ABSTRACT

Title of Thesis: MAGNETIC PROSPECTING IN ARCHAEOLOGY: A CALCULUS OF THE DETECTION OF PRECONTACT ARCHAEOLOGICAL FEATURES WITH MAGNETOMETRY

Isabelle Ortt, Master of Professional Studies, 2026

Thesis Directed By: Dr. Matthew M. Palus, Department of Anthropology

Magnetometers are sensitive to minute changes in the earth's magnetic field, allowing for the detection of, and mapping of, certain archaeological features. The ability of a magnetometer to detect and map an archaeological feature requires first that the feature be magnetic, but also that the magnetometer is recording data at a density appropriate to interact with the feature's anomalous magnetic field. The size of the feature's magnetic anomaly cannot be known prior to survey, as it depends on (among other things) the feature's actual size and composition, so some estimation is involved in the design of magnetometer surveys to allow for these unknowns. Decisions about data density, or survey resolution, may depend on *a priori* knowledge of the discipline, site type, and expected or sought-for features, but also may depend on state and federal archaeological standards. Data density standards in magnetometry are rarely standardized in formal State Historic Preservation Office documents, and despite the industrialization of many

facets of archaeological practice, archaeological geophysics has remained a 'craft' – the discipline lacks universally accepted density standards for data collection.

This lack of standardization provides opportunity for, and implies a need for, evaluation of existing recommendations for magnetometry in archaeology in specific scenarios, leading to the development of a research question: what are the effects of altered transect-interval widths on precontact, thermoremanent archaeological feature detection and location in magnetometry? This research represents an application of technical geophysical theory as a craft, in exploring the technical and subjective tension between the patchwork series of state recommendations, the academy, and the practitioner.

This research question is explored by evaluating if these features would have been detected and located across all 'recommended' transect intervals, regardless of applicable state guidelines. This thesis examines the effects of artificially reduced data density on two detected, ground-truthed, and mapped archaeological features originally detected during a high resolution survey with 0.25-meter transect interval separation. In this way, 0.50-meter and 1.0-meter transect interval separations are also evaluated, which in some states, literature, and scenarios, are accepted survey parameters. Additionally, detection of a feature has limited utility if the goal of the geophysical survey is to facilitate excavation of the feature (location) to assist with assessment of a site under the framework of the NRHP during work within the context of CRM. To this end, two levels of effort (soil core removal and shovel test excavation) are tested against the results of the interval analysis, in order to determine if the lower resolution survey results would result in the location of each feature during multiple ground-truthing scenarios.

MAGNETIC PROSPECTING IN ARCHAEOLOGY: A CALCULUS OF THE DETECTION OF PRECONTACT ARCHAEOLOGICAL FEATURES WITH MAGNETOMETRY

by

Isabelle Ortt

Thesis submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Master of Professional
Studies
2026

Advisory Committee:
Dr. Matthew M. Palus, Chair
Mr. Michael Konzak
Dr. Kathryn Lafrenz Samuels
Dr. Miguel Vilar

Acknowledgements

Thanks to my advisor, Dr. Matthew Palus, for providing encouragement and guidance for theoretical ideas of all sorts. Thanks to my committee, for their effort and time; and especially Mr. Michael Konzak, for providing feedback on the technical approaches and language utilized in this thesis. I would also like to broadly acknowledge the efforts of the faculty and staff of the distance-learning Cultural Heritage and Resource Management program at UMD – an inspiring effort to bridge the gap between industry and the academy and give us busy, working professionals a nontraditional path forward as graduate students. Additionally, this research would not have been possible without the support of my team and mentors within CRM at Stantec, particularly John Flood, Ryan Peterson, Matthew Pike, Katie Settle, and Duane Simpson, who have taught me everything I know about CRM archaeology. Thanks to Dr. Michele Buzon and the Anthropology Department at Purdue, who gave me my start in the field. Finally, a sincere thank you to our confidential client for generously allowing the use of anonymized geophysical data in pursuit of this research.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

AIA	Association of Iowa Archaeologists (Iowa SHPO)
ASCII	American Standard Code Information Interchange
CALTRANS	California Department of Transportation
CORS	Continuously Operating Reference Station
CRM	Cultural Resource Management
DAHP	Divisions of Archaeology and Preservation (Louisiana SHPO)
DoD	Department of Defense
EAC	European Archaeological Council
ER	Electrical Resistance, or Earth Resistance
FCR	Fire Cracked Rock
GNSS	Global Navigation Satellite System
GPR	Ground Penetrating Radar
SI	International System of Units
MA	anomalous positive magnetic gradient value field, AKA Magnetic Anomaly
MAE	Mean Absolute Error
MHT	Maryland Historic Trust (Maryland SHPO)
MISHPO	Michigan SHPO
NCOSA	North Carolina Office of State Archaeology (North Carolina SHPO)
NRHP	National Register of Historic Properties
nT	nanotesla, sub-unit of measurement of magnetic flux density
OHC	Ohio History Connection (Ohio SHPO)
RTK	Real-Time Kinematic
SD	Standard Deviation
SHPO	State Historic Preservation Office
STU	Shovel Test Unit
SQUID	Superconducting Quantum Interference Device
USACE	United States Army Corps of Engineers

List of Operators and Variables

$\| \|$	absolute value operator
α	confidence level
x	datapoint
$\frac{d[]}{dt}$	differential operator, time derivative
H_{ex}	external magnetic field
Q_i	*i*-th datapoint of example Dataset Q
R_i	*i*-th datapoint of example Dataset R
k_{th}	infinitesimally small (but larger than zero) subdivided segment of area
BA	magnetic flux (Scollar et al. 1990), in scenario described in text
N	number of datapoints in dataset
n	number of turns of wiring, within gradiometer (Equation 1; Scollar et al. 1990)
λ_N	Nyquist wavelength
μ	permeability (Equation 1; Scollar et al. 1990)
μ	mean (Equation 6)
P, P_k	probability of a successful search in (k_{th} region of) search area (Kryala 1964)
P_{hit}	probability of a successful search (Schmidt and Marshall 1997)
L_k	probability of identifying signature as potentially archaeological in (k_{th} region of) search area (defined here by author)
p, p_k	probability that object of search exists in (k_{th} region of) search area (Kyrala 1964)
F, F_k	probability that search (of k_{th} region) will result in discovery (Kyrala 1964)
A_k	search area (Kyrala 1964)
Σ	summation operator
A	surface area
a	survey interval spacing (Schmidt and Marshall 1997)
t	time
V_{sec}	induced voltage (Scollar et al. 1990)
λ	wavelength
w_z	zero-width (Schmidt and Marshall 1997)
σ	standard deviation
Z	standard score, or Z-score

Chapter 1: Introduction

Definition of Research Question

Geophysical prospection, magnetometry in particular, is increasingly employed as a tool in archaeological work in cultural resource management (CRM) contexts. Agency, academic, and industry standards all guide the ways archaeological work is conducted in CRM, but formal regulation of geophysical techniques is rare. In contrast with an "industrialized archaeology" characterized by "standardized procedures, evaluative criteria, and…routinized practice," an alternative conceptualization of archaeological work as 'craft' holds space for "subjectivity," "apprenticeship and mastery" and "experience" (Shanks and McGuire 1996: 80, 82, 86).

Geophysical work in CRM is often approached from a 'craft' basis in the absence of SHPO (State Historic Preservation Office) regulation, where practitioners develop their own standards and routines shaped by literature, their own lived experiences, and rarely, intersection with formal guidelines. These standards can and do differ, particularly as the amount of transects walked by during geophysical survey (or, survey resolution in the interval-dimension) is directly related to time required to complete the effort, and therefore cost of work. Survey density along those transects (or, in-line collection rate or density) is typically not a limiting factor in magnetometry, and therefore has less bearing on cost and schedule. A small minority of SHPOs regulate geophysics in CRM, including Ohio, which is the only state with published standards for magnetometry.

In 1997, researchers from the University of Bradford in the UK examined the effects of coarser (wider) sampling intervals on the interpretation of magnetometer data while conducting a study evaluating English Heritage guidelines for geophysics, which were intended to guide magnetometer usage in their region at the time (Schmidt and Marshall 1997). This work involved a discussion of sampling theory, using both mathematical derivations and practical experiments to define appropriate sampling intervals based on the size and depth of target objects. However, this early exploration of the effects of reduced survey intervals focused only on the magnetic detection of archaeological features, rather than the ability of this magnetic data to allow for their physical location using a defined method of excavation. To use magnetometry successfully as a tool of prospection, the magnetometer must detect the feature, and the feature must be able to be physically located using this data.

Magnetometer survey was completed with a dual fluxgate gradiometer at a Woodland and Archaic Period precontact site in central Ohio during an archaeological Phase II evaluation carried out under contract. This geophysical survey, followed by ground-truthing, accomplished the detection and documentation of two archaeological features, using higher resolution magnetometer data than is required by the Ohio SHPO. In this thesis, I will evaluate the detection of two thermal precontact archaeological features located on the Till Plains of Central Ohio, exploring the effects of survey resolution on the magnetic detection and location of these features from a 'geophysics as craft' perspective. Specifically: what are the effects of altered transect-interval widths on precontact, thermoremanent archaeological feature detection and location in magnetometry? To answer this question, I will explore whether these features would have been magnetically detectable at wider transect intervals, and whether they could have been subsequently located during ground-truthing excavations guided by these lower resolution magnetometry datasets. I will also explore the effects of human error and interpolation, which artificially increases the resolution of collected data by increasing the number of data points using mathematical algorithms, on the precision of feature location using these datasets.

Test Site Setting and Research Context in CRM

This thesis leverages data collected during an archaeological evaluation of a previously identified Archaic and Woodland Period surface scatter (n=129) of lithics, Fire Cracked Rock (FCR), and ground stone tools. This site (the Test Site) has been anonymized by request of the client on whose behalf this work was completed. Actual, objective locational information for the Test Site described above have been withheld on request of the confidential client. The Test Site was initially delineated by Stantec Consulting Services, Inc. (Stantec; then Cardno) during Phase I survey work in 2021 for a confidential client (Settle et al. 2022a, 2022b). The density of artifacts (129 artifacts across 1.12 acres) in the surface scatter at the site and presence of FCR suggested precontact occupation that may have been intensive rather than ephemeral. Following this survey, the site was determined by the Ohio SHPO, the Ohio History Connection (OHC), to be potentially eligible for listing in the National Register of Historic Places (NRHP) under Criterion D. Further work was carried out by Stantec at this site in 2025 during a geophysics-guided Phase II evaluation for that same confidential client. The magnetometry data analyzed in this thesis were collected during this archaeological evaluation, using a Bartington GRAD-601 dual fluxgate gradiometer with an in-line collection rate of 8 data points per meter along transects spaced 0.25 meter apart. This Phase II work (report in draft) led to the identification and excavation of two intact subsurface archaeological features, both thermal in nature and therefore magnetically detectable. These features are subsequently referred to as Feature 1 and Feature 2 and serve as test cases in this thesis.

The Test Site is situated on the eastern slope of a slight rise in rolling moraine in an agricultural field in central Ohio. Site soils are composed of Blount silt loam, end moraine, 2 to 4 percent slopes (Ble1B1), Glynwood clay loam, 6 to 12 percent slopes, eroded, and Wetzel silty clay loam (Wz). These soils are somewhat poorly drained, moderately well drained, and poorly drained, respectively, with clay/silty clay subsoil below 7 - 10 inches of plowzone. The agricultural field in which the site is located has been cultivated for at least 70 years. Both features were uncovered approximately 0.30 meter below the ground surface, truncated by the plowzone.

Data are utilized from one additional site delineated and investigated by Stantec over the course of the Phase I and II projects, for comparative purposes. This site (Comparison Site) is a moderate density (n=34) scatter of precontact artifacts with diagnostic elements dating to the Late Woodland and Proto-Historic Periods identified during Stantec’s Phase I work and further explored during Stantec’s 2025 archaeological evaluation (Settle et al. 2022a). This site was recommended for further work due to density of surface scatter and potential for more intensive excavation. The site is on the apex of a moderate rise in rolling moraine, within an agricultural field. Site soils consist of Blount silt loam, end moraine, 2 to 4 percent slopes (Ble1B1). Magnetometer data from this site were chosen as a comparative sample to those from the Test Site due to the similar descriptive statistics between the two datasets, which nonetheless led to the development of visually-distinct graphics. Locational information for the Comparison Site has also been omitted.

Summary of Background and Methods

Geophysical survey is a well-established method of preliminarily contextualizing an area or archaeological site without significantly disturbing any soil or subsurface archaeological deposits. However, employing geophysical instruments as tools of prospection requires 1) that archaeological features are both detectable (generally described in geophysical literature as 'anomalies') to those instruments and 2) that the instruments are capable of detecting the features (as 'anomalies') accurately enough to later locate them and facilitate their discovery during the planned level of effort allotted for confirmatory excavations. The level of effort allotted for ground-truthing may vary depending on schedule and budget constraints. A 'false negative' occurs when no feature is detected but is present, while a 'true negative' is a location absent of archaeological features. Essentially, it does not matter if features are detectable geophysically if these detections cannot be located using the methodology employed: if the relationship between survey resolution, level of excavation effort, and feature size is not fully understood, practitioners run the risk of mistaking 'false negatives' for 'true negatives' and ineffectively contextualizing archaeological sites. Both detection and location involve an interplay of the feature size, its chemical and physical makeup, its depth, and the survey resolution – a survey resolution sufficient for detection may not be detecting the feature accurately enough to lead to the feature's location later. What would the trade-off between efficiency and sufficiency be, in this case?

Survey resolution in magnetometry has both X-dimension (here, defined as in-line data collection along a transect) and Y-dimension (defined here as transect interval or separation) components. X-dimension resolution can be increased by increasing the frequency of datapoint recordation on the instrument traveling along a transect, which does not increase survey time. Practitioners typically maximize this resolution to what is practical for data storage and processing, as it does not increase the time required for fieldwork and therefore has limited effect on scheduling and budget. In contrast, Y-dimension resolution can only be increased by physically walking more transects, which directly affects time, and cost, of fieldwork. It is therefore the transect interval, here the Y-dimension of survey, which is the limiting factor for both budget and data density.

Geophysical data was collected at the Test Site using a Bartington Grad-601 dual fluxgate magnetometer in gradiometer configuration with an X-dimension resolution (X-resolution) of 0.125 meter and a Y-dimension resolution (Y-resolution) of 0.25 meter, a survey resolution in accordance with some industry and literature standards but above the recommended resolution for magnetometry survey published by the Ohio State Historic Preservation Office (SHPO). As the magnetometer survey at the test site did successfully facilitate the detection of and location of Features 1 and 2, it can be concluded that this survey resolution (0.125x, 0.25y) was sufficient to fulfill both detection and location conditions discussed above. As the feature dimensions and locations are known from the Test Site data, it is possible to examine the discrepancy between magnetic anomaly locations and feature locations across increasingly reduced resolutions, making this dataset uniquely suited to address this research question.

The simulated ground-truthing level of effort for the purposes of this thesis is four-tiered: 1) a single shovel test unit (STU) measuring 0.50 meter by 0.50 meter placed over the mathematical center (centroid) and 2) maximum value of the statistically-defined anomalous positive magnetic gradient value field (magnetic anomaly; MA) generated by the archaeological features, and to simulate a more ‘cost-effective’ means of evaluating an anomaly, 3) the placement of a single soil core test (of insignificant diameter) at the centroid and 4) maximum of each magnetic anomaly. The placement of STUs at magnetic anomaly centroids allowed for the discovery of both features during the archaeological evaluation at a Y-resolution of 0.25 meter, although the anomaly centroids targeted during Stantec’s fieldwork were visually identified rather than calculated, a key difference in methodology which did affect feature location.

In order to evaluate the potential effect of reduced resolution on feature detection in this case, the collected dataset was split into two test datasets of halved Y-resolution (to 0.50 meter), staggered by 0.25 meter. These datasets are technically acceptable by state standard, but are below what some in industry and literature recommend for prospection-level survey resolution. In order to obtain a comparative sample of known sub-standard resolution, an additional dataset of halved Y-resolution was extracted from each 0.50-meter Y-resolution dataset to produce two datasets of 1.0-meter Y-resolution. It is expected, based on the size of the anomalous values associated with both features, that the 1.0-meter datasets will be sufficient for one-dimensional detection but will not allow for the further characterization of either feature. These four datasets were then interpolated back to the Y-resolution of the collected dataset in order to allow for point-by-point quantitative comparison and visual comparison of graphics with identical resolution to the original collected dataset, as well as to explore the effects of interpolation on precision and accuracy of feature prospection in this case.

The effectiveness of lower-resolution datasets as prospection tools is evaluated by repeating the ground-truthing level of effort with the 0.50-meter and 1.0-meter Y-resolution datasets. If 0.50-meter Y-resolution is sufficient to detect and characterize both identified features in this case, the data and graphics developed from the two 0.50-meter Y-resolution datasets should not vary significantly between these datasets and the original dataset collected at a closer Y-interval. Similarly, if a 1.0-meter Y-resolution is sufficient, the data contained in those reduced Y-resolution datasets should direct practitioners to the detection and characterization of both identified features, just as the collected data did. While it may not be necessary to completely characterize an anomaly prior to ground truthing, the methodological cost of cheaper, faster work should be accounted for quantitatively in order to understand the shifting potential of mistaking a 'false negative' for a 'true negative'.

Thesis Outline

An overview of geophysical, and specifically magnetic, prospection theory can be found in Chapter 2. This chapter details the origins of geophysical prospection, the theory and practical applications of magnetometry, and the instruments and sensor configurations currently utilized in archaeological prospection. The discussion of instruments and sensor configuration focuses on magnetic fluxgate gradiometers, one of which is utilized for data collection in this thesis. Sources of error and resolution in magnetometry are briefly examined, and the process and common practices for processing magnetic data are then detailed.

Chapter 3 contains a review of literature relevant to establishing standards of resolution in magnetometry. Applicable state recommendations and industry standards for instrument and survey resolution are detailed. The body of research which establishes these best practices for in-line collection rate and transect interval separation in magnetometry explored, as well as the foundations in sampling theory used to define detection limits for archaeological features by size.

This chapter concludes with a brief description of the environment and physiology of the test site.

Data collection and processing methods are described in Chapter 4. Data collection involves geophysical grid setup, size, and placement; instrument resolution settings; and ground-truthing procedures. Data processing functions within the processing package, TerraSurveyor, are explained. The characteristics of the collected data, and dataset generation procedures for test Datasets 2-9, are introduced. Finally, the quantitative Mean Absolute Error (MAE) test is introduced, which will be used to evaluate the point-by-point differences between each dataset.

The results of the magnetometer survey and subsequent ground-truthing at the test site are presented in Chapter 5. Feature and magnetic anomaly plan view positions are displayed, and in-line and Y-dimension magnetic gradient traces are presented and used to calculate the limit of detection for each of the two anomalies attributed to an archaeological feature. Graphics of the two collected magnetometry grids are also displayed.

The following chapter, Chapter 6, contains a qualitative and quantitative analysis of the collected data (Dataset 1), as well as the generated test datasets of varying resolutions. Graphics were generated from each dataset using the TerraSurveyor geophysical data processing software package. Excel Visual Basic was used to carry out MAE calculations between Datasets 1-9, and calculate Z-value thresholds in order to provide a standard definition for ‘anomalous’ magnetic gradient values generated by archaeological features in a standard manner across all nine datasets. The physical locations, centroids, and maximum values of these magnetic anomalies are then compared across datasets and with the physical extents of archaeological features located during ground truthing. This allows for the determination of which resolutions are sufficient for detection vs. deeper characterization of magnetic anomalies as features, and how large a portion of each anomalous extent would need to be excavated during ground truthing in order to have a successful search (a feature detection) at that resolution.

The conclusions of this study, including recommendations for future work and a discussion of relevant technical and human errors, are presented in Chapter 7, followed by references.

Chapter 2: Theoretical Background

Introduction to Geophysical Prospection

Although geophysical prospection gained prominence throughout the 20th century, especially in England, it was not often used in North America even in the early years of the 21st century (Kvamme 2003: 132). Geophysical prospection uses machines that measure "various physical properties such as acoustic velocity, magnetic intensity, gravity, or electrical conductivity/resistivity" (Berezowski et al. 2021: 2881). Variations in these physical properties are caused by alterations to the physical subsurface environment, some of which can be attributed to human behavior. These alterations are detectable when they result in variations in temperature, moisture content, density, electrical conductivity, chemical and mineral properties, and magnetic susceptibilities (Davenport 2001: 89). These alterations, typically referred to as 'anomalies' in literature and practice, are detected with most accuracy from one to two meters below the surface (Sitters 2012:6). The advantages of utilizing several different methods of prospection are well-established (Kvamme 2003: 143). Historical documentation can also be leveraged as a guiding source of evidence when prospecting (Gramlich et al. 2024: 6517).

An understanding of the operation of prospecting instruments, including potential avenues of error generation, is key to a robust, precise, and accurate deployment of geophysics as a tool in archaeology. This research examines data resolution in magnetometry, one method of geophysical prospection. Magnetometry is useful to archaeologists because human behavior can generate features that are detectable magnetically. The following discussion will focus on the scientific principles underlying magnetometry, as well as operational details and sources of error relevant to the Bartington GRAD-601 fluxgate magnetic gradiometer, the vector magnetometry machine used to collect the data analyzed in this work. A complete review of magnetics is beyond the scope of this work, but further technical details are available in Scollar (1990), Aspinall (2008), and Clark (1990), among others.

The Physics of Magnetism

Magnetism can be described as "the force exerted between magnetic poles" (Scollar et al. 1990: 376). Magnetic fields are generated by the movement of charged particles, more specifically, the orbital motion and spin of electrons (Scollar et al. 1990: 378). Charged particles are omnipresent, and therefore most types of matter are consistently sensitive to, interacting with, and/or generating magnetic fields. The movement of metallic elements within the earth's outer core and the core-mantle boundary generates an ever-present magnetizing field of approximately 70 to 20 microtesla (μT; Aspinall et al. 2008: 5; Clark 1990: 64). This field exhibits time-dependent diurnal variations as well as spatial variations due to varying concentrations of ferrimagnetic material in the earth's crust (Aspinall et al. 2008: 18-19). The shape and strength of anomalies in geophysical data, produced by archaeological features, are influenced by this magnetizing field (Aspinall et al. 2008: 19). Magnetic flux density (a property related to, but not equivalent to, magnetic field strength) of archaeological features is often measured in SI subunits of nanotesla (nT), and one nT is equivalent to 10^{-9} tesla (T; Clark 1990: 64). One μT is equivalent to 10^{-6} T, to give an idea of the scale of the magnetic strength of archaeological features compared to the ambient magnetic field of earth (Aspinall et al. 2008: 5).

Magnetometry is a kind of near-surface geophysical prospection which measures "small changes in the earth's magnetic field". Composition of materials in the top few meters of soil determines their retained magnetism; ferrous (iron-containing) objects (commonly encountered on post-contact archaeological sites in North America) will produce "positive" or "positive and negative dipole" readings, while "negative" readings can be obtained from less magnetically susceptible materials (Conyers 2017:27). In this way, "local variations" of the strength of Earth's magnetic field are mapped, and unlike GPR and ER, which use induced or "artificial" fields to measure properties of the ground, magnetometry is a "passive method" and employs only the existing magnetic field of earth (Kvamme 2006: 206).

