Geographic Information Systems Correlation Modeling as a Management Tool in the Study Effects of Environmental Variables’ Effects on Cultural Resources

Brian Wallace
Boise State University
GEOGRAPHIC INFORMATION SYSTEMS CORRELATION MODELING AS A MANAGEMENT TOOL IN THE STUDY EFFECTS OF ENVIRONMENTAL VARIABLES’ EFFECTS ON CULTURAL RESOURCES

by

Brian N. Wallace

A Project
Submitted in Partial Fulfillment
Of the Requirements for the Degree of
Master of Applied Anthropology
Boise State University
June 2011
The project presented by Brian N. Wallace entitled Geographic Information Systems Correlation Modeling as a Management Tool in the Study Effects of Environmental Variables’ Effects on Cultural Resources is hereby approved:

Mark G. Plew  
Advisor  
Date

John Ziker  
Committee Member  
Date

Margaret A. Streeter  
Committee Member  
Date

John R. (Jack) Pelton  
Dean of the Graduate College  
Date
DEDICATION

To my parents

Jerry and Cheryl Wallace
ACKNOWLEDGMENTS

The author would like to express his appreciation to those individuals and entities who contributed to the completion of this thesis. Dr. Mark Plew is owed a great debt of gratitude for his invaluable advice, guidance, and friendship throughout my graduate experience. It is because of his helpful suggestions both for this paper and his mentorship over the last few years that have allowed me succeed beyond my own expectations. To the Cultural Resource and GIS staff of the Boise District of the Bureau of Land Management (BLM) whom allowed for access to the necessary data to run analyses successfully and for helpful guidance along the way. Chris Willson of BSU whom always offered helpful discussions related to GIS and how we apply them in today’s Archaeology. Dr. Chris Hill of BSU offered helpful ways to address all elements of this project and to think more critically of how to approach this undertaking. To the other members of my thesis committee, Dr. John Ziker and Dr. Margaret Streeter, a deep thanks for their support, guidance, and patience both during the writing of this and throughout the years of graduate school. I appreciate the friends and family who stood by and offered assistance in so many forms during the writing of this paper and through graduate school as a whole. A deep appreciation goes to my sister Angie, who throughout life has always been there and helped me in so many ways during this time. Finally, I would like to express a deep gratitude for my friend Danny whose expertise in GIS and a willingness to help was at times the only source which kept me on track to see this come to a conclusion.
ABSTRACT

Utilizing Geographic Information Systems (GIS) offers the field of Cultural Resource Management greater capacity in managing resources. New regression analysis tools recently released in ESRI ArcGIS software offer potential for determining more accurate statistical analyses of the relationships between cultural material and environmental variables. The contemporary trend of federal cultural resource managers and GIS analysts working with smaller budgets is to allocate fiscal resources for tools which will enable them to continue successfully managing their resource. ArcGIS software continues to be the industry standard in managing spatial data to accurately represent the existence, condition, and location of cultural material. With the recent inclusion of Ordinary Least Squares (OLS) regression and Geographic Weighted Regression (GWR) tools in the ArcGIS software package, the potential exists for GIS archival project to go beyond data collection and management to accurately analyze spatial relationships.

This focus of this project is twofold. First, to provide a list of operational processes, to assess the feasibility of applying these new analysis tools to create a statistical significant correlation modeling for better management of cultural resources by improving the ability to locate sites. Second, this study will focus on testing the use and applicability of existing data within the Great Basin physiographic region where sample areas have been constructed, and assess their value towards modeling projects.
# TABLE OF CONTENTS

PROJECT APPROVAL PAGE................................................................. i

DEDICATION.......................................................................................... ii

ACKNOWLEDGMENTS........................................................................ iii

ABSTRACT.......................................................................................... iv

TABLE OF CONTENTS......................................................................... v

LIST OF FIGURES............................................................................... vii

LIST OF TABLES................................................................................ ix

CHAPTER ONE: INTRODUCTION....................................................... 1
   Correlation Modeling Background................................................... 2
   Study Area......................................................................................... 5
   Environmental Setting ....................................................................... 5
   Study Goals....................................................................................... 9

CHAPTER TWO: GIS ANALYSIS.......................................................... 11
   Value................................................................................................ 11
   Regression Analysis......................................................................... 12
   Ordinary Least Squares Regression.................................................. 13
   Geographic Weighted Regression....................................................... 16

CHAPTER THREE: PROJECT METHODS............................................. 19
   Data Acquisition............................................................................... 19
LIST OF FIGURES

Figure 1. Map of General Study Area ................................................................. 6
Figure 2. Elevation Map of Study Areas .............................................................. 7
Figure 3. Study area grid of one-km-squared cells ................................................. 24
Figure 4. OLS tool window to run regression analysis ........................................... 29
Figure 5. OLS results for multivariate analysis ..................................................... 31
Figure 6. Multivariate results window of Spatial Autocorrelation Moran’s I tool .... 34
Figure 7. Scatter Plot Matrix window for one-km-squared multivariate analysis .... 36
Figure 8. OLS results for multivariate analysis with site locations ....................... 38
Figure 9. One-Kilometer-squared and one-acre-squared study areas topographic overlay ........................................................................................................ 45
Figure 10. OLS results for independent distance-to-water and dependent variables .................................................................................................................. 46
Figure 11. OLS results for independent elevation and dependent variables ............. 49
Figure 12. OLS results for independent slope and dependent variables ................. 52
Figure 13. OLS results for independent aspect and dependent variables ............... 55
Figure 14. One-acre-squared Moran’s I tool of distance to water and site variables .................................................................................................................. 60
Figure 15. One-acre-squared Moran’s I tool of aspect and site variables ............... 62
Figure 16. One-acre-squared Moran’s I tool of slope and site variables ............... 64
Figure 17. One-acre-squared Moran’s I tool of elevation and site variables ........... 66
Figure 18. One-acre-squared Moran’s I tool of soil and site variables ................. 68
Figure 19. One-acre-squared Moran’s I tool of multivariate variable analysis

Figure 20. Scatter Plot Matrix window for one-acre-squared multivariate Analysis
LIST OF TABLES

Table 1. Post-OLS process results window to indicate model performance………. 15
Table 2. OLS results table for one-kilometer-squared multivariate analysis .......... 32
Table 3. GWR results window for one-acre-squared multivariate analysis.......... 40
Table 4. OLS results table for one-kilometer-squared distance-to-water analysis…. 47
Table 5. One-kilometer-squared Moran’s I tool of distance-to-water and site variables……………………………………………………………….. 48
Table 6. OLS results table for one-kilometer-squared elevation analysis.......... 50
Table 7. One-kilometer-squared Moran’s I tool of elevation and site variables…… 51
Table 8. OLS results table for one-kilometer-squared slope analysis……………… 53
Table 9. One-kilometer-squared Moran’s I tool of slope and site variables……… 54
Table 10. OLS results table for one-kilometer-squared aspect analysis………….. 56
Table 11. One-kilometer-squared Moran’s I tool of aspect and site variables…… 57
Table 12. OLS results table for one-kilometer-squared soils analysis…………….. 58
Table 13. OLS results table for one-acre-squared analysis for distance-to-water…. 59
Table 14. OLS results table for one-acre-squared analysis for aspect and site variables………………………………………………………………….. 61
Table 15. OLS results table for one-acre-squared analysis for slope and site variables……………………………………………………………….. 63
Table 16. OLS results table for one-acre-squared analysis for elevation and site variables………………………………………………………………….. 65
Table 17. OLS results table for one-acre-squared analysis for soils and site variables………………………………………………………………….. 67
Table 18. OLS results table for one-acre-squared for multivariate analysis……….. 69
CHAPTER ONE: INTRODUCTION

In January, 2009, Environmental Systems Research Institute (ESRI) released two new spatial tools for the ArcGIS software package—Ordinary Least Squares Regression (OLS) and Geographical Weighted Regression (GWR). The introduction and integration of these spatial statistics tools into the software increases users’ ability to conduct spatial regression analysis by allowing research to be driven within a single commercialized platform. Prior to these tools, research relied on other aspatial statistical programs (SAS, SPSS) to conduct analysis before applying the results to a GIS spatial platform. This approach inevitably required a significant background in statistics, familiarity with statistical software programs, and access to expensive software to apply the research. Many federal, state, and private resource specialists lack the ability to adhere to these criteria.

