0.06%	0.00%	0.00%	-0.01%	0.01%
	3b N-X	-0.04%	0.04%	-0.14%	0.14%	-0.04%	0.04%	-0.04%	0.04%	-0.13%	0.13%	-0.02%	0.02%	0.00%	0.00%	-0.04%	0.04%
	3c N-X	-0.04%	0.04%	0.02%	-0.02%	-0.34%	0.34%	-0.05%	0.05%	-0.60%	0.60%	-0.17%	0.17%	0.00%	0.00%	-0.05%	0.05%

Table 14. Summary for differences between data manipulation schemes [addition and optimal exponent] in percentage accuracy in detecting targets across the control site survey area defined by thresholds. Resolution type contains three depths: (a) 0.33m, (b) 0.66m, and (c) 1.00m. Grouping are for (1) No Shift in grid positioning, (2) Shift along the X axis (North/South), (3) Shift along the Y axis (East/West), (4) Difference in accuracy between groupings 2 and 3, (5) Difference in accuracy between groupings 1 and 3, (6) Difference in accuracy between groupings 1 and 2.

		99% Threshold				95% Threshold				90% Threshold				85% Threshold			
	Resolution Type	*Actual 1, Calculated 0*	*Actual 1, Calculated 1*	*Actual 0, Calculated 1*	*Actual 0, Calculated 0*	*Actual 1, Calculated 0*	*Actual 1, Calculated 1*	*Actual 0, Calculated 1*	*Actual 0, Calculated 0*	*Actual 1, Calculated 0*	*Actual 1, Calculated 1*	*Actual 0, Calculated 1*	*Actual 0, Calculated 0*	*Actual 1, Calculated 0*	*Actual 1, Calculated 1*	*Actual 0, Calculated 1*	*Actual 0, Calculated 0*
1	1a	3.00%	-3.00%	-1.90%	1.90%	-13.00%	13.00%	-13.34%	13.34%	-7.00%	7.00%	-18.43%	18.43%	0.00%	0.00%	-11.67%	11.67%
	1b	0.51%	-0.51%	-0.08%	0.08%	0.25%	-0.25%	0.70%	-0.70%	0.77%	-0.77%	-0.12%	0.12%	0.00%	0.00%	0.69%	-0.69%
	1c	0.00%	0.00%	-0.22%	0.22%	-1.02%	1.02%	-1.37%	1.37%	-1.27%	1.27%	-3.24%	1.40%	0.00%	0.00%	-1.34%	1.34%
	2a	3.00%	-3.00%	-1.90%	1.90%	-13.00%	13.00%	-13.34%	13.34%	-7.00%	7.00%	-18.43%	18.43%	0.00%	0.00%	-11.67%	11.67%
	2b	4.00%	-4.00%	-0.06%	0.06%	-3.00%	3.00%	0.00%	0.00%	-1.00%	1.00%	-1.57%	1.57%	0.00%	0.00%	0.00%	0.00%
	2c	-1.00%	1.00%	-0.59%	0.59%	0.00%	0.00%	-3.09%	3.09%	-4.00%	4.00%	-6.21%	6.21%	0.00%	0.00%	-2.97%	2.97%
	3a	-0.30%	0.30%	-1.09%	1.09%	-13.24%	13.24%	-11.82%	11.82%	-5.58%	5.58%	-16.94%	16.94%	0.00%	0.00%	-10.51%	10.51%
	3b	0.29%	-0.29%	0.33%	-0.33%	0.29%	-0.29%	0.33%	-0.33%	0.29%	-0.29%	-0.15%	0.15%	0.00%	0.00%	0.32%	-0.32%
	3c	0.00%	0.00%	-0.10%	0.10%	-0.88%	0.88%	-1.49%	1.49%	-1.18%	1.18%	-3.13%	3.13%	0.00%	0.00%	-1.46%	1.46%
2	1aX	-1.75%	1.75%	-1.02%	1.02%	-8.16%	8.16%	-9.77%	9.77%	-8.45%	8.45%	-14.09%	14.09%	0.00%	0.00%	-8.84%	8.84%
	1bX	0.00%	0.00%	-0.05%	0.05%	-0.26%	0.26%	0.06%	-0.06%	0.00%	0.00%	-2.25%	-0.02%	0.00%	0.00%	0.05%	-0.05%
	1cX	0.00%	0.00%	-0.16%	0.16%	-1.75%	1.75%	-1.64%	1.64%	-0.88%	0.88%	-3.33%	1.68%	0.00%	0.00%	-1.60%	1.60%
	2aX	2.53%	-2.53%	-1.56%	1.56%	-10.13%	10.13%	-11.63%	11.63%	-8.86%	8.86%	-15.10%	15.10%	0.00%	0.00%	-10.27%	10.27%
	2bX	2.54%	-2.54%	-0.21%	0.21%	0.00%	0.00%	-1.18%	1.18%	-1.26%	1.26%	-1.09%	1.09%	0.00%	0.00%	-1.14%	1.14%
	2cX	0.00%	0.00%	-0.72%	0.72%	-1.27%	1.27%	-3.84%	3.84%	-2.53%	2.53%	-7.21%	7.21%	0.00%	0.00%	-3.64%	3.64%
	3aX	-1.35%	1.35%	-1.08%	1.08%	-10.13%	10.13%	-9.77%	9.77%	-6.76%	6.76%	-13.67%	13.67%	0.00%	0.00%	-8.84%	8.84%
	3bX	-0.34%	0.34%	0.46%	-0.46%	-1.35%	1.35%	-0.20%	0.20%	0.67%	-0.67%	-0.61%	0.61%	0.00%	0.00%	-0.20%	0.20%
	3cX	-0.34%	0.34%	-0.12%	0.12%	-1.35%	1.35%	-1.91%	1.91%	-0.34%	0.34%	-3.67%	3.67%	0.00%	0.00%	-1.86%	1.86%
3	1aY	-0.52%	0.52%	-1.10%	1.10%	-9.04%	9.04%	-10.46%	10.46%	-19.38%	19.38%	-17.33%	17.33%	0.00%	0.00%	-9.40%	9.40%
	1bY	0.00%	0.00%	-0.05%	0.05%	-0.26%	0.26%	0.05%	-0.05%	0.00%	0.00%	-2.25%	-0.02%	0.00%	0.00%	0.05%	-0.05%
	1cY	-0.52%	0.52%	-0.09%	0.09%	-0.26%	0.26%	-0.96%	0.96%	-1.29%	1.29%	-4.81%	1.01%	0.00%	0.00%	-0.94%	0.94%
	2aY	3.92%	-3.92%	-1.61%	1.61%	-10.78%	10.78%	-12.61%	12.61%	-14.71%	14.71%	-19.90%	19.90%	0.00%	0.00%	-11.07%	11.07%
	2bY	0.98%	-0.98%	-0.28%	0.28%	-2.94%	2.94%	-0.95%	0.95%	-0.98%	0.98%	-2.58%	2.58%	0.00%	0.00%	-0.92%	0.92%
	2cY	-0.98%	0.98%	-0.52%	0.52%	-2.94%	2.94%	-4.15%	4.15%	-6.86%	6.86%	-9.48%	4.40%	0.00%	0.00%	-3.96%	3.96%
	3aY	0.00%	0.00%	-1.18%	1.18%	-9.31%	9.31%	-10.46%	10.46%	-18.62%	18.62%	-16.94%	16.94%	0.00%	0.00%	-9.40%	9.40%
	3bY	0.00%	0.00%	0.33%	-0.33%	-0.60%	0.60%	-0.50%	0.50%	-0.30%	0.30%	-2.08%	2.08%	0.00%	0.00%	-0.49%	0.49%
	3cY	-0.60%	0.60%	0.01%	-0.01%	-0.30%	0.30%	-1.25%	1.25%	-1.20%	1.20%	-4.43%	4.43%	0.00%	0.00%	-1.22%	1.22%
4	1a X-Y	-1.24%	1.24%	0.08%	-0.08%	0.88%	-0.88%	0.69%	-0.69%	10.93%	-10.93%	3.24%	-3.24%	0.00%	0.00%	0.56%	-0.56%
	1b X-Y	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
	1c X-Y	0.52%	-0.52%	-0.07%	0.07%	-1.49%	1.49%	-0.68%	0.68%	0.41%	-0.41%	1.48%	0.66%	0.00%	0.00%	-0.65%	0.65%
	2a X-Y	-1.39%	1.39%	0.05%	-0.05%	0.65%	-0.65%	0.99%	-0.99%	5.84%	-5.84%	4.81%	-4.81%	0.00%	0.00%	0.79%	-0.79%
	2b X-Y	1.56%	-1.56%	0.07%	-0.07%	2.94%	-2.94%	-0.23%	0.23%	-0.28%	0.28%	1.49%	-1.49%	0.00%	0.00%	-0.22%	0.22%
	2c X-Y	0.99%	-0.99%	-0.21%	0.21%	1.67%	-1.67%	0.31%	-0.31%	4.33%	-4.33%	2.27%	2.81%	0.00%	0.00%	0.31%	-0.31%
	3a X-Y	-1.35%	1.35%	0.10%	-0.10%	-0.83%	0.83%	0.69%	-0.69%	11.86%	-11.86%	3.27%	-3.27%	0.00%	0.00%	0.56%	-0.56%
	3b X-Y	-0.34%	0.34%	0.13%	-0.13%	-0.75%	0.75%	0.30%	-0.30%	0.97%	-0.97%	1.47%	-1.47%	0.00%	0.00%	0.29%	-0.29%
	3c X-Y	0.26%	-0.26%	-0.13%	0.13%	-1.05%	1.05%	-0.66%	0.66%	0.87%	-0.87%	0.77%	-0.77%	0.00%	0.00%	-0.64%	0.64%
5	1a N-Y	3.52%	-3.52%	-0.80%	0.80%	-3.96%	3.96%	-2.89%	2.89%	12.38%	-12.38%	-1.09%	1.09%	0.00%	0.00%	-2.28%	2.28%
	1b N-Y	0.51%	-0.51%	-0.03%	0.03%	0.51%	-0.51%	0.65%	-0.65%	0.77%	-0.77%	2.13%	0.14%	0.00%	0.00%	0.64%	-0.64%
	1c N-Y	0.52%	-0.52%	-0.13%	0.13%	-0.76%	0.76%	-0.41%	0.41%	0.02%	-0.02%	1.57%	0.39%	0.00%	0.00%	-0.40%	0.40%
	2a N-Y	-0.92%	0.92%	-0.29%	0.29%	-2.22%	2.22%	-0.73%	0.73%	7.71%	-7.71%	1.48%	-1.48%	0.00%	0.00%	-0.61%	0.61%
	2b N-Y	3.02%	-3.02%	0.22%	-0.22%	-0.06%	0.06%	0.95%	-0.95%	-0.02%	0.02%	1.02%	-1.02%	0.00%	0.00%	0.92%	-0.92%
	2c N-Y	-0.02%	0.02%	-0.08%	0.08%	2.94%	-2.94%	1.06%	-1.06%	2.86%	-2.86%	3.27%	1.81%	0.00%	0.00%	0.98%	-0.98%
	3a N-Y	-0.30%	0.30%	0.09%	-0.09%	-3.93%	3.93%	-1.36%	1.36%	13.04%	-13.04%	0.00%	0.00%	0.00%	0.00%	-1.10%	1.10%
	3b N-Y	0.29%	-0.29%	0.00%	0.00%	0.89%	-0.89%	0.84%	-0.84%	0.59%	-0.59%	1.93%	-1.93%	0.00%	0.00%	0.81%	-0.81%
	3c N-Y	0.60%	-0.60%	-0.10%	0.10%	-0.58%	0.58%	-0.24%	0.24%	0.03%	-0.03%	1.30%	-1.30%	0.00%	0.00%	-0.24%	0.24%
6	1a N-X	4.75%	-4.75%	-0.87%	0.87%	-4.84%	4.84%	-3.57%	3.57%	1.45%	-1.45%	-4.34%	4.34%	0.00%	0.00%	-2.83%	2.83%
	1b N-X	0.51%	-0.51%	-0.03%	0.03%	0.52%	-0.52%	0.65%	-0.65%	0.76%	-0.76%	2.13%	0.14%	0.00%	0.00%	0.64%	-0.64%
	1c N-X	0.00%	0.00%	-0.07%	0.07%	0.73%	-0.73%	0.27%	-0.27%	-0.39%	0.39%	0.09%	-0.28%	0.00%	0.00%	0.26%	-0.26%
	2a N-X	0.47%	-0.47%	-0.34%	0.34%	-2.87%	2.87%	-1.71%	1.71%	1.86%	-1.86%	-3.33%	3.33%	0.00%	0.00%	-1.40%	1.40%
	2b N-X	1.46%	-1.46%	0.15%	-0.15%	-3.00%	3.00%	1.18%	-1.18%	0.26%	-0.26%	-0.47%	0.47%	0.00%	0.00%	1.13%	-1.13%
	2c N-X	-1.00%	1.00%	0.13%	-0.13%	1.27%	-1.27%	0.75%	-0.75%	-1.47%	1.47%	1.00%	-1.00%	0.00%	0.00%	0.67%	-0.67%
	3a N-X	1.06%	-1.06%	-0.01%	0.01%	-3.10%	3.10%	-2.05%	2.05%	1.18%	-1.18%	-3.27%	3.27%	0.00%	0.00%	-1.66%	1.66%
	3b N-X	0.64%	-0.64%	-0.13%	0.13%	1.65%	-1.65%	0.53%	-0.53%	-0.38%	0.38%	0.46%	-0.46%	0.00%	0.00%	0.52%	-0.52%
	3c N-X	0.35%	-0.35%	0.03%	-0.03%	0.47%	-0.47%	0.42%	-0.42%	-0.84%	0.84%	0.53%	-0.53%	0.00%	0.00%	0.40%	-0.40%

Table 15. Summary of accuracy in detecting targets across the control site survey area defined by data manipulation schemes (*see Tables 12 and 13*) when combining all depth slices of GPR together. Accuracies are also defined by type of shifting applied to grids. Example is given for Resolution Type 1 at a 95% threshold.

		No Shift				X Shift				Y Shift			
	Depth (m)	*Actual 1, Calculated 0*	*Actual 1, Calculated 1*	*Actual 0, Calculated 1*	*Actual 0, Calculated 0*	*Actual 1, Calculated 0*	*Actual 1, Calculated 1*	*Actual 0, Calculated 1*	*Actual 0, Calculated 0*	*Actual 1, Calculated 0*	*Actual 1, Calculated 1*	*Actual 0, Calculated 1*	*Actual 0, Calculated 0*
Addition (Lowest Rank)	0.33	0.00%	100.00%	0.47%	99.53%	1.75%	98.25%	0.33%	99.67%	1.03%	98.97%	0.38%	99.62%
	0.66	0.25%	99.75%	1.38%	98.62%	0.26%	99.74%	1.33%	98.67%	0.26%	99.74%	1.33%	98.67%
	1.00	0.00%	100.00%	0.32%	99.68%	0.29%	99.71%	0.37%	99.63%	0.00%	100.00%	0.34%	99.66%
	Combined	0.49%	99.51%	0.94%	99.06%	1.35%	98.65%	0.76%	99.24%	0.59%	99.41%	0.90%	99.10%
Exponential (Top Rank)	0.33	13.00%	87.00%	13.81%	86.19%	9.91%	90.09%	10.10%	89.90%	10.08%	89.92%	10.84%	89.16%
	0.66	0.00%	100.00%	0.68%	99.32%	0.52%	99.48%	1.27%	98.73%	0.52%	99.48%	1.27%	98.73%
	1.00	1.02%	98.98%	1.69%	98.31%	2.04%	97.96%	2.01%	97.99%	0.26%	99.74%	1.31%	98.69%
	Combined	1.02%	98.98%	0.91%	99.09%	0.78%	99.22%	0.78%	99.22%	0.59%	99.41%	0.90%	99.10%

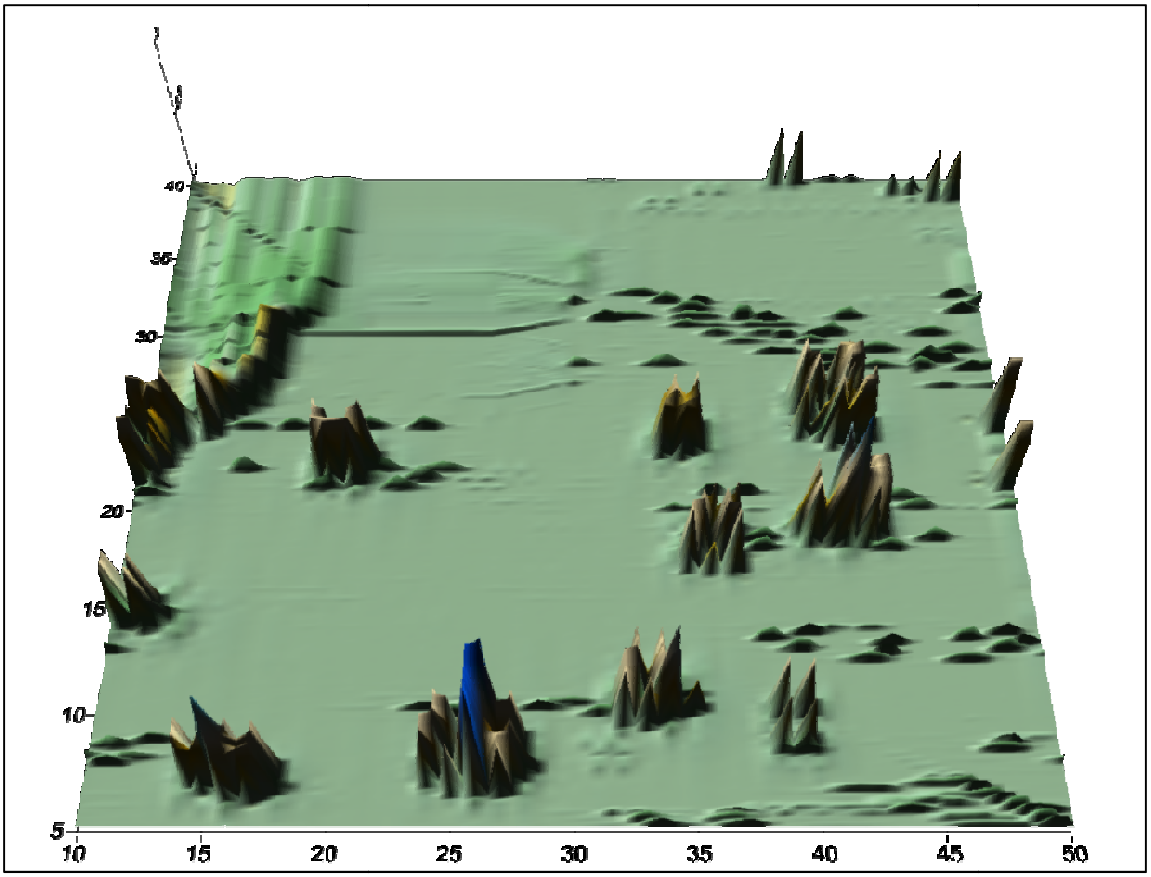


Figure 31 A. Original (raw) magnetic gradiometry data collected over the study site at the University of Tennessee Geophysical Research Station. Data were collected over a portion of the greater 50 meter by 40 meter area. Data in the top left quartile is not valid. Peaks are considered to be probable locations of buried targets.