Some materials, such as iron, exhibits magnetism in the absence of a magnetic field, while most others do not; at high temperatures (the Curie point, approximately 500-600 degrees Celsius but specific to each material), the magnetic domains of soils, clays, and rocks will become aligned parallel with the earth's magnetic field at that moment (Conyers 2017: 31; Kvamme 2006: 207). This alignment appreciates into a permanent remanent magnetic effect in these materials, as the magnetic domains of the component particles are no longer randomly oriented, thereby cancelling out (Kvamme 2006: 207). This magnetic effect is known as thermoremanent magnetism and can be detected by magnetometers, although it is worth noting that some degree of thermoremanence is produced at any elevated temperature, and can occur naturally as well – for example, during the formation of igneous rocks (Kvamme 2006: 208). Thermoremanence of kilns, clay ovens, thermal cooking pits, fired bricks, hearth fires, or razing will leave a magnetic signature (Aspinall et al. 2008: 21).

Due to accumulation of iron minerals and oxidation reactions in waterlogged topsoil, the presence of magneto-tactic bacteria, and human activity, topsoil is generally more magnetically susceptible than subsoil (Kvamme 2006: 208); this property of soils is most useful to a prospector seeking to use magnetometry to identify grave shafts. Shafts lined with bricks, which would carry thermoremanent magnetism, or marked with ferrous materials could also be detectable. Deposits of organic waste exhibit increased magnetic susceptibility relative to surrounding soils (Aspinall et al. 2008: 25). Certain rocks, such as limestone, are less magnetic than surrounding soils and may be visible as a negative signature in data (Aspinall et al. 2008: 28). In this way, both the contrasts between values and the objective values of magnetic field strengths (measured in nanoteslas) can be meaningful to interpretation (Conyers 2017: 28-30).

As it is a three-dimensional phenomenon, the earth's magnetic field can be broken down into X-, Y-, and Z-components. Readings of magnetic flux density can be taken by scalar or vector machines. Scalar magnetometers measure total magnetic field strength (a value representative of all directional components of the magnetic field, combined into a single value), while vector magnetometers measure one directional component of the field (Aspinall et al. 2008: 29). Single-sensor magnetic field readings vary due to solar wind, charged particles ejected from the sun which enter earth's atmosphere at high speeds, and geological variations, but this data variation can be reduced by utilizing two sensors. Dual sensor magnetometers (such as gradiometers) can filter out variation in field readings due to underlying geology and earth's non-static magnetic field, and allow for detection of more subtle and localized variations generated by archaeological features (Aspinall et al. 2008: 31). Gradiometers consist of two sensors which are mobile, but separated from each other by a fixed distance (typically less than 1 meter). This relatively short distance ensures both sensors are affected approximately and functionally equivalently by deep geology and earth's magnetic field (Aspinall et al. 2008: 33).

Data Collection in Magnetometry

Several different types of magnetometers are on the market and have utility for archaeological work, such as SQUID, alkali-vapor, electron spin resistance, proton free-precession magnetometers, and fluxgate magnetometers. As the latter type is the instrument utilized in this work, the discussion will focus on this machine. Fluxgate magnetometers were developed to detect unexploded ordinance during World War 2, and utilize several wires coiled around ferromagnetic toroid cores. Fluxgate magnetometers are designed to allow the cores to exist within what is approximately a zero field, only detecting field variations due to an external magnetic field and reducing the magnetic effects of the components of the magnetometer itself (Aspinall et al. 2008: 36). In the case of a magnetometer operating on earth within an archaeological site, this external magnetic field consists of the overlaid values of the ambient magnetic field of earth, geological variations, and variations due to archaeological features. Utilizing fluxgate magnetometers in a gradiometer configuration, the former two phenomena can be filtered out, allowing for the detection of near-earth anomalies. Additionally, the gradiometer configuration reduces error generated by tilting a vector magnetometer from the direction of its axis of measurement, something which is likely to occur during data collection due to changes in elevation, terrain, vegetation, and operator gait (Aspinall et al. 2008: 39).

Fluxgate magnetometers measure a voltage (V_{sec}) induced by an external magnetic field (H_{ex}) which produces a magnetic flux (BA) in a core of known surface area (A) as the permeability (or how much magnetic flux a material can support; μ) of the core alternates with changing flux, creating a voltage such that :

$$V_{sec} = nA\frac{dB}{dt} \tag{1}$$

Where *n* signifies the known number of turns of secondary wiring within the gradiometer, *t* is recorded time, and *dB/dt* is the first derivative of magnetic flux over time (Scollar et al. 1990: 456). So while the fluxgate magnetometer directly measures V_{sec}, this voltage can be related to magnetic flux density *B* at that same point in space in which that voltage is generated using the equation above. It is matrices of this value (in SI units of T; subunits nT) which are exported as data points from the magnetometer, processed, interpreted, and graphically presented in geophysical literature and reporting.

Magnetometry survey is time-efficient, as no contact with the ground is needed; vegetation is less of an issue for a magnetometer (Conyers 2017: 28) provided the prospector can still transport the magnetometer at the appropriate orientation and speed for data collection. Vector magnetometers, such as fluxgates, are typically configured to measure the vertical (or z) field component (Aspinall et al. 2008: 62). The machine is carried by an operator on transects, often in ‘zig-zags’. The operator begins at one end of the grid, travels down one transect along the y-axis to the end of the transect, moves some distance along the X-axis, and returns to that same end of the grid on the next transect (the zig-zag). As the operator moves, data is collected on a specific frequency and along lines spaced by chosen X-axis intervals. This process continues until the data grid has been surveyed. The frequency of data collection and interval width determine the data density of the collected data set along the X- and Y-axes.

Data Processing in Magnetometry

Once data is collected, it must be processed and interpreted. Plots and other graphical representations of magnetic data do not directly correspond to the shapes of the archaeological features which generate them (Aspinall et al. 2008: 64). The width of an anomaly at half its maximum reading has been said to roughly equate to the width or depth of the buried feature, whichever is greater (Clark 1990: 83). Processed magnetometer data is typically presented as a greyscale image, with most negative values shown as pure white and most positive values shown as pure black, and all other values somewhere in-between. Magnetically reactive materials contain a positive and negative pole – like a theoretical bar magnet. Strongly magnetic anomalies with 'dipolar', or more accurately bipolar, characteristics will generate areas of strong negative values which flip to strong positive values (Aspinall et al. 2008: 58-60).

Other anomalies may present as 'monopolar', although no theoretical monopole actually exists. Geophysical literature frequently references 'monopolar' anomalies, which may appear to exhibit only negative or positive values, but such anomalies are likely reflective of anomalies whose other pole is oriented further from the magnetometer and is too far, and/or too weak, to be detectable within the limits of the instrument used (Aspinall et al. 2008: 71). The strength of magnetic flux density of a dipole (approximated as an iron bar magnet) decreases exponentially (by a power of three) relative to distance – values differ in real-life scenarios, but distance to detector and magnetic flux density do share a relationship (Aspinall et al. 2008: 60). Anomaly shape can also be used to estimate depth (Aspinall et al. 2008: 72) although this is not considered an exact science.

Data Quality and Resolution

The resolution of a particular instrument can be defined as the ability to separate two closely spaced anomalies. Due to their data filtering abilities, gradiometer configurations are inherently of higher resolution than a comparable single sensor system (Aspinall et al. 2008: 76). Noise can be defined as all signals other than the signal generated by an archaeological feature (Aspinall et al. 2008: 77). Both random (instrument, cultural, and operator) and coherent (drift due to instrument, diurnal, soil, cultural, geologic, and operational) noise may affect magnetometer data (Aspinall et al. 2008: 80-82). Field procedures and processing can identify, ameliorate, and/or remove some but not all noise, and level of acceptable noise will vary site-by-site (Aspinall et al. 2008: 82). Additionally, signals tend to weaken and broaden with depth, and clarity decreases rapidly at depth beyond one meter (Clark 1990: 78).

Although it is common to increase the frequency of data collection to an interval much smaller than the transect interval, this only increases data resolution in one direction – the axis of travel (Clark 1990: 76). Increased frequency of data collection is preferred from a cost standpoint, as it does not increase survey time; halving the transect interval may nearly double the time for a field survey. A recording interval (or frequency) of two data points/meter is the smallest recommended frequency for detailed recording, although increasing to four data points/meter supplies "further gain" and eight data points per meter produced only "marginal improvement" in a survey of Roman pottery kilns in an area containing iron fragments (Clark 1990: 81).

While data points are often interpolated during processing of magnetometry data (often in the transect interval dimension) to improve image quality, these are not true data points and are calculated as most likely values based on the values which surround them (Aspinall et al. 2008:

133-134). All magnetic surveys involve sampling, and despite what processed images may suggest, coverage of an area is not smooth or complete (Aspinall et al. 2008: 110). To obtain one reading across an archaeological anomaly (identify presence/absence), the transect spacing must theoretically be the width of the expected feature size. Additional data are necessary to refine anomaly shape (Aspinall et al. 2008: 111). Put another way, the presence or absence cannot be determined for a feature which is smaller than the size of the transect interval. On the other hand, Aspinall et al. (2008: 112-113) argue that too much data density can introduce spurious soil noise which may be obstructive to larger archaeological features. While resolution is inherent to a particular machine, survey methods and data density should be considered on a site-by-site basis.

Summary of Theoretical Background

This chapter has examined the theory underpinning the operation of geophysical prospecting instruments, specifically those which relate to magnetism and data resolution in magnetometry and with focus on the specific instrument used to collect the data analyzed in this thesis: the Bartington GRAD-601 dual fluxgate gradiometer. Human behavior is among the factors which can introduce magnetic variation into the environment, and these variations in magnetic flux density, or in magnetic gradient values (if using a gradiometer) can be due to the presence of archaeological features. Therefore, detecting these magnetic variations can lead to the identification of archaeological resources.

Chapter 3: Historical Background

Archaeological Geophysics as 'Craft'

Moving beyond the physical science of magnetism, the practical application of magnetometry in archaeology is guided by accepted practice and institutional standards. In this chapter, this body of literature is reviewed with a focus on the determination of acceptable survey resolutions in magnetometry, with particular attention to transect interval width. The experimental foundations of this research are considered through the lens of modern CRM practice and the application of geophysical techniques in archaeology as a 'craft.'

The New Archaeology of the 1960s, processualism, prioritized the development of a positivistic archaeology, in which analysis of "adequate data" using "proper scientific methods" results in objective conclusions (Trigger 2006: 17). This approach generally acts in pursuit of unification, or standardization, of science (Trigger 2006: 29). Specifically, within archaeology, this approach led to the development of an industrialized approach emulating a "factory model of production" utilizing teams of specialists, managed by a single principal to "maximize efficiency and standardiz[e] the product", which is archaeological knowledge. Standardization can also be a response to negative results and perceived lack of quality control (Shanks and McGuire 1996: 85).

In contrast to this empirical standardization, Shanks and Hodder (1995: 20) argue that objectivity itself is a construct which is crafted. Subversion of this industrial standardization allows for the unification of the discipline without homogenization, under the banner of 'craft' (Shanks and McGuire 1996). Archaeology as craft acknowledges "subjectivity," allows for the transfer of knowledge within the “apprenticeship and mastery” system and through practitioners’ “experience” (Shanks and McGuire 1996: 82, 86).

A craft archaeology allows for the passing of “field techniques from one excavator to the next”, relying on the apprenticeship model to transmit knowledge between academic generations (Caraher 2019: 378). The alternative is standardized, professional qualification, and no state specifically regulates the qualifications of practitioners of archaeological geophysics, although the California Department of Transportation (CALTRANS) made overtures at regulating practitioners of geophysics within archaeology (CALTRANS 2014: 6). Archaeology as craft does not mean eliminating hierarchy and abandoning scholarly goals, but it does acknowledge and critique power inequities inherent to the organization of archaeological projects (Walker and Saitta 2002: 206). Critical archaeology and scientific archaeology can coexist, while “government-mandated” practice can lead to disenfranchisement (Atalay et al. 2014: 9, 13).

Shanks and McGuire (1996: 84-85) characterize CRM archaeology as entrenched in the “factory model,” embodied by the standardization of the discipline and accompanied by a focus on efficiency and profit. This standardization is literal, and written into the state, federal, and agency standards which direct the way archaeology is conducted in CRM. However, if lack of regulation signals craft, then the application of geophysical techniques in CRM can be viewed as craft itself.

State, Federal, and International Standards for Magnetometry

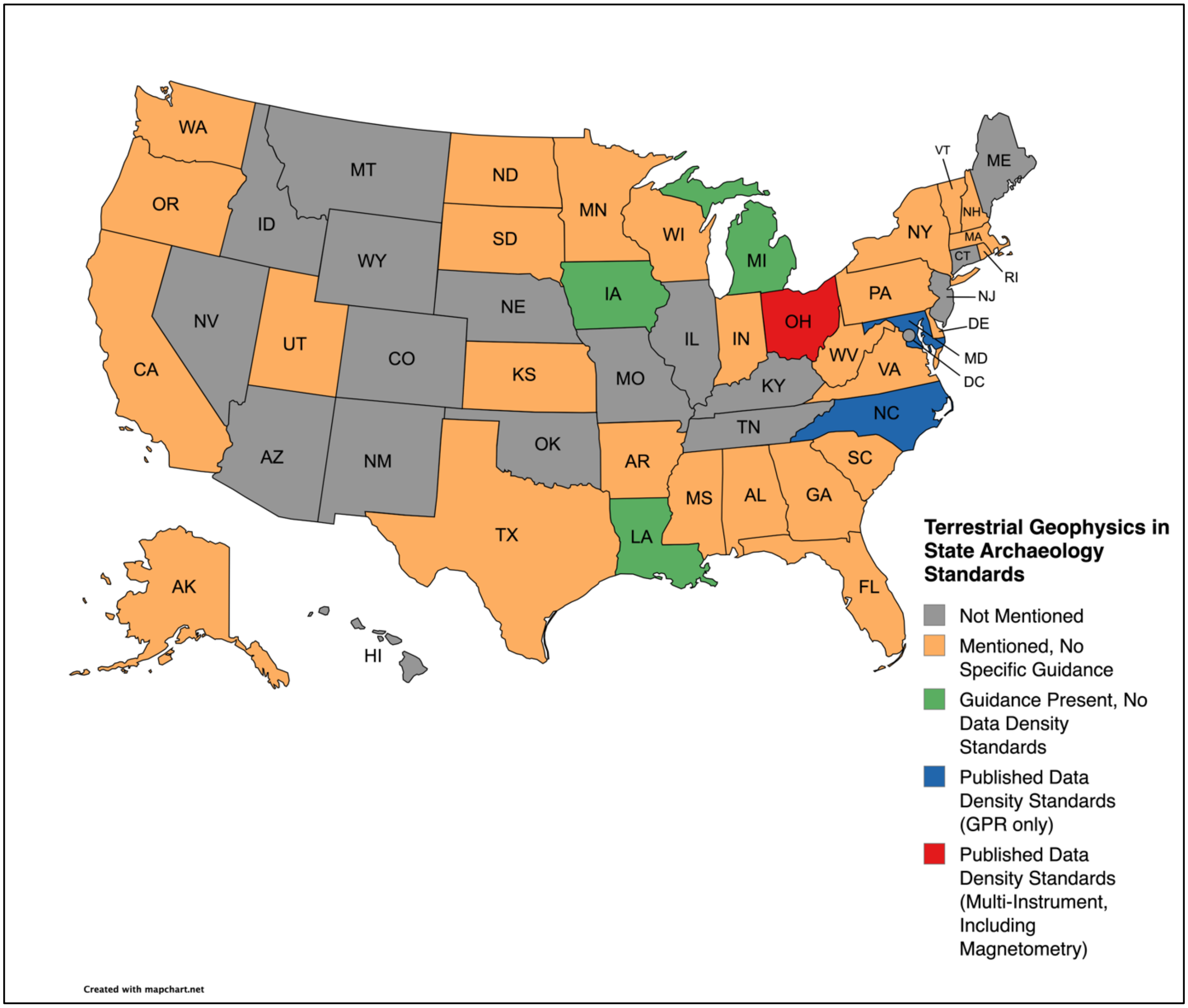


Figure 1. Coverage of Terrestrial Geophysics in State Archaeology Standards. Figure credit to author.

State standards, published by SHPOs and/or codified in state regulations, guide archaeological fieldwork in CRM, including fieldwork with a geophysical component. However, state standards which specifically regulate the way geophysical fieldwork is performed are largely absent.

State archaeological standards of 17 states[1] and the District of Columbia (DCSHPO 1998) do not mention geophysical fieldwork or remote sensing at all (Figure 1). An additional 27 states[2] do address geophysics and/or remote sensing in state archaeological guidelines, but do not offer any specific requirements or best practices pertinent to data density (Figure 1). Only six states, including Ohio – the state in which the data analyzed in this thesis was collected – address geophysical standards in greater detail, as seen in Figure 1 below.

Of these six state SHPOs, only Ohio has published data density standards relevant to magnetometry. According to the published Ohio History Connection (OHC, also known as the Ohio State Historic Preservation Office) archaeology guidelines, the minimal accepted standards for utilizing a magnetometer for archaeological work in Ohio require a data density of 8-10 readings per meter with a 0.50-meter transect spacing (OHC 2022:11; OHC 2023:17). A set of state geophysical guidelines, supplementary to and referenced by the 2023 OHC publication, state that "magnetometry is the best instrument for detecting [precontact] pits in Ohio, assuming data are collected at a rate of at least 8 samples per meter along transects spaced 50 cm apart" (OHC 2022: 11). This data density is also recommended for shaft type features (OHC 2022: 37).

Two states (North Carolina and Maryland) address data density, specifically transect width, but only for GPR data collection. Maryland recommends GPR data be collected along

[1] Alaska (AKDNR 2006), Arizona (AZSHPO 2001), Colorado (HC 2024), Connecticut (Poirier 1987), Hawaii (HISHP 2002a; 2002b), Idaho (ISHS 2026), Illinois (ILSHPO 2025), Kentucky (KHC 2017), Maine (MHPC 2019), Missouri (MSP n.d.), Montana (MHS 2024), Nebraska (NESHPO 2017), Nevada (NVSHPO 2019), New Jersey (NJHPO n.d.), New Mexico (NM 2006), Oklahoma (OHS 2023), Tennessee (TDEC 2024), Wyoming (WYSHPO 2026)

[2] Indiana (DHPA 2022), Alabama (AHC n.d.), Arkansas (AHPP 2010), California (SCA 2020), Delaware (DDHCA 2015), Florida (FDHR 2003), Georgia (GCPA 2019), Kansas (KSHS 2018), Massachusetts (MHC 2004), Minnesota (Anfinson 2005), Minnesota (MN OSA 2011), Mississippi (MDAH 2019), New Hampshire (NHDHR 2025), New York (NYAC 1994), North Dakota (SHSND 2026), Oregon (ORSHPO 2025), Pennsylvania (PASHPO 2021), Rhode Island (RIHPHC 2021), South Carolina (CSCPA 2005), South Dakota (SDSHS 2023), Texas (Borgens et al. 2024), Utah (UTSHPO 2024), Vermont (Robinson et a. 2017), Virginia (VDHR 2017), Washington (WADAHP 2018), West Virginia (Trader 2021), Wisconsin (Dudzik et al. 2024)

transects spaced 0.25 meter apart, while North Carolina accepts transect intervals less than or equal to 0.50 meter (MHT 2025: 22; NCOSA 2023: 13). Neither state provides specific recommendations for the implementation of other geophysical instruments.

Michigan has not published specific data density standards for geophysical data collection, but does reference the standards published by Ohio, the United States Army Corps of Engineers (USACE), and the Department of Defense in its SHPO manual (DoD; MISHPO 2023: 2). For reference, USACE guidelines define 1.0-meter transect intervals as the minimum acceptable standard for geophysical survey in general, while acknowledging that reduced intervals of 0.50 meter may be necessary to avoid "miss[ing] relatively small and/or low-contrast features” (Somers et al. 2003: 95). DoD standards state that a transect interval separation of 0.50 meter is “common” and acceptable for detection of “many” precontact and historic features magnetically (Ernenwein and Hargrave 2009: 12, 30). Also in the category of general recommendations, Louisiana calls for “appropriate” transects, but does not quantify this; Iowa state standards for archaeology acknowledge that wider intervals can increase “margin for error” but do not specify an acceptable such margin (AIA 2023: 53; DAHP 2021:12).

Looking internationally, the European Archaeological Council (EAC) has more stringent requirements for magnetometry data collection, depending on a pre-defined level of effort. Referencing earlier guidelines published by English Heritage, the EAC emphasizes the context-dependent nature of data density standards. For “detailed” magnetometer survey, a transect interval spacing of 0.25 meter is ideal, given Clark’s (1990) rule that the transect interval should not be wider than the feature is buried beneath the soil (approximately 0.30 meter for plowzone-capped features). If the goal is merely detection (called “prospection” by the EAC), coarser transect separation of 1.0 meter is acceptable, while 0.50-meter transect separation may represent "a good compromise for most investigations" (Schmidt et al. 2013: 64).

<u>Defining Survey and Instrument Resolution</u>

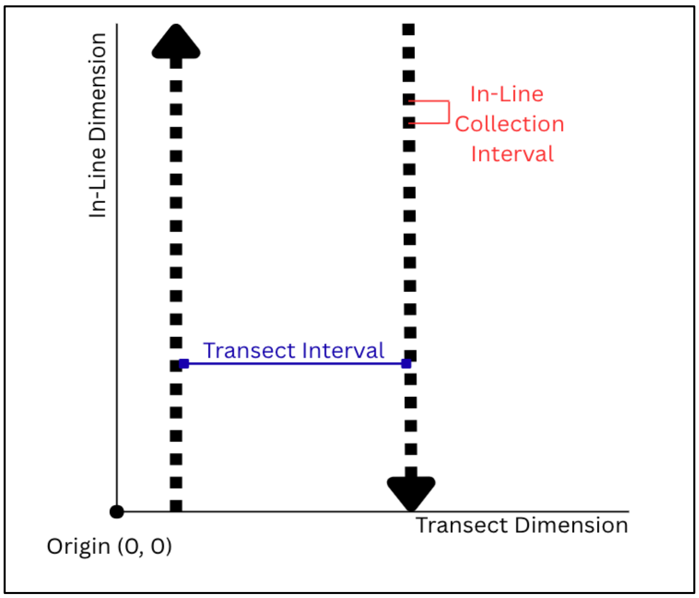


Figure 2. Dimensions of Sampling in Magnetometry. Figure credit to author.

Sampling, or data collection, in magnetometry exists in two physical directions: the direction of travel along each transect or traverse (in-line data), and the distance which separates each traverse from the other traverses (transect interval). Together, these parameters define the survey resolution, summarized in Figure 2 above. The in-line sampling interval does not typically impact the survey's spatial resolution (Aspinall et al. 2008: 110). In contrast, the transect interval can only be reduced by adding sensors or walking additional traverses, both of which have cost impacts. Because of this, transect interval defines the "perceived spatial resolution" of a survey. Feature size, feature depth, and risk of over-sampling should be considered when determining the optimum survey density (Aspinall et al. 2008: 111).

Some argue that more data is not always better. Soil noise may become apparent in data collected with a “high sensitivity” magnetometer at more dense intervals, defined by Aspinall et al. (2008: 212) as data collected on a 0.20-meter by 0.20-meter resolution. Aspinall et al. (2008: 112-113) present two arguments regarding increased data density: that it may more easily allow practitioners to distinguish between soil and instrument noise and features of archaeological interest; and that the spurious soil noise may make interpretation of “major archaeological trends” more difficult and therefore “over-sampling” should be avoided. OHC subscribes to the former of the two arguments regarding data density and survey resolution presented by Aspinall et al. (2008: 112-113): that “collecting too much data is never a bad thing” aside from scheduling and budgeting concerns (OHC 2022: 37). Inappropriate sampling density may lead to misinterpretation of features, or they may escape detection entirely (Schmidt and Marshall 1997: 343).