This study aims to induce the feasibility of applying OLS and GWR statistical tools toward understanding relationships between archaeological site location and the environment using existing data currently available to authorized researchers. To predict the location of cultural material with a higher degree of accuracy, while integrating both spatial and statistical processes into one software program offers value to the archaeological and scientific community as a whole. As the need of the cultural resource scientist to use spatial data trends towards mandatory practice, integrating tools and processes more efficiently managing resources becomes increasingly important.
Though this study has been designed with cultural resources in mind, the principles of OLS and GWR transcend all disciplines and can be effectively utilized to further understand any spatial relationships between properly specified dependent and independent variables thus extending usable application well beyond the scope of this project.

By evaluating the operational feasibility of applying these spatial tools will, as a byproduct, provide an operational list of processes. This “recipe” of processes has been designed to create, test, and apply results in analyzing statistical significance and can be replicated for use in future applications.

**Correlation Modeling Background**

Predictive models are tools for projecting known patterns or relationships into unknown times or places. As Robert Warren and David Asch point out, such models are useful in archaeology. To date only a fraction of the world’s archaeological sites have been discovered and documented on the landscape. One method of understanding and protecting cultural resources is to create formal models capable of predicting site locations (Warren and Asch 2000:6).

The use of Geographic Information Systems (GIS) in predictive correlation modeling has been available to archaeologists for over thirty years. During this time the availability of this powerful tool has significantly increased the ability of researchers to focus on geographical areas with a higher probability of cultural material. However, inevitably, since the introduction of GIS, varying opinion has existed within the
community regarding the methods necessary for successful modeling applications, and more importantly about what questions can be answered with modeling efforts.

The need for utilizing such a powerful tool as GIS goes undisputed; however, debate regarding the direction and manner in which such research should be applied does exist. The effectiveness of predicting past culture through the location of cultural material is the source of such debate. James Ebert and Timothy Kohler argue caution in undertaking correlation modeling: trying to generalize where prehistoric people lived from the location of discarded materials circumvents the explanatory framework of natural post-depositional processes, ecosystem variability, and consideration for the underlying systems which drive human behavior (Ebert and Kohler 1988:127). To a degree this argument has merit. To effectively understand and predict where humans once existed and consequently the cultural material remains of their presence, an archaeologist must keep in mind the effects of post-depositional environmental processes which can alter the location of material, and understand the underlying behavioral systems which drove people across the landscape. This is best understood by the study of cultural material within its context after it has been discovered. However, it is not necessary to apply all process and systems data to effectively locate cultural material.

It is argued by many (Kvamme 1990; Thoms 1988; Warren and Asch 2000), including the researcher, that although limited explanatory power exists in determining human systems through locating cultural material, the application of independent environmental variables to explain existence correlation with archaeological artifacts increases the probability of discovery. The search for spatial correlations has identified many key environmental variables useful in predicting site locations. Knowledge of
which environmental settings are likely to contain certain types of sites increases the probability of locating cultural material (Thoms 1988: 637).

Presently the need for federal and state agencies to continue effectively managing cultural resources under increasingly restrictive budgetary constraints has caused many to seek out innovative approaches. The use of GIS and more specifically the new OLS and GWR applications will effectively narrow the focus on areas with a higher likelihood of cultural activity. By narrowing the scope of field efforts to areas of increased probability of site location it will afford researchers more resources to further study the context of newly discovered site.

The scope of this project in its exploratory stages is to collect existing cultural and environmental data available to resource specialists and apply them to the OLS and GWR tools in ArcGIS. Perfecting a properly specified correlation model could take years where statistical significance and reliability from one study area to another can be trusted to return relatively similar results and in many cases requires data collection and processing beyond the scope of this project. The use of existing data for this study does have its limitations; however, in the initial stages of exploring these new tools it is advisable to begin with what is known. In its exploratory stage this study will defer the use of systems theory and environmental processes data for future study. The principal aim of this study is to create a foundation by designing a usable model with existing data, test its application using tools which recently have become available in ArcGIS, and provide an operational process “recipe” for further research into this approach.
Study Area

To analyze the effectiveness of OLS and GWR two study areas were compared. The first: a rectangle comprising 1400 one-kilometer-squared grid cells. The creation of this area was determined necessary to provide a macroscopic viewpoint for the relationship between the independent variable and dependent variables. The study area measures an arbitrary 35 cells horizontally by 40 cells vertically.

A second study area was created using three 640-acre section rectangles, each comprising one-acre squared grid cells. This study grid was conceived to offer a small-scale viewpoint of the correlation relationship permitting analysis of variable values at a higher rate over space. These three smaller topographic section-sized grids were arbitrarily created within the larger 1400 kilometer-squared sized sample area. All geographic computations were done in NAD 1983 Zone 11 geographic coordinate projection. Creating the two arbitrary separate study areas for analysis was deemed necessary since this project is exploratory and is not otherwise in conjunction with other analyses.

Environmental Setting

The study area boundary (see Figure 1-2) falls within what is regionally known as the Western Snake River Plain, part of the Columbia Intermontane Province and the High Lava Plains subprovince (Freeman, Forrester, and Lupher 1945). Major water features that intersect the study area include Battle Creek, Big Jacks Creek, Big Springs Creek, Blue Creek Camas Creek, Dickshooter Creek, Dry Creek, Little Jacks Creek, Lost Valley Creek, Pole Creek, Rattlesnake Creek, Rock Creek and Shoofly Creek. Many of the
features and depositional content within the geographical area are largely a result of the Bonneville Flood event which occurred roughly 14,500 years ago (Malde, 1968). This more recent geologic event was caused by the breaching of the Red Rock Pass divide near Pocatello, which released a monumental discharge of water-cutting canyons, stripped land surfaces, and deposited sand, gravel, and boulder bars within the canyon and other drainage systems (Jenks, Bonnichsen, and Godchaux, 1998).

Figure 1. General Study Area Location in southwestern Idaho.
Figure 2. An elevation overlay of both Study Area (1 KM, 1-acre) grids.
The geology of the Western Snake River Plain began to be formed around 12 million years ago through numerous eruptions of rhyolite and later basalt flows (Hackett and Bonnichsen 1995: 37-38). Evidence of these early eruptions can be found extensively throughout the region where rhyolite and basalt surface deposits dominate the landscape.

Vegetation within the study area includes many species which have adapted to the arid climate of the region. Dominant in the biosphere are varieties of small and large sagebrush (*Artemisia Tridentata*). Perennial grasses are common and include native bluebunch wheatgrass (*Agropyron spicatum*), Sandberg bluegrass (*Poa secunda*), needlegrass (*Stipa spp.*) and Indian Ricegrass (*Oryzopsis hymendoides*). Other flora which exist in the study area include willows (*Salix*), cottonwoods (*Populus*), sedges (*Carex*), rushes (*Juncus*), and chokecherries (*Prunus virginiana*) found in riparian areas. Other important vegetative species found regionally include camas (*Camassia quamash*) and balsamroot (*Balsamorhiza sagitata*) (Daubenmire 1952).

Other invasive vegetative species found within the study area were introduced more recently and have no prehistoric context but have spread rapidly over the biome affecting site visibility. These include cheat grass (*Bromus tectorum*), medusa head (*Taeniatherum caput-medusae*), leafy spurge (*Euphorbia esula*) and other invasive exotic species.