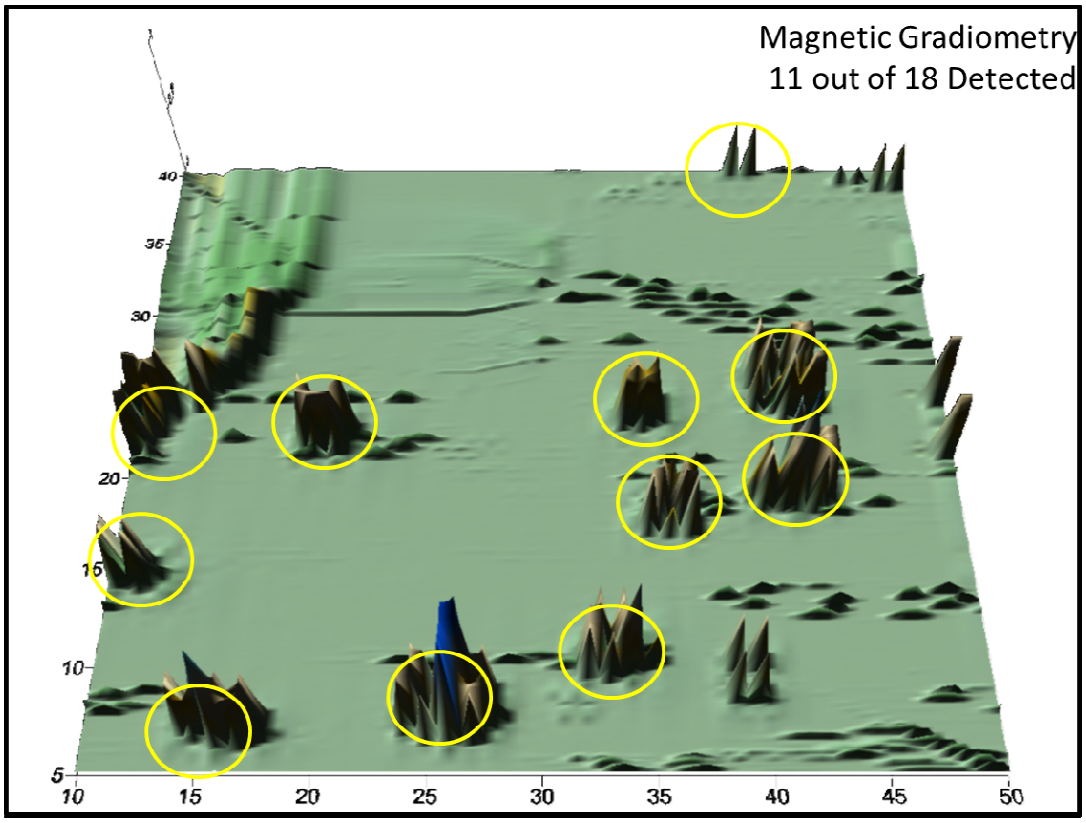


Figure 31 B. Interpretation of original (raw) magnetic gradiometry data collected over the study site at the University of Tennessee Geophysical Research Station. Yellow circles indicate locations of buried targets that also show a peak in the data. This interpretation allows for the calculation of Type I and Type II error presence.

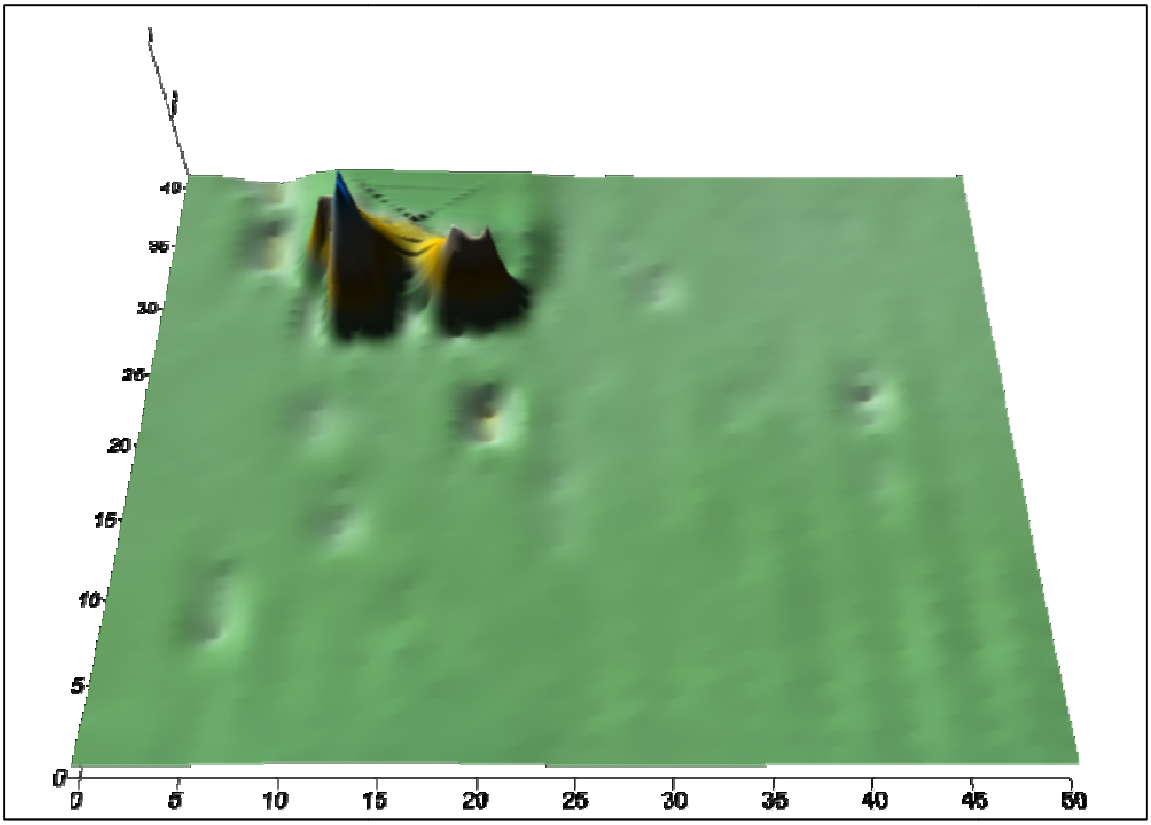


Figure 32 A. Original (raw) EM-31 ground conductivity data collected over the study site at the University of Tennessee Geophysical Research Station. Data were collected over the entire 50 meter by 40 meter area. Depressions in the surface are considered to be probable locations of buried targets.

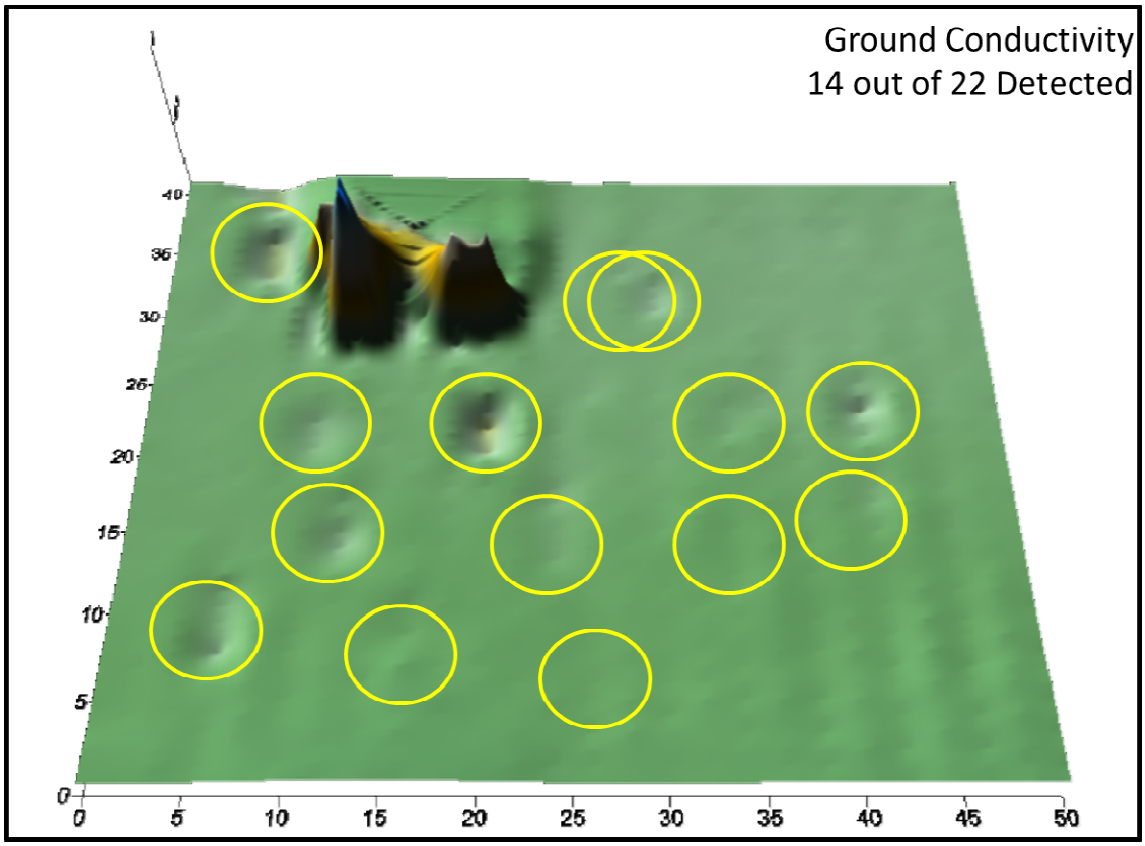


Figure 32 B. Interpretation of original (raw) EM-31 ground conductivity data collected over the study site at the University of Tennessee Geophysical Research Station. Yellow circles indicate locations of buried targets that also show a peak in the data. This interpretation allows for the calculation of Type I and Type II error presence.

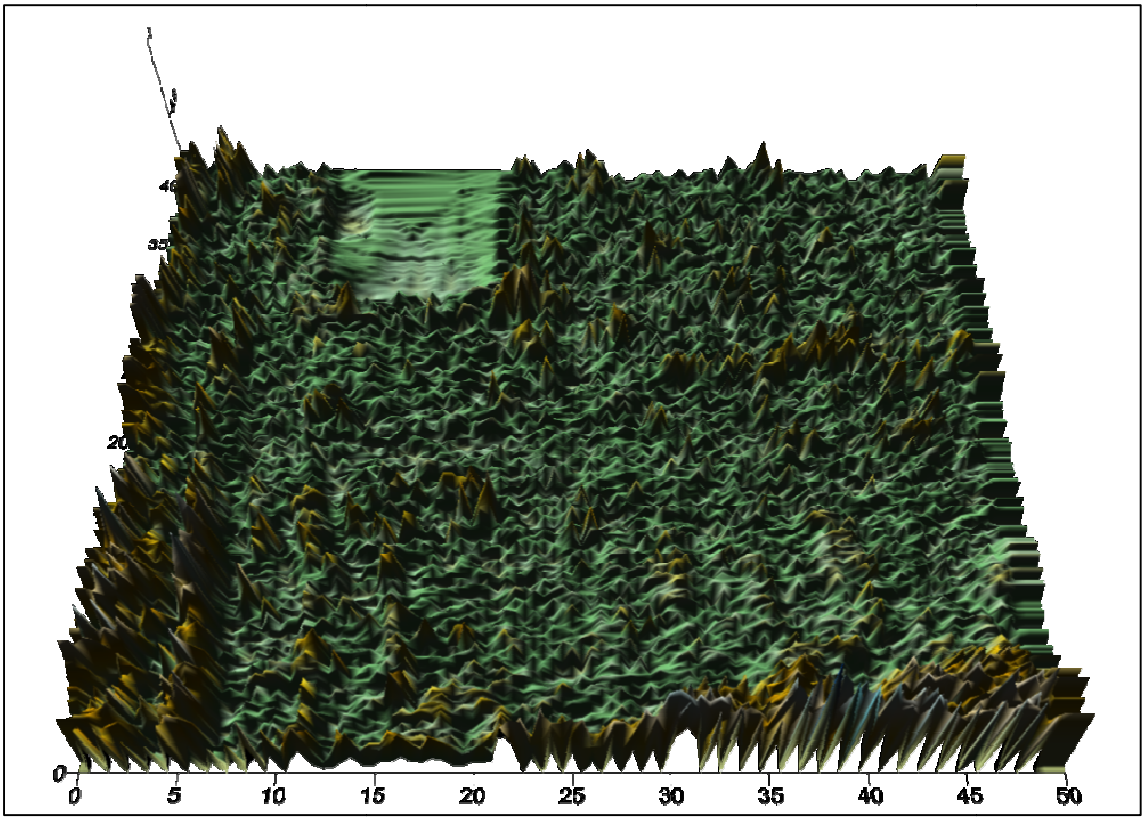


Figure 33 A. Original (raw) ground penetrating radar data collected over the study site at the University of Tennessee Geophysical Research Station. Data were collected over the entire 50 meter by 40 meter area. Peaks are considered to be probable locations of buried targets.

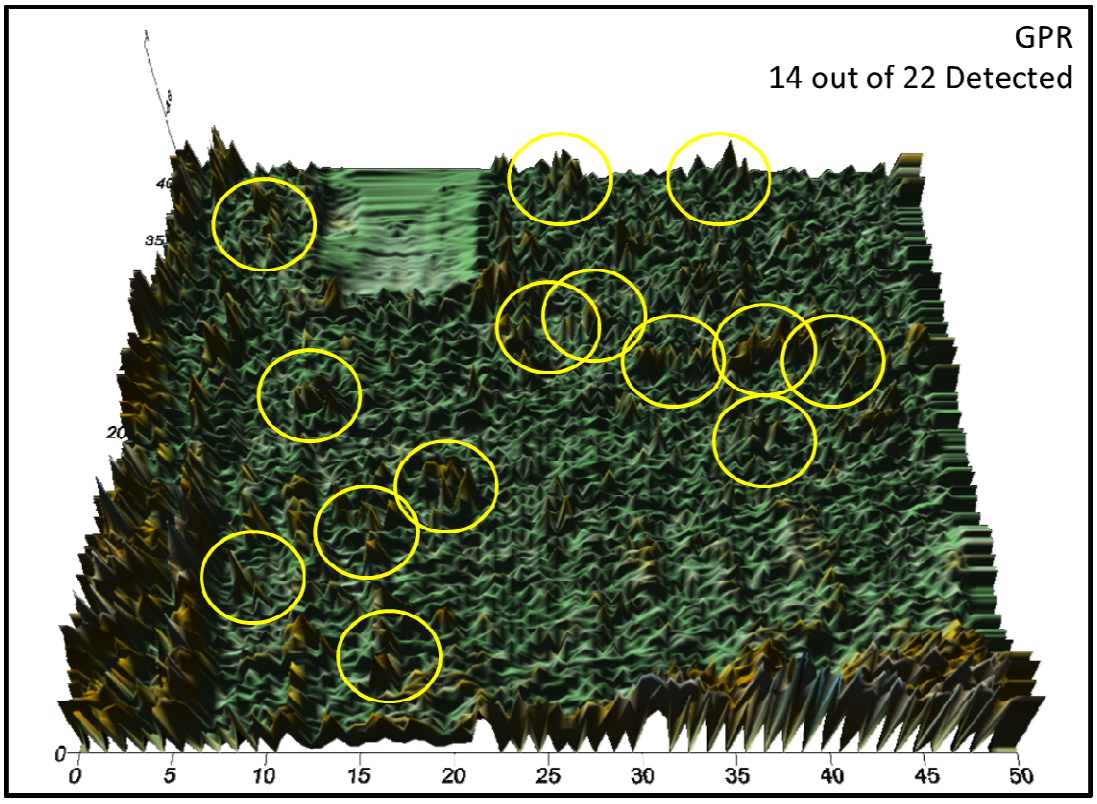


Figure 33 B. Interpretation of original (raw) ground penetrating radar data collected over the study site at the University of Tennessee Geophysical Research Station. Yellow circles indicate locations of buried targets that also show a peak in the data. This interpretation allows for the calculation of Type I and Type II error presence.