The OHC guidelines consider the question of data density to be the second-most important geophysical survey parameter, next to choice of appropriate instrument (OHC 2022: 37). The two are intertwined to a degree. Just as resolution is inherent to a machine, choice of instrument places a practical and/or literal upper limit on data density. For example, the Bartington Grad-601 dual fluxgate magnetic gradiometer is equipped with a Data Logger with the capacity to record in-line densities on multiples of two, between 1 and 8 readings per meter along each transect line. This parameter is limited to 4 readings/meter on larger [40 square meter] grids. The Data Logger can store up to 125,000 readings and can be operated at a sensitivity of 0.01 nT or 0.1 nT (Bartington Instruments Ltd. n.d.: 5). Therefore, the Data Logger for the Bartington GRAD-601 is capable of transect intervals of 0.25 meter, 0.50 meter, or 1.0 meter. An in-line data density of 8 readings per meter effectively collects data every 0.125 meter, or 12.5 centimeters, assuming the operator is traveling at a steady speed along the entire transect, something that is difficult to do in the field.

Though coarser than European recommendations, of all American states and agencies surveyed here, Ohio has by far the most robust archaeological standards relevant to data density in geophysics and magnetometry, specifically. But how were the parameters contained in the OHC guidelines developed?

Defining A Successful Search

All magnetometry surveys involve strategic sampling (Aspinall et al. 2008: 110). The sampling strategy is designed to increase the likelihood of detecting a feature (or, a successful search), which has been described statistically in general terms by Kyrala (1964). The probability (*P*) of a successful search depends on two separate probabilities: 1) the probability (*p*) that the object of the search exists within the search area *A* and 2) the conditional probability (*F*) that searching that area would result in the discovery of the object (Kyrala 1964: 105). This can be written as:

$$P = pF \tag{2}$$

or,

$$P = \sum_{k} p_k F_k \tag{3}$$

where subscript *k* indicates the probability of the object existing in the k^{th} region of smaller search region A_k (p_k) and the probability of finding the search object (F_k) taken as a sum for all *k*.

The probability F depends on sampling strategy, while the probability p is independent of field methods. Kyrala (1964: 105) explains that p (or p_k) must be estimated or determined by previous survey in order to quantify these relationships. For example, the probability of an archaeological feature existing in a particular search area could be estimated based on local site density data, previous research, or using probability models. While probability modeling is beyond the scope of this work, defining a successful search in mathematical terms guides a thoughtful and evidence-based determination of survey resolution.

Later scholars capitalized on this approach. Schmidt and Marshall (1997: 334) redefined Kyrala's (1964) probability of a successful search as:

$$P_{hit} = \frac{\pi w_z^2}{4a^2} \quad (4)$$

where w_z (zero-width – the distance from the center of a Schmidt and Marshall's cubic-meter test anomaly at which the anomaly is first detected) is equal to twice the size of the feature and a is the spacing of the survey grid. By this definition, to ensure a successful search (allowing P_{hit} = 1), for an anomaly of size 1 meter (allowing w_z,= 2) the grid spacing a (or transect interval and in-line sampling rate) must be smaller than 1.8 meters (Schmidt and Marshall 1997: 344). Schmidt and Marshall used Kyrala's (1964) paper and a series of test transects at various intervals over a theoretical ferrous, cubic anomaly with sides 1 meter long to determine this grid spacing and define zero width – but this generalization does not approach the complexity of a real-world scenario (Schmidt and Marshall 1997: 334).

To determine an adequate survey resolution, Schmidt and Marshall (1997: 344-345) also examined the "sampling theorem." This generalized quantitative approach leverages the presentation of magnetic in-line data as traces, with some depiction of magnetic flux along the

Y-axis and physical position on the transect (or in space, relative to the center of the anomaly, etc.) along the X-axis. To effectively sample a frequency, the sampling rate must not exceed half of the smallest wavelength λ_{min} – also known as the Nyquist wavelength ($\lambda_N = 0.5\lambda_{min}$). If the magnetic anomaly is graphed as a trace and the goal of the sampling strategy is to detect the "peak" value of the anomaly (and the trace is assumed to approximate a half-sine wave), collecting data points at λ_N will ensure detection of anomalies of this 'frequency.' Therefore, the minimum sampling rate would need to be equivalent to the size of the object in order to meet the definition of a successful search (Schmidt and Marshall 1997: 344-345).

However, in both of these situations, the feature would only be detected in one data point at P = 1. Additional readings would be required to further define the shape or extent of the anomaly and feature, which are not necessarily or often equivalent. Following this approach, practitioners acknowledge that transect intervals spaced one meter apart do not collect enough data to detect pit features smaller than one meter in diameter, though it is expected that "smaller features can be detected at this data density if their fill includes burned earth" (Burks and Cook 2011: 672). In other words, when a feature contains burned earth, it is expected that the anomalous readings will extend beyond the physical boundaries of the feature and allow for detection at wider intervals.

Recommended Transect Intervals

Aspinall et al. (2008:112) suggest 0.50 meter by 0.50 meter as a sufficient survey resolution, if aiming to exceed neither the size nor depth of features at that measurement. In making this recommendation, Aspinall et al. (2008) follow the work by Schmidt and Marshall

(1997), who found that a magnetic gradiometer survey collecting data on a 0.50-meter by 0.50-meter grid was more effective at identifying Iron Age archaeological features than a grid of 0.25 meter by 1.0 meter. Specifically, Schmidt and Marshall (1997: 343) sought to evaluate guidelines published by English Heritage concerning best practice for magnetometry surveys. At the time, these guidelines held that readings should be recorded at a density of not less than 1.00 meter by 0.50 meter (David 1995:18). By 2008, English Heritage suggested that a survey resolution of 0.25 meter by 0.50 meter was necessary for characterization of an area, while a more relaxed resolution of 0.25 meter by 1.0 meter was permissible for evaluation (English Heritage 2008: 8).

Schmidt and Marshall (1997: 345-346) surveyed an Iron Age enclosed farmstead at resolutions of 0.50 meter by 0.50 meter and compared these results with data extracted at a wider interval of 1.0 meter by 1.0 meter. At the "coarser" interval, "certain features are missing" and the overall image was less crisp. The authors also isotropically (in-line and transect intervals are equivalent) surveyed an Iron Age ditch enclosure at 0.25-meter by 0.25-meter, 0.50-meter by 0.50-meter, and 1.0-meter by 1.0-meter resolutions, and extracted a dataset representative of a non-isotropic (asymmetrical) 0.25-meter by 1.00-meter resolution from lower resolution datasets, using linear interpolation to fill in the wider transect interval (Schmidt and Marshall 1997: 346).

Ultimately, Schmidt and Marshall's study found that the highest resolution revealed features unseen in lower resolution datasets, including "linear feature connecting pits" not visible in the dataset with the identical in-line sampling rate but wider transect interval. In fact, the non-isotropic dataset most closely resembled the dataset with the least resolution with a transect interval and in-line data sampling rate of 1.0 meter (Schmidt and Marshall 1997: 346). This highlights the importance of transect interval on overall resolution, and illustrates that an insufficient transect interval cannot be compensated for by denser in-line sampling. Despite the above conclusion and the clarity of their 0.25-meter interval data, the authors stopped short of recommending transect intervals smaller than 0.50 meter. The authors did celebrate recent successful deployments of arrays of multiple gradiometers spaced 0.20 meter apart, arrays which made available denser datasets “still recorded with a good survey rate” (Schmidt and Marshall 1997: 348). This implies the authors’ preference for transect intervals smaller than 0.50 meter, were reduced survey speed not a factor.

Chapter 4: Methods

Introduction to Field and Analytical Methods

This section describes the field efforts undertaken to survey a portion of the test site with a magnetometer and the subsequent ground-truthing methods which confirmed the presence of two archaeological features. In order to more fully understand the effect of resolution on the detection, locations, and appearances of the anomalies generated by these features and detected by the magnetometer in the original survey, additional datasets of lower-resolution are constructed. A framework is presented that tests these datasets with the original, collected dataset in order to examine the consequences of decreased resolution, and artificially increased resolution, on the theoretical rediscovery of the two identified archaeological features.

Magnetometry Data Collection

Artifact distribution data from a previous Phase I archaeological pedestrian survey of the test site was analyzed to identify areas of highest surface artifact concentration. Using this information, a grid layout with data points every 10 meters (32.8 feet) was superimposed over the site density map in the Emlid Flow land survey app, allowing for flexible in-field placement of suitable grid corners based on both microtopography and artifact density. Once established in the field, grid corners were recorded using an Emlid Reach RS3 GNSS receiver, which provided real-time kinematic (RTK) corrections via the Ohio Continuously Operating Reference Station (CORS) satellite-tracking network. Positioning accuracy with this setup is typically within three centimeters. These grid corners were used to georeference collected magnetometry data and guide subsequent subsurface testing.

Grid corners were marked in-field with wooden stakes, and fiberglass pull tapes were used to delineate every other meter along the northern and southern grid edges. These points were marked with plastic stakes, which anchored a single fiberglass tape (metal components removed) with in-line collection intervals marked with electrical tape. This marked-interval tape was used to define each transect interval and for operator reference to steady pace of in-line data collection. Magnetometry data was collected at the test site using a Bartington GRAD-601 dual fluxgate gradiometer. The gradiometer was configured to support an in-line data density of 8 data points per meter (or every 0.125 meter) along the Y-axis on transects spaced 0.25 meter apart along the X-axis, with a sensitivity of 0.01 nT. The gradiometer has no vertical control but the depth can be assumed to extend approximately one to two meters below the ground surface; therefore the captured reading is a summation of the magnetic flux densities within this range of depths.

Gradiometer data at the test site was collected in two square grids (Grid 1 and 2) measuring 30 meters by 30 meters. These grids were placed to overlap with the site landform, a slight rise in rolling moraine, as well as denser portions of the previously recorded surface scatter in order to increase the likelihood of detecting subsurface features. Following data collection and processing, both grids were georeferenced, and promising anomalies marked with points in the Emlid Flow land survey map. In the field, an Emlid Reach RS3 GNSS Receiver was used to relocate these areas of anomalous magnetic signals for ground-truthing. Shovel test units (STUs) measuring 0.50 meter by 0.50 meter were centered on the visually-identified, approximate centroid of the visually-identified magnetic anomaly, and expanded as necessary to follow any discovered irregularities in the soil profile – determined by previous survey of the test site and corroborated by published soil surveys – which may be indicative of an archaeological feature. Two out of seven tested magnetic anomalies (referred to here as Magnetic Anomalies [MA] 1 and 2) proved to be archaeological features, and both (Features 1 and 2) are used as test cases to evaluate various actual transect intervals below.

Magnetometry Data Processing

The raw, collected gradiometer data (Dataset 1) of (in-line, transect interval) resolution (0.125, 0.25) were processed using TerraSurveyor Program Version 3.0.35, developed by DW Consulting Geophysical Data Services. Processing of real data for actual interpretation was limited to use of the DeStripe and Clip functions. Additional processing was used to create polished graphics to aid in discussion and demonstrate the obfuscating capabilities of interpolative functions in TerraSurveyor. Processing steps for production of polished graphics include the DeStripe, Low Pass Uniform (Median) Filter, Interpolation (to isotropy), and Clip functions. The Interpolation function was also used in the development of isotropic test data sets to compare to unclipped Dataset 1.

The DeStripe function orients each grid or traverse around the mean/median/or mode, as specified, by subtracting the mean/median/or mode of each grid or traverse from each data point. Threshold values can be set to limit the dataset if in 'mean' mode, to mitigate effects of extreme magnetic gradient values (generated by ferrous metal, for example) on each grid or traverse. DeStripe typically produces bipolar data oriented around zero (DW Consulting 2019: 73). This serves to remove the "striping effect within a grid" which can be created by directional effects, operator habit, drift, or sensor differences (DW Consulting 2019: 10).

The Low Pass Filter removes high frequency components and has a 'smoothing' effect that can be used to generate cleaner images for clients or final reports. Using the Uniform (Median) weighting, the function calculates the median of all values within a specified window in the X- and Y-dimensions. The value at the center of each window is replaced with the Median (DW Consulting 2019: 93). This function was used with a window size of 3 in both dimensions. This function was only applied on data following interpretation, for the production of cleaner graphics, as necessary.

Interpolation is used to increase or decrease the resolution of the data by generating new datapoints on a specified interval by axis using a cubic spine polynomial – a mathematical means of calculating new datapoints in an existing series by determining a curve on which existing datapoints lie and pulling new values on the specified interval from this curve. Excessive interpolation can create artifacts, and cannot not be used to actually enhance data, only resolution – "any improvement is artificial" (DW Consulting 2019: 75). Interpolation must take place after other processes, which depend on the original data for calculations. The Interpolation function cannot be used to physically reduce the resolution of a dataset while leaving a certain desired interval data intact; it will always calculate new data points, on an increased or decreased resolution, using a cubic spline algorithm. TerraSurveyor places each datapoint in the center of the 'pixel' it samples, so reducing or increasing resolution also shifts the physical location of the datapoint relative to the grid corners and zero-axes (DW Consulting 2019: 25).

The Clip function replaces all values outside a specified minimum and maximum values with those values. Acceptable minimum and maximum values can be defined by inputting an absolute value or by specifying a range defined by increments of standard deviations calculated from the dataset. The Clip function removes extreme values, which can create greater contrast between less-extreme values and highlight anomalies of archaeological interest (DW Consulting 2019: 72).

The Clip function was also applied after other processing steps, as the removal of extreme data points is predominantly used to physically highlight more subtle variations in the data to aid in the human perception of the graphic itself. This function replaces values outside the retained range with the highest or lowest permitted value, introducing artificially flattened data into the dataset which may affect functions (Interpolate, Low pass Filter) which depend on original data for calculations. All non-visual analyses were conducted on unclipped datasets, to avoid the inadvertent reduction of variation within datasets. Dataset 1 was quiet prior to processing (values approximately within +/- 25 nT), with most values far from the extreme edges of this distribution.

Generation of Test Dataset Series

The collected gradiometer data of resolution (0.125, 0.25) were first subjected to initial, minimal processing (DeStripe and Clip to 3 SDs) in order to guide the ground-truthing phase of the geophysical investigation. This dataset and results of ground-truthing provide a baseline understanding of the relationship between geophysical data and actual archaeological and environmental features at the test site. This thesis examines the effects of altered transect-interval widths on precontact, thermoremanent archaeological feature detection and location in magnetometry – allowing for the evaluation of the current OHC recommendations of 0.50-meter transect intervals for magnetometry against simulated comparative datasets of a higher and lower resolutions.

Surveying the same area multiple times at different interval widths each time is one way to address this question, but this requires additional time and effort. To streamline the process – and reduce the introduction of additional error due to instrument drift over time – these comparative reduced-resolution datasets can be created by splitting apart one real dataset of the highest (0.125, 0.25) resolution in the study (Figure 3). As such, the two grids of collected gradiometer data were re-processed minimally (subjected only to DeStripe, to reduce sensor-related errors) and then exported out of TerraSurveyor as an ASCII .csv file. This file (Dataset 1) contains 57,600 z-value (magnetic gradient) datapoints across 480 columns (corresponding to the 480 datapoints collected at 0.125-meter intervals along two 30-meter transects) and 120 rows (corresponding to the number of 0.25-meter transects which fit in a 30-meter grid).

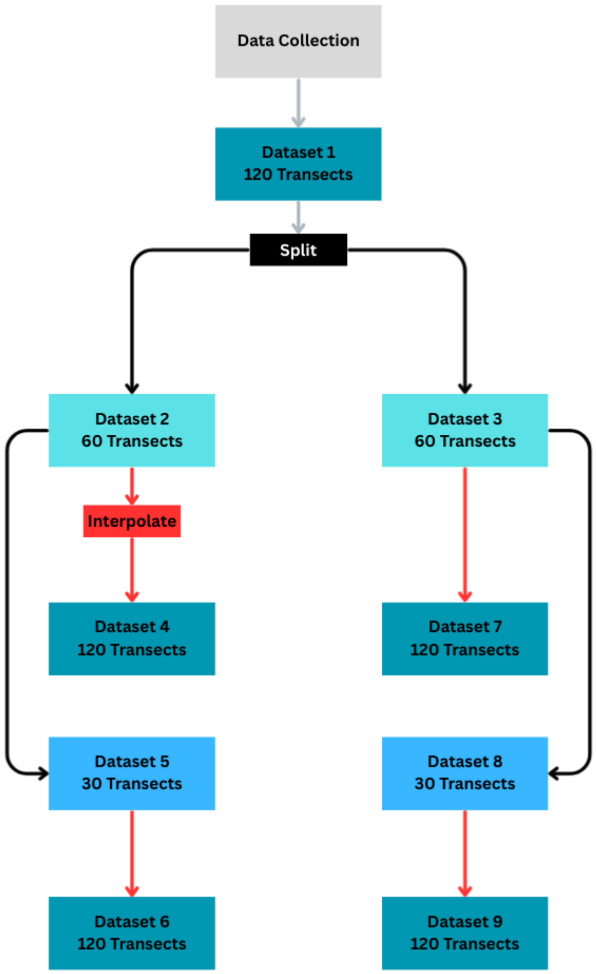


Figure 3. Workflow for creation of Datasets 1-9 from Collected Data

In order to evaluate the effect of halving the Y-resolution and effectively surveying the test site twice at a transect interval of 0.50 meter, the Dataset 1 .csv file was first split in half, with every odd row going into one new file (Dataset 2) and every even row going into another (Dataset 3). Datasets 2 and 3 allow for two comparisons to be made. The first is between each other – Datasets 2 and 3 are two (0.125, 0.5) resolution datasets staggered by 0.25 meter. If either of the identified features could not be detected, or located according to the ground-truthing methodology described above, this would indicate that the reduced-resolution (0.125, 0.5) data is not sufficient to function as a prospecting tool in this case.

Further statistical and visual comparison with the higher-resolution (0.125, 0.25) collected Dataset 1 may help elucidate the amount of precision lost in this case when Y-resolution is halved. As Datasets 2 and 3 each contain half the datapoints of Dataset 1, direct point-by-point comparisons cannot be carried out without artificially doubling the Y-resolution of these datasets back to 0.25 meter using the Interpolation function in TerraSurveyor (producing Datasets 4 and 7, respectively). Comparing these datasets with the collected Dataset 1 allows for the effect of interpolation to be assessed on the precision and accuracy of the appearance and positions of Anomalies 1 and 2.

To further evaluate this analytical framework, two coarser (0.125, 1) resolution datasets (Datasets 5 and 8) were extracted from half the rows of Datasets 2 and 3, respectively, beginning with the first line of each dataset. Based on a review of literature and current state and practitioner recommendation and guidelines, this resolution is not necessarily expected to be suitable as a prospecting tool. Further interpolation in TerraSurveyor to produce two datasets (6 and 9) isotropic with the collected Dataset 1 allows for further consideration of the effects of additional interpolation (quadrupling Y-resolution) on magnetometer data as a prospecting tool in this case.

Table 1. Summary of Properties of Datasets 1-9

Dataset	Character	Resolution (X,Y)	Origin (X, Y)	Parent Dataset
1	Real	0.125, 0.250	(0, +0.125)	N/A
2	Real	0.125, 0.500	(0, +0.125)	1
3	Real	0.125, 0.500	(0, +0.375)	1
4	Y-Interpolation	0.125, 0.250	(0, 0)	2
5	Real	0.125, 1.000	(0, +0.125)	2
6	Y-Interpolation	0.125, 0.250	(0, -0.250)	5
7	Y-Interpolation	0.125, 0.250	(0, +0.250)	3
8	Real	0.125, 1.000	(0, +0.375)	3
9	Y-Interpolation	0.125, 0.250	(0, 0)	8

All datasets are summarized above in Table 1. The 'character' refers to whether the dataset is composed of actual, collected datapoints or interpolated datapoints using the TerraSurveyor function described above. The 'origin' of each dataset refers to the position of the first collected datapoint in each set. The use of the Interpolation function on Datasets 2, 3, 5, and 8 to produce Datasets 4, 7, 6, and 9 required each .csv file to be re-imported into TerraSurveyor, where the datapoint locations were shifted as a set to the position required by the native data organization scheme of TerraSurveyor. For example, while the first datapoint of Dataset 2 occupies a position of (0, +0.125) meters along the X- and Y-axes of the collection grid, because the dataset has resolution (0.125, 0.5), in TerraSurveyor, the first datapoint in this set exists at (X, Y) coordinate position (0, 0.25). One round of Interpolation to double the Y-resolution replaces this datapoint with two calculated datapoints at (0, +0.125) and (0, +0.375) meter within

TerraSurveyor, and so on. But since the first datapoint actually occupied the (0, 0.125) meter position, and was only re-allocated the (0, 0.25) meter position in TerraSurveyor due to the data organization framework of the software, the *actual* position of the first datapoint of Dataset 4 is (0, 0). This process is visually summarized below in Figure 4.

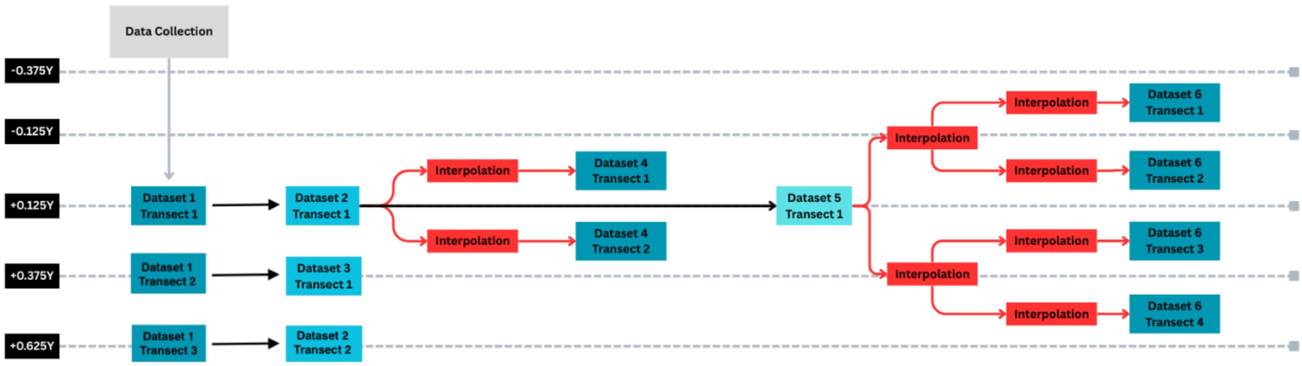


Figure 4. Data Organization Framework in TerraSurveyor and the Interpolation function, including effect on dataset origin

Analysis of the data organization framework within TerraSurveyor is necessary to facilitate calculation of Magnetic Anomaly 1 and 2 locations in real-world coordinates, in order to compare to the actual locations of Feature 1 and 2. The position of the first datapoint in each dataset is significant, because if the shifting relative locations due to TerraSurveyor and the process of physically splitting datasets to reduce resolution are not accounted for fully, analyses of the calculated locations of Magnetic Anomalies 1 and 2 relative to the actual GIS-recorded locations of Features 1 and 2 will be unreliable. Only the interpolated datasets are affected by this shift. A small amount of transects are 'located' outside of the original survey area from Dataset 6, but these data were not excluded from this analysis, as the complete graphics were utilized for visual analysis, which does not require in-depth calculation of magnetic anomaly locations.