Prior to EuroAmerican settlement, the vegetative biosphere was dominated by big sagebrush (*Artemisia tridentate*) and other various desert shrubs and grasses covering the Lower Snake River Plain (Yensen, 1982).
A diverse fauna population is found in and around the study area. Principal large mammal species include mule deer (*Odocoileus hemionus*), bighorn sheep (*Ovis Canadensis*), pronghorn antelope (*Antilocapra Americana*), mountain lion (*Puma concolor*), coyote (*Canis latrans*), badger (*Taxidea taxus*), river otter (*Lutra canadensis*), beaver (*Castor canadensis*), and elk (*Cervus canadensis*), along with other smaller mammal species such as the desert black-tailed jack rabbit (*Lepus californicus*), Townsend’s ground squirrel (*Citellus townsendii*), and Townsend’s pocket gopher (*Thomomys townsendii*). Prominent bird species include greater sage-grouse (*Centrocercus urophasianus*), prairie falcon (*Falco mexicanus*), golden eagle (*Aquila chrysaetos*), Brewer’s sparrow (*Spizella breweri*), and sage thrasher (*Oreoscoptes montanus*) (Bureau of Land Management, 2010). The study area is a diverse community of mammals, birds, and reptile species.

**Study Goals**

The primary goals of this analysis are to determine the possibility and the feasibility of applying existing resource and environmental data to newly released spatial statistical tools within ArcGIS to:

- utilize existing spatial and cultural resource data in ArcGIS to identify what, if any, independent environmental variables can be shown to have a relationship with the dependent site location variable using;
- use these existing data to determine what degree of relationship the newly released Ordinary Least Squares Regression and Geographic Weighted Regression tools can provide if any;
○ determine degree existing data can effectively be used to run regression analysis. And if “manipulation” of existing data is deemed necessary, what steps need to be done to meet this goal;

○ ascertain what degree of statistical significance the results from OLS and GWR modeling provide a useable quantitative correlation between the environment and site location;

○ and, to discover if the application of OLS and GWR as spatial statistical tools is a feasible option for modeling efforts.
CHAPTER TWO: GIS ANALYSIS

Value

The question inevitably arises as to the value of correlation “predictive” modeling and its place in the field of archaeology. Considering the nature of both highly varied approaches and highly varied results, this is a valid question. Agencies may be willing to support development of modeling applications but not new research into methods or interests in interpreting results. Many efforts which are applied do not get published, and there exists a similarity to approaches which suggests a lack of innovation beyond procedures established in earlier efforts (Kvamme, 2006:4).

Integrating these new regression tools into the heavily utilized ArcGIS software suite aims to address this issue by offering a new-found capacity for spatial correlation and predictive analysis; however, the results of these analyses must be treated carefully during initial exploratory stages. Applying OLS and GWR methods, though powerful in calculating data quantitatively, cannot be considered an alternative to conducting actual fieldwork.

This study produces a foundation for work in the field of spatial correlation modeling. Considering the budgetary and logistical challenges many field efforts face, it is believed these types of spatial modeling efforts have wide application within the community.

The scope of this project is targeted to help federal and state cultural resource managers explore innovative approaches in meeting their program needs. With
decreasing fiscal appropriations it has become increasingly more challenging to meet the management criteria expected to effectively manage a resource. Given this trend of shrinking operational resources it is necessary to apply new methods of research to anticipate future needs. In this aim this project is believed to have value.

Rarely do federal agencies employ full-time statisticians in district and field offices responsible for properly managing environmental and cultural resources. However, as needs and laws continue requiring more accurate data collection, use, and representation, agencies have increasingly invested in GIS personnel.

A benefit of incorporating these new regression tools within ArcGIS is the expanded capacity of the well-trained analyst to apply their experience and perform similar regression studies. The value of such applications extends beyond the aims of cultural resource management but is applicable to a number of sciences, disciplines, and subjects.

Though this application represents the latest in a long line of regression analysis techniques, the presence of OLS and GWR within a spatial platform such as ArcGIS will help permit further study. An approach which makes regression analysis accessible and usable in ArcGIS and offer statistical and quantitative access to test data without the need for an extensive background in statistics or the use of other statistical software offers research value well beyond this study.

**Regression Analysis**

Regression analysis is the backbone of predictive correlation modeling. Regression is used to analyze relationship variables. Regression, simply defined,
describes the linear relationship between a dependent variable and independent variables. Whereas the independent or predictor variables were plotted on the x-axis and the response or dependent variable plotted on the y-axis line, be it a positive or negative, would indicate such a relationship (Gotelli and Ellison, 2004: 239). Regression analysis allows one to understand how the value of the dependent variable changes when any one of the independent variables is altered.

**Ordinary Least Squares Regression (OLS)**

In a simple bivariate regression analysis, Ordinary Least Squares regression, as used in ArcGIS, shows the spatial relationship between variables- the independent variable, “x,” the predicting variable, and the dependent variable, “y,” the variable to be predicted. If the relationship between y and x can be determined using observed values, then it is possible to get the predicted value of y for any value of x (Mitchell 2005: 212). Multivariate regression analysis assesses the relationship between two or more independent variables and the dependent variable. This particular study is an example of a multivariate regression in which multiple independent variables (slope, aspect, distance to water etc.) are measured against the dependent variable (site location).

Questions regarding which influence OLS regression include the following:

- Does a linear relationship exist between the dependent and each independent variable?
- Do the residuals have a mean of “0”, or do the overestimates balance out the underestimates?
Do the residuals have a constant variance at all locations in the study area, or, rather, homoscedasticity?

Are the residuals randomly arranged along the regression line, rather than creating any pattern?

Are the residuals normally distributed?

Are the independent variables highly correlated? A high correlation between two or more independent variables it is called multicollinearity. When this occurs the analysis will return unreliable coefficient estimates and large standard errors or high variation in results between one sample and the next (Mitchell, 2005: 217).

Ordinary Least Squares (OLS) regression is the primary regression tool within the ArcGIS Spatial Statistics Toolbox. It is also the proper starting point for all spatial regression analyses. It provides a global model of the variable or process being researched and is capable of analyzing the bivariate relationship between the dependent, “y”, variable and each dependent, “x”, variable or the multivariate relationship between multiple independent variables and the dependent variable.

After the OLS tool is run, a results window indicates specific criteria to assess the model’s performance (Table 1).
Table 1: Post-OLS process results window to test model performance

In assessing the model’s performance, six indicators allow the researcher the necessary feedback to determine whether or not the analysis was successful:

1. Model performance. Both the Multiple R-Squared and Adjusted R-Squared values are measures of model performance. Possible values range from 0.0 to 1.0. The Adjusted R-Squared value is always a bit lower than the Multiple R-Squared value because it reflects model complexity (the number of variables) as it relates to the data and consequently more accurately measures model performance.

2. Explanatory variables in the model: Coefficient, Probability or Robust Probability, and Variance Inflation Factor (VIF). The coefficient for each explanatory variable reflects both the strength and type of relationship the explanatory variable has to the dependent variable (positive vs. negative correlation).
3. Model significance. Both the Joint F-Statistic and Joint Wald Statistic measures overall model statistical significance. The Joint F-Statistic is trustworthy only when the Koenker (BP) statistic is not significant. If the Koenker (BP) statistic is significant, one should consult the Joint Wald Statistic to determine overall model significance. The null hypothesis for both of these tests is that the explanatory variables in the model are NOT effective. To reach a 95 percent confidence level, a p-value (probability) smaller than 0.05 indicates a statistically significant model.

4. The Koenker (BP) Statistic is a test to determine whether the explanatory variables in the model have a consistent relationship to the dependent variable both in geographic space and in data space. When the model is consistent in geographic space, the spatial processes represented by the explanatory variables behave the same everywhere in the study area (the processes are stationary). When the model is consistent in data space, the variation in the relationship between predicted values and each explanatory variable does not change with changes in explanatory variable magnitudes (there is no heteroscedasticity in the model). The null hypothesis for this test is that the model is stationary. To reach a 95 percent confidence level, a p-value (probability) smaller than 0.05 indicates statistically significant heteroscedasticity and/or nonstationarity. Regression models with statistically significant nonstationarity are often good candidates for Geographically Weighted Regression analysis.