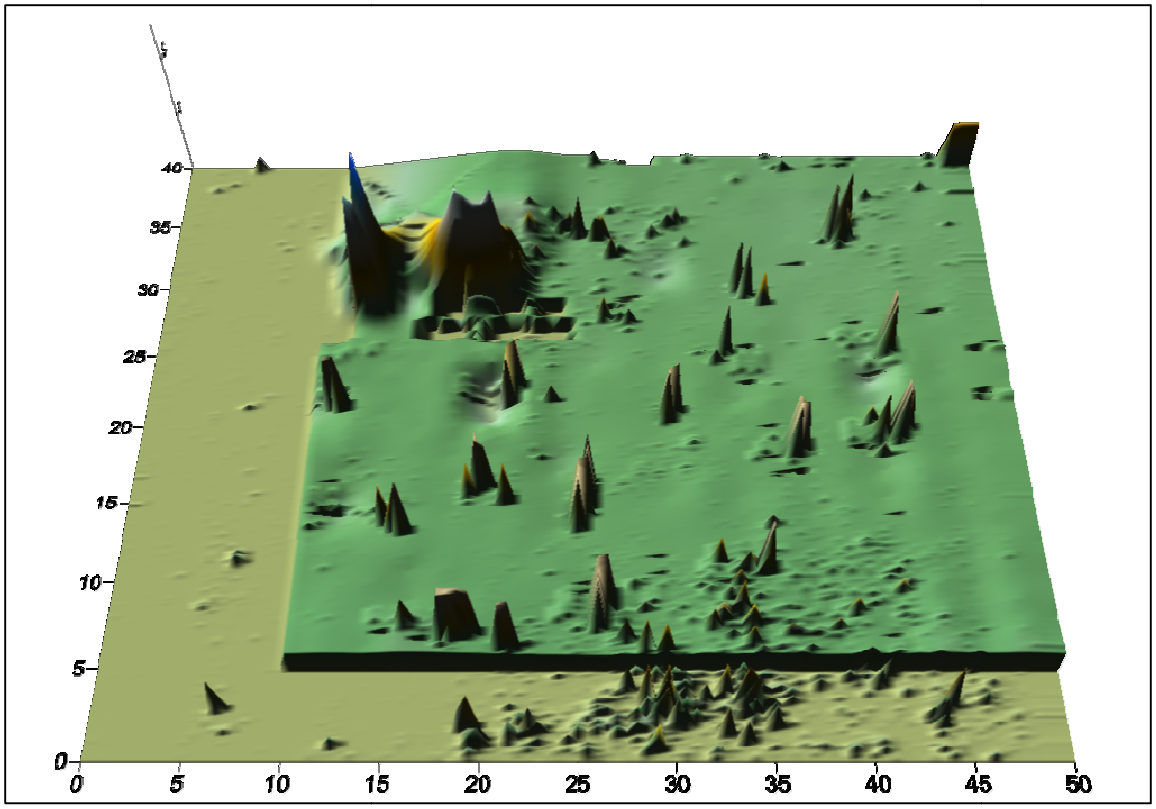


Figure 34 A. Representative example of all datasets integrated together. Data have been manipulated by raising each normalized value to the optimal exponent. This example is at a 1.00 meter depth slice, Resolution Type 1, and has not been shifted. Peaks are considered to be probable locations of buried targets.

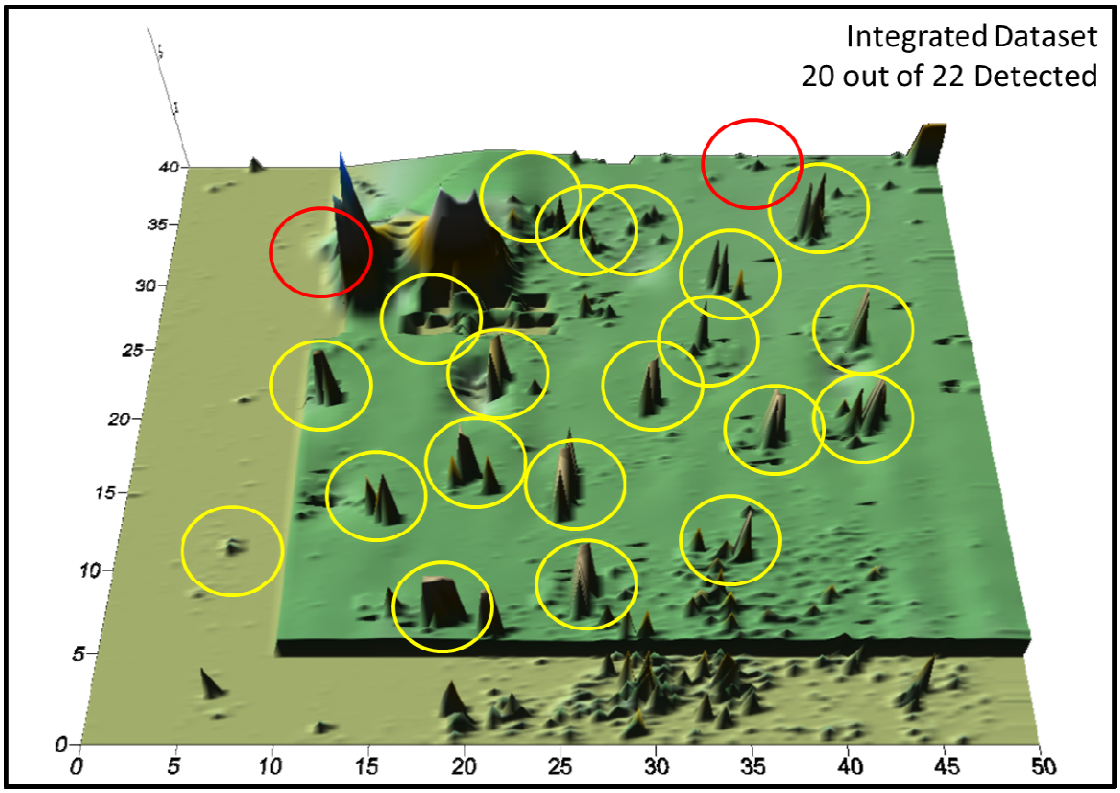


Figure 34 B. Interpretation the integrated dataset from all three techniques collected over the study site at the University of Tennessee Geophysical Research Station. Yellow circles indicate locations of buried targets that also show a peak in the data. Red circles indicate the locations of targets that were not detected by any of the techniques. This interpretation allows for the calculation of Type I and Type II error presence.

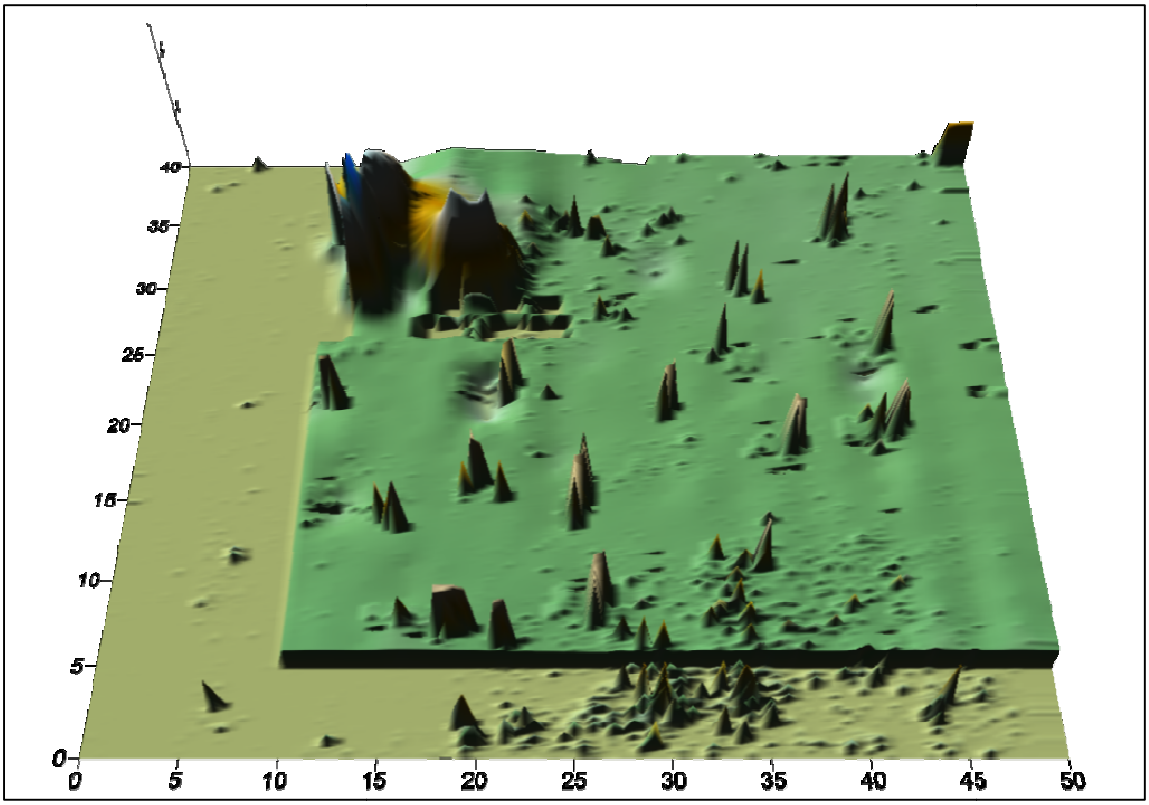


Figure 35. Representative example of all datasets integrated together. Data have been manipulated by raising each normalized value to the optimal exponent. This example is at a 1.00 meter depth slice, Resolution Type 1, and has shifted along the X axis (North/South) by 2.5 meters. Peaks are considered to be probable locations of buried targets.

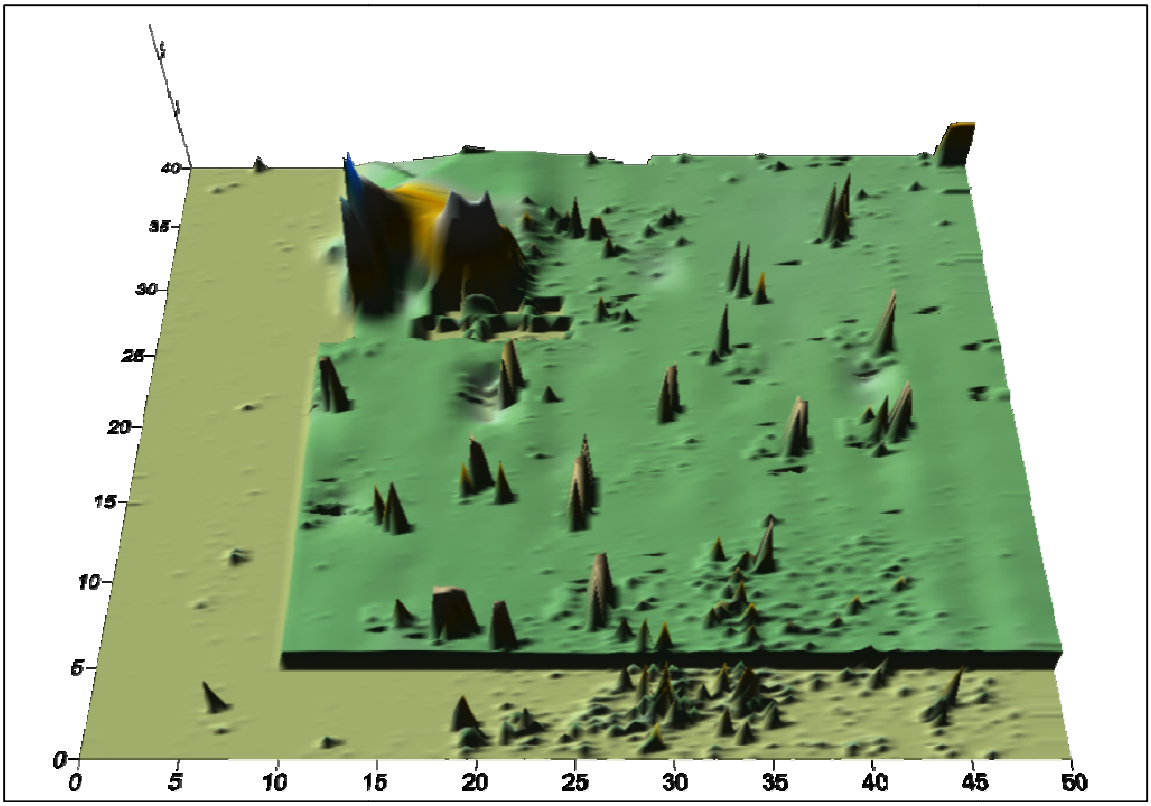


Figure 36. Representative example of all datasets integrated together. Data have been manipulated by raising each normalized value to the optimal exponent. This example is at a 1.00 meter depth slice, Resolution Type 1, and has shifted along the Y axis (East/West) by 2.5 meters. Peaks are considered to be probable locations of buried targets.

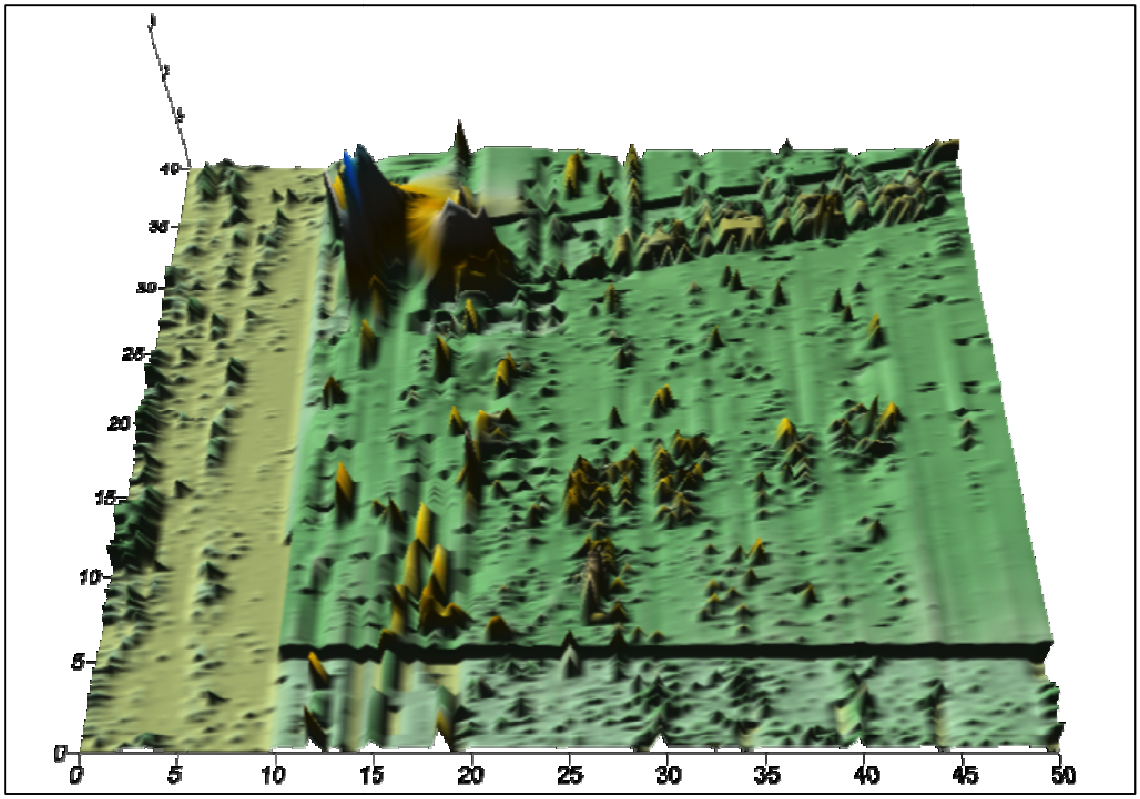


Figure 37. Representative example of all datasets integrated together. Data have been manipulated by raising each normalized value to the optimal exponent. This example combines all depth slices (0.33m, 0.66m, and 1.00m), Resolution Type 1, and has shifted along the X axis (North/South) by 2.5 meters. Peaks are considered to be probable locations of buried targets.

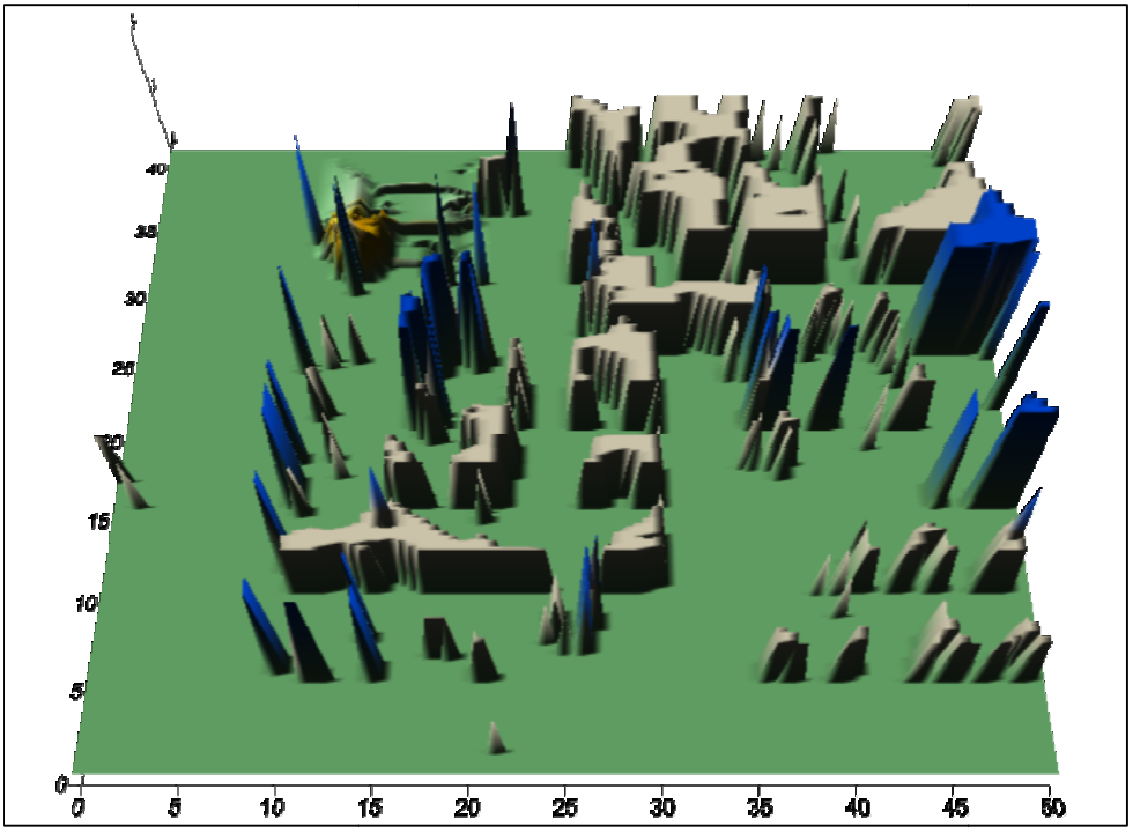


Figure 38. Representative example of all datasets integrated together and plotted as a binary response of presence (1) or absence (0) of target for each data point in the grid. Data have been manipulated by raising each normalized value to the optimal exponent. This example combines all depth slices (0.33m, 0.66m, and 1.00m), Resolution Type 1, and has shifted along the Y axis (East/West) by 2.5 meters. Peaks are considered to be probable locations of buried targets. Blue peaks are targets detected by two or three techniques. Block shapes give a general sense of shape, size, and orientation of the buried objects.