Visual and Statistical Comparative Methods

As interpretation of magnetometry results can be both quantitative and visual, graphics were generated in TerraSurveyor with both minimal processing (DeStripe and Clip to 3 SDs only) and more intensive processing (DeStripe, X-and-Y Matching Interpolation, Low Pass Filtering, and Clip to 3 SDs) was used to created 'polished' graphics. Both the minimally processed and highly processed graphics were compared visually in order to evaluate the effect of interpolation and/or reduced survey intervals on visual appearance of survey results.

Datasets 1-9 contain a series of z-values (magnetic gradient values) corresponding to specific locations defined by X (distance along transect line) and Y (transect interval) coordinates. These z-values are displayed graphically by TerraSurveyor and function essentially as pixel intensity values in an image. One metric of measuring how much one image, or dataset, differs from another is Mean Absolute Error (MAE). MAE is calculated using the following equation as a summation of the absolute difference between each point in two Datasets (for example: Datasets 'Q' and 'R') of equivalent size and resolution, each containing N datapoints:

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|Q_i - R_i| \tag{5}$$

Due to the size of high-resolution datasets (57,600 datapoints), this calculation was performed on exported Datasets 1, 4, 6, 7, and 8 using an iterative Visual Basic script in Microsoft Excel. The flow of MAE comparisons is described in Figure 5 below.

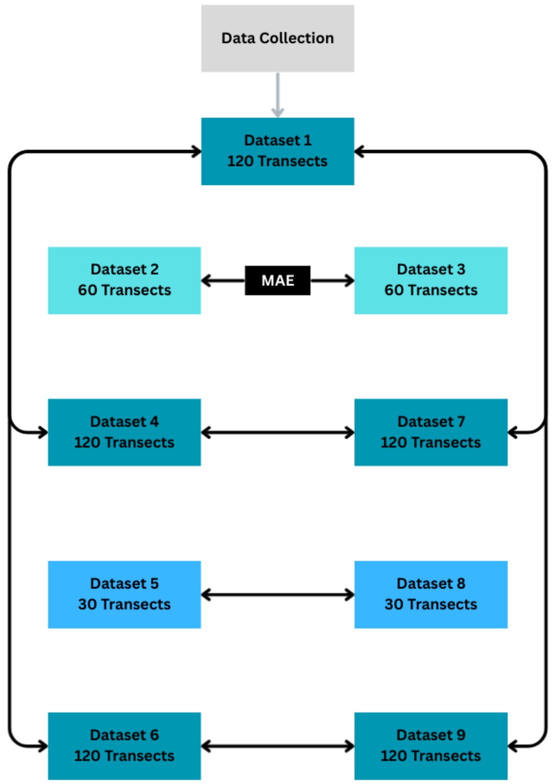


Figure 5. MAE Comparisons across Datasets 1-9

Finally, the Magnetic Anomaly 1 and 2 footprints across all datasets will be compared with the actual locations of Features 1 and 2, whose coordinate locations have been converted to represent the feature locations relative to the coordinate system of the magnetometry grids, which occupy a rectangular portion of space with origin (0, 0) meters located at the grids' southwest corner, and extend to position (30, 60) meters at the grids' northeast corner. No actual, objective coordinates for these resources are utilized in this thesis due to the sensitive nature of these resources; all coordinates of actual archaeological resources are relative, only.

While anomaly extent and centroids were determined via visual approximation in the original field effort which confirmed the presence of Features 1 and 2, for the purposes of this analysis, an objective and statistical definition of ‘anomalous’ was warranted in order to keep the definition of anomaly consistent across changing datasets. Rather than relying on visual identification, standard scores were calculated for each dataset in order to objectively develop a threshold of determining which magnetic gradient values qualified as ‘anomalous’ across Datasets. These standard scores (Z values) were calculated for each data point in each Dataset using an iterative application of the following formula in Excel Visual Basic:

$$Z = \frac{(x - \mu)}{\sigma} \tag{6}$$

where the difference between each datapoint (x) and the dataset mean (μ) was divided by the dataset standard deviation, or σ. An alpha (α) of 0.025 was chosen as a benchmark (critical Z=+1.96) in order to capture values which fall within the top 2.5% of the dataset in the definition of ‘anomalous’. This definition of ‘anomalous’ was chosen to roughly correspond to the visually-identified anomaly extents in the collected Dataset. This provided a reliable and standard threshold for ‘anomalous’ which could be repeated across Datasets 1-9 without introducing error to the analysis due to inconsistent visual delineation of anomaly extents and centroids.

Application of Methods Described Above

Results from the archaeological evaluation of a precontact archaeological site in central Ohio, located in an agricultural field subject to cultivation for decades, are presented in the following chapter. These results follow the employment of the data collection and processing methods described above, which allowed for the initial identification of Features 1 and 2 at the Test Site. The remaining analytical methods (dataset generation and visual and statistical comparisons) are discussed in the Analysis chapter to follow the Results, and were not directly employed during Stantec's fieldwork or the generation of deliverables pertinent to the archaeological evaluation conducted under contract in 2025. These methods were devised by this author for research purposes.

Chapter 5: Results

Introduction to Results

This chapter presents the results of the Bartington GRAD-601 dual fluxgate magnetic gradiometer survey of two 30-meter by 30-meter grids (Grids 1 and 2) on an Archaic and Woodland Period precontact site (the Test Site) in central Ohio during an archaeological Phase II evaluation. This survey was completed with a resolution of 0.125 meters (8 readings per meter) in the X-dimension, with a transect interval of 0.25 meters in the Y-dimension (Dataset 1) at a sensitivity of 0.01 nT. Ground truthing following this survey detected two subsurface precontact archaeological features (Features 1 and 2). The extents of Features 1 and 2 were mapped and georeferenced, allowing for the comparison of anomalous magnetic gradient values and actual feature extents. Magnetic gradient in-line (X-dimension) and transect interval (Y-dimension) traces and gradiometer data maximum readings are also presented for both features.

Geophysical Survey and Ground Truthing Results

The geophysical survey at the Test Site resulted in the collection of 57,600 datapoints, in two grids each measuring 30 meters by 30 meters. After DeStriping along transects to remove the effect of consistent differences in readings between the two gradiometer sensors due to sensor error, this dataset contained values ranging from +22.17 nanotesla (nT) to -24.25 nT, with a mean of 0.9 nT, a standard deviation (SD) of 1.265679, and a survey resolution of 0.125 meter

(X-dimension) by 0.25 meter (Y-dimension). Although the dataset does contain some more extreme values, most of the data is centered around the mean.

This data is visually depicted in Figure 6 and Figure 7 below. Figure 6 depicts Grids 1 and 2 with minimal processing – only a Clip at 3 SDs to improve visibility of more subtle variations in magnetic gradient values – while the data in Figure 7 were processed more intensively to improve the appearance of the graphic. The two figures are indistinguishable visually, likely due to the high resolution possessed by the raw data leaving little opportunity for interpolation of values. The minimally processed and polished graphics for generated Datasets 2-9, along with their statistics, are presented and compared in the following chapter.

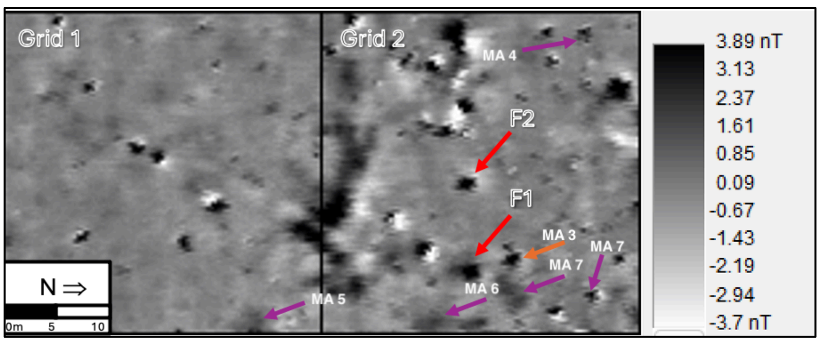


Figure 6. Minimally processed collected data (Dataset 1), with scale of magnetic gradient values (nT). Features 1 and 2 (F1; F2; red arrows), non-cultural feature (orange arrow; MA 3) and non-feature (purple arrow) investigated magnetic anomalies (MA 4-7) also depicted. Photo courtesy of Stantec Consulting Services, Inc.

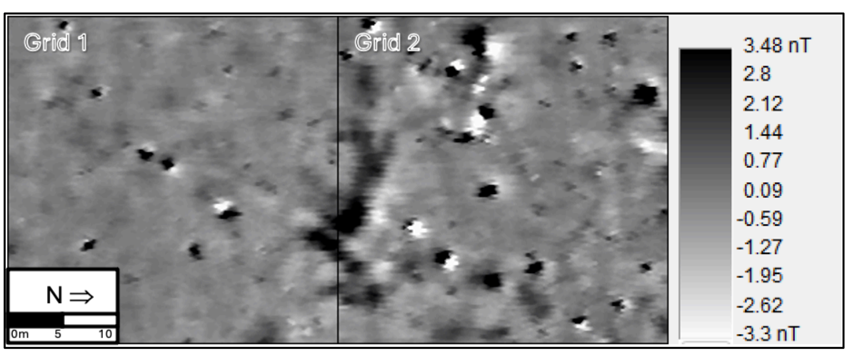


Figure 7. Polished graphic of gradiometer survey results, with scale of magnetic gradient values (nT). Photo courtesy of Stantec Consulting Services, Inc.

After collection and processing of magnetic gradiometer data, seven Magnetic Anomalies (MAs 1-7) were chosen for further investigation via physical excavation, or ground-truthing. Anomalies were chosen based on their visual similarity to expected signatures of archaeological features (i.e. rounded, circular or oval, or otherwise pit-shaped; no linear features were noted in this data) and their magnetic gradient values (approximately within +/-10 nT in unclipped data). Additionally, the team sampled various types of magnetic anomaly, per Ohio archaeological guidelines – leading to the identification of Feature 2 despite the higher than typical (+18.52 nT) maximum value for an apparently thermoremanent archaeological feature. The character of MAs 1-7 is summarized in the table below. Magnetic anomaly sizes are visually-determined, as no standard magnetic anomalous extents were calculated for MA 3-7 considering their exclusion from this analysis. As depth and shape and physical makeup of potential archaeological features corresponding to MA 1-7 are unknown at this point, it is not possible to calculate potential feature sizes based on magnetic anomaly sizes at this stage, prior to ground-truthing.

Table 2. Summary of Tested Magnetic Anomalies

MA Number	Maximum Magnetic Gradient Value (+ nT)	MA Size N-S (m; visual)	MA Size E-W (m; visual)	Shape
1	7.27	2	2	Circular
2	18.52	2	2	Circular
3	7.73	1.75	1.75	Oval
4	6.70	1	1	Circular
5	3.05	3	3	Amorphous
6	2.09	2.5	1.5	Amorphous
7	2.52	3	5	Amorphous

Following ground truthing, the magnetic anomalies (MA 1 and 2) marked with red arrows in Figure 6 were found to be related to cultural features and are marked as such in Figure 6: Feature 1 consists of a circular, basin-shaped thermal pit approximately 1.2 meters in diameter (Figure 8) while Feature 2 is an approximately 0.25-meter oval concentration of Fire Cracked Rock (FCR) and thermally altered soils (Figure 9). The magnetic anomaly generated by Feature 1 is an approximately circular region of positive (black) magnetic gradient values with diffuse edges, and a slightly negative (very light gray) halo to the north, although this could be environmental. The maximum magnetic gradient value of this anomaly is +7.27 nT. A positive value in this range and lack of a strong negative bipole suggests thermoremanent rather than inherent magnetism, and the shape and size of the anomalous value field in general corresponds to the size and shape of a precontact thermal pit feature, although size and shape of anomaly can be misleading. Interpretation of the anomaly corresponding to Feature 2 is less straightforward – the anomalous values are slightly more concentrated, and the anomaly has a stronger and clear negative pole to the north, corresponding to the stronger (maximum +18.52 nT) positive gradient values present. This degree of positive response evidently still corresponds to thermoremanent, rather than inherent, magnetism given the apparent thermal nature of this feature.

Magnetic Anomalies 3-7 were also ground-truthed according to the methods described in the previous chapter. MA 3 yielded a non-cultural feature, a shallow area of deeper plowzone which was not definitively cultural. As MA 3 appeared to be environmental, and not cultural, it was excluded from this analysis. Upon ground-truthing, no cultural or environmental features of interest were found to correspond to MA 4-7. These magnetic anomalies were therefore considered, although as complete, 100 % excavation of the surveyed area was not carried out, it is not certain whether MA 4-7 are 'true negatives' or 'false negatives.' As this analysis required georeferenced, cultural feature extents to cross-reference with geophysical data, these tested Magnetic Anomalies were also excluded from this analysis. Nothing specific in the geophysical data promoted MA 1 and 2 as archaeological over MA 3-7; all anomalies (with the exception of MA 2) fit the approximate nT range and general expected appearance (rounded) of potential thermoremanent precontact archaeological features – hence, the importance of ground-truthing when evaluating geophysical data for archaeological potential.

Features 1 and 2 were located approximately 30 centimeters below ground surface, and had likely been truncated by historical agricultural activity. Five additional anomalies were explored via excavation, resulting in the discovery of one non-cultural, environmental feature (orange arrow; Figure 6). No natural or cultural features were detected for the remaining four anomalies that were investigated, indicated with purple arrows above (Figure 6).

Figure 8. Feature 1, which generated the Anomaly 1 signature, in planview. Photo courtesy of Stantec Consulting Services, Inc.

Figure 9. Feature 2, which generated the Anomaly 2 signature, in planview. Photo courtesy of Stantec Consulting Services, Inc.

Z-scores were calculated for every datapoint in Dataset 1, using the method described in the previous chapter. An alpha (α) of 0.05 was chosen (critical Z=+1.96) to provide a threshold to capture values which fall within the top 2.5% of the dataset. For the purposes of this thesis, a

Z-score of greater than or equal to +1.96 characterizes a magnetic gradient value as 'anomalous' within this dataset. This definition provides an objective, but tailored, delineation of 'anomalous' areas which roughly correspond to the visually distinct, magnetically responsive areas shadowing Features 1 and 2 and which can be applied to generated Datasets 2-9 in a consistent way (Figure 10; Figure 11).

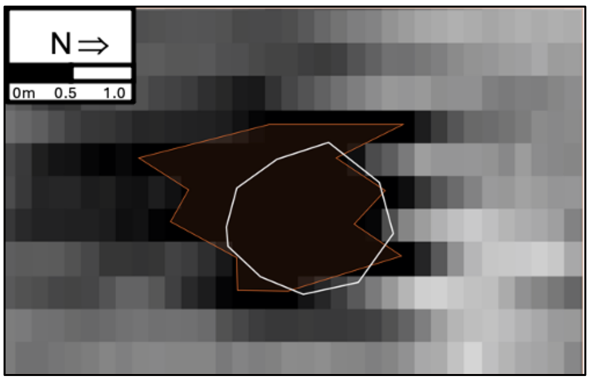


Figure 10. Extent of Feature 1 (in white) and Anomaly 1 (in orange) over clipped gradiometer data. Photo courtesy of Stantec Consulting Services, Inc.

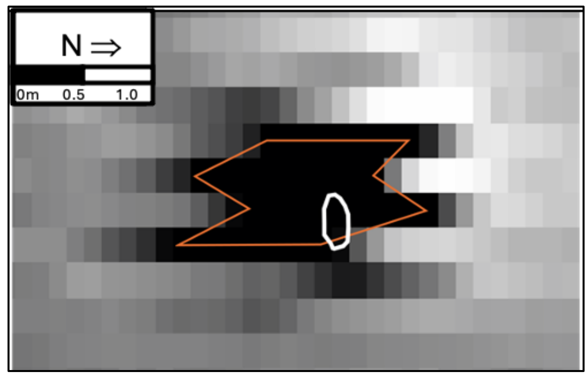


Figure 11. Extent of Feature 2 (in white) and Anomaly 2 (in orange) over clipped gradiometer data. Photo courtesy of Stantec Consulting Services, Inc.

Both MA 1 and 2 overlap with Features 1 and 2, but not precisely. The relatively greater magnetic gradient values generated by Feature 2 (maximum +18.52 nT) has resulted in a greater size discrepancy between Anomaly 2 and Feature 2, while the locational discrepancy between MA 1 (maximum +7.27 nT) and Feature 1 is less extreme. In other words, Feature 1 is relatively less magnetic than Feature 2, and therefore its magnetic field has a more limited 'reach' than the field generated by Feature 2. When mapped to scale (in meters), the maximum magnetic gradient value and magnetic anomaly centroids can be seen to lie within the boundaries of Feature 1 (Figure 12), but not Feature 2 (Figure 13). Qualitatively, this suggests a more precise detection of Feature 1 than Feature 2 at the (X, Y) resolution (0.125, 0.25) of the collected dataset. Both features were detected by the excavation of a STU over the anomaly centroids during fieldwork.

Feature 1, but not Feature 2, could have been detected by more expedient ground-truthing (a soil core) centered over its magnetic gradient maxima or centroid of its anomaly. This suggests that, qualitatively, the survey resolution employed during fieldwork was more than sufficient for Feature 1 at all levels of effort considered here, but only sufficient for certain levels of effort for identification of Feature 2.

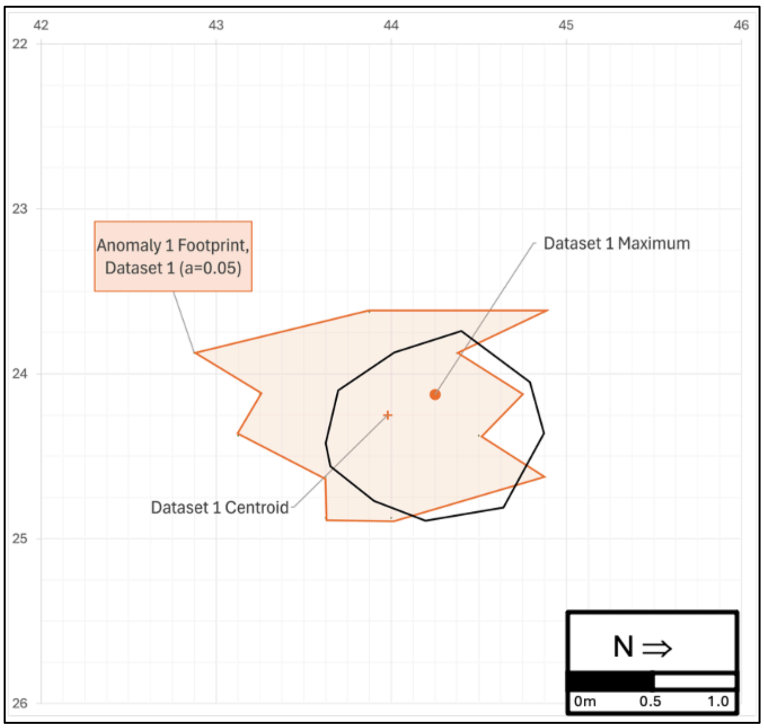


Figure 12. Extent of Feature 1 (in black) and Anomaly 1 (in orange), along with Anomaly 1 maximum magnetic gradient value and centroid location. Plotted to scale (in meters). Graphic courtesy of the author.

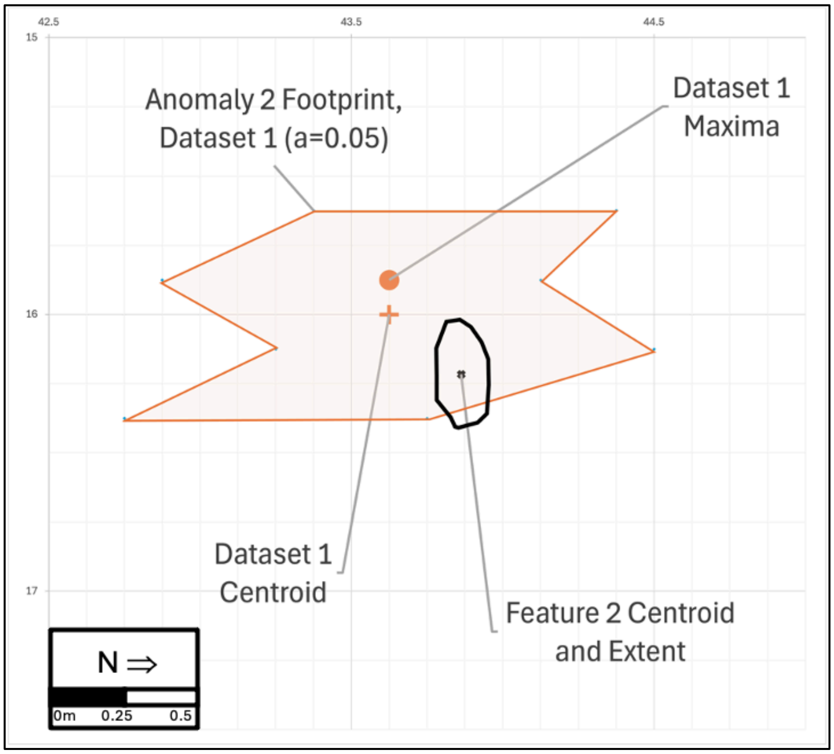


Figure 13. Extent and centroid of Feature 2 (in black) and Anomaly 2 (in orange), along with Anomaly 2 maximum magnetic gradient value and centroid location. Plotted to scale (in meters). Graphic courtesy of the author.

<u>*Limits of Detection of Signatures Corresponding to Confirmed Features*</u>

The one-dimensional traces of magnetic gradient values (in nT) graphed against distance (in meters) across the traverse displaying the greatest magnetic response, or peak, are displayed below for MA 1 and 2. The traces depict magnetic gradient values from a single transect in the in-line, or X-, dimension (Figure 14 and Figure 15) and transect, or Y-, dimension (Figure 16 and Figure 17) for MA 1 and 2, respectively. While the distance values displayed along the X-axis are arbitrary (all magnetic gradient traces begin at an in-line position of 0.0 when exported from TerraSurveyor) the scale is not; these trace graphs and the raw data they represent can be related to frameworks for estimating feature size described by Clark (1990: 83) and ensuring feature detection (Schmidt and Marshall 1997: 334-345).

According to Clark (1990: 83), a feature's size can be estimated as half the width of the trace peak. Schmidt and Marshall (1997: 344-345) calculated that, in order to detect a feature, the transect interval must not exceed half the width of the signature peak generated by the feature. The magnetic gradient value traces for Features 1 and 2 measure approximately 2.0 meters in the X-dimension (Figure 14 and Figure 15). In the Y-dimension, or transect dimension, the magnetic gradient value trace for Feature 1 measures approximately 2.0 meters, while Feature 2 measures 1.5 meters (Figure 16 and Figure 17). Therefore, both features should be detectable at intervals of 1.0 meter in the X-dimension, and 1.0 meter (Feature 1) and 0.75 meter (Feature 2) in the Y-dimension.

Clark's (1990: 83) size estimation method is fairly accurate for Feature 1, but misses the mark on Feature 2, which was much smaller than the corresponding anomaly – not unexpected, as the shape and size of magnetic anomalies do not necessarily correspond to the shape and size of the feature generating the magnetic signature. The feature detection limits based on Schmidt and Marshall's (1997) calculations will be compared with the various tested resolutions in the following chapter.

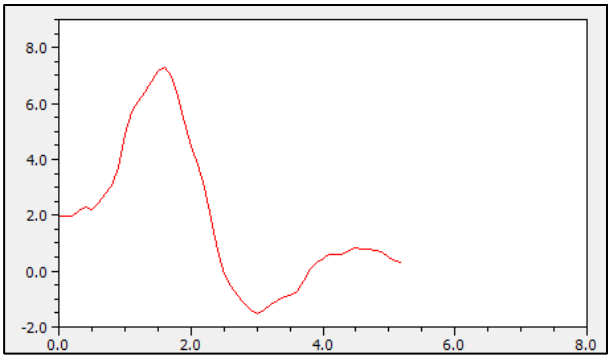


Figure 14. Magnetic gradient (nT) trace of Anomaly 1 graphed against arbitrary distance (m) in the in-line, or X-, dimension, Feature 1, Dataset 1, along transect 97. Graphic courtesy of Stantec Consulting Services, Inc.