5. Model bias. The Jarque-Bera statistic indicates whether or not the residuals (the observed/known dependent variable values minus the predicted/estimated values) are normally distributed. The null hypothesis for this test is that the residuals are normally distributed, so if a histogram of those residuals were constructed, it would resemble the classic bell curve, or Gaussian distribution. When the p-value (probability) for this test is small (smaller than 0.05 for a 95 percent confidence level, for example), the residuals are not normally distributed, indicating model misspecification (a key variable is missing from the model).

6. Spatial autocorrelation. The Moran’s I tool located in the Spatial Statistics toolbox assesses whether the residuals are not spatially autocorrelated (ESRI, 2010).

**Geographic Weighted Regression (GWR)**

After running the OLS tool to determine a “global” relationship value between variables, the next step is to apply these dataset outputs in GWR. Applying GWR analysis to regression analysis is important in further deducing the “local” relationship between the dependent and independent variables. As previously discussed, global statistics offer insight into a correlative relationship by deducing the average relationship between variables. However, this average relationship may not be representative of the
analysis in any specific geographical area and could possibly mask interesting and important local differences in the relationship between variables in what is being studied (Fotheringham, Brunsdon, and Charlton, 2002: 2). Models calibrated with data equally weighted across the study area constitute a global perspective which consequently would yield global estimates on the relationship, requiring explanation of the effect of explanatory variables on the variable which you are trying to explain.

Local statistics offer a better view of this relationship and therefore a statistical approach beyond a study in which variables are represented by a single value. The use of a GWR approach can show what is actually happening in different parts of the study area by exploring how certain spatial data varies from one location to the next (Fotheringham, Brundson, and Charlton, 2002: 7). In short, analysis using local statistics can offer a better indication of more specified geographical study areas by analyzing the variance within the study area itself versus providing a global statistical approach which analyzes the study area without considering variability, a process which inevitably results in a less accurate indicator of variable relationship. Local forms of spatial analysis provide much more information on spatial relationships in developing models and in better understanding spatial processes (Fotheringham, Brundson, and Charlton, 2002: 25).

The integration of OLS and GWR within the heavily commercialized ArcGIS software offers the capacity of researching both the relationship of variables and the relationship between them in a very precise geographical sense. This application ultimately offers a powerful tool in beginning to understand the correlation between locations of phenomena but also with the use of local statistics to begin answering the “why” questions of said phenomena. No longer does research have to depend on
multiple and sometimes complex software packages to conduct analysis. Though it must
be said once again that to build a properly specified model to accurately tell the “story” of
correlative relationships between variables is a difficult undertaking, the ability to test
and retest within a familiar framework is indeed a powerful option for cultural,
environmental, and other analysis.
CHAPTER THREE: PROJECT METHODS

Data Acquisition

All correlation models require great effort and care in data preparation. GIS analysis is data-driven, meaning data is required to complete successful analysis. It is apparent that data must always exist for analysis, but for the results to be reliable, such analysis must also be properly conditioned with the proper attributes, projected in the same geographical coordinate system, represent the same scale, and adhere to strict data accuracy standards. In fact, because spatial statistical tools run most of the calculations relatively quickly, the real work lies in preparing the datasets to be used effectively in ArcGIS.

One important consideration when running analysis with OLS is that both the variables being sought (dependent) and those used to find the relationship (independent) must be interval or ratio values (Mitchell, 2005:211). The tools cannot correctly specify relationships using nominal or ordinal data. This challenge becomes increasingly evident when attempting analysis with existing spatial and resource data.

With regard to Archaeological analysis acquiring existing datasets is relatively a simple; however, preparing such datasets is an involved process, the steps of which will later be described. To acquire necessary archaeological site data, depending on the location of analysis, one must contact the appropriate agency managing the land, acquire permission, and work with both that agency and, as is the case in Idaho, the State Historical Preservation Office. For this analysis, a Cultural Resource Use Permit (CRUP)
was issued by the Bureau of Land Management’s State Archaeologist and acquisition of the sensitive archaeological site spatial dataset was coordinated with the regional BLM Boise District office. For the sake of these exploratory analyses, this project’s scope is limited to the use of cultural resource sites which have been classified as prehistoric. The socio-economic and cultural lifeways between indigenous Native peoples and those of EuroAmericans varies in regard to their existence and use of the surrounding landscape. Due to the exploratory nature of this project, it was believed prudent to limit the scope.

In using environmental data to construct datasets representing independent variables, it is imperative when running all spatial analysis to ensure they have met strict accuracy standards. To accomplish this very important task, it is advisable to verify that each dataset created has metadata explaining the source of its creation, its geographic projection, and origin. For analysis in Idaho, a good place to start to acquire spatial data which has met data creation standards is [www.insideidaho.org](http://www.insideidaho.org).

All data gathered for this analysis was compiled from many sources, analyzed, and properly reprojected into the same geographical coordinate system (North American Datum of 1983 Zone 11) and extent (Study Area) before application. For elevation, slope, and aspect data the United States Geologic Survey (USGS) National Elevation Dataset (NED10m zone11) were used for fullest accuracy. Soil data was acquired from the United States Department of Agriculture Natural Resources Conservation Service (NRCS). River and stream data were acquired from the USGS National Hydrography Dataset.

Once the necessary datasets are analyzed and determined to meet any and all accuracy criteria, work begins on preparing the data for analysis. Before model
preparation begins, however, it is advisable to properly construct the GIS to ensure
analysis operates efficiently and without error. The importance of doing so can not go
understated since a wealth of data will be constructed, used, and at times discarded.
Without taking the steps to ensure the GIS is set up properly, the potential for risk of
losing computational outputs, data becoming either incorrectly projected (its global
location) or not projected at all, and most importantly the inability to effectively manage
data, run high.

Since these analyses are exploratory, and do not occur in conjunction with an
ongoing project, arbitrary study areas were constructed within the BLM’s Boise District
Bruneau Field Office. To effectively measure model performance of dependent and
independent variables utilizing existing data, the study area was constructed to include a
sizeable sample of cultural resource sites (1 km-squared n= @ 1158, one-acre-squared n=
@128) to ensure model operation.

Data Preparation

With necessary suggestions, advisories, and warnings effectively noted,
preparation of both GIS and data can begin. Though the steps listed here are not the only
means of reaching the end goal (ArcGIS was designed to operate with operational
diversity) they should offer a proper intuitive approach (recipe) to conduct analysis.

GIS preparation:

◦ With the recent release of ArcGIS 10 new functionality has been integrated to
  help the user run multiple operations not available previously in ArcGIS 9x.
  To start open up an ArcInfo licensed version of ArcGIS 10 and ensure that the
  Spatial Analyst toolbox and Editor Tools are operational.
◦ With ArcGIS 10, ArcCatalog and Search functions have been integrated into
  the ArcMap session. The usefulness of this will become increasingly evident
as the project progresses. Open the ArcCatalog screen into the ArcMap session and dock it to allow seamless transition between the Map and Catalog functions. This is advisable to ensure proper data management and efficiency.

- With Catalog open, browse to the working folder-> right-click new-> scroll down and create a new File Geodatabase. This will in effect become the spatial file storage system where all spatial computations, file creations and deletions, and outputs will be housed. It is advisable to make this geodatabase the default geodatabase so outputs of spatial operations will automatically be housed here. To do this-> right-click on the working geodatabase-> and select “Make Default Geodatabase”. This ensures all output files will be stored in one place, allowing for much more efficient and accurate data management.