4.4 Discussion

We begin by discussing how the issues of data sampling heterogeneities within traditional data integration methods are resolved within the proposed methodology. Statistical analysis of which resolution type, depth of investigation, and shifting direction (or no direction) of the data points has the most significant impact on accurately locating the buried targets are also discussed. Additionally, a brief commentary on why certain targets might not have been detected within the study site is given, along with an assessment of the general occurrence of Type I (false positives) and Type II (false negatives) errors are present in the data.

4.4.1 Data Resolution

It is crucial to set up appropriate parameters in both data acquisition and processing to ensure the highest degree of certainty in the integrated data models. This research has addressed the challenge of data sampling heterogeneities inherent in geophysical databases (as mentioned in the introduction section) by (1) setting up consistent parameters within the geophysical techniques utilized in this study; (2) developing a comprehensive model for integration of datasets; and (3) utilizing appropriate statistical measures (i.e. kriging) to transform and interpolate data points within grids when constructing certain datasets prior to integration. As shown in **Table 10**, each of the three geophysical techniques utilized in this research collected data at very different sample spacing (i.e. resolutions). The very first step in the data integration process is to ensure that all data points being merged have exact XY locations, otherwise there will be

interpolations of interpolated data points, which increase the amount of error in the final dataset.

Summaries for the percentage accuracy in detecting targets across the control site, as defined by threshold and resolution type for the optimal exponent scheme are shown in **Tables 12;** the traditional addition scheme is in **Table 13.** The variable of shift direction when dividing the survey area into 5m by 5m grids does not have a statistical significance on the resolution type. This means that the grids can be shifted in either the North/South or East/West direction with little effect on the accuracy of target detection. However, when considering there is a difference [in some cases as much as 13.36%] in accuracy between the grids that are shifted and the grids that are not shifted, it appears as though there is some type of edge effect on the data, so it is advantageous [for the purposes of this case study] to shift the grids in either direction. This is true for all resolution types and both data manipulation schemes being highlighted.

4.4.2 Statistical Analysis

Highlighted in **Tables 12 to 13** is how the resolution types and shift directions are affected by the threshold levels applied to the data. Part one of this article concluded that the 95% threshold would be most appropriate to be applied to the authentic data. This was not found to be consistent with the results of the statistical analysis performed during part two of this research. For authentic data, it is consistently shown in the statistical analysis that an 85% threshold is most appropriate (see **Tables 12, 13 and 14**); However, there is a risk that should the 85% threshold be applied, the shape and size of the target

can become distorted, which may inadvertently cause errors in discriminating certain characteristics of the buried target. Conversely, when looking at each survey's *geoData EMG Accuracy Summary*, it is found that there is no statistical difference in the percentage accuracy across all 5m by 5m grids (n=76 for the not shifted grids; n=86 for the shifted grids) of the survey area, suggesting that a 95% threshold is still appropriate.

The summary of the difference between the two represented data manipulation schemes is shown in **Table 14**. There is a statistical difference between the three different depths that were extracted from the GPR data and their percentage accuracies. Across all resolution types, depth "a" (0.33m) consistently has the highest amount of errors in detecting the location of targets. Additionally, when evaluating the results displayed in **Table 14**, the difference between the highlighted data manipulation schemes is an order of magnitude higher than for the other two depths (b- 0.66m and c- 1.00m). This is somewhat expected, as the survey site is a part of The University of Tennessee's Agricultural Extension Center and the land is repeatedly used for farming experiments where pieces of equipment or other miscellaneous objects could have been buried at a shallow depth and detected with the geophysical instruments.

These discrepancies between resolution shift type and depth of investigation, when comparing the two highlighted data manipulation schemes, are minimized significantly once all depth slices are combined into one dataset. **Table 15** gives a summary of how the combined datasets improve the uncertainty level in discrete anomaly detection. When comparing the combined accuracy percentages, differences between both data manipulation schemes are minimal, with a maximum difference at 0.57%. Based on this

analysis, it is suggested that any data manipulation scheme could be applied to the original, normalized data values with the stipulation that all depth slices be integrated together. Should only one depth be included in the site investigation, all variables previously discussed need to be evaluated prior to interpretation of where targets are located.

4.4.3 Data Visualization

All presented graphical depictions of the data are representative examples of the results of this research's proposed methodology for quantitative data integration. **Figures 31 to 33** display the original, raw data from the survey area with data values normalized; no interpolation of data points has been conducted. For each subsequent figure, resolution type 1 at a 95% threshold is shown and comparisons are made visually between the two highlighted data manipulation schemes. These visual representations are provided to give supporting evidence to the underlying premise of this research that a quantitative approach must be taken to determine target location, not a qualitative one as is traditionally done (i.e. once the data is mapped out, anomalies are identified visually which may lead to a large number of errors).

Data is mapped in the program Surfer, with peaks generally indicating the location of buried objects, as seen in **Figures 34 to 37**. In each case, there are additional peaks where, according to **Figure 19**, there should not be a target. This is particularly true for the 0.33m depth slice (**Figure 34**), due to reasons previously mentioned in section 4.4.2. Once at the 1.00m depth slice (**Figure 36**), those additional peaks are nearly all removed.

Any remnant anomalies that are not expected in the data may be due to soil saturation [assumed from the large pond of water at the edge of the survey area] in the Northeast quadrant of the survey area and a metal fenced area in the Southwest quadrant interfering with the signal responses. The integration of all depth slices with all geophysical techniques, while quantitatively identifies the location of targets with a high level of confidence, does not prove to be as clear to decipher in a graphical context, as shown in **Figure 37**. However, when this type of graphical representation of the data is interpreted side by side with a plot of the data points transformed into binary responses (**Figure 38**), this task becomes more clear cut. A major advantage of this plot diagram is that the data points become pixilated and the resulting block shapes give a general sense of target shape, size, and orientation. Additionally, as in the case of **Figure 38**, it is easy to identify which targets were detected with multiple geophysical techniques.

4.4.4 Error Analysis

There was one target that, while expected to be detected in the survey site, was not detected by any of the geophysical techniques utilized in this research (*Target Map ID 14, **Figure 19***). This target is composed of Styrofoam, and while quite large at 9'x2'x4" either (1) did not have any differences in the physical properties each geophysical technique is dependent upon detecting, or (2) the object through time has biodegraded to an insignificant amount and the surrounding soil had filled in the voided space, thus rendering it as though it had never existed. It is not expected to have been missed due to survey design, despite the width of the object being less than the sample spacing of 0.5m

because it is dipping at N45E, which would cause it to be detected with the 10cm sample interval along each survey line.

As expected with every geophysical investigation, there is a slight occurrence of Type I (false positives) and Type II (false negatives) errors present in the data from the survey area. When interpreting **Figure 38**, there are a number of false positives, which have been identified as those shapes on the plot that come to a point because those high data values are only contained within a couple of the 10cm by 10cm cell of the entire survey area; all buried objects extend beyond one or two of those cells. These false positives are being caused by small metal objects of some kind. Blocked shapes are indicative of the true targets. There was only one false negative that can be accurately calculated with this study site and has been previously discussed.

4.5 Conclusions

The methodology employed in this part of the research illustrates how crucial it is for there to be consistent and appropriate parameters set in place for both acquisition and processing of geophysical data, especially in cases where more than one data set will be integrated together; this ensures that the highest degree of certainty is obtained. The hypothesis, "Integration of two or more geophysical techniques will result in an improved subsurface image and reduce uncertainty in discrete anomaly detection," was proven correct through the careful implementation and expansion of the methodology developed in Part One of this research.

Geophysical data (GPR, ground conductivity, and magnetic gradiometry) were strongly correlated to the known discrete targets of the control site, with a number of

variables (resolution of data, shift direction of data, depth of investigation, threshold applied to data, and data manipulation scheme) assessed for statistical significance ($p<0.05$) toward being able to accurately identify the location of the targets. Statistical methods employed during this stage of the research were logistical regression, proc GLM, proc frequency, univariate, and kriging to prepare the data for use in the testing of the developed data integration methodologies and aid in the quantitative assessment of target identification.

The issue of data sampling heterogeneities inherent in geophysical databases has been resolved and it has been shown that a threshold of 95% [as suggested in part one] is an appropriate threshold to apply to all data points along with a series of different data manipulation schemes. Additionally, the variables that proved to be most significant in the accurate detection of discrete anomalies was depth of investigation, which is consistent with the initial findings from part one. Other variables evaluated, while having the potential to alter interpretation and identification of target location, were not found to be statistically significant; however, they should not be discounted when designing how to systematically interpret datasets.

Discrepancies between resolution shift type and depth of investigation, when comparing the data manipulation scheme. It should be emphasized that while the integration of all depth slices with all geophysical techniques quantitatively identifies the location of targets with a high level of confidence, it does not prove to be as clear to decipher the location of discrete targets in a graphical context. The end result of this research, discussed within this article, provides supporting evidence that the integration

of two or more geophysical techniques will result in an improved subsurface image and reduce uncertainty in discrete anomaly detection.

5. CONCLUSIONS

5.1 Summary

This dissertation research presents a novel and systematically tested, purely quantitative methodology for the integration of two or more data sets collected using near-surface geophysical techniques. It has met the increasing need for this type of methodology in the fields of archaeological prospecting, environmental sciences, and forensics with a detailed and refined approach. The underlying concepts of this project are that single geophysical methods are typically not able to detect all discrete target types, and that utilizing multiple techniques - and the integration of multiple technique data - can produce significant improvements in data quality and target detection. Both hypotheses tested throughout this research were proven to be correct:

- Certain targets, given multiple variables and parameters, will be detected with a greater degree of certainty than others when a specified combination of processing and merging of data is implemented.
- Integration of two or more geophysical techniques will result in an improved subsurface image and reduce uncertainty in discrete anomaly detection.

In essence, the proposed methodology has led to an understanding of the importance in the development of a quantitative data integration methodology for improving subsurface imaging and reducing uncertainty in discrete anomaly detection.

5.2 Objectives Met

The primary objective of this research was to improve success rates as defined through data quality and visualization techniques within geophysical surveys for discrete

anomaly detection (e.g. locating clandestine underground tunnels, locating buried objects, mapping historical features). Each hypothesis tested incorporated different goals to meet this objective:

- Develop a comprehensive model for creating typical signal responses for various materials of objects found at The University of Tennessee's Geophysical Research Station
- Create synthetic data sets for each geophysical techniques utilized in study
- Determine how resolution of data and data sampling heterogeneities affects integration of data sets and the resulting ability to discriminate targets with a higher degree of certainty
- Determine which variables involved with a geophysical survey are most significant in the discrimination of targets (e.g., target depth, target size, composition of target)
- Complete a statistical analysis to evaluate the effectiveness of the methodologies to quantitatively merge different geophysical data sets (e.g., addition, multiplication, exponential)
- Correlate geophysical data with known discrete "targets" by utilizing an integration of multiple geophysical techniques

Each of these goals was met to satisfy the primary objective of this research. The methodology employed illustrates how crucial it is for there to be consistent and appropriate parameters set in place for both acquisition and processing of geophysical data, especially in cases where more than one data set will be integrated together; this

ensures that the highest degree of certainty is obtained. The end result of this research provides supporting evidence that the integration of two or more geophysical techniques will result in an improved subsurface image and reduce uncertainty in discrete anomaly detection.

5.3 Overall Impact

While this research was originally designed to enhance archaeological geophysical surveys by incorporating improved multi-tool geophysics, the resulting methodology for quantitatively merging different types of geophysical data together shows potential for being utilized in many different areas of interest. These may include, but are not limited to, environmental monitoring (i.e. contaminant transport, groundwater studies), UXO detection, clandestine underground tunnel detection (national security), locating stratigraphic features of the subsurface (e.g. geologic formation), and mining/exploration (e.g. minerals and natural resources location). The development of methods for quantitatively merging different types of geophysical data will allow for the enhancement of structures and features in the data through S/N enhancement of features detectible through more than one technique. By improving visualization methods of the data, the interpretations may be seen more clearly and will be more convincing to the scientist conducting the investigation, helping to quickly and accurately meet the objectives of the individual project. Of particular interest for this dissertation, the data integration methodologies will give archaeologists additional tools in their planning and choosing of locations and methods of excavations, saving project managers valuable time, money,

and/or other resources. The methodologies developed through this research satisfy the ever growing need within the private sector and scientific community for a powerful time- and cost-effective approach for integrative analysis of multiple geophysical data sets.

5.4 Future Work

There are a number of additional investigative components that were not able to be conducted within the scope of this project, of which could be expanded for future projects.

1) Analysis of the errors associated within each geophysical technique during data collection. All methods involve some element of natural human error because each survey line is rarely collected in a completely straight line. Additionally, in cases of the magnetic gradiometry and EM-31 ground conductivity data, the sample spacing of the data points along each survey line are dependent upon the human carrying the instrument walking at the exact same pace throughout the entire survey; this is nearly impossible to do.
2) Conducting additional statistical analysis like Principal Component Analysis (PCA) and Cluster Analysis. PCA can also account for determining amounts of variance between data sets (Snyder et al 2001); when data is standardized, each variable within the statistical method contributes a variance of unity (Davis 2002). Another advantage of PCA is that as correlations of variables are identified and removed from the equation, noise from the combined data sets can become

isolated, which may help in readjusting pre-processing procedures to further enhance the signal-to-noise ratio. Cluster analysis describes the process whereby multivariate data is analyzed for the presences of natural groups or clusters that possess certain properties. For the purposes of this research, the presence of clusters will give an indication that multiple geophysical data sets have identified a target or subsurface feature of interest with a high degree of certainty.

3) Apply the developed methodology to additional survey sites (i.e. case studies) where there are unknown discrete targets. An example of this type of area is the Cherokee Farm research site, located at a separate University of Tennessee Agricultural Extension Center than what was used to develop the Geophysical Research Station. Geologic setting is similar, making this site an ideal transition area for further refining the methodology proposed in this research.

REFERENCES

Allen, D., Schuurman, N., Deshpande, A. and Scibek, J. Data integration and standardization in cross-border hydrogeological studies: a novel approach to hydrostratigraphic model development. Environmental Geology, February 2008, v. 53, no. 7, p. 1441-1453.

Ambrose, H. 2005. Quantitative Integration and Three-Dimensional Visualization of Multi-tool Archeological Geophysical Surveys. State University of New York at Buffalo, MS Thesis, 217 p.

Ansmann, A., et al. (2010), The 16 April 2010 major volcanic ash plume over central Europe: EARLINET lidar and AERONET photometer observations at Leipzig and Munich, Germany, *Geophys. Res. Lett.*, *37*, L13810

Baker, G.S., Jordan, T.E., and *Talley, J., 2007, An introduction to ground penetrating radar (GPR), *in* Baker, G.S., and Jol, H.M., eds., Stratigraphic Analyses Using GPR: *Geological Society of America Special Paper 432*, p. 1–18

Baradello, Luca, Carcione, Jose, and Gei, Davide. Fast monostatic GPR modeling. Geophysics. Volume 69. No. 2. 2004. Pages 466-471.

Billings, S., Pasion, L., and Oldenburg, D, 2003. Discrimination and classification of UXO using magnetometry: Inversion and error analysis using robust statistics. SAGEEP Proceedings.

Bishr, Y. 1998. *Overcoming the semantic and other barriers to GIS interoperability*. International Journal of Geographic Information Science. 12(4):299–314

Blakely, RJ. 1995. Potential Theory in Gravity and Magnetic Applications. Cambridge University Press.

Böniger, U. and Tronicke, J.,On The Potential of Kinematic GPR Surveying Using a Self-Tracking Total Station: Evaluating System Crosstalk and Latency. IEEE Transactions on Geosciences and Remote Sensing, 2010, v. 48, no. 10, p. 3792-3798

Casey, KF and Baertlein, BA. 1999. An overview of electromagnetic methods in subsurface detection: *in* Detection and identification of visually obscured targets, Taylor and Francis, Oxford, 9-46

Coleou, Thierry, Poupon, Manuel, Azbel, Kostia.*Unsupervised seismic facies classification: A review and comparison of techniques and implementation*
The Leading Edge 2003 22: 942-953

Colombo, D. and De Stefano, M., Geophysical modeling via simultaneous joint inversion of seismic, gravity, and electromagnetic data: Application to prestack depth imaging. The Leading Edge, March 2007, p. 326-331

Cormack, R. M. 1971. A review of classification. J. Roy. Statist. Soc. Ser. A, 134: 321–353.

Gaines, D.P., Baker, G.S., Hubbard, S.S., Watson, D, Brooks, S., and Jardine, P. 2010. Detecting perched water bodies using near-surface seismic time-lapse travel-time tomography in Advances: *in* Miller, R.D., Bradford J,H, Holliger, K., eds., *Advances in Near Surface Seismology and Ground-Penetrating Radar*, Society of Exploration Geophysicists, Tulsa, OK, ISBN 9781560802242.