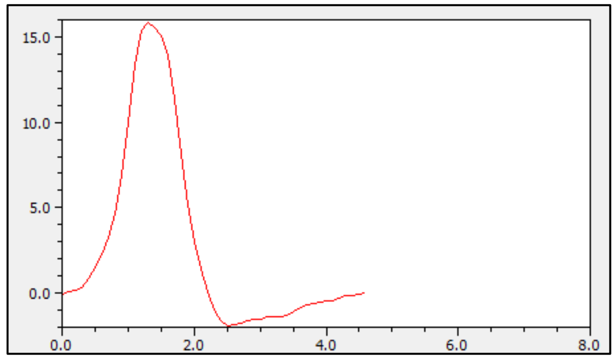


Figure 15. Magnetic gradient (nT) trace of Anomaly 2 graphed against arbitrary distance (m) in the in-line, or X-, dimension, Feature 2, Dataset 1, along transect 64. Graphic courtesy of Stantec Consulting Services, Inc.

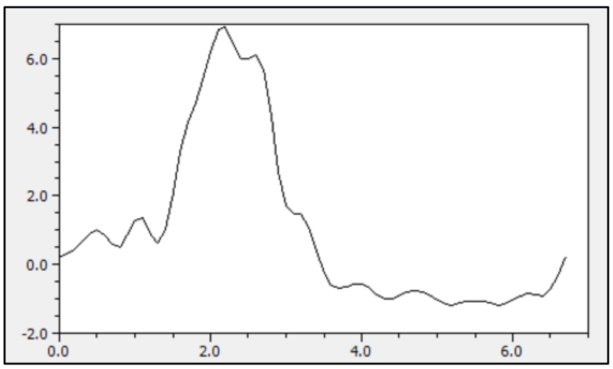


Figure 16. Magnetic gradient (nT) trace of Anomaly 1 graphed against arbitrary distance (m) in the transect, or Y-, dimension, Feature 1, Dataset 1, along in-line position 354. Graphic courtesy of Stantec Consulting Services, Inc.

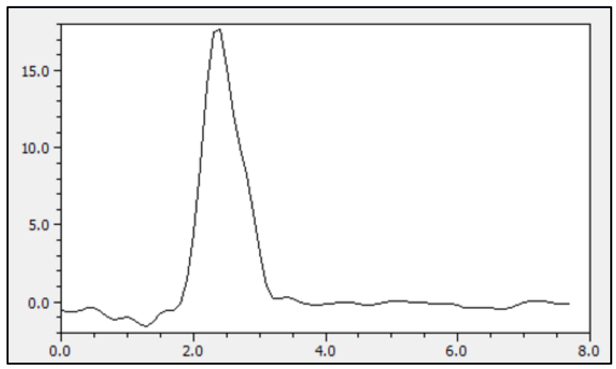


Figure 17. Magnetic gradient (nT) trace of Anomaly 2 graphed against arbitrary distance (m) in the transect, or Y-, dimension, Feature 2, Dataset 1, along in-line position 351. Graphic courtesy of Stantec Consulting Services, Inc.

Conclusion of Results

This chapter has presented the results of magnetic gradiometer survey at a precontact archaeological site in central Ohio, located in an agricultural field subject to cultivation for decades. The survey resulted in the detection of numerous magnetic anomalies. After ground-truthing of seven anomalies, two intact archaeological features were identified. Physical, relative locational data and minimally processed gradiometer data from the survey are presented. In the following chapter, these real results will be compared with a series of Datasets generated via physical slicing (for reduced resolution) and interpolation (for artificially bolstered resolution) in order to examine the significance of this survey resolution in the Y-dimension (0.25 meters) for the detection of Features 1 and 2, which display magnetic gradient trace widths of two meters and physical measurements of one meter and 0.25 meter, respectively.

Chapter 6: Analysis

Introduction to Analysis

In this chapter, the collected magnetometry data (Dataset 1) is split into two datasets of halved resolution in the Y-dimension (Datasets 2 and 3). Two datasets of halved resolution in the Y-dimension (Dataset 5 and Dataset 8) were then extracted from Datasets 2 and 3. All four of these datasets were then interpolated in the Y-dimension to a Y-resolution of 0.25 meter, to allow for direct datapoint-to-datapoint comparisons with Dataset 1, which was collected along transects spaced 0.25 meter apart (Figure 18). While Figure 18 represents a streamlined reminder of this workflow, a more detailed illustration is available in Chapter 4 (Figure 3).

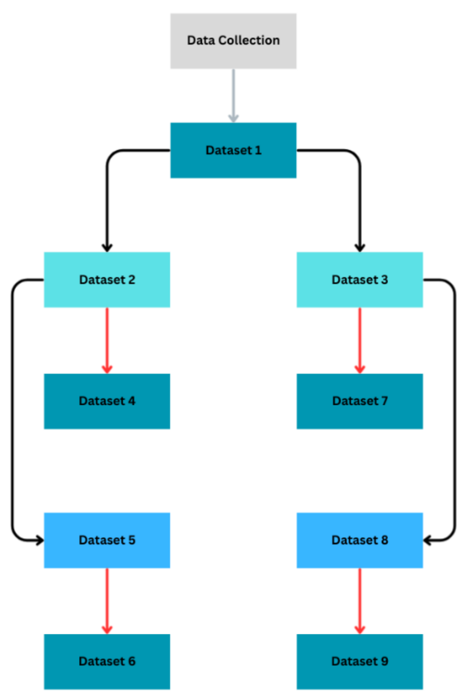


Figure 18. Summarized Workflow for Test Dataset Generation (recall Figure 1)

The content of datasets of equivalent resolution can be compared via calculated Mean Absolute Error (MAE) values. The lightly processed and polished dataset graphics are visually compared. Finally, the locations of maximum magnetic gradient value and Anomaly 1 and 2 extents and centroids are compared graphically with the actual locations of Feature 1 and 2 across each dataset. This series of comparisons suggests several points about broadening of intervals during field collection (simulated here by a physical reduction of resolution in the Y-dimension) and the effects of Interpolation on dataset content. An analytical summary of Datasets 1-9 is presented below in Table 3.

Table 3. Resolution, Creation, and Origin of Datasets 1-9

Dataset	**Resolution (X, Y)**	**% Collected Data in Dataset, Prior to Processing**	**Interpolation Yes/No**	**Origin (X, Y)**
1	(0.125, 0.25)	-	-	(0, +0.125)
2	(0.125, 0.50)	100%	No	(0, +0.125)
3	(0.125, 0.50)	100%	No	(0, +0.375)
4	(0.125, 0.25)	50%	Yes	(0, 0)
5	(0.125, 1.0)	100%	No	(0, +0.125)
6	(0.125, 0.25)	25%	Yes	(0, -0.250)
7	(0.125, 0.25)	50%	Yes	(0, +0.250)
8	(0.125, 1.0)	100%	No	(0, +0.375)
9	(0.125, 0.25)	25%	Yes	(0, 0)

<u>Intra-site and Inter-site Mean Absolute Error (MAE) Values</u>

As MAE values are a measure of relative variation between sets of data, their values are theoretically only limited by the dataset values themselves. In order to relate MAE values to relative dataset similarity for a set of values, it is useful to establish a threshold for what value can be expected for ‘different’ images. This can be done by calculating MAE values for images known to be visually distinct, in order to provide perspective on MAE values from Dataset 1-9 comparisons. Comparisons were first carried out between each grid of Dataset 1 and a dissimilar dataset in order to provide perspective of how magnetic gradient values collected in a similar environment and general area but in different physical locations in space appear in MAE values. The sample dataset is known to be dissimilar because it comes from a different archaeological site (Comparison Site), but comparative as it comes from the same region and possesses roughly the same descriptive statistics as Dataset 1. Additionally, the two grids from Dataset 1 are visually distinct, and a MAE value calculated between Grids 1 and 2 also serves as a threshold value for 'visually distinct.'

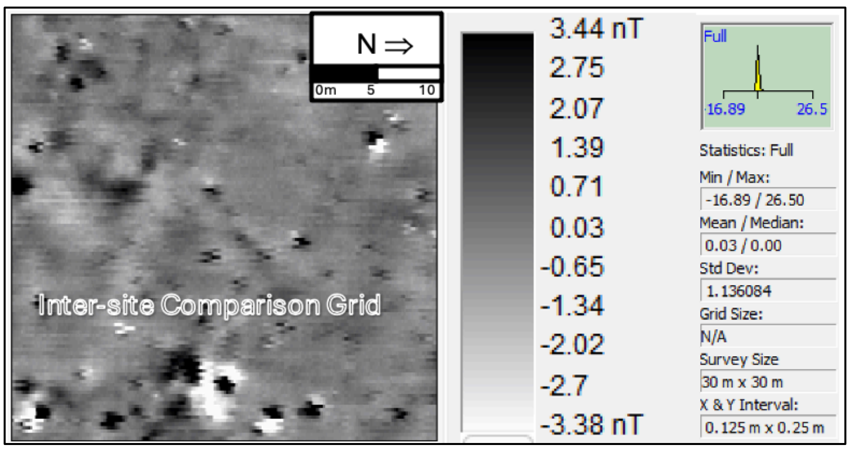


Figure 19. Minimally processed grid from additional precontact archaeological site (Comparison Site), collected on nearby parcel under similar environmental conditions.

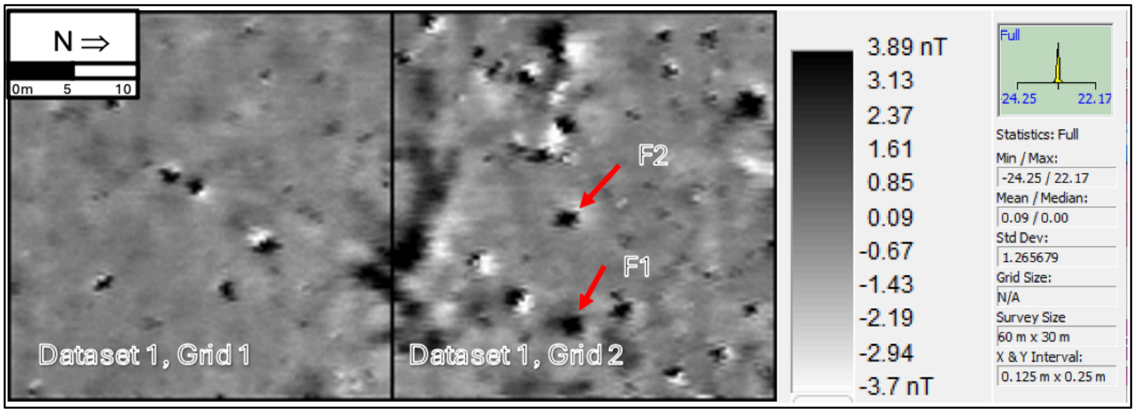


Figure 20. Minimally processed collected data (Dataset 1) from the Test Site, with scale, and appearance of Anomalies/Features 1 and 2 indicated with red arrows.

Figure 19 shows the minimally-processed graphic produced from one 30 meter by 30 meter grid of magnetometer data collected from the Comparison Site, along with scale and descriptive statistics. The data collected from the Comparative Site is similar in range and distribution to the data from Grids 1 and 2 of the Test Site (Dataset 1; Figure 6). These three datasets can be compared via MAE, in order to provide a standard from which to judge if a MAE value could be produced by two visually distinct images. As these datasets were collected from different physical locations, and the images produced by them are visually distinct, the calculated MAE values in Table 4 below represent 'dissimilar' images for the approximate range and distribution of data in Datasets 1-9. All MAE values were calculated using unclipped datasets, while all graphics shown are composed of clipped data to aid in the identification of more subtle differences in magnetic gradient values.

Table 4. Intra-site and Inter-site MAE values, for perspective

Type of Comparison	Datasets Being Compared	MAE
Intra-site	Grid 1 vs Grid 2	1.02282743
Inter-site	Grid 1 vs Inter-Site Comparison Grid	0.419175
Inter-site	Grid 2 vs Inter-Site Comparison Grid	1.15728681

The MAE values displayed in Table 4 can provide a means from which to judge the similarity between Datasets 1-9 by providing example values calculated between similar but not identical datasets which display as visually distinct images. These MAE values, which are expected to be ‘high’ as the datasets and images themselves are similar but not identical, range from just over 0.4 to over 1 for the three non-identical and non-related datasets compared below. Understanding the point-by-point differences in Datasets 1-9, variations of the same data transformed in some way, provides a means of quantitatively assessing the variation introduced by reducing resolution and employing interpolation in TerraSurveyor, even if these actions do not necessarily result in visually distinct images between datasets. This is a means of exploring variation between datasets that cannot be seen by the naked eye.

Table 5. Datasets 1-9 Comparison

Comparison	**Resolution**	**Datasets Being Compared**	**MAE**	**Offset between datasets (m)**
Real vs. Interpolated, Down Lineage	0.125x0.25	1 v. 4	0.39386944	0.125
Real vs. Interpolated, Down Lineage	0.125x0.25	1 v. 7	0.24438837	0.125
Real vs. Interpolated (2x), Down Lineage	0.125x0.25	1 v. 6	0.63497066	0.375
Real vs. Interpolated (2x), Down Lineage	0.125x0.25	1 v. 9	0.57322743	0.125
Real vs. Real (Split), Across Lineage	0.125x0.5	2 v. 3	0.52898472	0.250
Interpolated vs. Interpolated, Across Lineage	0.125x0.25	4 v. 7	0.52347483	0.250
Real vs. Real (Split), Across Lineage	0.125x1.0	5 v. 8	0.53713819	0.250

Comparison	Resolution	Datasets Being Compared	MAE	Offset between datasets (m)
Interpolated vs. Interpolated, Across Lineage	0.125x0.25	6 v. 9	0.53859844	0.250

MAE values were calculated among pairs of Datasets (1-9) possessing equivalent resolutions in order to quantitatively evaluate the point-by-point variation from one dataset to the other following various splits and interpolations of the original, collected data (Table 5). MAE values provide one means of judging the effect of broader intervals and/or interpolation on the precision and accuracy of resulting magnetic datasets. Interpolated datasets contain 100% interpolated values, and the 'Offset' values above are calculated with the interpolative process, or physical split of datasets, in mind. These values are relevant to the comparison of anomaly positions within each dataset, as calculating accurate physical locations for Anomalies 1 and 2 requires the calculus of the relative origin coordinates of each Dataset.

Dataset 1 is most similar to Datasets 4 and 7, which each contained (prior to interpolation) more (50%) collected datapoints (from Dataset 1) than any other generated Dataset of equivalent resolution and do not share any collected datapoints in common. Datasets 4 and 7 are more similar to Dataset 1 (MAE=0.39…, 0.24…) than they are to each other (MAE=0.52…), and are about as similar to each other as their un-interpolated parent Datasets 2 and 3 (MAE=0.52…).

Datasets 6 and 9 were generated via Y-dimension interpolation from Datasets 5 and 8, which were cut and assembled from Dataset 1. Datasets 6 and 9, which each contained only 25% collected values from Dataset 1 prior to interpolation, are less similar to Dataset 1 (MAE=0.63…, 0.57…) than they are to each other (MAE=0.53…). Datasets 5 and 8 are composed of 100% collected data, and each hold 25% of the content of Dataset 1. Datasets 5 and 8 displayed a comparable similarity (MAE=0.53…) to the other 'cross-lineage' dataset comparisons (4 v. 7, 6 v. 9).

This series of comparisons suggests several points about broadening of intervals during field collection (simulated here by a physical reduction of resolution in the Y-dimension) and the effects of interpolation on dataset content. As the proportion of collected data points used to generate interpolated data points, to interpolated points, shifts to favor the latter, MAE increases between generated datasets and Dataset 1. This result is not unexpected, as it is known over-interpolation in TerraSurveyor can reduce data quality (DW Consulting 2019: 75). However, in contrast, MAE values between datasets from across the Dataset 2 and Dataset 3 lineages (2 v. 3, 4 v. 7, and 6 v. 9) remains relatively constant. Additional interpolation between these datasets does not seem to affect MAE, or, put another way, the degree of similarity between the datasets. Rather, the initial split of Dataset 1 into half-Y-resolution Datasets 2 and 3 seems to set the tone for dataset similarity or dissimilarity.

This would suggest that the factor causing the increasing MAE values as comparisons move from Y-interpolated 0.50 meter interval datasets (Datasets 4 and 7) to comparisons with Y-interpolated 1.0-meter interval datasets (Datasets 6 and 9) is due not to the Y-interpolation itself, but the shifting *proportion* of collected data in the dataset prior to interpolation. Collecting data on 1.0-meter intervals and 0.50-meter intervals, staggered by 0.25 meters, results in the assembly of datasets with different values. Is this difference significant enough to affect interpretation?

Interpretation of geophysical results involves both quantitative evaluation and visual comparison. In the subsection to follow, graphical representations of Datasets 1-9 are visually compared in order to explore the practical, rather than statistical, differences between the datasets. Do these quantitative results align with visual differences that might lead a practitioner to make a different decision in terms of anomaly character, or ground-truth for an anomaly in a different location, when faced with Dataset 2 or 3, or Dataset 1 or 9?

Visual Comparison

Collected data (Dataset 1) is compared with graphics produced from minimally-processed, and more extensively processed ('polished'), test Datasets 2-9 in Table 6, Table 7, and Table 8 below. Minimal processing indicates data were DeStriped (Median) along Traverses and Clipped to 3 Standard Deviations [SDs]). 'Polishing' indicates data were DeStriped (Median) along Traverses, Interpolated to Match X- and Y-axes resolutions, Low-Passed (Uniform; Median) and Clipped to 3 SDs. While no quantitative analyses were performed on 'polished' datasets, these more processed graphics serve to show the generative power of Interpolation and the ability of processing software to mask low-resolution datasets. The visual degree of difference between the minimally- and highly-processed datasets is related to the ratio of dataset starting and ending resolutions. In other words, the more datapoints generated during data processing, the greater the degree of visual difference in the final graphical product. This is evidenced by the lack of visual difference between the minimally-processed Dataset 1, which has resolution (0.125, 0.25) and the 'polished' graphic from this Dataset, which was interpolated to double the Y-resolution only one time to match the X-dimension resolution, for a final resolution of (0.125, 0.125). Magnetic Anomalies 1 and 2 are marked in each dataset with red arrows.

Table 6. Visual Comparison of Collected Data (Dataset 1) with Lowest Resolution (0.125x, 1.0y) Generated Datasets (5 and 8)

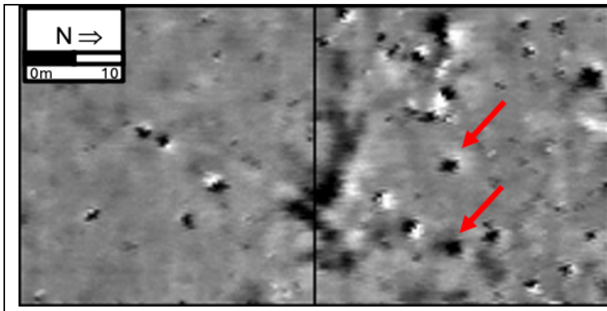
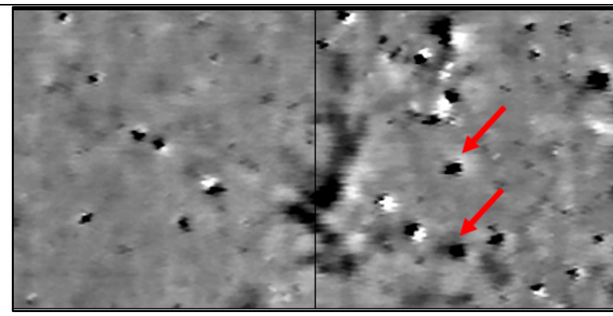
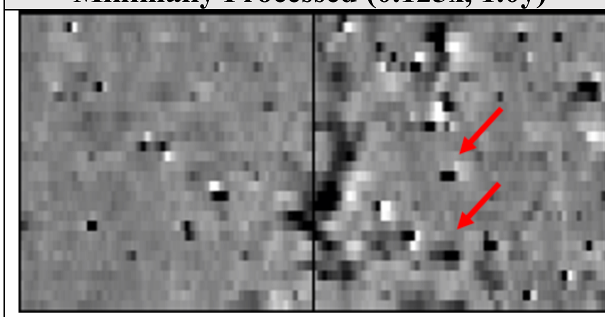
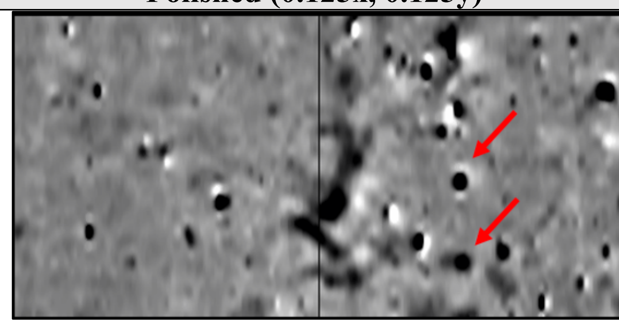
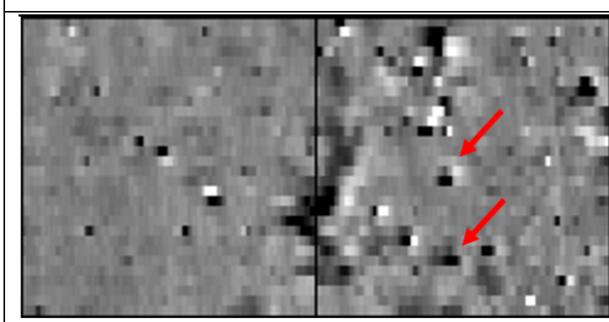
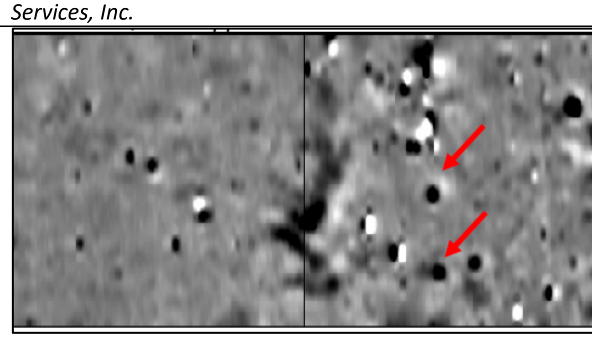
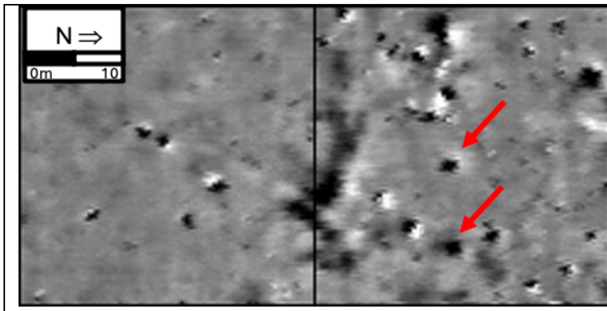
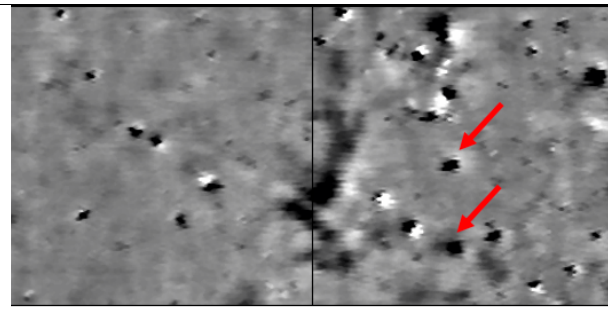
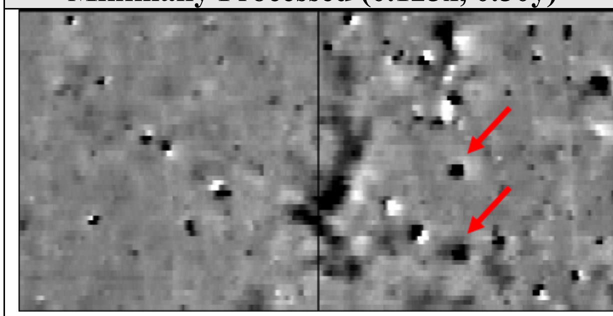
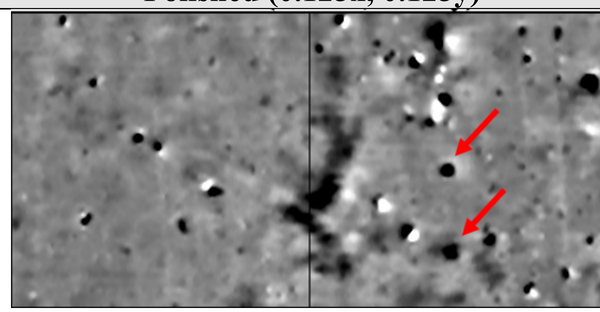
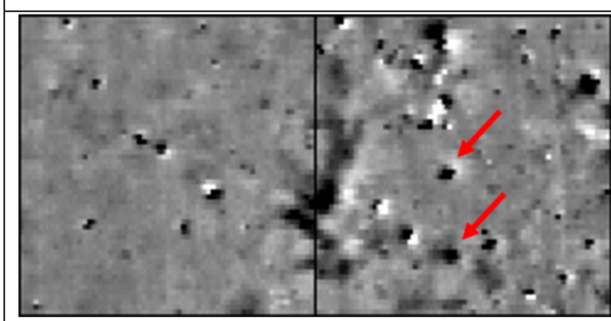
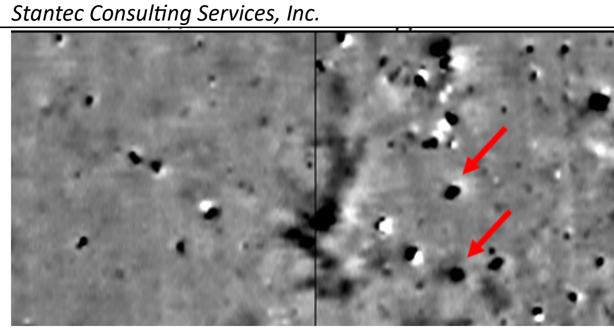
 *Dataset 1, Minimally Processed after Collection. Graphic courtesy of Stantec Consulting Services, Inc.*	 *Dataset 1 DeStriped, Interpolated to 0.125m resolution on X and Y axes, Low-Passed (Uniform; Median) on Window Size 3x3, and Clipped. Graphic courtesy of Stantec Consulting Services, Inc.*
Minimally Processed (0.125x, 1.0y)	**Polished (0.125x, 0.125y)**
 *Dataset 5, DeStriped and Clipped. Graphic courtesy of Stantec Consulting Services, Inc.*	 *Dataset 5, DeStriped, Interpolated to match X and Y axes, Low Passed, and Clipped. Graphic courtesy of Stantec Consulting Services, Inc.*
 *Dataset 8, DeStriped, and Clipped. Graphic courtesy of Stantec Consulting Services, Inc.*	 *Dataset 8, DeStriped, Interpolated to match X and Y axes, Low Passed, and Clipped. Graphic courtesy of Stantec Consulting Services, Inc.*