- Bring in all recently acquired data to be used for OLS and GWR analysis and store it within the default geodatabase and ensure such data are properly projected in the same (NAD 1983 Zone 11) geographic projection. The map session is now ready for analysis.

Though regression analysis is conducted to measure the relationship between variables and the effectiveness of OLS in two different study areas, data preparation in principle is the same. To avoid repetition, the following steps represent those which were taken in preparation for the one-kilometer-squared analysis. They are also indicative of final processes deduced after a lengthy period of trial and error to reach the most effective solution in utilizing existing data for proper model operation.

The analysis geographic projection is USGS NAD 1983 Zone 11. All data not correctly projected in this datum will effectively be transformed using the ArcGIS transformation tool to ensure accuracy and function of analysis.

### Spatial Data Preparation

Open up ArcMap and dock ArcCatalog into viewer screen.

Useable datasets for OLS and GWR regression analysis located in default geodatabase:

1. NED 10-meter NAD 1983 Zone 11 raster dataset.
   - Aspect, Elevation, Slope calculations within the Study Area.
2. NRCS Soil vector dataset.
   - Soil classifications within the Study Area.
3. USGS National Hydrography vector dataset.
   ◦ River and stream classifications within the Study Area.
4. Archaeological site database created by Idaho State Historical Preservation Office in conjunction with state and federal agencies.
   ◦ Cultural resource site location and classification.

Other datasets in the analysis which will be used for reference and scale (provided by Boise District BLM):

2. BLM Boise District Bruneau Field Office boundary vector dataset.

Vector Grid Creation:

Import third-party Hawth’s Geospatial Sampling Tool and dock into toolbar settings. This is necessary to properly create a vector grid since ArcGIS lacks the capability to do so effectively. It is necessary to create a spatial grid of cells for the regression tools to analyze the relationship between the dependent and independent variables within what Ken Kvamme (1995:3) has called “decision surfaces” or in essence an encapsulated polygon of measurement. Without the use of the vector grid, the regression tools have no frame of reference for what to measure.

◦ Open up Hawth’s sampling tool -> Create Vector Grid with a cell output measure of 1000m (1 km). A spatial grid has been formed comprising of 1400 cells or effectively a study area = 1400-squared kilometers. The output will be called “SA_Analysis_Grid.”
◦ It is now necessary to create a centroid point for each cell to calculate the values of each independent variable. To do this, search for and open the Feature to Point tool in the data management toolbox with the following parameters: Input feature: The newly created Vector Grid. Output feature class, which by default will be stored in the working File Geodatabase and will be named “SA_Analysis_Grid_centroid.”

The purpose behind creating a centroid point for each cell is to measure the variable values within each one. It is apparent that measuring the values of each one-km-squared cell does not precisely depict the relationship between dependent and
independent variables; however, during the initial exploration after extensive testing of both studies to limit analysis to this method. Future research will inevitably ascertain a more accurate statistical cell size; however, for the purposes of this project the scope remains exploratory.

Figure 3: Study area grid of one-km-squared cells.

- With the extent of the Study Area established, the following step is optional; however, it is advisable to create a Study Area boundary identical to the extent of the study grid for further analysis and future map representation.
- Right-click on the working geodatabase and select new feature class. In the geographic projection select projected coordinate systems-> NAD 1983-> Zone 11. This ensures that the new polygon feature class is correctly projected. Add new Study Area boundary feature class to the map session.
- Open the Editor tool and select the new Study Area feature class as the template you wish to edit. At the bottom choose to create a new rectangle polygon and with
the “snapping” feature “On”, create a rectangle polygon to match the extent of the Study Area grid. Exit editor and save changes with the newly created Study Area boundary.

Preparation of Dependent and Independent Variable Data:

It is now necessary to effectively confine the variable datasets to the Study Area boundary for analysis. Omitting establishing the geographic extent for the data would cause complications for future analysis as there would be no defined area for which to compute the relationship between variables. Other concerns include the size and extent of the “original” datasets. Attempts to analyze the entire soils data, for example, cause the software to work unnecessarily hard to reach an end conclusion and reduces efficiency.

Dependent Variable:

- Search for and open the Clip Tool in the Spatial Analyst Toolbox. As your input feature class, input “original” sites and clip them to the extent of the Study Area. This will output a new feature class “SA_Sites.” When prompted to add the layer to the current map session say “Yes.”
- Right-click on the new feature class and select open “attribute table.” Search through the attributes for the Prehistoric classification field. This will be used to further constrict the analysis to only sites with a Prehistoric classification.
- Open up Select by Attributes Tool and select the “SA_Sites” feature class as the input. Go to the SQL box and enter the Prehistoric = (Go to Unique Field in upper-right hand corner and double click on “1” -> press “Apply,” then close. The prehistoric sites within the “SA_Sites” feature class are now selected (highlighted).
- Right-click on the “SA_Sites” feature class and scroll down to the Selection option. Open the Selection option and scroll down to “Make a layer from Selection” and left-click. A new layer of prehistoric sites within the Study Area has now been created for further analysis. At this point the layer can only be used in this data frame; it is necessary for future work to create its own feature class. Right-click on the dataset; scroll down to Export. Left-click on Export and select the option to “Export into a Feature Class.” Because the working geodatabase has been made the default, it will automatically choose to go to the right location for storage. Name the new feature class “SA_Sites_Prehistoric.”
- An advisable option to keep the workspace clear of unnecessary data layers is to scroll down and right click on previously used datasets which are no longer of use and remove them from the map session. This does not erase the datasets in case they are needed again; it simply takes them out of the current map session.
Independent Variables:

Soils:
- Search for and open Clip Tool in the Spatial Analyst Toolbox. Input feature class is the soils dataset and clip them to the extent of the Study Area. This will output a new feature class “SA_Soils.” When prompted to add the layer to the current map session, say “Yes”.

Rivers and streams (hydro):
- Search for and open Clip Tool in the Spatial Analyst Toolbox. Input feature class is the hydrologic dataset and clip them to the extent of the Study Area. This will output a new feature class, “SA_Hydro.” When prompted to add the layer to the current map session, say “Yes”.
- As bodies of water inevitably change as the environment forces change the landscape, it was necessary to further specify the relationship potential of sites and distance to water. With the extensive presence of intermittent streams throughout the Study Area which may or may not have existed during the Archaic period and beyond the focus was further set on the use of perennial bodies of water.
- Right-click on the new “SA_Hydro” feature class and select open “attribute table.” Search through the attributes for the Hydrograph classification field. This will be used to further restrict the analysis to only bodies of water with a Perennial classification.
- Open up Select by Attributes Tool and select the “SA_Hydro” feature class as the input. Go to the SQL box and enter the equation: Hydrograph = (Go to Unique Field in upper-right hand corner select Hydrograph and double click on the unique field “Perennial” -> press; “Apply;” then close. The perennial water bodies within the “SA_Hydro” feature class are now selected (highlighted).
- Right-click on the “SA_Hydro” feature class and go down to the Selection option. Open the Selection option and scroll down to “Make a layer from Selection” and left-click. A new layer of perennial water bodies within the Study Area has now been created for further analysis. At this point the layer can only be used in this data frame; it is necessary for future work to create it’s own feature class. Right-click on the dataset scroll down to Export. Left-click on Export and select the option to Export into a Feature Class. Because the working geodatabase has been made the default, it will automatically choose to go to the right location for storage. Name the new feature class “SA_Hydro_Perennial”.

Slope, Aspect, Elevation:
- Turn on the NED 10-meter Zone 11 raster dataset in the ArcMap session. Search for and open the Mask Raster Tool using the NED 10-meter as the input raster dataset, set the mask (the data extent which you want to effectively clip the data to) to the Study Area Boundary, and name the output raster dataset “SA_NED_Z11.” When prompted to add the new layer to the map session, say “Yes”.
- This will effectively “mask” all data outside of the Study Area and ensure all datasets have the same spatial extent.
Search for and open the Slope Tool in Spatial Analyst. Use the newly masked raster dataset “SA_NED_Z11” as the input and name the output “SA_NED_Z11_Slope” as the output. A new raster dataset will be created with the degrees of slope calculated and added to the attribute field.