Dannemiller, N. and Li, Y., A new method for determination of magnetization direction. Geophysics. Volume 71. No. 6. 2006. Pages L69-L73

Davis, J.C., 2002, *Statistics and Data Analysis in Geology,* 3rd edn. John Wiley, New York

Debeglia, N. and Corpel, J., 1997. Automatic 3D interpretation of potential field data using analytical signal derivatives. Geophysics 62. 87-96

Ferré, T., von Glinski, G., and Ferré, L. *Monitoring the Maximum Depth of Drainage in Response to Pumping Using Borehole Ground Penetrating Radar* Vadose Zone J., Nov 2003; 2: 511 - 518.

Galicia, Jose. *The offshore Golden Lane: New outline of opportunities from the integration of geologic and geophysical data.* The Leading Edge, July 2001; 20: 763 - 782.

Giannoploulos, A. Modelling ground penetrating radar by GprMax. Construction and Building Materials. Volume 19. 2005. Pages 755-762

Hastie, T.J. and Tibshirani, R.J. (1990), Generalized Additive Models, New York: Chapman and Hall.

Henderson, Jonathan, Purves, Stephen J., Fisher, Gaynor, Leppard, Chris *Delineation of geological elements from RGB color blending of seismic attribute volumes.* The Leading Edge 2008 27: 342-350

Herron, Donald A. *Problems with too much data*. The Leading Edge 2001 20: 1124-1126

Hill, Ian. *Resolution and quality assurance in near-surface geophysical surveys: A multimethod field trial and results.* The Leading Edge 2008 27: 1516-1524

Honarmand, M., Ranjbar, H., Moezifar, Z. *Integration and Analysis of Airborne Geophysics and Remote Sensing Data of Sar Cheshmeh Area, Using Directed Principal Component Analysis* Exploration and Mining Geology 2002 11: 43-48

Hornby, Brian and Herron, Don. *Introduction to this special section: Borehole geophysics/VSP* The Leading Edge, Jun 2007; 26: 731.

Kim, W. and Seo, J. 1991. *Classifying schematic and data heterogeneities in multidatabase systems.* IEEE 24(12):12–18

Kirkham, G., Falck, H., Siddorn, J. *An Integrated 3D GIS Model of the Yellowknife Camp: A Tool for Efficient Development* Exploration and Mining Geology 2003 12: 49-59

Kreisberg, V., Bogdanov, A., Solomatin, A., Kozlov, E. *Geophysical data processing on parallel computers in our experience*. Geophysics February 1992. Pg 45

Kvamme, K.L. 2007. Chapter 14: Integrating Multiple Geophysical Datasets. In: J. Wiseman and F. El-Baz (eds.), Remote Sensing in Archaeology, Springer, New York, p. 345-374.

Lado, L. Rodriguez, Polya, D. A., Hegan, A. *A logistic regression method for mapping the As hazard risk in shallow, reducing groundwaters in Cambodia*. Mineral Mag 2008 72: 437-440

McNeill, J.D., Electromagnetic Terrain Conductivity Measurements at Low Induction Numbers, Geonics Limited Technical Note TN-6, October 1980.

Metternicht, G.I., Zinck, J.A., Remote sensing of soil salinity: potentials and constraints, Remote Sensing of Environment, Volume 85, Issue 1, 25 April 2003, Pages 1-20

Milici, Robert and Smith, James. Stratigraphy of the Chickamauga Supergroup in its Type Area. State of Tennessee Department of Conservation: Division of Geology. 1969.

Mishchenko, Michael I., and Coauthors, 2007: Accurate Monitoring of Terrestrial Aerosols and Total Solar Irradiance: Introducing the Glory Mission. *Bull. Amer. Meteor. Soc.*, **88**, 677–691.

Morris, Bill, Pozza, Matt, Boyce, Joe, Leblanc, George *Enhancement of magnetic data by logarithmic transformation.* The Leading Edge 2001 20: 882-885

Pasion, L., Billings, S., Kingdon, K,, Oldenburg, D., Lhomme, N., and Jacobson, J. Cooperative Inversion of Time Domain Electromagnetic and Magnetometer Data for the Discrimination of Unexploded Ordnance. JEEP, September 2008, Volume 13, Issue 3, pages 193-210

Pettersson, Jared and Nobes, David. Environmental geophysics at Scott Base: ground penetrating radar and electromagnetic induction as tools for mapping contaminated ground at Antarctic research bases. Cold Regions Science and Technology 37 (2003) 187– 195

Powers, Michael. Modeling frequency-dependent GPR. The Leading Edge. November 1997. Pages 1657-1662.

Rahman, A., Frank, T., Tsai, C., White, C., Carlson, D. and Willson, C. *Geophysical data integration, stochastic simulation and significance analysis of groundwater responses using ANOVA in the Chicot Aquifer system, Louisiana, USA*. Hydrogeology Journal. Vol. 16, no. 4, pp. 749-764. Jun 2008.

Salem, A., Ravat, D., Gamey, T., and Ushijima, K., Analytical signal approach and its applicability in environmental magnetic investigations. Journal of Applied Geophysics. Volume 49. 2002. Pages 231-244

Schowengerdt, RA, 1997, *Remote Sensing: Models and Methods for Image Processing*. Academic Press, San Diego.

Sheffield, Tatum M., Meyer, Doug, Lees, Jack, Payne, Barton, Harvey, Elizabeth Lorenzetti, Zeitlin, Michael J., Kahle, Gail. *Geovolume visualization interpretation: A lexicon of basic techniques.* The Leading Edge 2000 19: 518-522

Singh, Yeshpal. *Lithofacies detection through simultaneous inversion and principal component attributes*. The Leading Edge 2007 26: 1568-1575

Snyder, Scott W., Mauger, Lucy L., Ames, Dorothea *Benthic Foraminifera and Paleoecology of the Pliocene Yorktown and Chowan River Formations, Lee Creek Mine, North Carolina, USA* Journal of Foraminiferal Research 2001 31: 244-274

Song, L., Oldenburg, D., Pasion, L., and Billings, S., Adaptive Focusing for Source Localization in EMI Sensing of Metallic Objects: A Preliminary Assessment. JEEP. September 2008, Volume 13, Issue 3, pages 131-145

Stanley, Clifford R. Numerical transformation of geochemical data: 1. Maximizing geochemical contrast to facilitate information extraction and improve data presentation *Geochemistry: Exploration, Environment, Analysis*; February 2006; v. 6; no. 1; p. 69-78

Stock, K. and Pullar, D. 1999. *Identifying semantically similar elements in heterogeneous spatial databases using predicate logic.* Interoperating geographic information systems, second international conference, INTEROP'99 Zurich, Switzerland, 1999. Springer, Heidelberg, pp 231–252

van der Kruk, J., Vereechken, H., and Jacob, R. Identifying dispersive GPR signals and inverting for surface wave-guide properties. The Leading Edge, October 2009, v. 28; no. 10; p. 1234-1239

Witten, Alan. Geophysica: A MATLAB-based software for the simulation, display, and processing of near-surface geophysical data. Computers and Geosciences. Volume 28. 2002. Pages 751-762

APPENDIX

A Framework for Building Quantitative Skills and Field Experience in Near-Surface Geophysics by Incorporating Multiple Techniques and Instructional Methods

Megan E. Carr and Gregory S. Baker
Earth and Planetary Sciences Department, The University of Tennessee, Knoxville, TN

"The students will probably never realize just how much self-discipline was required of the instructor to let the class work freely with the expensive equipment…"
~Nicholas Tibbs, 1994

1. Abstract

The need for geoscience curricula that emphasize both quantitative skills and knowledge of how to carry out field work effectively has become increasingly apparent in today's job market. This paper presents the framework of The University of Tennessee's contribution towards meeting these needs. The Tennessee Intensive Near-surface Geophysics Study (TINGS) program is a three week Monday-thru-Friday (9am-5pm) course that introduces multiple near-surface geophysical techniques and allows the student to become familiar with the theory behind each technique, gain experience operating the geophysical equipment, and to train in the software packages specific to each technique by processing their own data. Emphasis is placed on proper survey design, maintaining proper quality control, and working in a team environment to implement the plans successfully.

Key Words: Geophysics, Quantitative skill building, University curriculum, Field-based studies.

2. Introduction

Introductory geology courses often give few glimpses into the kind of technical and quantitative problems routinely tackled by Earth scientists working in industry or academia (Shea, 1990). Once a student has begun coursework at the junior and senior level, courses may include some field time. However, the course experiences typically do not adequately prepare them for employment after graduation, and rarely will these courses expand upon the quantitative problem solving skills acquired earlier in their undergraduate career. Many students do not feel confident about their ability to use mathematics to solve problems or to make well-informed decisions (Macdonald et.al., 2000), with many students feeling uncomfortable and unpracticed at "reading" equations (Kruse, 1995).

Dentith and Trench (1992) were among the first to address the problem of teaching mathematically diverse geophysics classes, emphasizing semi-quantitative data interpretation at the start of each academic term; exams were primarily essay based and did not involve higher math skills to attempt leveling the playing field for the students with a less robust mathematical background. Kenyon (2000) found this approach to be unacceptable, as all students should be able to handle mathematical calculations and geophysics and other quantitative courses in a geology curriculum are to help to achieve that end. A comprehensive review of research studies dealing with the impact of fieldwork (Rickinson et. al., 2004) concluded that well planned and delivered fieldwork provides experiences that cannot be duplicated in the classroom (Nugent et. al., 2008); it also positively impacts attitudes, leading to reinforcement between affective and

cognitive domains of learning and higher level learning. Other research has shown that field experiences not only permit but actually encourage perception of the integrated whole, not just the individual parts (Kern & Carpenter, 1986).There have been a number of courses that have shown that an instructional approach that emphasizes investigative, research-based exercises increases student interest in coursework and improves comprehension and retention of fundamental scientific concepts (Smith, 1995). Incorporating cooperative learning activities within the classroom, instead of focusing on the traditional lecture-only style of learning environment, has been shown to have a positive effect on students' comprehension and attitudes, setting the stage for higher cognitive thinking skills to be employed throughout the semester. Additionally, it is important that students obtain competence with equipment comparable in complexity and application to that regularly used in the industry to become more attractive to employers (Tibbs and Cwick, 1994). Despite this understanding, there are few courses specifically designed for near-surface (upper 200 meters of the subsurface) geophysical instruction described within the literature that guide instructors in incorporating this subject area within their existing programs.

We believe it is the responsibility of the college/university to evolve with the demands of industry to ensure their graduates are not only knowledgeable of the subject matter, but also have proficiency in marketable skills. It is our hope that upon completion of the presented Tennessee Intensive Near-surface Geophysics Study (TINGS) course, the students are better prepared to begin an introductory level geophysics job, conduct their own investigations as part of a senior project, and/or have a sufficient understanding

of the field of geophysics to prepare for graduate school. Motivation for this paper is to provide a framework of TINGS and illustrate how it meets industry needs. Objectives of the course are to: (1) incorporate multiple near-surface geophysical techniques and instructional methods in the classroom and in the field; (2) develop methods for teaching geophysics to students of diverse academic backgrounds; and (3) build student skill sets (quantitative, qualitative, communication) that aid in interpreting geophysical data sets. It is expected that these skill sets will help them in other aspects of their careers.

3. Developing an Effective Field-Based Course

The overriding objective of the TINGS program is for the students to experience effective field and laboratory exercises in near-surface geophysics that can be adapted to use elsewhere. The framework presented here may be adapted and expanded by other instructors into a traditional 15-week semester course.

3.1 Course Outline

Participants in the TINGS course are exposed to a sequence of pre-existing half to three-day short courses by industry partners that involve various software and/or hardware sections knitted together via a series of lectures and laboratory exercises by University of Tennessee personnel. These industry instructors are representatives of the manufacturers of the equipment for all geophysical techniques discussed in the course, providing a unique perspective on the instrumentation, software packages, and scientific theory (i.e. physics) behind each technique. This aspect of the course is designed to prevent “burn out” from the instructors by each of them only having to present

information for a short period of time. In addition, students also are safe-guarded from experiencing the phenomenon—common in intensive courses—of having to be exposed to the same presentation style for the full three weeks of the course. The multiple studies conducted by Linek et. al. (2003) support this model of having multiple instructors for any given course. Their findings further support the call for collaborative planning, collaborative implementation, and ongoing collaborative assessment processes as key components in shifting the educational from "how teachers teach" to "how children learn" (Goodlad, 1994).

The TINGS program has been offered at the university for three different summer mini-term sessions (2007, 2008, and 2010) that fall between the Spring and regular Summer terms, with the same representatives from each industry partner teaching their portion of the course. This has allowed the material being discussed and the instruction on how to run the equipment and process data to be consistent each field season. Due to the intensity level of the course (3 credit hours) and amount of time spent during the day in the field and the lab, students are not allowed to enroll in any other courses in this term.

The structure of each section within the TINGS course consists of 3 days. Day 1 will have a morning session of lecturing, followed by an afternoon of demonstration of the equipment and collection of data over the field site. Day 2 will have either the morning collecting data and the afternoon with lecture, or vice versa (weather dependent). In some instances, it is more appropriate to have a portion of the afternoon to introduce the associated software for processing data. Day 3 typically involves all follow up lectures,

and the remainder of the time for student work on processing the data and writing reports of their findings for that geophysical technique. By following this format as a guide, traditional lectures made up approximately 30% of course time, while other activities (labs, surveys, etc.) made up the other 70%. All sections within the TINGS course are taught by industry partners except the seismic refraction and reflection sections, which are taught by one or two faculty members of The University of Tennessee. It should be noted that the industry partners voluntarily spent their time and resources to provide instruction for TINGS, including the use of equipment and transport of any associated course materials, allowing the program to be diverse with very little costs absorbed by the university.

Throughout the course, the objectives outlined in the introduction section will be met by the following activities being key aspects of the curriculum:

1) Implement cooperative-learning activities.
2) Short assignments of increasing difficulty to improve quantitative skills; some can be individual in nature, although it is best if done in groups.
3) Field work designed and conducted by students.
4) Correlations made of class data (individually) to identify anomalies and determine locations of subsurface features within field site.

Assessment of this framework for building quantitative skills and field experiences in near-surface geophysics is conducted by evaluating the performance of each student.

Each student's individual assessment is designed to reflect the importance of class research and a final report.

3.2 Student Demographics

The course is limited to 15 participants each term it is offered. This is due to:

1) Logistics of transporting students from the university to the field site.
2) The idea that intensive, fast-paced courses are better served with a low enrollment number for instructors to better assess student performance.
3) Limited access to the field equipment; If the enrollment is too high, the risk of students not getting adequate training on the instrumentation and/or software processing is high.

Historically, student demographics have consisted of geology majors and minors and engineering students. The course is offered as a senior level, but will occasionally have students just finishing their sophomore year with physics and mathematics requirements still remaining (e.g. Calculus) to be completed. Additionally, the vast majority of the TINGS students have not had any prior formal geophysics courses. However, the course is open to industry partners and therefore many students have come from the environmental consulting field, some of which have been removed from an academic setting (and subsequently any mathematics) for 10 to20 years.

3.3 Field Site Description

Geophysical surveys served as the main activity outside the traditional lectures and were all conducted on the University of Tennessee's Experimental Agricultural Research

Station. Located between Alcoa Highway 129 and the Tennessee River (approximately two miles to the south of the University of Tennessee main campus in Knoxville, Tennessee), this site is also referred to as the Environmental Hydrology and Geophysics Teaching and Research Site (**Figure 39**).

Given the climate of East Tennessee during the Spring months, both wet and dry conditions are expected, which can affect results and variables are discussed during the course. Consequently, relative vadose zone saturation and water table elevations will likely vary among tests conducted on different days, possibly affecting the relative times of refracted first-arrivals among the seismic profiles (e.g., Gaines, 2010). Soil conditions across the site vary from residual soils developed directly on sedimentary bedrock (near the highway) to loamy soils developed on alluvial terraces at elevations above the river. Silt or sandy silt dominates the top 6.1 m of strata, which overlies approximately 0.9 – 1.5 m of fine to medium sand and cemented sand. The lowest portion of the stratigraphic section is comprised of fractured shale till and limestone until reaching bedrock at a depth of approximately 11.6 m. Bedrock is Ottossee Shale, which is a Middle Ordovician member of the Chickamauga Group. As a whole, it is generally characterized by fine-grained calcareous shale with some interbedded limestone.