Table 6 above compares the lowest resolution generated datasets (5 and 8) with the collected Dataset 1. These datasets are completely composed of real, collected datapoints on a 1.0-meter interval, beginning at the first traverse (Dataset 5) and the second traverse (Dataset 8) of the collected dataset. These graphics confirm the expected limit of detection for Features 1 and 2, calculated in the previous chapter; both features are only visible in a single traverse in the un-interpolated dataset graphics, corresponding to only one ‘hit’ in the Y-dimension. In the case of Features 1 and 2, magnetometry data collected on a one meter interval is sufficient to establish only presence/absence of these anomalies/features at this survey interval. The resolution of initial datasets is so low, interpolation to match the in-line collection interval of 0.125 meter visually distorts both features. Three rounds of Y-dimension interpolation to ‘double’ resolution has stretched both Anomalies 1 and 2 when compared with their appearance in the collected data. The minimally-processed graphics for Datasets 5 and 8 are visually similar (recall MAE=0.53… for these two datasets).

Table 7. Visual Comparison of Collected Data (Dataset 1) with Lower Resolution (0.125x, 0.50y) Generated Datasets (2 and 3)

 *Dataset 1, Minimally Processed after Collection. Graphic courtesy of Stantec Consulting Services, Inc.*	 *Dataset 1, Polished. Graphic courtesy of Stantec Consulting Services, Inc.*
Minimally Processed (0.125x, 0.50y)	**Polished (0.125x, 0.125y)**
 *Dataset 2, DeStriped and Clipped. Graphic courtesy of Stantec Consulting Services, Inc.*	 *Dataset 2, DeStriped, Interpolated to 0.125m resolution on X and Y axes, , Low-Passed and Clipped. Graphic courtesy of Stantec Consulting Services, Inc.*
 *Dataset 3, DeStriped and Clipped. Graphic courtesy of Stantec Consulting Services, Inc.*	 *Dataset 3, DeStriped, Interpolated to 0.125m resolution on X and Y axes, Low-Passed, and Clipped. Graphic courtesy of Stantec Consulting Services, Inc.*

Datasets 2 and 3 are compared visually with the collected data in Table 7. Datasets 2 and 3 each contain half the collected data present in the higher resolution (0.125, 0.25) Dataset 1, beginning with Traverse 1 (Dataset 2) and Traverse 2 (Dataset 3), respectively. These datasets are the experimental equivalent of collecting magnetometer data on the test site with transects spaced 0.50 meters apart. The graphics for both datasets are roughly visually quite similar (recall MAE=0.52…). Features 1 and 2 appear slightly distorted between the two datasets, when compared with the collected Dataset 1, but neither feature is visually unrecognizable. This slight distortion is exaggerated by the double Y-interpolation for the ‘polished’ figures.

Table 8. Visual Comparison of Collected Data (Dataset 1) with Equivalent Resolution (0.125x, 0.25y) Generated Datasets 4 and 7, interpolated from 0.50-meter transect interval Datasets 2 and 3

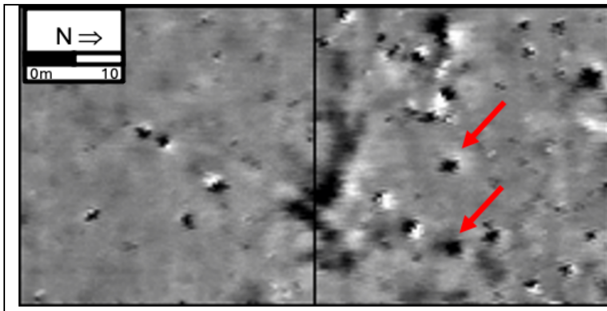
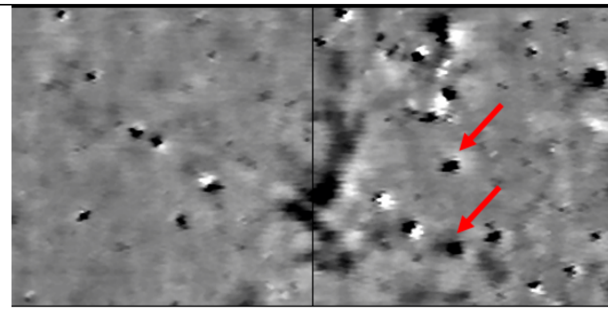
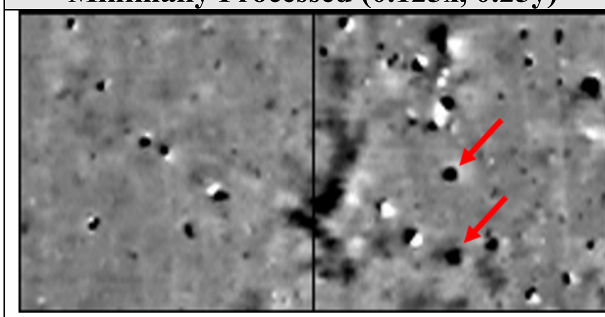
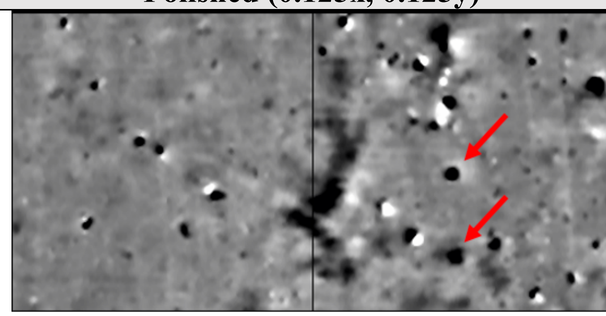
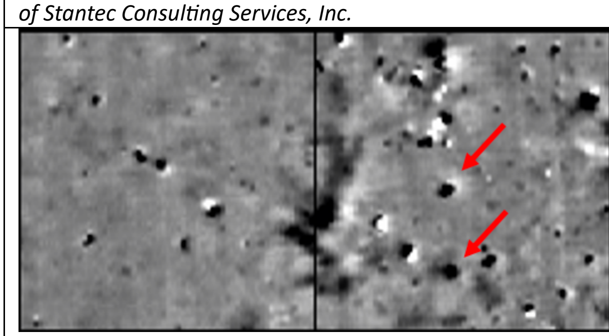
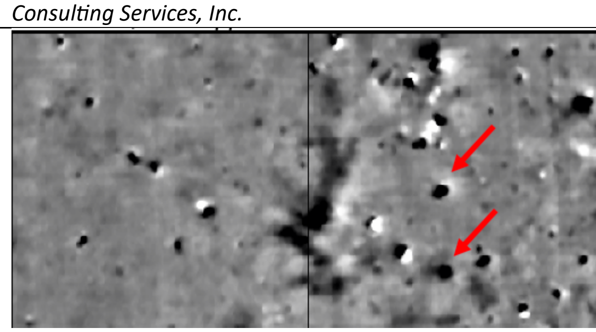
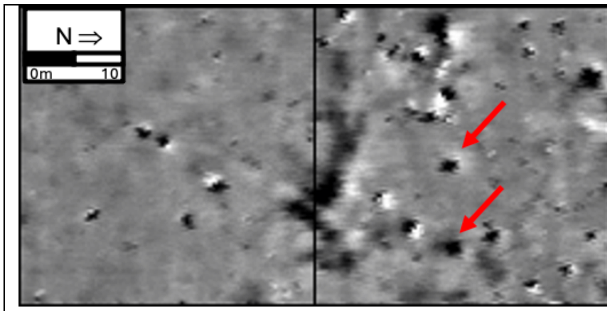
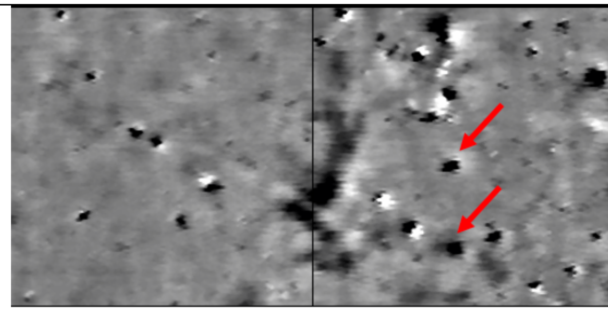
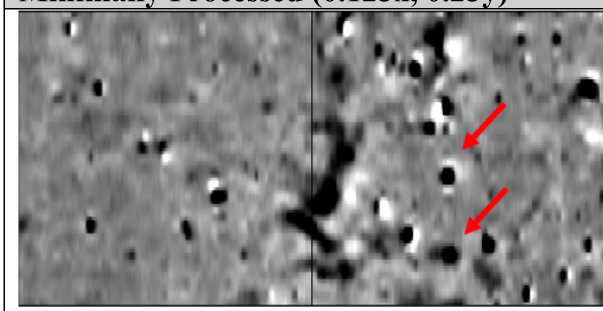
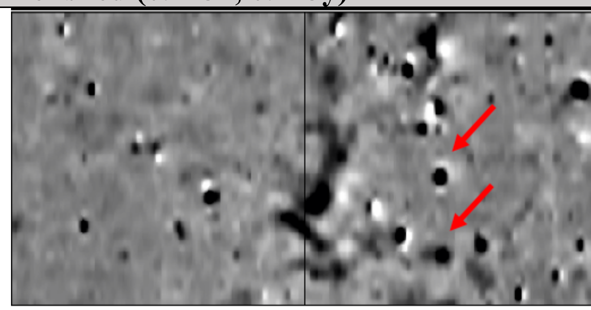
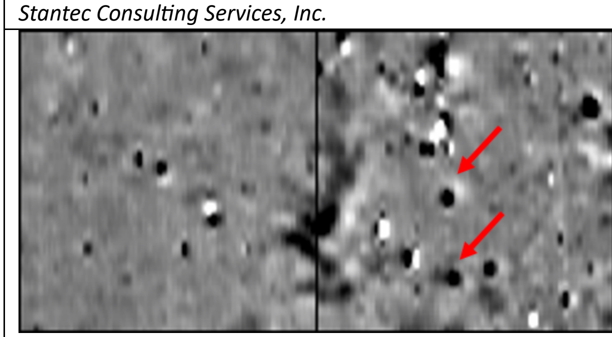
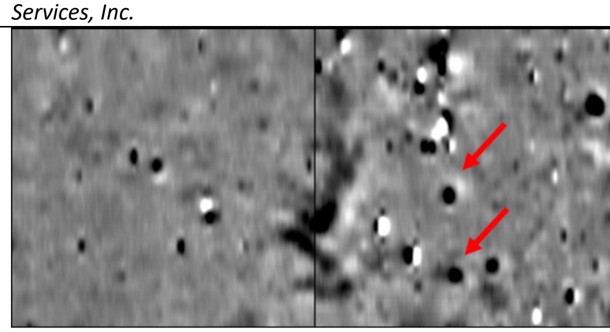
 *Dataset 1, Minimally Processed after Collection. Graphic courtesy of Stantec Consulting Services, Inc.*	 *Dataset 1, Polished. Graphic courtesy of Stantec Consulting Services, Inc.*
Minimally Processed (0.125x, 0.25y)	**Polished (0.125x, 0.125y)**
 *Dataset 2 Interpolated to 0.25m in the Y dimension to produce Dataset 4, DeStriped, and Clipped. Graphic courtesy of Stantec Consulting Services, Inc.*	 *Dataset 4, DeStriped, Interpolated to match X and Y axes, Low Passed, and Clipped. Graphic courtesy of Stantec Consulting Services, Inc.*
 *Dataset 3 Interpolated in Y axis to produce Dataset 7, DeStriped, and Clipped. Graphic courtesy of Stantec Consulting Services, Inc.*	 *Dataset 7, DeStriped, Interpolated to match X and Y axes, Low Passed, and Clipped. Graphic courtesy of Stantec Consulting Services, Inc.*

Table 9. Visual Comparison of Collected Data (Dataset 1) with Equivalent Resolution (0.125x, 0.25y) Generated Datasets 6 and 9, interpolated from 1.00-meter transect interval Datasets 5 and 8

 *Dataset 1, Minimally Processed after Collection. Graphic courtesy of Stantec Consulting Services, Inc.*	 *Dataset 1, Polished. Graphic courtesy of Stantec Consulting Services, Inc.*
Minimally Processed (0.125x, 0.25y)	**Polished (0.125x, 0.125y)**
 *Dataset 5 Interpolated Twice in the Y axis to produce Dataset 6, DeStriped, and Clipped. Graphic courtesy of Stantec Consulting Services, Inc.*	 *Dataset 6, DeStriped, Interpolated to match X and Y axes, Low Passed, and Clipped. Graphic courtesy of Stantec Consulting Services, Inc.*
 *Dataset 8 Interpolated Twice in the Y axis to produce Dataset 9, DeStriped, and Clipped. Graphic courtesy of Stantec Consulting Services, Inc.*	 *Dataset 9, Interpolated to Match X and Y axes, DeStriped, Low Passed, and Clipped. Graphic courtesy of Stantec Consulting Services, Inc.*

Unlike the datasets being visually compared in Table 6 and Table 7, those presented in Table 8 and Table 9 are not completely composed of collected datapoints; in order to compare the collected data directly and point-by-point with the lower-resolution datasets (2, 3, 5, and 8) some level of interpolation was required in order to create the same number of datapoints per dataset. Interpolation to double resolution in the Y-dimension of Datasets 2 and 3 produced Datasets 4 and 7, respectively. Interpolation to quadruple resolution in the Y-dimension of Datasets 5 and 8 produced Datasets 6 and 9, respectively.

The additional interpolation is evident in the smoothness of both the minimally-processed and 'polished' graphics for Datasets 4, 7, 6, and 9. Visually, these datasets appear to resemble one another fairly well, despite some distortion and stretching in Datasets 6 and 9, which were subject to the most interpolation to create even the minimally-processed graphics. Recall MAE values between Datasets 1 and 4 (MAE=0.39…), Datasets 1 and 7 (MAE=0.24…), Datasets 1 and 6 (MAE=0.63…), and Datasets 1 and 9 (MAE=0.57…), Datasets 4 and 7 (MAE=0.52…), Datasets 6 and 9 (MAE=0.53…). All but the former two MAE values exceed the most 'similar' MAE value to come from a MAE test between grids of equivalent size and resolution, and approximately equivalent range and statistics, which were collected on two different archaeological sites (MAE=0.41…). The grids which produced an MAE=0.41… when compared produced visually distinct graphics. How can this be?

The answer lies in what the MAE is measuring. A MAE comparison is examining the average difference between two datapoints at the same physical location in the grid, calculated across every datapoint in each grid. While the graphics produced by Datasets 1-9 do not visually differ much, their MAE values are certainly not 0. The difference is simply invisible to visual examination at the scale presented here – the MAE values in concert with the visual comparison would suggest that generated Datasets 2-9 may differ significantly but very slightly. The datapoint values themselves, and therefore the anomaly footprints themselves, are shifting slightly from dataset to dataset. Some shifting in the Y-dimension is expected due to the interpolative process in TerraSurveyor, which centers each datapoint in the center of its sampled 'tile' and then replaces each existing datapoint with two new datapoints at half the new transect interval above and below the replaced datapoint. However, offset between datasets cannot completely explain some MAE values; the comparison between Datasets 1 and 9, while only offset by 0.125 meter, has a greater MAE than the Datasets with offsets of 0.25 meter.

<u>*Comparison of Anomaly and Feature Footprints in Datasets 1-9*</u>

Magnetic Anomaly 1 and 2 Centroids and Maximum Magnetic Gradient Values, All Datasets

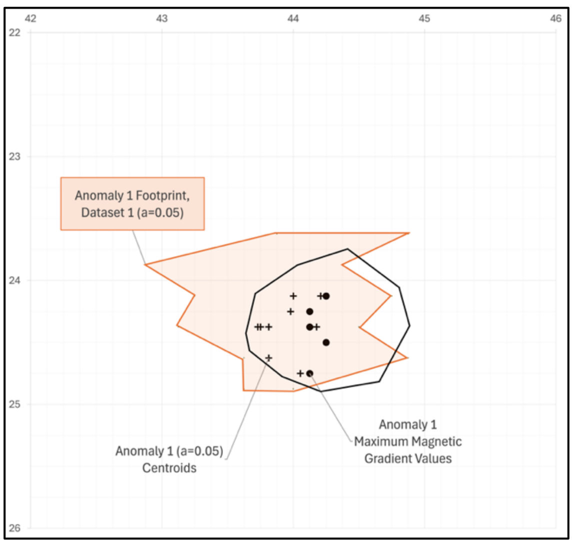


Figure 21. Anomaly 1 Centroids and Maximum Magnetic Gradient Values, All Datasets. Graphic courtesy of the author.

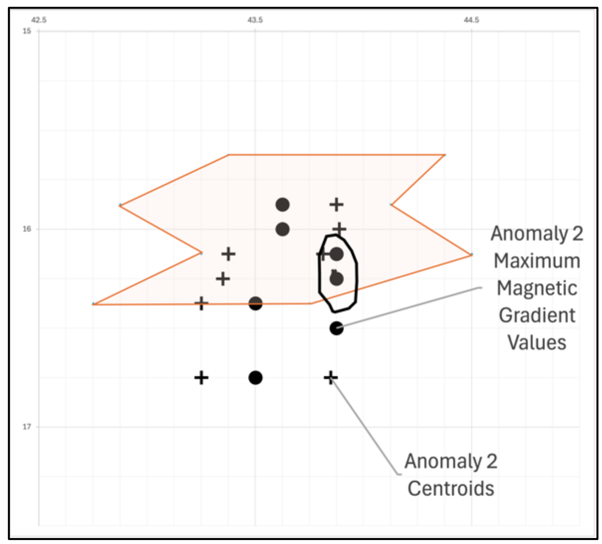


Figure 22. Anomaly 2 Centroids and Maximum Magnetic Gradient Values, All Datasets. Graphic courtesy of the author.

Features 1 and 2 (in black) overlap with the anomalies they generate (in orange; Figure 21; Figure 22). All Magnetic Anomaly 1 dataset maxima fall within the boundaries of Feature 1 (~1.2 meter diameter circle), although there is a departure of up to +0.75 meters (Dataset 9 maximum) from the original collected data. The eastward displacement of reduced-resolution and interpolated datasets is due to the way TerraSurveyor organizes and interpolates data. Datapoints are mapped in the 'center' of each tile sampled by the datapoint, and interpolation halves the size of this tile in the dimension of interpolation. The first transect of data collected at (0X, 0.125Y) and containing datapoints on transects every one meter (Dataset 5) would be imported into TerraSurveyor with the first data point at an origin point (0X, 0.50Y) and interpolations and graphics reflect this. Hence, the maxima of the interpolated datasets (6 and 9) created from the lowest-resolution clipped datasets (5 and 8) are the furthest east-ward, with the interpolated datasets (4 and 7) created from the mid-resolution (0.25Y) datasets (2 and 3) occupy positions between the collected data (Dataset 1) and Datasets 6 and 9.

Feature and Magnetic Anomaly 2 demonstrate the effect of a smaller feature and more expanded area of anomalous signature. Most dataset maxima (including that of the collected data) fall outside the physical boundaries of Feature 2 (Figure 22). Eastward displacement from interpolation is also visible in the Anomaly 2 data. If attempting to locate Features 1 and 2 by targeting anomaly maxima only, Feature 1 would likely be located even at reduced survey intervals, while Feature 2 likely would not be. The maxima of Dataset 2 falls within Feature 2, while the maxima of Dataset 3 does not – indicating that a Y-resolution, or transect interval separation, of 0.50 meter or 1.0 meter would not necessarily lead to identification of Feature 2 based on maxima targeting alone.