Search for and open the Aspect Tool in Spatial Analyst. Again, use the newly masked raster dataset “SA_NED_Z11” as the input and name the output “SA_NED_Z11_Aspect” as the output. A new raster dataset will be created with the degrees of aspect calculated and added into the attribute field.

Search for and open the Contour Tool in Spatial Analyst, one more time using the raster dataset “SA_NED_Z11” as the input and name the output “SA_NED_Z11_Elevation” as the output. A new raster dataset will be created with the elevational data calculated and added to the attribute field. For all three tools when prompted to add output rasters to the map session, say “Yes.” All output raster datasets and feature classes, setting the working geodatabase as the default will in fact automatically be stored in the working geodatabase unless otherwise specified. For all analysis keep the default cell size as is.

It is now necessary to convert the new raster datasets into usable vector data for future analysis. With the goal of converging all the necessary attribute data into one polygon feature class (SA_Analysis_Grid) for OLS and GWR to run, it is necessary to have all data in a vector format.

Search for and open the Raster to Polygon Tool in the Conversion Toolbox. The newly created raster datasets “SA_NED_Z11_Slope, SA_NED_Z11_Aspect, SA_NED_Z11_Elevation” will be the three inputs each time the tool is run. For each output feature class dataset, name them “SA_Slope, SA_Aspect, SA_Elevation” respectively.

At this point all data to be used in the regression analysis model are correctly projected geographically, all share the same extent, and all are properly converted to a usable vector format. The next process will include joining all variable data into one dataset (SA_Analysis_Grid) to be run in the model. This will require joining the data both spatially and by tabular joining methods.

Spatial Joining the Data:

Before spatially joining the data, retrieve the values for each independent variable at each grid cell centerpoint (SA_Analysis_Grid_centroid). If you do not complete this step prior to the joining the SA_Analysis_Grid feature class will lack any value to run regression analysis.

To do this search for and open, the Extract Values to Points tool located in the Spatial Analyst Toolbox. The input feature class is: SA_Analysis_Grid_centroid, to extract elevation data the input raster will be: SA_NED_Z11_Elevation, and the output feature class is SA_Grid_Centroid_Elevation. Ensure to check the “Interpolate values at the point locations” box inside the tool window. Run the tool.

Repeat this procedure for both slope and aspect values using the same SA_Analysis_Grid_centroid for the input feature class, SA_NED_Z11_Slope, and SA_NED_Z11_Aspect rasters, with SA_Grid_Centroid_Slope and SA_Grid_Centroid_Aspect output feature classes respectively.
Elevation, Slope, and Aspect have now been effectively joined into new vector feature classes which will be used shortly for further analysis.

- To calculate values for the soils feature class, right-click on SA_Analysis_Grid_centroid; select Joins and Relates. In the tool window select Join data from another layer based on location which will output a new feature class which will be named SA_Grid_Centroid_Soils. Join the centroid layer to the soils feature class using a tabular join on the ObjectID attribute. The output will be a new feature class containing values for soils at each cell centroid.

- To calculate values for rivers and streams (hydro) feature class, right-click on SA_Analysis_Grid_centroid; select Joins and Relates. Again in the tool window select Join data from another layer based on location which will output a new feature class which will be named SA_Grid_Centroid_Hydro. Join the centroid layer to the hydro feature class using a tabular join on the ObjectID attribute. The output feature class will contain values for rivers and streams at each cell centroid.

- Finally it is necessary to calculate values for the dependent site location feature class (SA_Sites_Prehistoric). Again right-click on SA_Analysis_Grid_centroid layer; select Joins and Relates. In the tool window select Join data from another layer based on location, which will output a new feature class to be named SA_Grid_Centroid_Sites. The output feature class will contain a site count value for each cell centroid. For all cells which have a null value it is important to reclassify this statistic to 0 for future analysis. To do this run, the field calculator on the site count column with the following expression: null = 0.

- It is now time to join all dependent and independent variables to the Study Area grid. Prior to undertaking this task it is advisable to delete any and all extraneous attribute fields which were created from the previous joins to avoid confusion when joining the final data to the final SA_Analysis_Grid. To do this right click on each newly created Centroid feature class and open the attribute table. From the attribute table carefully right-click on all data columns which won’t be used in the final analysis including all fields which were created during the joins. The end product will show the ObjectID, ObjectFID, values, and other data deemed pertinent for further analysis.

- When both the dependent and independent variable data are properly created, it is necessary to spatially join these feature classes to the SA_Analysis_Grid, which will then be further applied to OLS and GWR tools. To do this, right-click on SA_Analysis_Grid, select Joins and Relates, and join each of the SA_Grid_Centroid_Elevation, SA_Grid_Centroid_Slope, SA_Centroid_Aspect, SA_Grid_Centroid_Soils, SA_Grid_Centroid_Hydro, and SA_Grid_Centroid_Sites feature classes using a tabular join to final SA_Analysis_Grid polygon feature class using, with the exception of the dependent site values, the following expression: (OBJECTID = OBJECTFID). To calculate the site values, enter the following expression (OBJECTID = OBJECTID).

- The SA_Analysis_Grid feature class polygon now is complete with all necessary values for OLS and GWR analysis. It is advisable, however, to “clean” the data by renaming the attribute columns within the the attribute field to names which won’t confuse further output analysis. To do this, right-click each column heading,
properties, and rename the fields “Elevation, Slope, Aspect, Soils, Hydro, and Sites” respectfully.
◦ The SA_Analysis_Grid feature class polygon is now prepared for OLS and GWR regression analysis.

**Model Preparation and Analysis**

Preparing OLS and GWR is a relatively simple task after properly preparing the data for analysis. Within the ArcMap session, search for and open the Ordinary Least Squares Regression tool in the Spatial Statistics Toolbox (Figure 4).

![OLS tool window](image)

**Figure 4.** OLS tool window to run regression analysis.
As indicated in the figure, the input feature class will be the one-kilometer-squared analysis grid containing values for all variables. The Unique ID field is created in the attribute table to identify every grid calculation. If the attribute table has yet to be populated with a Unique ID, open the attribute and calculate a new field using a double integer and name it Unique ID. In the field calculator run the following expression: UniqueID = OBJECTFID. This is a necessary step as OLS will not successfully run with an OBJECTID or OBJECTFID attribute designated as the UniqueID. In this one-kilometer-squared analysis, there will be 1400 different attribute cells.

The output feature class output will automatically default to the working geodatabase, which will be called SA_Analysis_Grid_OLS. The dependent variable, or variable for which a relationship is being determined with the environmental variables will be the (cultural resource) site count. Check the boxes next to the independent variables being run in the analysis (Elevation, Slope, Aspect, Hydro, and Soils). Once these fields have been successfully populated and checked, press OK.

Depending on the size of the study area and the amount of attribute information it needs to calculate, the tool will run in a matter of moments. The output of the SA_Analysis_Grid_OLS will be automatically included within the map session and will resemble the results overlaid on a topographical representation of the Study Area after modifying the newly created cells which show no relationship to clear and making those cells which do have a relationship transparent by 40% (Figure 5).
Figure 5. OLS results from multivariate analysis using all dependent (site location) and independent variables (elevation, slope, aspect, hydro, and soils). The results are categorized by a standard deviation of calculations (green cells = areas less likely to find sites), red areas = areas more likely to find sites over a null hypothesis (clear cells). Results will be further explained in Chapter 4.
To determine how well the model worked, open the results window after the OLS tool has finished its calculations. To do this, find and open the drop-down Geoprocessing menu. Choose Results, expand Current Session and Ordinary Least Squares, right-click Messages and choose View. This will open a results window from your analysis shown in Table 2.