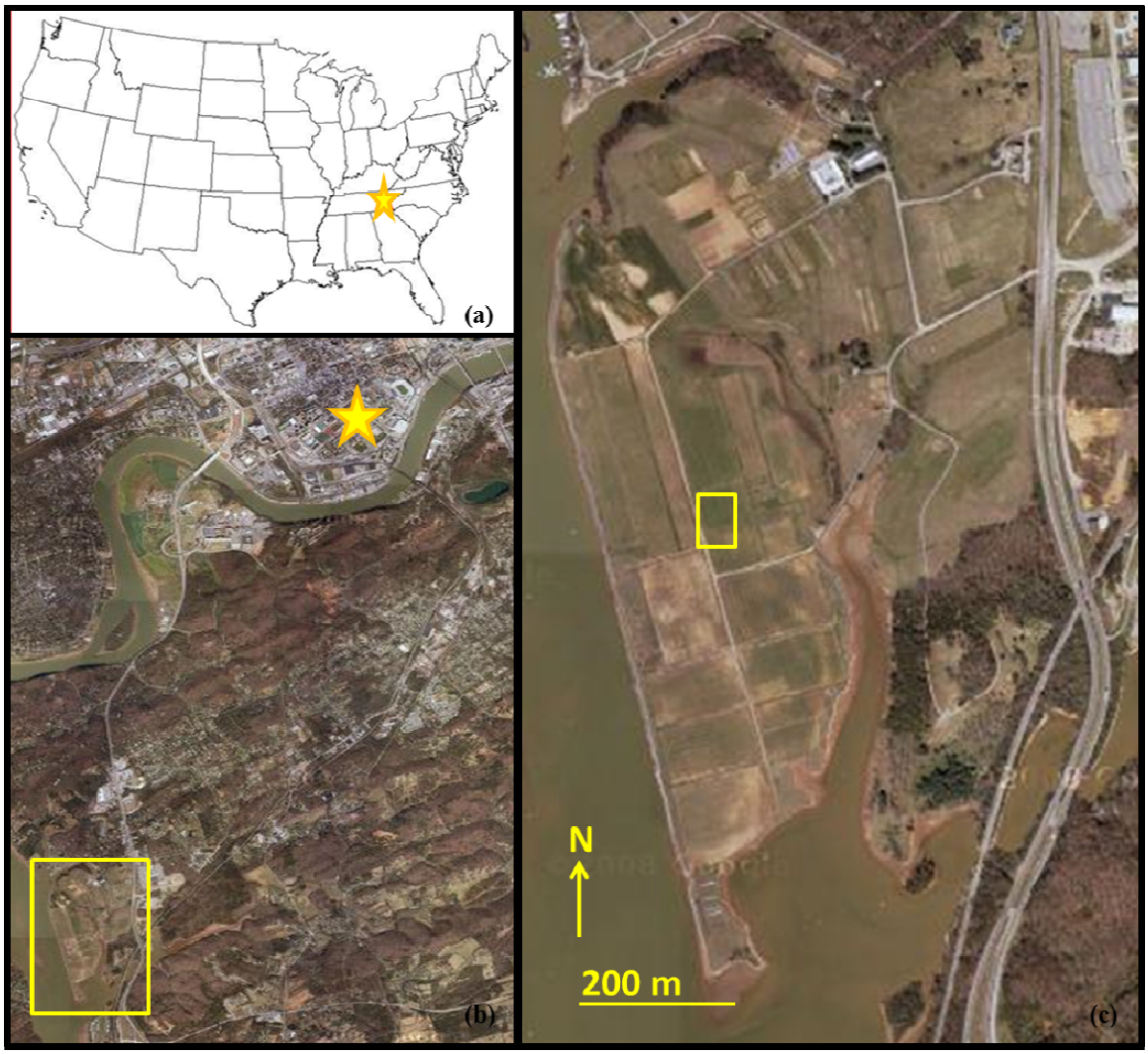


Figure 39. (a) Map showing location of where TINGS course is offered (star indicates Knoxville, TN); (b) Close-up view of Knoxville, TN with yellow box designating The University of Tennessee Geophysical Research Station and star showing Knoxville, TN; (c) Close up view of yellow box from (b), with smaller yellow box indicating general location of where field work for TINGS course is conducted.

The field site additionally contains known targets that were buried in the spring of 1999 having detailed positioning given by latitude, longitude, and depth within the sub-surface (**Figure 40**). Information including size, shape, composition material and orientation is given in **Table 16**. It is assumed that there has been sufficient time for the ground to settle, soils to begin to develop, and any disturbance to the subsurface (and resulting signal in the geophysical data) to be minimized. This has been assessed by noting that data of various types collected over back-filled holes where no object was buried yield no significant anomalies compared to the background.

Although fairly "quiet" from a geophysical noise perspective, the site is susceptible to some background noise from various sources. A relatively large water pump is used intermittently to supply a portion of the agricultural site is located about 200 meters ENE of the plot's NE corner. However, for the purposes of the TINGS course, this noise is not significant and provides an extra teaching situation in what to pay attention to for a geophysical study (i.e. survey design and data processing practices). In addition, the surrounding agricultural plots are occasionally mowed or plowed and the vehicle traffic can cause some issues for seismic techniques.

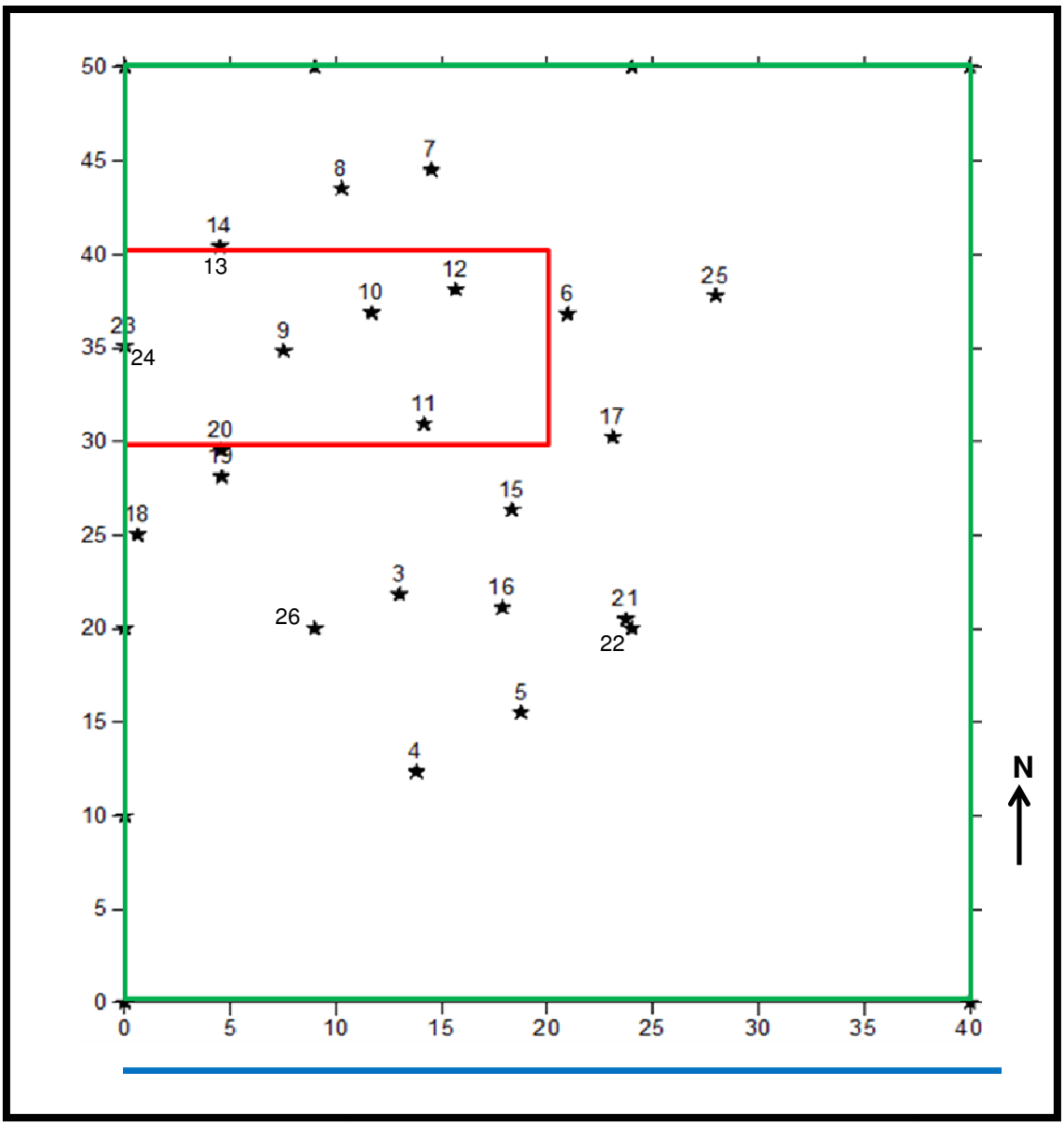


Figure 40. Map displaying locations of buried targets. Grid is measured in meters. Green box is the areal extent of the ground penetrating radar, magnetic, EM-31, and EM-61 surveys. Red box is the areal extent of the 3-D electrical resistivity survey. The blue line is the location of the seismic survey, taken outside the field area for teaching purposes only (data is not correlated with other techniques). Target descriptions for each number on the map is given in Table 16; stars along boundary of survey grid denote markers for the survey and are not targets.

Table 16. Target Descriptions for TINGS Field Site

Map ID	Description*
3	Vertical 55 gal drum
4	Horizontal 55 gal drum, N-S
5	Horizontal 55 gal drum, N-S
6	Horizontal 55 gal drum, E-W
7	Steel scrap, 3 pcs 3-4 feet long
8	Vertical 55 gal drum
9	Plastic 55 gal drum, freshwater and gravel filled
10	Vertical 55 gal drum
11	Plastic 55 gal drum, saltwater and gravel filled
12	Iron pipe, 3" diameter, 42" long
13,14	2 pcs styrofoam, 9'x2'x4", dipping N45E
15	Cement blocks, 1.5 cu feet pea gravel
16	Aluminum gutter; 5 pcs, 6.5-8 feet long
17	Coil of 12/3 copper wire
18	Solid iron rod, ~41" long, 1" diameter
19	Iron Pipe, 4" diameter, 80"? long
20,21	Two vertical drums, 33" center to center along N-S line
22	Iron Pipe, 4" diameter, 64" long
23,24	Two horizontal drums, 19" separation end to end, N-S
25	Styrofoam block, 1 yard cube
26	2 pcs galvanized pipe, 5.5 and 8 feet long

* For location of the buried targets in relationship to each other, refer to Figure 40

3.4 Factors to Consider in Curriculum Design

3.4.1 Geophysics for Non-Geophysics Students

Clearly "geophysics" and "geophysicist" are loose terms encompassing a diverse range of techniques, and individuals with vastly different backgrounds. Teaching geophysics—that is, training geophysicists—poses unique difficulties due to this diversity because any class is likely to include various types of students who will have acquired knowledge in areas relevant to the subject, and who have vastly different expectations of the course (e.g., Dentith and Trench, 1992). It is also expected that students in this course will have a variety of traditional science students and non-traditional students of varying backgrounds, as agreed upon by Bluth and Young (1997), which creates a challenge to present scientific material in ways that spark and hold the interests of science and engineering majors without alienating or frustrating those removed from the academic setting. Geoscience educators have maintained that field work is "critical to the development of spatial reasoning, to the ability to create integrated mental visualizations of Earth processes, and to developing facility with analyzing the quality and certainty of observational data supporting geoscience theories" (Manduca et. al. 2002).

One simple way to remedy this problem is to incorporate cooperative-learning methods throughout the course. Because the TINGS course is designed to be hands-on and team-oriented for students to become comfortable in a multitude of problem-solving environments, these methods are critical. However, the success of collaborative work pivots on all participants assuming responsibility for the process and product of learning

(Perumal, 2008). Cooperative-learning proves particularly helpful once students are processing data collected earlier in the day; working cooperatively sped up the process of debugging spreadsheets and cross-checking field notes for a more complete set of observations prior to interpretation of data. Students, when exposed to this type of learning environment, increase their ability to express themselves in terms of the science they are studying (McManus, 1995). Another advantage of this method is that those students not yet comfortable with the technical nature of geophysics felt at ease asking for assistance or clarification throughout the problem-solving process, resulting in a higher quality in their final project. What is essential to collaborative work, though, is a positive interdependence among students, an outcome to which everyone contributes, and a sense of commitment and responsibility to the group's preparation—for the learning process and product (Peruma, 2008).

Various forms of collaborative learning have been described in college level statistics courses. Instructors employing these techniques reported greater student satisfaction with the learning experience, reduction of anxiety, and a belief that student performance was greater than students could have achieved working independently (Delucchi, 2006). Martin Nikirk (2012) encourages the integration of technology in the classroom when the students are working collaboratively together. This is particularly true in the students he refers to as "Millenial Students" (under the age of 35) because of their comfort level with the use of computers and interactive learning.

3.4.2 Building Quantitative Skills

The geosciences continue to become more quantitative, and in turn geoscience educators need to consider a variety of issues regarding the development of quantitative skills required for geosciences courses at all levels of curriculum (Macdonald et.al., 2000). Throughout the TINGS course students are not required to sit down and derive the equations behind the theory of each geophysical technique they are exposed to; however, they do develop a number of quantitative skills. Students are forced to think analytically and communicate their interpretations of the geophysical data.

As the course progresses, there are small assignments of increasing difficulty required of the students, designed to build upon the skills learned in previous units. Examples of these assignments can be found in Table 17, although this is by no means an inclusive list of the types of assignments given to the students. This approach follows that of Kenyon (2000), in which the "stepped homework" method includes assignments having an increased mathematical difficulty to incorporate the diversity of the student's quantitative backgrounds. The question of the reliability of the student's interpretation of the data (i.e. how reliable are their answers) is an important one to address, as many variables within the field site are known only approximately. This allows students to improve their skills by critically thinking about the project objectives and how their survey design might be affected by those variables.

Table 17. Examples of Assignments Given to Students for TINGS Course

Level of Difficulty	Assignment
1	Given the dimensions of the targets you are trying to locate, design a survey that will minimize any false positives and false negatives in the data, and reduce the edge effect. Calculate the depth of each target given the estimated velocity of the subsurface geology.
2	Examine the two seismic records. What are the p-wave propogation velocities of each layer? Do you think the interface is dipping? Why or why not? What is the thickness of the upper layer?
3	Using the seismic data, identify how many layers are present and label the refractions associated with each layer. What do you think are the strong reflections that are zero-intercept times of 6.4, 9.6 and 12.8 seconds? Use the "slope-intercept" method to calculate the depths of the boundaries and draw a simple sketch of the boundary depths and associated velocities below.
4	With the provided raw data from transects perpindicular to and centered on a hypothetical fracture trend, appropriately process the data, plot it, and interpret the hydrologic significance of the plot. You will also need to calculate a linear regression equation to obtain a residual potential value for each X location. Examine the pattern shown with respect to the setting and the hypothetical resistivity interpretation, and explain it. Does it validate or contradict the interpretation?

3.4.3 Teaching Effectively in the Field

There are a number of examples of integrating teaching with field research in the literature, all of which provide essential guidelines in providing an ideal forum for teaching and cooperative learning. When designing a field project for this course, it was particularly important to follow the logic of Anderson and others (1999) and incorporate multidisciplinary, student-led research that introduces and uses a variety of field methods, fosters interaction between undergraduate and graduate students, addresses provocative scientific questions, and develops a sense of *esprit de corps* among the participants. As Leo Smith (1995) states, these investigative, research-based exercises have improved the learning experiences for students by engaging them in the process of solving local geologic problems using data which they have collected themselves. The timeframe of the TINGS course (Monday-Friday, 9am-5pm for three weeks) results in more contact hours with the students than a traditional 15-week semester allowing more time to be spent teaching in a hands-on, field-methods environment. Combined with observations of the students while engaged in the cooperative-learning exercises, the instructors are able to better assess and effectively distribute the students' talents and maximize the effectiveness of each group. This is particularly true when students are presenting charts, data, and information to their peers (Nikirk 2012).

Conveying to the students that the skills acquired during this course will help them in future careers is vital to sustain a positive and encouraging attitude. These students may become professional geophysicists, geoscientists, or engineers, all of which (as noted by Klasner et al. 1992) will be exposed to literature and reports that contain geophysical

data. In addition to the theory and background of each geophysical technique discussed in this course, their experiences in the field will give them a good understanding of how geophysical data are gathered and interpreted. It is important to point out to the students that each technique has limitations as well as advantages so they are able to critically evaluate the science (or lack of science) within journal articles and discern between good and bad results and conclusions.

3.5 Geophysical Methods Discussed

The structure of the TINGS course included the following geophysical techniques. **Table 18** gives a brief outline of which techniques were included in the course, the industry partner associated with each session, the equipment that the students used for their investigations, and the length of each session. Each term that TINGS was offered the order in which the geophysical techniques were discussed varied due to scheduling constraints of the instructor's time, but the same basic structure was adhered to.

3.5.1 Electromagnetics

3.5.1.1 Ground Penetrating Radar (GPR)

Ground penetrating radar (GPR) utilizes propagating electromagnetic (EM) waves to detect changes in the electro-magnetic properties of the shallow subsurface. This technique responds to changes in wave propagation velocity as a wave travels through the subsurface and generates reflected energy detectable at the surface (Baker et. al., 2007). The propagation velocity of EM waves (i.e., the controlling factor on the generation of

Table 18. Outline of Course Topics and Industry Partners (Instructors)

Geophysical Technique	Industry Partner	Equipment Used	Length of Session (Days)
Seismic Reflection	University of Tennessee	Geophones, Hammer source	3
Seismic Refraction	University of Tennessee	Geophones, Hammer source	1
Magnetics	Geometrics, Inc.	Bartington 601 Gradiometer	2
Electrical Resistivity	AGI, Inc.	SuperSting R8	3
Ground Penetrating Radar	Sensors and Software, Inc.	Pulse EKKO Pro	3
Electo-Magnetics/Ground Conductivity	Geonics, Inc.	EM-31, EM-61	3

reflections) is determined by the dielectric permittivity contrast between the background material and the target. Dielectric permittivity is defined as the ability of a material to store and then permit the passage of EM energy when a field is imposed on the material (Baker et al., 2007) and can be measured in the lab or in situ.

A GPR unit consists of transmitting and receiving antennae, where the transmitting antenna generates an EM pulse in the subsurface that travels into the subsurface, reflects off an interface or scatters off point sources (both caused by a contrast in dielectric permittivity). This reflected/scattered energy then travels back to the surface where it is recorded by the receiving antenna. The time it takes for the wave to travel down to an interface and back up to the surface is called the travel time, and is used to determine the in situ propagation velocity of the subsurface material (Baker et al. 2007) and subsequently the estimated depth to the feature.

3.5.1.2 Ground Conductivity

Ground conductivity refers to the electrical conductivity of the subsurface of the Earth. Terrain conductivity measurements are made by inducing (or generating) an electromagnetic (EM) current into the ground from a transmitter coil, and recording the resulting secondary electromagnetic field at a receiver coil a fixed distance away. The EM-31 and EM-61 ground-conductivity meters are one-person devices containing both transmitter and receiver coils and use an electromagnetic inductive technique that allows measurements without electrodes or ground contact. With this method, surveys can be carried out under most geologic conditions including those of high surface resistivity.