Transect Interval Spacing of 0.50 Meter

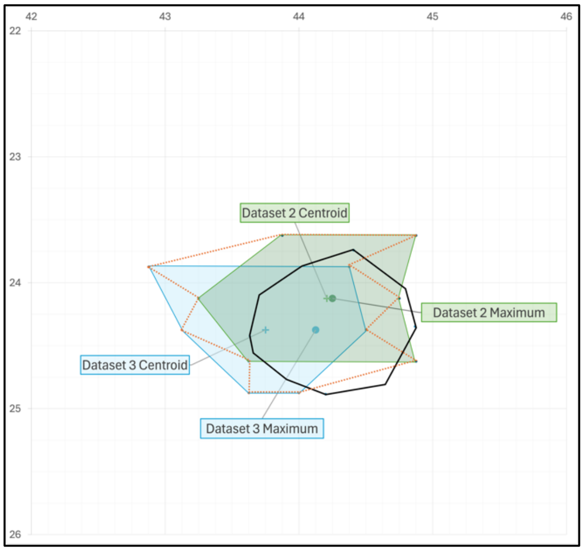


Figure 23. Anomaly 1 in Datasets 2 and 3, against Feature 1 extent and Anomaly 1 footprint from collected data, with anomaly centroids and maximum magnetic gradient values. Graphic courtesy of the author.

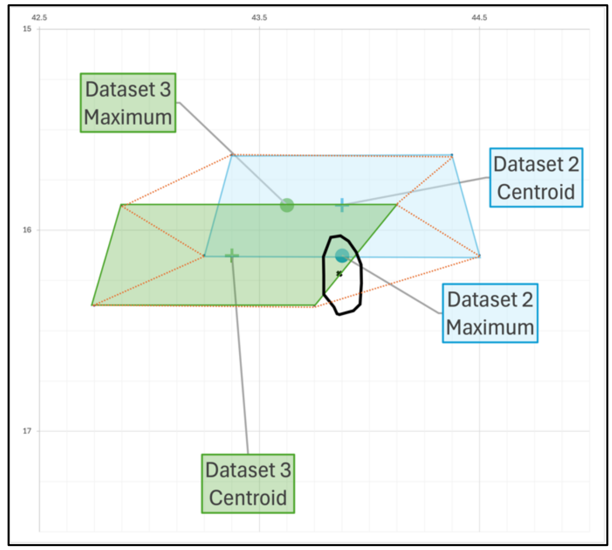


Figure 24. Anomaly 2 in Datasets 2 and 3, against Feature 2 extent and Anomaly 2 footprint from collected data, with anomaly centroids and maximum magnetic gradient values. Graphic courtesy of the author.

Here, the boundaries of Features 1 and 2 (in black) are plotted against the Dataset 1 Anomaly 1 and 2 footprints (orange dotted line) and the footprints of Magnetic Anomalies 1 and 2 contained within the 0.50-meter transect interval Datasets 2 and 3 (Figure 23; Figure 24). Datasets 2 and 3 were created by collating every other line of Dataset 1 into separate Excel/.csv files and re-importing these datasets into TerraSurveyor for graphical display and interpolation (Datasets 4 and 7) in order to provide direct point-by-point comparatives for collected Dataset 1. Magnetic Anomalies 1 and 2 appear offset from one other at 0.50-meter Y-resolution. Removing a soil core from both dataset maxima would identify Feature 1, but not necessarily Feature 2; removing a soil core at both anomaly centroids may not result in the identification of either feature, given the considerable southerly displacement of Dataset 3's Magnetic Anomaly 1 and the displacement of both dataset magnetic anomalies from Feature 2. Removing the plowzone from half of Dataset 2 and 3 magnetic anomalies would result in the positive identification of Feature 1, while testing of the entirety of Dataset 2 and 3 anomalies would be necessary to identify Feature 2.

Transect Interval Spacing of 1.0 Meter

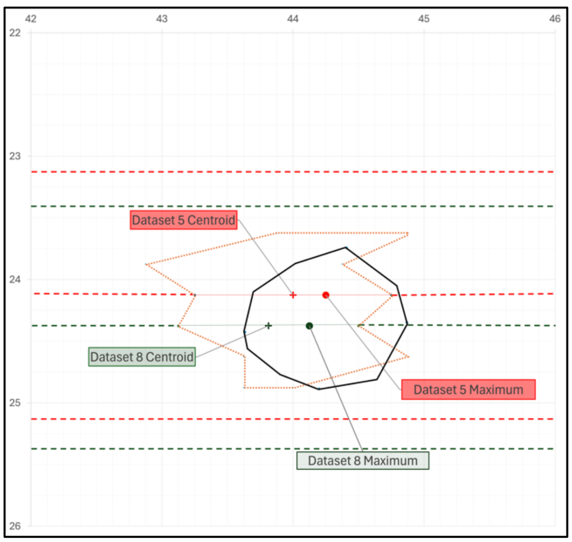


Figure 25. Anomaly 1 in Datasets 5 and 8, against Feature 1 extent and Anomaly 1 footprint from collected data, with anomaly centroids and maximum magnetic gradient values. Graphic courtesy of the author.

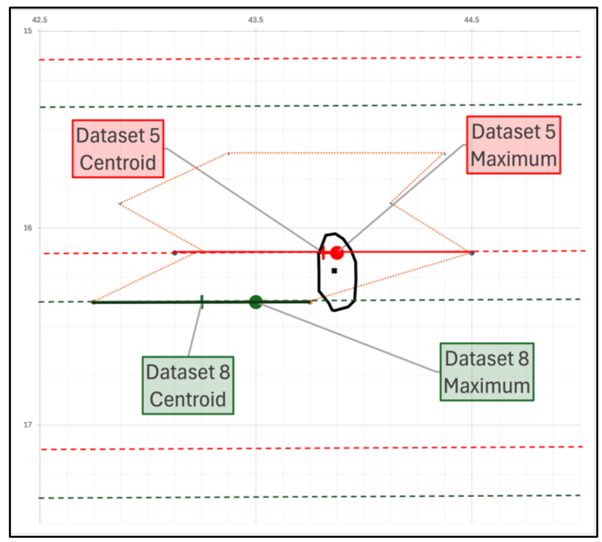


Figure 26. Anomaly 2 in Datasets 5 and 8, against Feature 2 extent and Anomaly 2 footprint from collected data, with anomaly centroids and maximum magnetic gradient values. Graphic courtesy of the author.

The boundaries of Features 1 and 2 (solid black line) and Dataset 1 Magnetic Anomalies 1 and 2 (dotted orange line) are plotted above against the detection points of both anomalies in the 1.0-meter transect interval Datasets 5 and 8 (Figure 25; Figure 26). Datasets 5 and 8 were created by collating the first rows of data from Datasets 2 and 3, plus successive rows every other row, to produce two 1.0-meter Y-resolution datasets beginning at 0.125-m-Y and 0.375-m-Y, respectively. As previously established, the Y-dimension breadth of Magnetic Anomalies 1 and 2 suggest that a 1.0-m resolution would be sufficient only for one detection; this is confirmed by one dimensional Dataset 5 and 8 detections depicted above. Testing along the entire detection period of Magnetic Anomaly 1 would be necessary in order to detect Feature 1, while testing along the detection periods of Magnetic Anomaly 2 would only result in detection in one case (Dataset 5). The one-dimensional detection gives no indication of the shape of either feature.

Transect Interval Spacing of 0.25-Meter, Interpolated

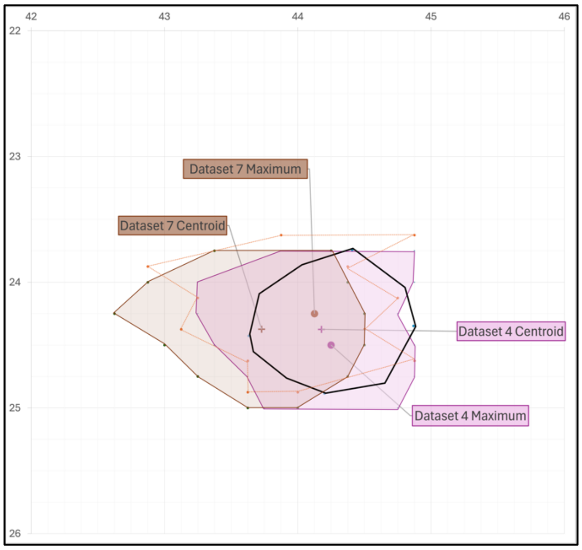


Figure 27. Anomaly 1 in Datasets 4 and 7, against Feature 1 extent and Anomaly 1 footprint from collected data, with anomaly centroids and maximum magnetic gradient value. Graphic courtesy of the author.

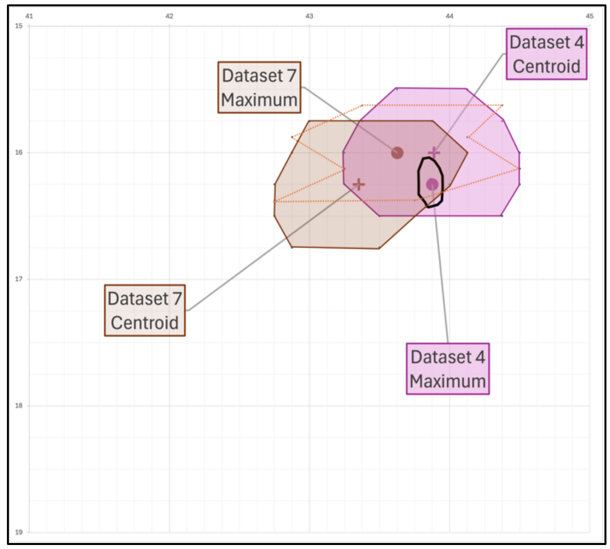


Figure 28. Anomaly 2 in Datasets 4 and 7, against Feature 2 extent and Anomaly 2 footprint from collected data, with anomaly centroids and maximum magnetic gradient values. Graphic courtesy of the author.

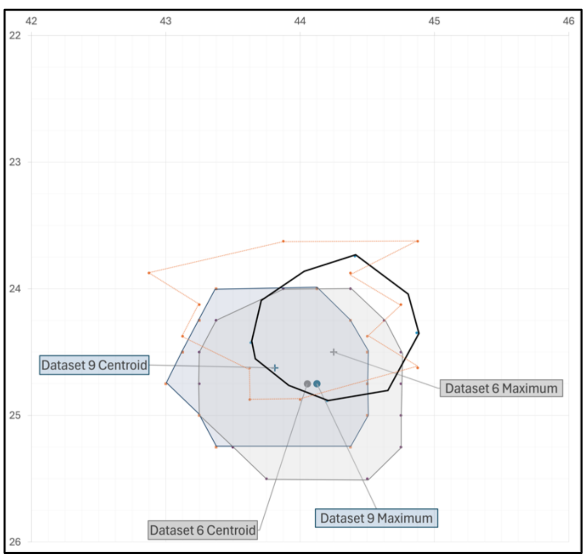


Figure 29. Anomaly 1 in Datasets 6 and 9, against Feature 1 extent and Anomaly 1 footprint from collected data, with anomaly centroids and maximum magnetic gradient values. Graphic courtesy of the author.

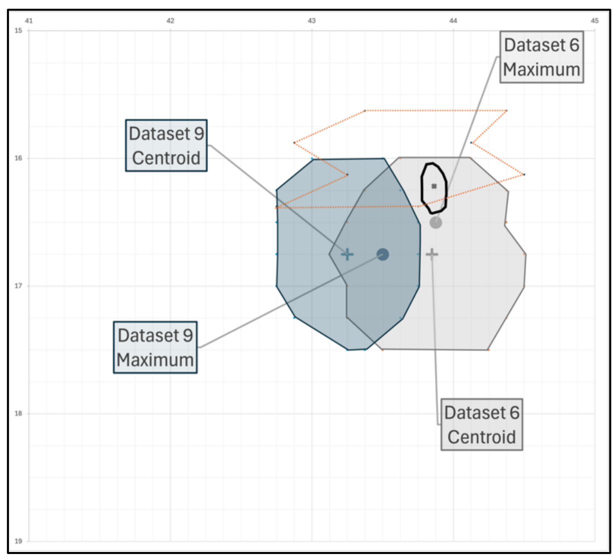


Figure 30. Anomaly 2 in Datasets 6 and 9, against Feature 2 extent and Anomaly 2 footprint from collected data, with anomaly centroids and maximum magnetic gradient values. Graphic courtesy of the author.

Datasets 4, 6, 7, and 9 are compared with Dataset 1's Magnetic Anomaly 1 and 2 footprints (orange dashed line) as well as the Feature 1and 2 extents (solid black line; Figure 27; Figure 28; Figure 29; Figure 30). These datasets were created by doubling (Datasets 4 and 7) and quadrupling (Datasets 6 and 9) the Y-resolution, or interval width, of 0.50-m-interval Datasets 2 and 3 and 1.0-m-interval Datasets 5 and 8, respectively. These datasets were created for two reasons: to provide datasets of equivalent resolution with the collected dataset in order to allow for datapoint-to-datapoint comparative evaluation (see MAE test above) and to explore the distortive effects of interpolation on lower resolution datasets. Interpolation has more distortive effects on Magnetic Anomaly 2, which is a more extreme departure from surrounding values than Magnetic Anomaly 1. The 'smoothing' effects of interpolation are also evident in the interpolated datasets, particularly when compared with the Magnetic Anomaly 2 footprint in Dataset 1.

While the easterly displacement is partially due to the data organization and interpolative processes within TerraSurveyor, the extent of its effect appears to be related to particulars of the anomalous signatures themselves. The dataset maxima experience roughly the same offset, but the anomalous footprints are stretched further east in the Feature 2 assemblage than in the Feature 1 comparison. It is not necessary to explore in detail the relationship between the interpolated dataset footprints and the feature extents, as these datasets contain entirely simulated data. However, it is worth noting that while all Feature 1 is overlapped by some portion of all four Anomaly 1 footprints, coverage is uneven. Additionally, Magnetic Anomaly 2 in Dataset 9 misses Feature 2 entirely, and due to the small size of this feature and the expanded footprint of its anomalous value range, most of the 'anomalous' area does not correspond to the Feature at all. This greatly increases the likelihood of a 'false negative' during detection efforts.

Summary of Discussion

In this chapter, the collected magnetometry data were split and/or interpolated to create eight additional datasets in order to probe the relationship between dataset Y-dimension resolution, anomaly extent and location, and archaeological feature location. These datasets and the collected data (Dataset 1) were evaluated qualitatively and quantitatively. This included location levels of effort consisting of theoretical soil cores and STUs placed over the magnetic gradient maxima or anomaly centroids. The results of these location thresholds are displayed below in Table 10 and Table 11. Note that "Barely Located" refers to detection within 10 centimeters of the anomaly edge, or detection requiring accuracy of less than 10 centimeters. *Italics* indicate interpolation was utilized to artificially bolster the transect interval resolution. As Features 1 and 2 were detected consistently across datasets, these summary tables only include results relevant to simulated location during ground-truthing.

Table 10. Summary of Anomaly 1 Location

Anomaly 1		**Soil Core**		**STU**	
Dataset	**Transect Interval**	**Maxima**	**Centroid**	**Maxima**	**Centroid**
1	0.25 m	Located	Located	Located	Located
2	0.50 m	Located	Located	Located	Located
3	0.50 m	Located	Located	Located	Located
4	*0.25 m*	*Located*	*Located*	*Located*	*Located*
5	1.00 m	Located	Located	Located	Located
6	*0.25 m*	*Located*	*Barely Located*	*Located*	*Located*
7	*0.25 m*	*Located*	*Barely Located*	*Located*	*Located*
8	1.00 m	Located	Located	Located	Located
9	*0.25 m*	*Located*	*Barely Located*	*Located*	*Located*

Table 11. Summary of Anomaly 2 Location

Anomaly 2		**Soil Core**		**STU**	
Dataset	**Transect Interval**	**Maxima**	**Centroid**	**Maxima**	**Centroid**
1	0.25 m	Not Located	Not Located	Barely Located	Located
2	0.50 m	Located	Not Located	Located	Located
3	0.50 m	Not Located	Not Located	Barely Located	Not Located
4	*0.25 m*	*Located*	*Not Located*	*Located*	*Located*
5	1.00 m	Located	Located	Located	Located
6	*0.25 m*	*Not Located*	*Not Located*	*Located*	*Not Located*
7	*0.25 m*	*Not Located*	*Not Located*	*Barely Located*	*Not Located*
8	1.00 m	Not Located	Not Located	Not Located	Not Located
9	*0.25 m*	*Not Located*	*Not Located*	*Not Located*	*Not Located*

The results above indicate that maximum magnetic gradient values are in general more predictive of feature locations than magnetic anomaly centroids (for this particular definition of anomalous magnetic gradient value). As could have been supposed, more excavation (STU vs. soil core) increases the chances of exposing any feature. The larger feature, Feature 1, which was a circular thermal pit approximately 1.30 meters in diameter, was detected at coarsest resolutions of transect intervals; it was additional interpolation which put the feature at risk classification as a 'false negative'. In this case, the feature is detected, but at wider survey intervals (0.50 meter and 1.0 meter) additional interpolations in the interval plane have the ability to compromise the precision of the detection of the feature by the least reliable method (soil core placed over magnetic anomaly centroid).

On the other hand, both interpolated and un-interpolated datasets missed the detection of Feature 2. This smaller feature, an oval-shaped thermal feature measuring 0.30 meter at its widest point, was missed by un-interpolated transect intervals wider than 0.25 meter, and the interpolated datasets confirm this pattern as expected. Even the narrowest transect interval (0.25 meter) is not precise enough for pinpoint-level detection via soil core.

Visual analysis, Mean Absolute Error (MAE) between datasets, and a comparison of feature locations and anomaly footprints for Features 1 and 2 show that using magnetometry as an archaeological prospection tool is requires consideration of planned level of effort for ground-truthing. Additionally, although this study examined only two archaeological features, these features do illustrate that feature size and magnetic strength both have effects on detection, and precise detection, limits. As feature size is unknown prior to ground-truthing, this may require more intensive levels of geophysical survey to allow for the detection and location of smaller, more magnetic features with anomalies technically detectable at lower resolution surveys.

This analysis suggests that Feature 1 should be easily located and detected even at broader resolutions than that of Dataset 1. However, this analysis also builds in no uncertainty due to machine or human error. In reality, Feature 1 was barely located using the visually-identified centroid during Stantec's initial fieldwork (Figure 31). While the STU (in red) was excavated only about 0.25 meter from the calculated magnetic anomaly centroid, it nearly missed the feature (outlined in black) entirely.

Figure 31. STU Excavation to Detect Feature 1. Photo credit to author.

This STU placement (in red, above) illustrates the importance of controlling potential error when attempting to locate magnetic anomalies while minimizing ground disturbance. It is apparent that human error can effectively counteract precision gained by increasing magnetic data density. Although Feature 1 was consistently detected and located in this analysis, even the high-resolution Dataset 1 did not guarantee its *actual* location. Conversely, Feature 2 was consistently detected, but not consistently location. It is evident that magnetic detection of magnetic anomalies and physical location of features should not be equated, as both geophysical limitations (i.e. survey resolution – Feature 2), field methods (i.e. Feature 2) and human error (i.e. Feature 1) can all impact location, even with high-resolution detection.

Chapter 7: Conclusion

Summary of Research

While some aspects of modern archaeological practice have become institutionalized through standardization, the application of geophysical techniques in CRM remains a 'craft,' as defined by some theorists (i.e. Shanks and McGuire 1996). Only one state, Ohio, provides specific guidelines for data collection using a magnetometer (OHC 2022; OHC 2023). In the absence of institutional regulation, the 'craft' of geophysics has proliferated in the US, guided by the lived experience of practitioners and academic research (Aspinall et al. 2008; Clark 1990; Conyers 2017; Kvamme 2003, 2006; Schmidt and Marshall 1997; Scollar et al. 1990).

The subversive 'craft' of geophysics has some advantages: for one, allowing for the tailoring of methods to specific scenarios, as seen here. However, lack of standardization can lead to inconsistent and unreliable methods and results, and transference of knowledge depends on individual practitioners to transfer it, something which can be complicated by age, access to individuals, heavy workloads, and the 'institutionalized' nature of CRM archaeology. Additionally, transmission of knowledge happens at the behest of the practitioner, and may therefore be transmitted unequally or incompletely. Just as standardization can lead to disenfranchisement, so can the individualized approach. Similarly, just as standardized methods may be unsuitably 'one-size-fits-all', unevaluated 'crafted' methods may be insufficient, under-tested, or incorrect.

The 'craft' nature of the sub-field provides motivation to evaluate current accepted practices in specific scenarios, not for the purposes of 'unification' but to put forward a greater and more nuanced understanding of how best to employ this technology responsibly. To this end, this thesis evaluated the effect of reduced survey resolution and subsequent interpolation on the theoretical detection of two confirmed precontact thermoremanentally-magnetized archaeological features, recorded during Phase II archaeological work carried out under contract in central Ohio. How does altered transect interval-dimension resolution affect the simulated detection and location of actual, precontact, thermal archaeological features? In a sense, this research itself exists within and furthers this 'craft' of geophysics, as it considers a technical scenario beyond what is required by current Ohio recommendations.

Both features were located via excavation of 0.50-meter square shovel test units (STUs) placed over the approximate (visually determined) centroids of the visually defined anomalous magnetic areas generated by each feature. A phased approach to ground-truthing, involving the modeled placement of soil cores and STUs, over both the maximum detected magnetic gradient value in each anomaly and the centroid of each anomaly, was simulated on the collected dataset and a series of test datasets of varying resolution.

This work showed that, in general, maximum values are more predictive of feature locations than anomaly centroids, and that the ability of a feature to be both detected and located varies due to the size of the feature when compared with the size of its corresponding field of anomalous magnetic values. Both features evaluated here had similarly sized magnetic anomalies, but different physical extents. Feature 1, which when ground-truthed measured approximately 1.2 meters and rather closely corresponded to the footprint of its anomaly, was detected and located (albeit barely in some cases) across every resolution and ground-truthing approach applied in this study. Feature 2, which was significantly smaller (0.30 meter in the Y-dimension) was only detected and located about half the time. More excavation (a STU rather than a soil core) increased the likelihood of its detection, as did focusing the ground-truthing efforts on its maximum magnetic value rather than its anomaly centroid. And while Feature 1 would likely have been detected if a lower-resolution (0.50-meter transect interval) survey was performed, Feature 2 could not have been reliably detected and located at this interval. In this case, the institutional standards of the OHC do not appear to be sufficient for the detection of and location of all archaeological features at the Test Site, particularly given human error.

Finally, this work demonstrated that the ability to geophysically detect a feature's magnetic anomaly should not be equated with the potential for physically locating an archaeological feature via excavation, using that geophysical data. The simulated location of Feature 2 was not consistent, despite its consistent detection, while the location of Feature 1, during Phase II fieldwork, was nearly compromised by human error despite high-resolution (and therefore relatively precise) magnetic data which was used to guide ground-truthing. Estimations and predictions of detection, even mathematical ones (Kryala 1964; Schmidt and Marshall 1997) cannot accurately account for the *location* of features once detected. The detection of a feature, without the ability to physically locate it, has limited utility; a feature which is only detected but not located is actually a 'false negative,' but will most likely be perceived as a 'true negative.' In equating detection with location, and only accounting for detection, the practitioner risks missing the point (literally).

Sources of Error and Assumptions

This thesis is limited by the very small sample size (n=2) of confirmed archaeological features against which to evaluate survey resolution. Further work to run comparative analyses on any additional features located via this methodology would allow for the development of more robust conclusions. Several assumptions were made during the course of this research which must be acknowledged here. First, it was assumed that georeferencing, satellite-accessing locational devices, and human error did not contribute significantly to locational error in the field, and that the in-ground locations of Features 1 and 2 are accurate as recorded and depicted.