Table 2. OLS Results for multivariate analysis. As indicated in the upper-left hand assessment the Soils values, though reclassified, failed to calculate and were omitted from the analysis. An asterisk next to both the elevation and distance to rivers and streams indicates the two variables which have statistical value in this particular analysis.

![Table 2: OLS Results](image)

Though the R-squared value is only accounting for < 2% of the story, other indicators show the model meets other necessary p-values of model significance. When overall good model significance occurs it is necessary to account for the residuals in the
data from the OLS analysis. To do this ArcGIS integrates a Spatial Autocorrelation Moran’s I tool to test for spatial autocorrelation in the Spatial Statistics toolbox.

As the software eludes in ArcGIS Help, this tool is designed to measure spatial autocorrelation (feature similarity) based on both feature locations and feature values simultaneously. Given a set of features and an associated attribute, it evaluates whether the pattern expressed is clustered, dispersed, or random. The tool calculates the Moran's I Index value and both a Z score and p-value evaluating the significance of that index. In general, a Moran's Index value near +1.0 indicates clustering while an index value near -1.0 indicates dispersion. However, without looking at statistical significance you have no basis for knowing if the observed pattern is just one of many, many possible versions of random.

In the case of the Spatial Autocorrelation tool, the null hypothesis states that "there is no spatial clustering of the values associated with the geographic features in the study area". When the p-value is small and the absolute value of the Z score is large enough that it falls outside of the desired confidence level, the null hypothesis can be rejected. If the index value is greater than 0, the set of features exhibits a clustered pattern. If the value is less than 0, the set of features exhibits a dispersed pattern (ESRI, 2010).

When run, the tool will provide either a text box or a HTML window indicating whether or not the patterns in the data are random, clustered, or dispersed. Because the HTML is easier to understand open it. To do so, find and open the drop-down Geoprocessing menu, choose Results, expand Current Session and Moran’s I tool, right-click Messages, and choose View. The operation will open the window (Figure 6).
Figure 6. The results of running the Spatial Autocorrelation Moran’s I tool to test the residuals of the multivariate analysis indicate the variable attribute data are clustered.
To better understand which variables are useful in the model, ArcGIS has also integrated a Scatterplot Matrix tool to analyze the relationship each variable has with one another. To create a Matrix, go to the View menu→and select Graphs→Create Scatterplot Matrix. For the Layer/Table, choose the SA_Analysis_Grid_OLS. The order in which variables are chosen doesn’t matter, but it is easier to view relationship strength if the dependent variable is chosen first and the explanatory variables afterward. Click the empty cell under the Field name heading to get a list of fields; then choose variables accordingly. The rest of the properties won’t affect how the variables are plotted so click Next, and then Finish.

By adding the Site Count variable first, the column on the far left can be viewed and the relationship between the dependent and the other variables can be seen (Figure 7). It is clear the regression relationship between variables is weak or non-existent. As with any analysis, having the proper data is vital. The Scatter Plot Matrix showing the analysis value of the existing data is marginal at best in the one-kilometer-squared grid OLS model.

Although the results of the OLS regression tool do represent a properly built and robust model, it is apparent in this analysis there are problems using existing environmental data to analyze site locations. The results of the Spatial Autocorrelation Moran’s I tool indicate a clustering of residuals in the data attribute values. The results of the Scatter Plot Matrix also fail to show either a strong negative or positive relationship between the dependent variable and the other independent variables, a relationship which is essential for regression analysis.
After viewing relationship qualities between variables in the matrix and assessing the OLS results window, it is necessary to analyze each bivariate combination between the dependent and each independent variable in order to include the best combination of variables capable of telling the “story” as possible. Results of these analyses are located in the Appendix (see Figures 9-20 and Tables 4-18).

Figure 7. Scatter Plot Matrix representing each variable with one another. It is advisable to create a matrix for every attempt in using OLS to fully understand the relationship between variables. The Scatter Plot Matrix created for the 1 kilometer-squared analysis is showing little or no relationship between the dependent site count variable and the independent environmental variables.
CHAPTER FOUR: CONCLUSIONS

The OLS results indicators for the one-kilometer-squared analysis tell a pretty clear story with this modeling application. Though the output grid analysis does offer colored indicators as to whether certain areas have a greater or lesser potential for locating cultural resource sites when compared to a null-hypothesis, the results are not statistically significant using existing data and at this scale.

To further clarify, green cells represent areas where it would be less likely to find prehistoric sites based on the relationship between the dependent and independent variables compared to a null hypothesis; the red cells indicate areas where it would be more likely to find prehistoric sites based on this correlation analysis. However, as the results indicate, returned R-squared values account for less than 2% of the relationship between variables, the residuals are clustered, and the model lacks any strong positive or negative relationship between the dependent and independent variables. Adding known cultural resource site locations (Figure 8) helps illuminate these shortcomings by showing a significant relationship within cells which contain sites versus those which do not.

Extensive attempts to further specify both one-kilometer-squared bivariate analysis, and one-acre-squared bivariate and multivariate analysis using Ordinary Least Squares regression (see Appendix) also failed to return statistically significant results.

Addressing the goals of this exploratory research, it is believed, though minimally during these initial tests, that using environmental data does in fact have a relationship
Figure 8. OLS multivariate analysis including known cultural resource site location.
with cultural resource site location. It will, however, require a continued effort beyond these initial stages to develop statistically significant correlations; however, it is both possible and feasible to apply these tools in correlation modeling.

The use of existing data during these initial analyses proved to indicate very little or no relationship values between the dependent and independent variables. However, during this research other possible approaches for future modeling have presented themselves. In addition to using data available, it is believed that calculating other cells within the study grid as having a no-site or 0 site value will help increase the R-squared value of this approach. To effectively accomplish this, a study grid must be designed over a study area where pedestrian survey has deduced that no sites exist. It is believed after a ratio is established within the same environment that the calculation of grid cells containing sites versus those without any can be applied to further specify the model.

Further reclassification of the soils data will also aid in returning a more accurate representation of their relationship to the dependent variable. To accomplish this, research would be required to analyze the ratio of sites located within specific soil types to offer a rating of significance between one soil type and another.

As shown in Table 3, Geographic Weighted Regression failed to execute utilizing existing data within both study areas. It is understood, that the one-kilometer-squared approach failed due to the model’s failing to meet the necessary tests of significance. And although the second approach using one-acre-squared grid cells met all required tests of significance, GWR failed to execute because it lacked enough dependent variables to fully compute cell-based relationships. Future research would be directed toward finding a more specified grid size which passes all necessary tests of significance in OLS, while
also containing enough dependent variables to effectively run the more accurate regression.

Table 3. GWR analysis on 1-acre multivariate analysis. Failure of GWR to run properly is believed in part to a lack of a large enough sample of dependent variables. Again future research could address this problem by selecting a Study Area with a larger population of the dependent site variable.

After these initial approaches using OLS and GWR, it is clear more work must be done to further classify the data, to properly construct a grid where values can be more accurately calculated within each cell, and to include data which has yet to be created before statistically significant relationships between site location and the environment can be induced. The model itself has proven to be statistically significant. Data will need to be further developed and classified beyond these initial tests to successfully show significant relationship values. With these results, work can continue towards building and testing hypotheses to better understand the relationship between the environment and cultural material.

The possibilities for quantitatively addressing relationships between site locations and other independent variables using OLS and GWR do exist. The results of these initial efforts make it clear more work must be done to properly define the model;
however, it is clear the principles of this approach and of regression analysis in general
do offer an increased probability of locating cultural resources using independent
variables. It is necessary to further develop this approach before it can be considered an
effective tool in significantly modeling relationships. However, it is believed with
continued research and effort this approach will provide an effective and feasible option
in managing cultural resources.
REFERENCES

Bureau of Land Management

Daubenmire, Rexford F.