Abrupt changes in the conductivity measurements across the surveyed area are indicative of locations of the desired targets within the field site. Additionally, the electromagnetic response of the target will be primarily dipolar (Casey and Baertlein 1999) for the target/sensor geometries of metallic objects. One disadvantage of this technique is that determination of target depth is difficult, although the effective exploration depth for these instruments is about six meters.

3.5.2 Electrical Resistivity

Electrical resistivity (ER) studies in geophysics may be understood in the context of current flow through a subsurface medium consisting of layers of materials with different individual electrical resistance. The dependence of electrical conductivity (reciprocal of electrical resistivity measured in ER) on soil moisture and fluid salinity and the potential to monitor various subsurface are major reasons ER has become popular within the hydrologic science community (Jayawickreme et. al. 2010). Bulk electrical resistivities in the shallow subsurface are controlled largely by electrolytic conduction in aqueous fluids that are either distributed across grain boundaries or contained in pores, fractures, and faults (Ward 1990).

Conventional electrical-resistivity techniques have the added benefit of being relatively inexpensive (Stummer et. al. 2004). This technique provides a relatively low cost, noninvasive and rapid means of generating spatial models of physical properties of the subsurface. It is especially beneficial for contaminated land investigations where it is generally desirable to minimize ground disturbance (Chambers et. al. 2006). It should be noted that the total resistivity measured at the ground surface in field studies of multilayer

systems (like at the field site for the TINGS course) is not the true resistivity of the underlying material, but a weighted average of the whole system.

3.5.3 Magnetics

A magnetometer measures magnetic field strength at a specific location. For purposes of the TINGS course, we are interested in measuring how much the strength of a magnetic field changes between two specific points, or the "gradient" of the field. The features we are hoping to detect may have magnetic characteristics that cause a disturbance in the Earth's magnetic field in an area around the object. The appeal of this geophysical technique, when dealing with very shallow targets (upper 5m), is that the locations and depths of the sources are found with only a few assumptions about the nature of the source bodies, which are usually assumed as 2D magnetic sources. Salem (2002) explains that for these geological models, the shape of the amplitude of the analytic signal is a bell-shaped symmetric function located directly above the source body. When examining the data set, it is apparent that remnant magnetization-magnetism which remains in a body after the magnetizing force is withdrawn-is present when the target anomaly is relatively negative compared to the background. A normally magnetized body-magnetism vanishes if the external magnetizing force is removed-would produce a positive anomaly (Dannemiller and Li 2006). If the data set shows a dipole (i.e. both a positive and negative response) magnetic anomaly, it is presumed that the source body is metallic in nature.

3.5.4 Seismics

Seismic techniques generally involve measuring the travel time of certain types of seismic energy from sources (i.e. an explosion or weight drop) through the subsurface to arrays of ground motion sensors or geophones. As the energy generated from the sources travels throughout the subsurface, it spreads out as a hemispherical wavefront, eventually arriving at a geophone. Seismic refraction involves measuring the travel time of the component of seismic energy which travels down to the bedrock surface (or other distinct density contrast), is refracted along the boundary, and returns to the surface as a head wave along a wave front similar to the bow wake of a ship. The shock waves which return from the boundary are refracted waves, and for geophones at a distance from the shot point, always represent the first arrival of seismic energy.

Seismic reflection uses field equipment similar to seismic refraction, but field and data processing procedures are employed to maximize the energy reflected along near vertical ray paths by subsurface density contrasts (e.g., Baker 1999). Reflected seismic energy is never a first arrival, and therefore must be identified in a generally complex set of overlapping seismic arrivals - generally by collecting and filtering multi-fold or highly redundant data from numerous shot points per geophone placement.

3.6 Student Final Project

The final project is conducted on an individual basis and all students have the same problem to address. For students projects data are collected using all techniques discussed previously. For the GPR, Magnetics, EM-31 and EM-61, the data were correlated in the same grid space (40x50m plot). For electrical resistivity and seismic techniques, data

was over a smaller areal extent due to logistical constraints and the time involved in collecting and processing those data sets.

Once the data were processed with the respective software packages, images were created to display potential location of the buried targets using the program Surfer. **Figures 41-43** are examples of student work showing correct interpretation of the data for each geophysical technique. The students were then able to create layers within Google Earth over the field study site by incorporating GPS coordinates of the grid corners, with each layer displaying a different geophysical dataset. The scale of the surface map provided by Google Earth was adjusted using permanent landmarks within the survey area, ensuring that the location of the data maps were correct.

Once all layers are integrated into Google Earth, students were able to easily correlate the various anomalies with each other across techniques, predicting where they believed the buried targets were located. **Figure 44** gives a step-by-step visualization of how the final product was put together. **Figure 45** is a collection of photographs showing student involvement within the TINGS course. Additionally, students were required to produce a written report of their findings including a discussion of the types of errors associated with geophysical surveys. The final report is designed to provide the students with an understanding of the importance in data integration methodologies (e.g., Baker et al. 2001) for improving subsurface imaging and reducing uncertainty in discrete anomaly detection.

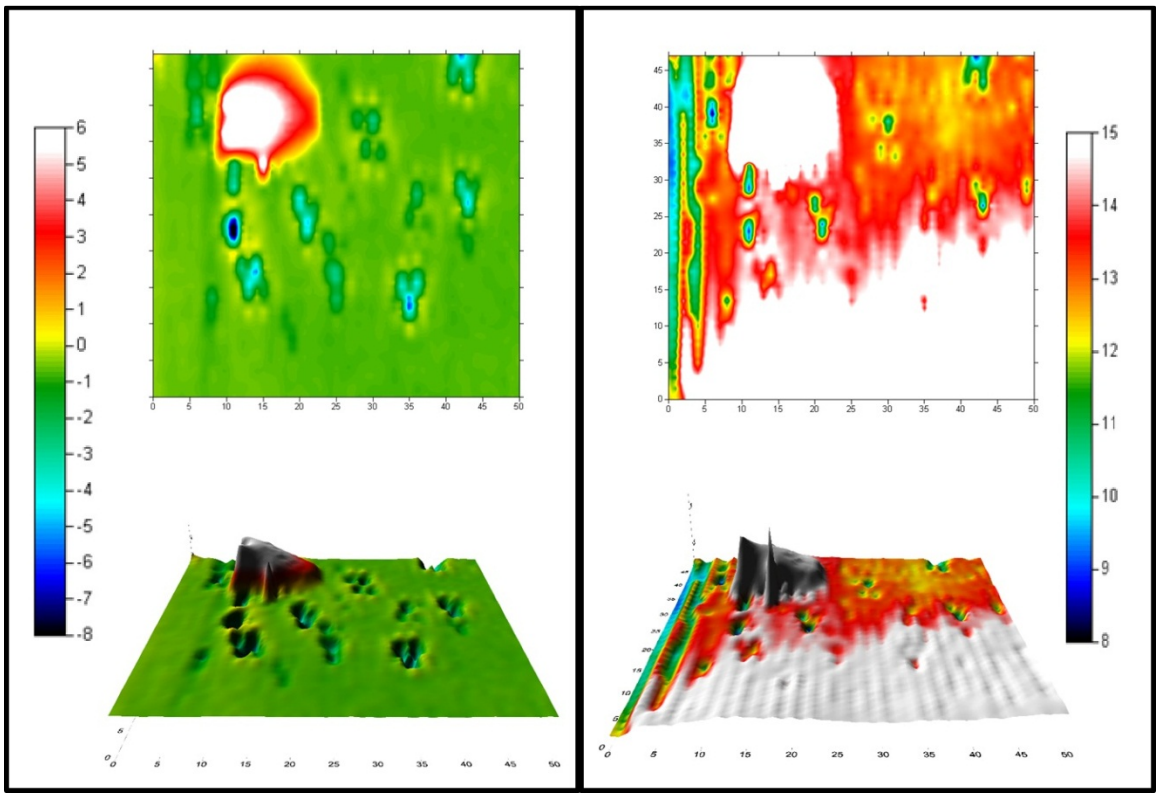


Figure 41. Example of a student's data map for conductivity. North is to the right, parallel to the x-axis. Maps on the left are for the inphase component, with maps on the right for the quadrature component. Units are in mS/m, with each tick mark denoting 5 m of distance across the grid. North is towards the right of the images. Target locations in are interpreted as the depressions in the lower maps. The anomaly in the southwest corner of the lower maps is related to a metal fence and not the location of a target.

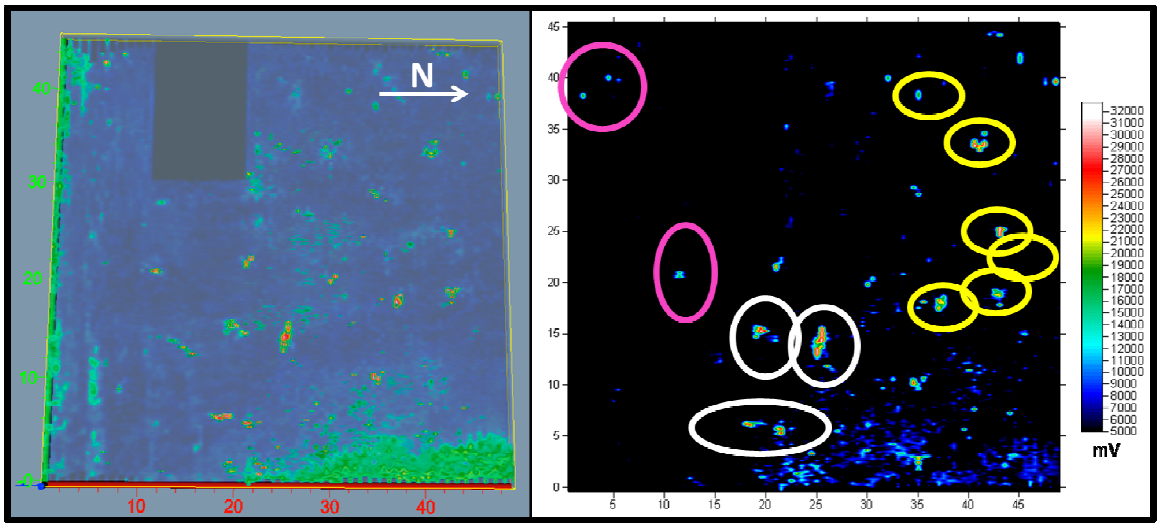


Figure 42. Example of student ground penetrating radar data interpretation. White circles are non-point source targets. Yellow circles are point source targets. Pink circles are monitoring wells. The "highlighted" area in the northeast corner of the grid suggests a high soil contrast, perhaps due to saturated soils (standing water) compared to the other sections of the grid. The blanked data region in the southwest corner of the lower maps is related to a metal fence and not the location of a target.

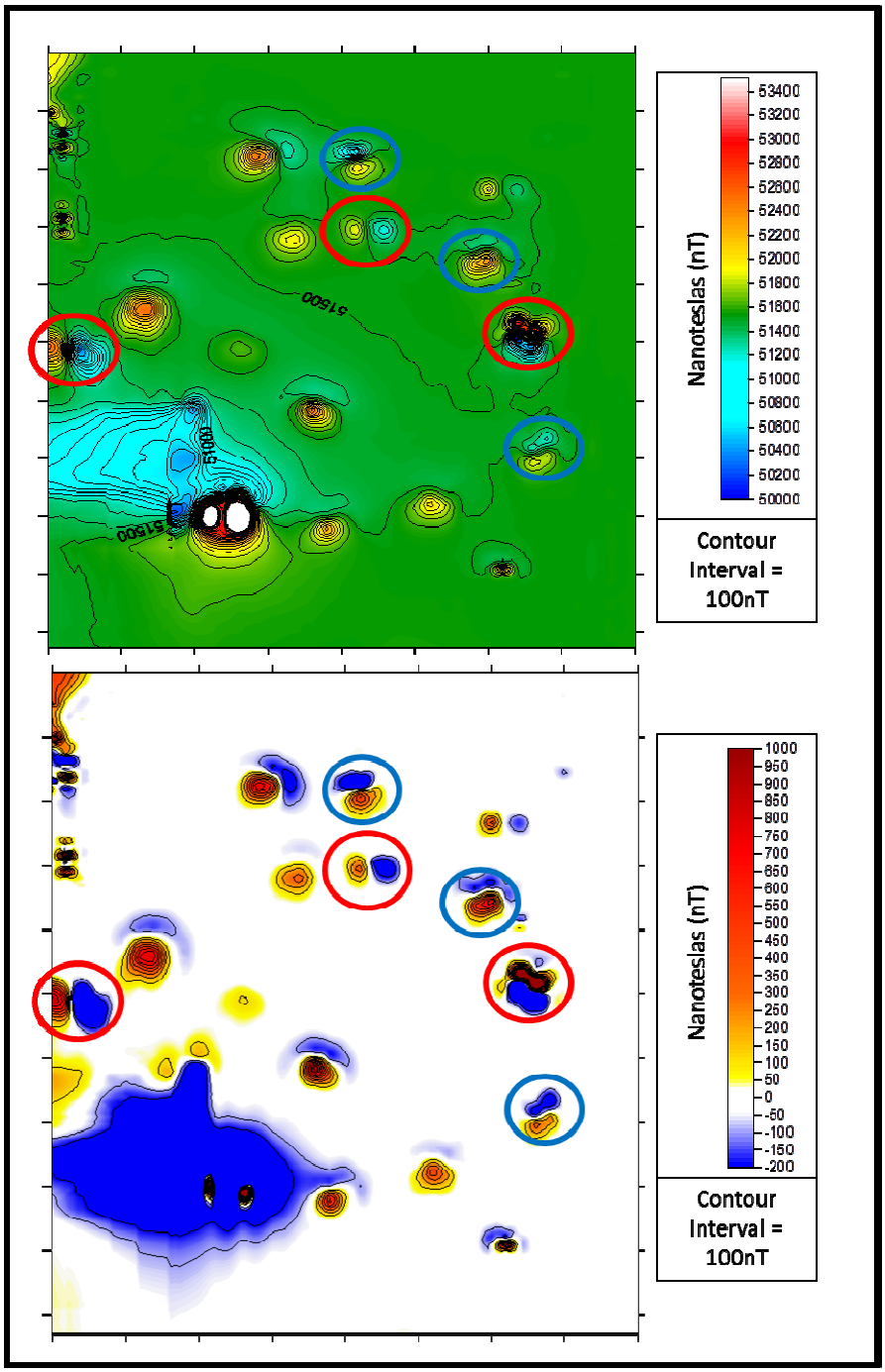


Figure 43. Example of student interpretation of magnetic data. North is up, parallel to the y-axis. The top map is from a cesium vapor gradiometer and displays data from the top sensor only. The bottom map is from the same instrumentation, but gives a gradient reading. Red circles are where targets are thought to have remnant magnetism. Blue circles indicate induced magnetic bodies. The anomaly in the southwest corner of the lower maps is related to a metal fence and not the location of a target.

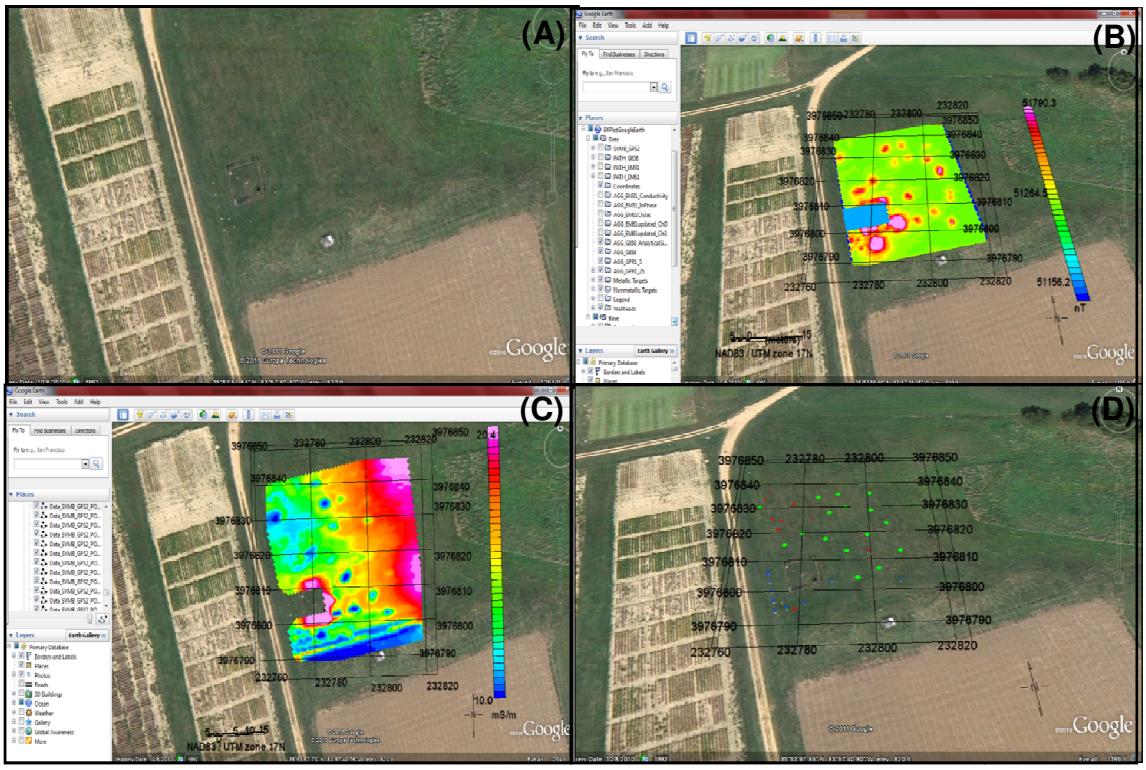


Figure 44. Step-by-step display of student's final project: (A) The location of the field site using Google Earth; (B) All data sets incorporated together, each as its own layer that can be selected separately or in groups to correlate the locations of anomalies in the data to better discriminate the targets; (C) Magnetic data as a layer, complete with GPS coordinates, a scale bar, north arrow, and scale for units measured; and (D) The interpreted location of the buried targets (red-metallic objects, green-nonmetallic objects, blue-well heads).