Second, it was assumed that defining an anomaly by an alpha (α) value of 0.05 is sufficient to establish a reliable, repeatable, and objective anomalous magnetic gradient value field and centroid to guide ground-truthing efforts. This level of selection considers magnetic gradient values anomalous if they differ from the mean by more than 1.96 times the dataset standard deviation. A different threshold for 'anomalous' values would change the area covered by the anomaly footprints, which would in turn change the centroid locations but not the maximum magnetic gradient values. Centroid positions appear to be more sensitive to interpolation, and maximum magnetic gradient values appear to be slightly better at facilitating feature locating than centroids. And as this thesis has shown, detection of anomalies does not guarantee the features themselves will be located, especially in cases where the feature size is significantly smaller than the spread of anomalous values.

Third, it was assumed that the visible 'staggering' on opposing transects did not significantly affect the anomaly centroid and maxima locations; although the centroid locations should be more insulated from operator error, maxima locations may be shifted due to characteristics of the machine operator handling or gait leading to the 'staggered effect, which was not counteracted with TerraSurveyor's DeStagger function to limit overall processing. In fact, effects of operator error (uneven pacing, tilting, veering from the transect path) on the collected data was not quantified, and cannot be accounted for; it has been assumed to be insignificant for the purposes of this analysis, but this may not be the case.

Fourth, due to the high density (eight readings per meter) of in-line data collection compared with the likely error in operator pacing, the X-dimension data point locational coordinates were plotted as if each transect began at 0.0 meter, although TerraSurveyor places datapoints in the center of their survey tiles. Therefore, all X-coordinates aside from the actual locations of Features 1 and 2 are offset south by 0.0625-meter from their 'actual' locations, for simplicity, as this analysis focused on the Y-interval resolution. As this offset in the X-dimension affects every collected datapoint, the relative locations of anomaly extents are not affected, but every anomaly extent is offset south by 0.0625 meter from the actual locations of Features 1 and 2.

Finally, it should be stated that as complete excavation of the surveyed area (defined here as 100% removal of plow zone) has not been completed at the Test Site, existing 'false negatives' could persist in the dataset. Without completely excavating the Test Site, 'false negatives' cannot be entirely ruled out, and the presence of additional archaeological features which we failed to locate during fieldwork may suggest that increased data density, and/or a different ground-truthing method, would provide a more accurate characterization of the Test Site.

Anomaly Detection Is Not Feature Location

It is evident that the problem of prospecting archaeological features via magnetometry is not simply a problem of magnetic anomaly detection, which relies on data density, but of physical feature location, which depends on data precision. Methods of developing appropriate sampling strategies in magnetometry have historically relied on anomaly size, rather than feature size, to establish detectability thresholds, worsening the divide between detection and location.

Recall that Kyrala's (1964:105) expression of the probability of a successful search, meaning detection, is written as the summation:

$$P = \sum_{k} p_k F_k \tag{3}$$

where subscript k indicates the probability of the object existing in the k^{th} region of smaller search region A_k (p_k) and the probability of finding the search object (F_k) taken as a sum for all k. Kyrala does not explicitly constrain this probability to feature or anomaly, but as this analysis demonstrated, and as many have likely observed in the field, feature extent and anomaly extent are not necessarily equal. Schmidt and Marshall (1997: 334) redefined Kyrala's (1964) probability of a successful search as:

$$P_{hit} = \frac{\pi w_z^2}{4a^2} \tag{4}$$

which explicitly refers to what the authors label "zero-width" (w_z), or the width of the magnetic gradient trace when mapped in a single dimension across the most magnetically-detectable point of the feature. Schmidt and Marshall reported a 1:2 relationship between feature breadth and anomaly extent when the feature in question was a buried metal cube of side length 1.0-meter. It was this 1:2 relationship and definition of w_z which was used to determine their detection interval of less than 1.8 meters for an anomaly peak width of 2.0 meters. Schmidt and Marshall are writing of both feature and anomaly size, but only in the context of this specific feature-to-anomaly-size relationship. These scholars suggested a tighter sampling rate for an anomaly of this size after consideration of the Nyquist rate – carrying a sampling rate of twice the maximum desired frequency to ensure its detection – but this again is based on the anomaly peak, not the feature size itself, and cannot account for great discrepancies between the latter and the former. As most archaeological features are not the metal cubic object Schmidt and Marshall (1997) experimented with, their 1:2 relationship between feature and anomaly size for detection is further destabilized.

Both features used as test cases in this thesis generated magnetically anomalous trace peaks measuring approximately 2.0 meters in the transect interval (Y) dimension. Following Schmidt and Marshall's (1997) approximation, both anomalies should be detectable at the widest transect interval (1.0 meter) evaluated here, and both were detected at that Y-resolution. However, if a magnetometer is deployed as a tool of prospection, it must be able to do more than detect a feature's magnetic anomaly, it must do so in a way that allows for the feature to be located during subsequent excavation activities. As this thesis has demonstrated, one-dimensional detection may not be sufficient as a locational tool depending on planned level of effort for ground truthing and the size of the feature which generates the anomaly. Further work to record feature extents and compare these extents to detected anomalies could supply enough data to refine and rework the Schmidt and Marshall (1997: 334) formula to one predictive of location, not just detection.

Sampling intervals can be designed with these constraints in mind, and while further work is necessary to fully define the relationship between feature size, shape, makeup, and anomaly size and appearance, sampling strategies can and should explicitly define the feature profiles which are likely mathematically or physically impossible to detect at a given level of effort. This is especially key when geophysical survey is the predominant guide of archaeological evaluative efforts.

'False negatives' will occur if:

- features are not magnetically detectable;
- features are magnetically detectable but not detected;
- features are detected but not located precisely enough to be confirmed as features during ground-truthing;
- features are precisely detected, but are not located and confirmed as features during excavation because the excavation strategy (i.e. soil cores, STUs, mechanical stripping) is insufficient for the survey resolution and/or feature size;
- the feature is precisely detected, but its anomaly is not recognized as a potential feature during data interpretation;
- errors in georeferencing or geolocation lead the investigator to miss the feature, even if all of the above are circumvented;
- All of the above are circumvented, but due to human error (i.e. imprecision in fieldwork), the feature is not located during ground-truthing.

Close attention to error is necessary to avoid introducing unnecessary, controllable error during data collection and processing, but we cannot control every source of error, and practitioners must be vigilant to error hiding between ostensibly objective datasets, graphics, and statistics.

Like geophysical sampling strategies, ground-truthing methods can also be tailored: specifically, toward each project's data density and anticipated error. Rather than targeting anomaly centroids or maximums, particularly with more relatively time- and labor-intensive excavations such as the employment of an STU, anomalies could be tested radially at smaller intervals centered around the centroid or maxima using soil cores, and, if coring is not possible, STUs or augers. Testing radially from the center in all four cardinal directions allows for flexibility for user error, displacement due to placement of transects, and distortion of anomalies. Testing a single location per anomaly will increase the odds of inadvertently creating a 'false negative' in the dataset.

All fieldwork, regardless of its affiliation with the CRM industry or academia, is limited in some way by time and/or budget. Therefore, practitioners may and likely will not have time to ground-truth every detected, archaeologically-promising anomaly in geophysical datasets. This introduces subjectivity and another source of uncontrolled error into the process. Additionally, human error, instrument error, and sampling strategy have the ability to adversely affect one's ability to both detect and locate archaeological features. As such, Kyrala's (1964) definition of a successful search could be amended to theoretically account for the separate probabilities of detecting *and locating* an object (or anomaly; Kyrala's F_k) when using magnetometry as a prospecting tool such as:

$$P = \sum_k p_k F_k L_k \tag{7}$$

where subscript k indicates the probability of the object existing in the k^{th} region of smaller search region A_k (p_k) and the probabilities of detecting the search object (F_k) and locating the search object (L_k) are taken as a sum for all k.

The probability (L_k) of locating a feature is affected by making the subjective decision to test a feature, possessing the data resolution and precision to locate the feature based on its detected signature, designing a method of testing for the feature sufficient for the survey resolution (known) and feature size (unknown) and successfully controlling instrument and human error sufficiently. The inclusion of L_k above (and the example of the near-miss of Feature 1) can be taken as a reminder that while geophysical tools have the ability to produce data which appears to be highly precise, and these methods can aid us greatly in efficient and responsible archaeological fieldwork, the probability of success will be affected by factors we control (sampling strategy) and factors we can't (unexpected human error and judgement). In this sense, the probability of success is both bolstered by and hindered by the 'craft' nature of geophysics in archaeology.

Bibliography

Alabama Historical Commission (AHC)
n.d. Archaeological Investigations. Alabama Historical Commission Administrative Code, Chapter 460-X-9.

Alaska Department of Natural Resources (AKDNR)
2006 Laws and Regulations Relating to Archaeology and Historic Preservation in Alaska (Excerpts Related to Permits Only).

Anfinson, Scott
2005 SHPO Manual for Archaeological Projects in Minnesota. Minnesota Historical Society, St. Paul, MN.

Aspinall, Arnold, Chris Gaffney, and Armin Schmidt
2008 Magnetometry for Archaeologists. Geophysical Methods for Archaeology, Vol. II. AltaMira Press, Lanham.

Arizona State Historic Preservation Office (AZSHPO), Arizona State Museum (ASM), and Arizona State Land Department (ASLD)
2001 Arizona Reporting Standards for Cultural Resources.

Arkansas Historic Preservation Program (ARHPP)
2010 Appendix B of the Arkansas State Plan: Guidelines for Archaeological Fieldwork and Report Writing in Arkansas. Arkansas Historic Preservation Program. Little Rock, AR.

Association of Iowa Archaeologists (AIA)
2023 Guidelines for Archaeological Investigations in Iowa.

Atalay, Sonya, Lee Rains Clauss, Randall H. McGuire, and John R. Welch
2014 Transforming Archaeology, in *Transforming Archaeology: Activist Practices and Prospects,* Sonya Atalay, Lee Rains Clauss, Randall H. McGuire, and John R. Welch (eds.), pp. 7-28. Left Coast Press, Albuquerque.

Bartington Instruments
n.d. GRAD601 Magnetic Gradiometer System. Bartington.com.

Berezowski, Victoria, Xanthé Mallett, Justin Ellis, and Ian Moffat
2021 Using Ground Penetrating Radar an Resistivity Methods to Locate Unmarked Graves: A Review. *Remote Sensing* 13: 2880-2903.

Borgens, Amy, Russ Brownlow, Andrea Burden, Emily Dylla, Amy Goldstein, Maggie Moore, Abby Peyton, and Ann Scott
2024 Guidelines and Standards for Reports. Report Guidelines Committee, Texas Council of Archaeologists.

Burks, Jarrod and Robert A. Cook
2011 Beyond Squier and Davis: Rediscovering Ohio's Earthworks using Geophysical Remote Sensing. *American Antiquity* 76(4): 667-689.

California Department of Transportation (CALTRANS)
2014 Exhibit 5.10: Geophysical Guidance for Archaeological Resources, in *Volume 2 – Standard Environmental Reference,* pp. 1-11.

Caraher, William
2019 Slow Archaeology, Punk Archaeology, and the 'Archaeology of Care.' *European Journal of Archaeology* 22(3): 372-385.

Clark, Anthony
1990 Seeing Beneath the Soil: Prospecting Methods in Archaeology. B.T. Batsford Ltd., London.

Conyers, Lawrence B.
2017 Ground Penetrating Radar and Magnetometry for Buried Landscape Analysis. Springer.

Council of South Carolina Professional Archaeologists (CSCPA)
2005 South Carolina Standards and Guidelines for Archaeological Investigations.

Davenport, Clark G., D.L. France, T.J. Griffin, J.G. Swanburg, L.W. Lindemann, V. Tranuell, C.T. Armburst, B. Kondrateiff, A. Nelson, K. Castellano, and D. Hopkins
2001 A Multidisciplinary Approach to the Detection of Clandestine Graves. *Journal of Forensic Sciences* 37(6): 1445-1458.

David, A.

1995 *Geophysical survey in archaeological field evaluation.* English Heritage Research and Professional Services Guideline No 1.

D.C. State Historic Preservation Office (DCSHPO)
1998 Guidelines for Archaeological Investigations in the District of Columbia.

Delaware Division of Historical and Cultural Affairs (DDHCA)
2015 Archaeological Survey in Delaware. State Historic Preservation Office. Dover, DE.

Division of Historic Preservation and Archaeology (DHPA)
2022 Guidebook for Indiana Historic Sites and Structures Inventory - Archaeological Sites. Indianapolis, Indiana.

Divisions of Archaeology and Historic Preservation (DAHP)
2021 Fieldwork Guidelines for Cultural Resource Investigations. Louisiana State Historic Preservation Office.

Dudzik, Mark J., Joseph A. Tiffany, and Katherine P. Stevenson
2024 Guide for Public Archaeology in Wisconsin (Revised by: Heather Walder, Randy Dickson, and Elizabeth Leith). The Wisconsin Archaeological Survey.

DW Consulting Geophysical Data Services
2019 TerraSurveyor User Manual. Program Version 3.0.35. DW Consulting.

English Heritage
2008 Geophysical Survey in Archaeological Field Evaluation. English Heritage.

Ernenwein, Eileen G. and Michael L. Hargrave
2009 Archaeological Geophysics for DoD Field Use: a Guide for New and Novice Users. Funded by the Environmental Security Technology Certification Program. Department of Defense (DoD).

Florida Division of Historical Resources (FDHR)
2003 Module Three Guidelines for Use By Historic Preservation Professionals.

Georgia Council of Professional Archaeologists
2019 Georgia Standards and Guidelines for Archaeological Investigations.

Gramlich, V.A., C.M. Fairfield, O.Z. Mao, and B.M. Wallen

2024 Mapping Heritage: Using Ground Penetrating Radar to Rediscover Revolutionary War Heroin Margaret Corbin's Resting Place. Presented at the 2024 IEEE International Geoscience and Remote Sensing Symposium, pp. 6517-6521.

History Colorado (HC)
2024 Historical, Prehistorical, and Archaeological Resources. 8 CCR 1504-7.

Idaho State Historical Society (ISHS)
2026 SHPO Guidance. Idaho State Historical Society. Electronic document, https://history.idaho.gov/shpo/idaho-shpo-guidance/, accessed March 22, 2026.

Illinois State Historic Preservation Office (ILSHPO)
2025 Archaeological Guidelines. Interim Phase I Field Methodology and Reporting Guidelines. Illinois Department of Natural Resources (DNR).

Kansas State Historical Society (KSSHS)
2018 Kansas SHPO's Guide to Archaeological Survey, Assessment, and Reports. Kansas State Historic Preservation Office. Topeka, KS.

Kentucky Heritage Council (KHC)
2017 Specifications for Conducting Fieldwork and Preparing Cultural Resource Assessment Reports. Sanders, Thomas N. (ed). Frankfort, KY.

Kvamme, Kenneth L.
2003 Multidimensional Prospecting in North American Great Plains Village Sites. *Archaeological Prospection* 10: 131-142.

2006 Magnetometry: Nature's Gift to Archaeology. In *Remote Sensing in Archaeology: An Explicitly North American Perspective,* Jay K. Johnson (ed.), pp. 205-235.

Kyrala, A.
1964 Basic probabilistic methods in geological search. *Geophysics,* 29: 105-108.

Maine Historic Preservation Commission (MHPC)
2019 Archaeological Survey Guidelines.

Maryland Historical Trust (MHT)
2025 Standards and Guidelines for Archaeological Survey and Excavation in Maryland (Draft). Maryland Department of Planning.

Massachusetts Historical Commission (MHC)
2004 950 CMR 70.00.

Michigan State Historic Preservation Office (MISHPO)
2023 Geophysical Standards Brief Fact Sheet; in Archaeological Standards: Technical Guidance for Archaeologists in the State of Michigan.

Mississippi Department of Archives and History (MDAH)
2019 Mississippi Standards for Archaeological Practices. MDAH, Historic Preservation Division, Archaeology and Historic Sites Section.

Missouri State Parks (MSP)
n.d. Guidelines for Phase I Archaeological Surveys and Reports. State Historic Preservation Office, Missouri Department of Natural Resources.

Montana Historical Society (MHS)
2024 Montana SHPO Consultation Guide: A Handbook for Cultural Resource Review and Compliance in Montana. Montana State Historic Preservation Office. Helena, MT.

Nebraska State Historic Preservation Office (NESHPO)
2017 National Historic Preservation Act, Archeological Properties, Section 106 Guidelines. Nebraska SHPO. Lincoln, NE.

Nevada State Historic Preservation Office (NVSHPO)
2019 Nevada Site Stewardship Program Basic Training Manual.

New Hampshire Division of Historical Resources (NHDHR)
2025 Standards and Guidelines for Archaeological Investigations in New Hampshire. Department of Natural and Cultural Resources.

New Jersey Historic Preservation Office (NJHPO)
n.d. Archaeological Requirements.

New Mexico, State of
2006 New Mexico Administrative Code (NMAC) Title 4: Cultural Resources, Chapter 10: Cultural Properties and Historic Preservation; Part 16: Standards for Excavation and Test Excavation.

New York Archaeological Council (NYAC)

1994 Standards for Cultural Resource Investigations and the Curation of Archaeological Collections in New York State.

North Carolina Office of State Archaeology (NCOSA)
2023 Archaeological Investigation Standards and Guidelines for Background Research, Field Methodologies, Technical Reports, and Curation.

Office of the State Archaeologist (OSA)
2011 State Archaeologist's Manual for Archaeological Projects in Minnesota. Office of the State Archaeologist, Ft. Snelling History Center, St. Paul, MN.

Ohio History Connection (OHC)
2022 Archaeology Guidelines Supplement: Geophysical Survey.
2023 Archaeology Guidelines Section: Fieldwork.

Oklahoma Historical Society (OHS)
2023 Fact Sheet #16: Guidelines for Developing Archaeological Survey Reports in Oklahoma and Report Components. State Historic Preservation Office.

Oregon State Historic Preservation Office (ORSHPO)
2025 Oregon Archaeological Guidelines: State and Federal Laws, Research, Field Methods and Analyses, National Register of Historic Places Evaluations, and Documentation Standards (Draft).

Pennsylvania State Historic Preservation Office (PSHPO)
2021 Guidelines for Archaeological Investigations in Pennsylvania. Pennsylvania Historical and Museum Commission. Harrisburg, PA.

Poirier, David A.
1987 Environmental Review Primer for Connecticut's Archaeological Resources. Connecticut Historical Commission, State Historic Preservation Office. Hartford, CT.

Rhode Island Historical Preservation and Heritage Commission (RIHPHC)
2021 Performance Standards and Guidelines for Archaeology in Rhode Island. Historical Preservation and Heritage Commission. Providence, RI.

Robinson, Jess, Laura V. Treischmann, Scott Dillon, and Yvonne B. Basque

2017 Guidelines for Conducting Archaeology In Vermont. Vermont Division of Historic Preservation, Advisory Committee.

Schmidt, A. and A. Marshall
1997 Impact of Resolution on the Interpretation of Archaeological Prospection Data. In *Archaeological Sciences: 1995,* A. Sinclair, E. Slater, and J. Gowlett, eds. Oxford: Oxbow Books.

Schmidt, Armin, Paul Linford, Neil Linford, Andrew David, Chris Gaffney, Apostolos Sarris, and Jörg Fassbinder
2015 EAC Guidelines for the Use of Geophysics in Archaeology: Questions to Ask and Points to Consider. International Society for Archaeological Prospection (ISAP), Aerial Archaeology Research Group (AARG), ArcLand, European Archaeological Council (EAC).

Schmidt, A. and N. Sheen
1995 Depth Estimation of Archaeological Features by Power Spectrum Calculations of Geomagnetic Surveys. In *Proceedings of Archaeometry '94* (ed. A. M. Ozer), Ankara, May 1994.

Scollar, Irwin, A. Tabbagh, A. Hesse, and I. Herzog
1990 Archaeological Prospecting and Remote Sensing. Topics in Remote Sensing, Vol. II. Cambridge University Press, Cambridge.

Settle, Kathleen, Rebecca O'Sullivan, Kathleen Bindley, Kaye Grob, and Isabelle Ortt
2022a Phase I Archaeological Reconnaissance for the [Confidential] Project, Bokes Creek, Rush Creek, and Perry Townships, Logan County, Ohio. Cardno, Inc. Prepared for [Confidential Client], Contract No. E320301703. Copies available from Cardno, now Stantec, Indianapolis, Indiana.
2022b Addendum Phase I Cultural Resources Investigation for the Proposed [Confidential] Project in Bokes Creek, Rush Creek, and Perry Townships, Logan County, Ohio. Cardno, now Stantec. Prepared for [Confidential Client], Contract No. E320301703. Copies available from Stantec, Indianapolis, Indiana.

Shanks, Michael and Ian Hodder
1995 Processual, Postprocessual, and Interpretive Archaeologies, in *Interpreting Archaeology: Finding Meaning in the Past,* Ian Hodder, Michael Shanks, Alexandra Alexandri, Victor Buchli, John Carman, Jonathan Last, and Gavin Lucas (eds.), pp. 1-34.

Shanks, Michael and Randall H. McGuire
1996 The Craft of Archaeology. *American Antiquity* 61(1): 75-88.

Sitters, Julian A.
2012 Remote Sensing: A Multi-Methodological Approach to the Detection of Graves. M.A. Thesis, The University of Wyoming.

Society for California Archaeology (SCA)
2020 Fieldwork and Reporting Guidelines for Cultural Resources.

Somers, Lewis E., Michael L. Hargrave, and Janet E. Simms
2003 Geophysical Surveys in Archaeology: Guidance for Surveyors and Sponsors. USACE Engineer Research and Development Center. SRDC/CERL SR-03-21.

South Dakota State Historical Society (SDSHS)
2023 South Dakota Guidelines for Complying with Federal and State Preservation Laws. State Historic Preservation Office, South Dakota State Historical Society. Pierre, SD.

State Historical Society of North Dakota (SHSND)
2026 North Dakota SHSND/SHPO Guidelines. Bismarck, ND.

State of Hawaii State Historic Preservation (HISHP)
2002a Rules Governing Standards for Archaeological Inventory Surveys and Reports.
2002b Rules Governing Standards for Archaeological Data Recovery Surveys and Reports.

Tennessee Department of Environment and Conservation (TDEC)
2024 Tennessee SHPO Standards and Guidelines for Archaeological Resource Management Studies.

Trader, Patrick
2021 Guidelines for Phase I, II, and III Archaeological Investigations and Technical Report Preparation. Joanna Wilson, ed. West Virginia State Historic Preservation Office.

Trigger, Bruce
2006 Studying the History of Archaeology, in *A History of Archaeological Thought,* Bruce Trigger, ed., pp. 1-39. Cambridge University Press, Cambridge.

Utah State Historic Preservation Office (USHPO)

2024 Cultural Resource Compliance Guidance. USHPO, Department of Cultural and Community Engagement, State of Utah.

Virginia Department of Historic Resources (VHDR)
2017 Guidelines for Conducting Historic Resources Survey in Virginia. VDHR, Richmond, VA.

Walker, Mark, and Dean J. Saitta
2002 Teaching the Craft of Archaeology: Theory, Practice, and the Field School. *International Journal of Historical Archaeology* 6(3): 199-207.

Washington State Department of Archaeology and Historic Preservation (WADAHP)
2018 Washington State Standards for Cultural Resources Reporting. Olympia, WA.

Wyoming State Historic Preservation Office (WYSHPO)
2026 Class I Report Guidelines. Wyoming SHPO. Cheyenne, WY.