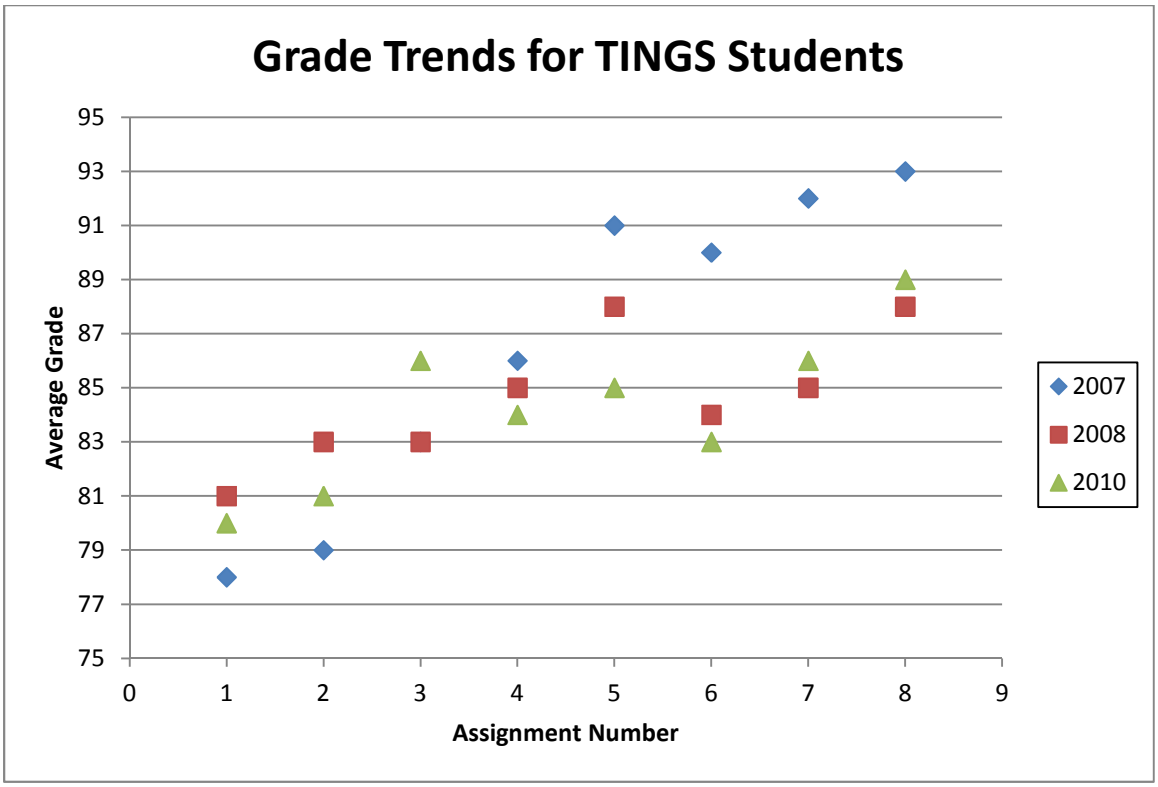


Figure 45. Grade trends for students for each year TINGS course was offered. Data points are the average grade among all students for each assignment, with assignment 9 being the final project; assignments increased in quantitative complexity throughout the course. The overall increase in grades throughout the term despite the difficulty level increasing give evidence that the pedigogy behind the TINGS course is effective.

4. Evaluation of Course

4.1 Objectives Met?

Through the presented framework of the TINGS course, the objectives of the program are satisfied. This is shown by the active participation of students throughout the three weeks, observations that student's critical thinking skills and observation skills improving in a short amount of time, and the final project for the students being of high quality despite academic background. The following is a short synopses of how each objective was met:

1. Incorporate multiple near-surface geophysical techniques and instructional methods in the classroom and in the field.

As highlighted in previous sections, the TINGS course is taught by a number of different instructors, each bringing their own experiences and particular style to the classroom. This gives students a well-rounded exposure to the field of geophysics, from academia to industry. The hands-on nature of the course forces the students to be involved with each step of the project, from survey design to data collection to data processing to interpretation. Additional challenges to the instructors when teaching the theory behind each technique, computer literacy, and general field etiquette were overcome with a small enrollment. Instruction alternated nicely between field time, class time, and computer lab time, ensuring the various learning styles of the students were all satisfied, as well as preventing each component of the course from being too mentally or physically exhausting for the instructors and students.

2. Develop methods for teaching geophysics to students of diverse academic backgrounds.

The TINGS program focuses on techniques, not problems, so participants may include geologists/geophysicists, engineers, archaeologists, agricultural scientists, or other industry professionals. Through the course of the program, students and professionals are introduced to survey design, acquisition, processing/visualization, and interpretation of various near-surface geophysics data. The cooperative-learning activities that take place throughout each phase of the course offers approaches that complement any working environment where team work and interdisciplinary approaches to solving problems are crucial. This is especially true in terms of increasing student comfort with and enthusiasm for quantitative questions (Kruse 1995). The group-learning approach enhances traditional lecture style classroom settings and has proven to have a positive effect on student comprehension and attitudes, which carries over to other academic disciplines throughout school and further into their workplace. The pedagogy behind this methodology may also be implemented with any course where field work is necessary.

3. Build student skill sets (quantitative, qualitative, communication) that aid in interpreting geophysical data sets.

The background of the student strongly influenced how involved they were at various stages of the class, particularly their participation in activities that required a high level of

quantitative skills. Written reports throughout the course improved scientific-writing and critical thinking skills. Once the skill level of each student was assessed, groups were created for smaller assignments or cooperative-learning activities to include as many different majors as possible. This allowed each student the opportunity to act as a secondary instructor, and, depending on the subject being highlighted in lecture, different students acted as the tutor for the group. It was proven effective to assign exercises of increasing mathematical difficulty or higher-order critical thinking, suggesting that the quantitative skills of the students were improved. **Figure 45** summarizes the overall trend of grades on assignments throughout the course; although the assignments increased in quantitative complexity, grades improved which suggests the presented pedagogy is effective. Additionally, it was emphasized where geophysics fits into each person's major field of study, helping with motivation in the learning process and resulted in a higher quality final project.

4.2 Advantages of Course

There are many advantages to the TINGS course curriculum and framework. Primarily, this is one of the few classes that exist that emphasizes both quantitative skills and trains students with the knowledge of how to carry out field work effectively, which has become increasingly apparent in today's job market. Because there is more contact hours with students compared to a traditional semester, there is more hands-on learning and all students are integrated in the process of planning and executing data collection and processing stages of each project (survey design, data acquisition, data processing, interpretation, reporting results). Multiple instructors allow varying view points on field

of geophysics and instructional style and methods appeal to multiple learning styles. With the course open to professionals, students have a strong potential to learn about jobs in the field of geophysics. Additionally, this course is unique as being the only course known to have industry partners giving the instruction, which provides students a more technical and detailed approach to understanding the software and equipment (from the experts).

4.3 Limitations

Limitations to this type of curriculum are mainly due to time constraints. As anyone who is familiar with conducting fieldwork knows, particularly in the realm of geophysics, it is rare to have both weather and equipment cooperating at the same time. A short timeframe for the course meant that there was little room for errors in data collection. Luckily, we have not experienced significant delays with equipment failure or software malfunctions, but it is important to note that should that occur, it would cause some difficulties for students to have time to correlate all geophysical data sets sufficiently.

Another limitation is that short time frame does not allow for complete synthesis of material before moving on to next technique, particularly for those students not previously familiar with geologic terminology and geophysical principles. Compared to a traditional semester, instruction on various geophysical techniques was not as detailed. As enrollment has increased through the years, student performance during course has become somewhat limited because class size does not allow all participants to be actively engaged during data collection phase of course. This may be overcome by splitting class into different task groups that rotate, so everyone is engaged at all times while in the

field; however, during the class sessions and data processing sessions, all students are actively engaged. **Figure 46** shows the students in "action" during the TINGS course.

4.4 Assessment of Pedagogical Impact

Overall student response (n=14) to the class in 2010 was positive. There was no quantitative assessment form for the students to fill out, but a general questionnaire was given two weeks after the final report was turned in to assess the effectiveness of the program. Students from the engineering and geology fields felt they were academically prepared for the course; students in other fields expressed an interest in a one-week optional course prior to the start of TINGS that would give them more background on the field of geophysics before being exposed to the equipment and data processing components of the course.

Students reported that they had a high level of enthusiasm throughout the course for the material discussed, and believed the instructors were equally interested in the students understanding the material and how the equipment worked. Additionally, a number of students mentioned an interest in learning about equipment maintenance, which might have eliminated some down-time in the field. However, they felt the data



Figure 46. Students in “action” during the TINGS course. From the top-right image and moving clockwise: processing ground penetrating radar data; receiving instruction on how to set up an electrical resistivity survey; learning how to operate the EM-31; Surveying with the magnetometer. Photographs courtesy of Noah McDougall.

processing sessions and team-based assignments and field sessions were a great benefit to their understanding of geophysics and the applicability of near-surface techniques would enhance hydrogeologic and geotechnical problem solving.

All students that completed the evaluation stated they would recommend the course to others and that their favorite part of the course was being able to work with the geophysical equipment.

5. Conclusion

The TINGS course has proven to be an effective and crucial aspect to the curriculum offered at the University of Tennessee, as well as the general scientific community. Since its inception in 2007, enrollment has consistently increased, suggesting that the knowledge and skills students gain by participating is beneficial and the course's importance is gaining recognition. Students are able to build the quantitative skills that industry desires, along with gaining the experience in carrying out field work effectively, making them more marketable upon graduation. The students were able to use complex equipment and learn how to plan and implement project designs, building communication skills. Having students from a range of academic backgrounds proved useful in training for effective participation in the work force and building team-oriented skills that are desirable for future employers. It is the suggestion of the authors that this course (or one like it) be incorporated into the general curriculum of geoscience departments to ensure that they are fulfilling their responsibility to evolve with the demands of industry and provide their graduates with not only knowledgeable subject matter, but proficiency in the skills demanded of them.

6. Acknowledgements

The authors would like to thank The University of Tennessee Agricultural Extension for cooperation during data collection and access to the field site. We would also like to thank the students of the TINGS 2007, 2008, and 2010 field season for data collection and use of their reports, particularly Noah McDougall, Aubrey Modi, and Rachel Storniolo. Lastly, we would like to thank Sensors & Software, Inc. (especially Greg Johnston), AGI, Inc. (especially Brad Carr), Geonics Inc. (especially Mike Catalano), Parallel Geosciences, Inc., and Geometrics, Inc. (especially Doug Groom) for their generous donation of instructors and use of equipment throughout the scope of the course.

7. References

Anderson, K., Hickson, T., Crider, J. and Graham, S. 1999. Integrating Teaching with Field Research in the Wagon Rock Project: Journal of Geoscience Education, v. 47, p. 227-235.

Baker, G.S. 1999. Processing near-surface seismic-reflection data: A Primer: edited by R.A. Young, *Society of Exploration Geophysicists Publications*. [ISBN 1-56080-090-9]

Baker, G., Jordan, T. and Pardy, J. 2007. An introduction to ground penetrating radar (GPR): The Geological Society of America Special Paper 432, p. 1-18.

Baker, G.S., Steeples, D.W., Schmeissner, C., Pavlovic, M., and Plumb, R. 2001. Coincident imaging with seismic and GPR techniques: *Geophys. Res. Lett.*, *28*, no 4, 627-630.

Bluth, Gregg and Young, Charles. 1997. An Applied Geophysics Laboratory for Introductory Environmental-Geology Courses: Journal of Geoscience Education, v. 45, p. 447-450.

Casey, K. and Baertlein, B. 1999. An overview of electromagnetic methods in subsurface detection: *in* Detection and identification of visually obscured targets, Taylor and Francis, Oxford, p. 9-46.

Chambers, Jonathan E., Kuras, Oliver, Meldrum, Philip I., Ogilvy, Richard D., Hollands, Jonathan. 2006. Electrical resistivity tomography applied to geologic, hydrogeologic, and engineering investigations at a former waste-disposal site: Geophysics v.71, p. B231-239.

Dannemiller, Neal and Li, Yaoguo. 2006. A new method for determination of magnetization direction: Geophyics, v. 71(6), p. 69-73.

Delucchi, Michael. 2006. The Efficacy of Collaborative Learning Groups in and Undergraduate Statistics Course, College Teaching, v. 54, p. 244-248.

Dentith, Michael and Trench, Allan. 1992. Teaching Geophysics To Students with Various Scientific Backgrounds: Journal of Geological Education, v. 40, p. 324-328.

Gaines, D.P., Baker, G.S., Hubbard, S.S., Watson, D, Brooks, S., and Jardine, P. 2010. Detecting perched water bodies using near-surface seismic time-lapse travel-time tomography in Advances: *in* Miller, R.D., Bradford J,H, Holliger, K., eds., *Advances in Near Surface Seismology and Ground-Penetrating Radar*, Society of Exploration Geophysicists, Tulsa, OK, ISBN 9781560802242.

Goodlad, J. 1994. Educational Renewal: Better Teachers, Better Schools, San Francisco: Jossey-Bass.

Jayawickreme, Dushmantha H., Van Dam, Remke L., Hyndman, David W. 2010. Hydrological consequences of land-cover change: Quantifying the influence of plants on soil moisture with time-lapse electrical resistivity: Geophysics v. 75, p. WA43-50.

Kenyon, Patricia. 2000. Building Quantitative Skills in Geoscience Courses Using Homework Assignments of Increasing Difficulty: Journal of Geoscience Education, v. 48, p. 418-422.

Kern, E. L., & Carpenter, J. R. 1986. Effect of field activities on student learning. *Journal of Geological Education*, v. *34*, p. 180-183.

Klasner, J., Crockett, J., Horton, K., Poe, M., and Wollert, M. 1992. Hands-On Teaching Through a Student Field Project in Applied Geophysics: Journal of Geoscience Education, v. 40, p. 53-61.

Kruse, Sarah. 1995. Using Cooperative-Learning Methods to Teach Quantitative Material in an Undergraduate Geophysics Course: Journal of Geological Education, v. 43, p. 357-360.

Linek, Wayne, Fleener, Charlene, Fazio, Michelle, Raine, I. LaVerne, Klakamp, Kimberly. 2003. The Impact of Shifting From “How Teachers Teach” to “How Children Learn,” Journal of Educational Research, v. 97, p. 78-89.

Macdonald, R., Srogi, L. and Stracher, G. 2000. Building the Quantitative Skills of Students in Geoscience Courses: Journal of Geoscience Education, v. 48, p. 409-412.

Manduca, C., Mogk, D., & Stillings, N. 2002. *Bringing research on learning to thegeosciences.* Retrieved July 26, 2006, from http://serc.carleton.edu/files/research_on_learning/ROL0304_2004.pdf

McManus, Dean. 1995. Changing a Course in Marine Geology from Lecture Format to a Cooperative-Learning Format: Journal of Geological Education, v. 43, p. 327-330.

Perumal, Juliet. 2008. Student resistance and teacher authority: the demands and dynamics of collaborative learning, Journal of Curriculum Studies, v. 40, p. 381-398.

Nikirk, Martin. 2012. Teaching STEM to Millennial Students, TechDirections, February Special STEM Issue, p. 13-15.

Nugent, Gwen, Kunz, Gina, Levy, Richard, Harwood, David, Carlson, Deborah. 2008. The Impact of a Field-Based, Inquiry-Focused Model of Instruction on Preservice Teachers' Science Learning and Attitudes, Electronic Journal of Science Education, v. 12, p. 94-111.

Rickinson, M., Dillon, J., Teamey, K., Morris, M., Choi, M. Y., Sanders, D., & Benefield, P., A Review of Outdoor Learning, Retrieved September 6, 2007 from http://www.nfer.ac.uk/research-areas/pims-data/summaries/fsr-a-research-reviewof-outdoor-learning.cfm

Salem, A., Ravat, D., Gamey, T. and Ushijima, K. 2002. Analytic signal approach and its applicability in environmental magnetic investigations: Journal of Applied Geophysics, v. 49, p. 231-244.

Shea, J.H. 1990. Mathematics in physical-geology textbooks: Journal of Geological Education, v. 38, p. 138-148.

Smith, George. 1995. Using Field and Laboratory Exercises on Local Water Bodies to Teach Fundamental Concepts in an Introductory Oceanography Course: Journal of Geological Education, v. 43, p. 480-484.

Stummer, Peter, Maurer, Hansruedi, Green, Alan G. 2004. Experimental design: Electrical resistivity data sets that provide optimum subsurface information: Geophysics v. 69, p. 120-139.

Tibbs, Nicholas and Cwick, Gary. 1994. An Equipment-Intensive Field Methods Course in Environmental Geoscience: Journal of Geological Education, v. 42, p. 455-460.

Ward, S. H. 1990. Resistivity and induced polarization methods, in Ward, S. H., Ed., Geotechnical and environmental geophysics: Society of Exploration Geophysicists, Investigations in Geophysics v. 5, p. 147–189.

VITA

Megan Estelle Carr was born at Pensacola Naval Air Station in Florida to parents Daniel and Dina Carr. After an active school career of academics, musical theater, church and school choirs, varsity tennis, and church youth group, she received her Advanced with Honors Diploma in 1998 from Clear Brook High School, Friendswood, Texas. Megan continued her education at Baylor University, Waco, Texas, earning a Bachelor of Arts in 2002 with a double major in Earth Science and Environmental Studies under the advisement of Dr. Susan Bratton. In 2004 she earned a Master of Science in Environmental Science under the advisement of Dr. Joe Yelderman. Before returning to academia to pursue a PhD in near-surface geophysics at The University of Tennessee with Dr. Gregory S. Baker, she taught 7th grade Science for two years in her home school district. Megan had the unique experience of teaching on the same campus with her mother and Mrs. Kroll, her intermediate school Earth Science teacher, to whom this dissertation is dedicated. Megan is the proud mother of two Jack Russell terriers, Zabo and Legolas, and will be continuing her career in Anchorage, Alaska after graduation.