Ebert, James I. and Timothy A. Kohler

ESRI
2010  “Interpreting OLS Results.” ArcGIS 10 Help Database, Redlands, CA. ESRI.

Gotelli, Nicholas J. and Aaron M. Ellison

Fotheringham, A. Stewart, Chris Brundson, and Martin Charlton
2002  Geographically Weighted Regression: the analysis of spatially varying relationships. Chichester, West Sussex, England; Hoboken, NJ. John Wiley and Sons Ltd.

Fowler, Catherine S. and Sven Liljeblad

Freeman, Otis W., J.D. Forrester, and R.L. Lupher
Hackett, Bill and Bill Bonnichsen  

Jenks, Margaret D, Bill Bonnichsen, and Martha M. Godchaux  

Kvamme, Kenneth L.  

Malde, Harold  

Mitchell, Andy  

Thoms, Alston V.  

Warren, Robert E. and David L. Asch  

Yensen, Dana  
APPENDIX
Figure 9. One-Kilometer-squared and one-acre-squared study areas overlay on topographic relief.
The following tables and figures represent bivariate OLS analysis with the dependent variable and each independent environmental variable for the 1KM-squared Study Area:

Figure 10. OLS results one-kilometer-squared analysis indicating the relationship between the dependent variable – independent distance-to-water variable.
Table 4. OLS Results one-kilometer-squared analysis window determining model significance on the relationship between the dependent variable and the independent distance-to-water variable. The results show model significance under all mandatory classifications making this relationship a good relationship indicator.
Table 5. Moran’s Tool Results one-kilometer-squared analysis text window indicating a high z-score for residual values of the OLS analysis between the dependent site variable and independent distance-to-water variable. A high z-score value indicates the residuals of the data are clustered.
Figure 11. OLS results one-kilometer-squared analysis indicating the relationship between the dependent variable – independent elevation variable.
Table 6. OLS Results one-kilometer-squared analysis window determining model significance on the relationship between the dependent variable and the independent elevation variable. The results show a lack of model significance under all mandatory classifications with the exception of the Jarque-Bera Statistic. This is indicative of a bad relationship indicator between the two variables.

Summary of OLS Results

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>StdError</th>
<th>t-Statistic</th>
<th>Probability</th>
<th>Robust SE</th>
<th>Robust t</th>
<th>Robust Pr</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.274520</td>
<td>0.059503</td>
<td>0.133385</td>
<td>0.900100</td>
<td>0.274520</td>
<td>0.059503</td>
<td>0.900100</td>
</tr>
<tr>
<td>ELEVATION METERS</td>
<td>0.000260</td>
<td>0.001180</td>
<td>0.218650</td>
<td>0.826969</td>
<td>0.000260</td>
<td>0.001180</td>
<td>0.826969</td>
</tr>
</tbody>
</table>

OLS Diagnostics

- Number of Observations: 1400
- Number of Variables: 2
- Degrees of Freedom: 1398
- Multiple R-Squared [2]: 0.000054
- Adjusted R-Squared [2]: 0.000001
- Joint F-Statistic [3]: 0.047939
  Prob(>F), (1,1398) degrees of freedom: 0.900690
- Joint Wald Statistic [4]: 0.209830
  Prob(>chi-squared), (1) degrees of freedom: 0.646900
- Moenker (BP) Statistic [5]: 0.006542
  Prob(>chi-squared), (1) degrees of freedom: 0.835689
- Jarque-Bera Statistic [6]: 190227.995002
  Prob(>chi-squared), (2) degrees of freedom: 0.000000

Notes on Interpretation

- Statistically significant at the 0.05 level.
- [1] Large VIF (> 7.5, for example) indicates explanatory variable redundancy.

WARNING 000561: Use the Spatial Autocorrelation (Moran's I) Tool to ensure residuals are not spatially autocorrelated.
Completed script OrdinaryLeastSquares...
Succeeded at Wed Apr 06 11:44:35 2011 (Elapsed Time: 11.00 seconds)
Table 7. Moran’s Tool Results one-kilometer-squared analysis text window indicating a high z-score for residual values of the OLS analysis between the dependent site variable and independent elevation variable. A high z-score value indicates the residuals of the data are clustered.
Figure 12. OLS results one-kilometer-squared analysis indicating the relationship between the dependent variable – independent slope variable.
Table 8. OLS Results one-kilometer-squared analysis window determining model significance on the relationship between the dependent variable and the independent slope variable. The results show model significance under 5 of 6 of the mandatory classifications making this relationship a good relationship indicator. The Koenker (BP) Statistic is just above the .05 p-value at .086.
Table 9. Moran’s Tool Results one-kilometer-squared analysis text window indicating a high z-score for residual values of the OLS analysis between the dependent site variable and independent slope variable. A high z-score value indicates the residuals of the data are clustered.
Figure 13. OLS results one-kilometer-squared analysis indicating the relationship between the dependent variable – independent aspect variable.
Table 10. OLS Results one-kilometer-squared analysis window determining model significance on the relationship between the dependent variable and the independent aspect variable. The results show a lack of model significance under all mandatory classifications with the exception of the Jarque-Bera Statistic. This is indicative of a bad relationship indicator between the two variables.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Standard Error</th>
<th>t-Statistic</th>
<th>Probability</th>
<th>Robust SE</th>
<th>Robust t</th>
<th>Robust Pr</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.522726</td>
<td>0.164889</td>
<td>3.170747</td>
<td>0.001568*</td>
<td>0.124765</td>
<td>2.878232</td>
<td>0.000120*</td>
</tr>
<tr>
<td>ASPECT</td>
<td>0.001119</td>
<td>0.000316</td>
<td>3.971487</td>
<td>0.170355</td>
<td>0.009709</td>
<td>1.577307</td>
<td>0.114966</td>
</tr>
</tbody>
</table>

### OLS Diagnostics

- **Number of Observations:** 1400
- **Number of Variables:** 2
- **Degrees of Freedom:** 1398
- **Akaike's Information Criterion (AIC) [2]:** 6852.914000
- **Adjusted R-Squared [2]:** 0.000629
- **Joint F-Statistic [3]:** 1.686876
  - **Prob(F), (1,1398) degrees of freedom:** 0.170443
- **Joint Wald Statistic [4]:** 2.487696
  - **Prob(>chi-squared), (1) degrees of freedom:** 0.114725
- **Koenker (BP) Statistic [5]:** 0.706039
  - **Prob(>chi-squared), (1) degrees of freedom:** 0.460762
- **Jarque-Bera Statistic [6]:** 190229.150226
  - **Prob(>chi-squared), (2) degrees of freedom:** 0.000000*

**Notes on Interpretation**

* Statistically significant at the 0.05 level.

[1] Large VIF (> 7.5, for example) indicates explanatory variable redundancy.

**WARNING:** Use the Spatial Autocorrelation (Moran's I) Tool to ensure residuals are not spatially autocorrelated.

Completed script: OrdinaryLeastSquares...

Succeeded at Wed Apr 06 12:14:27 2011 (Elapsed Time: 10.00 seconds)
Table 11. Moran’s Tool Results analysis text window for one-kilometer-squared cell size indicating a high z-score for residual values of the OLS analysis between the dependent site variable and independent aspect variable. A high z-score value indicates the residuals of the data are clustered.

![Image of Moran's Tool Results]

Executing: SpatialAutocorrelation FA_SA_OLS_Aspect Residual NO_REPORT
INVERSE_DISTANCE EUCLIDEAN_DISTANCE NONE # #
Start Time: Wed Apr 06 12:16:26 2011
Running script SpatialAutocorrelation...
WARNING 000853: The default neighborhood search threshold was 1000.1.
  Global Moran's I Summary
Moran's Index: 0.632440
Expected Index: -0.000715
Variance: 0.000351
z-score: 33.777512
p-value: 0.000000
Writing html report....
E:\Grad School\THESIS_WORKING_2011\DATA\Boise_work_2011\MoransI_Result6.html
Completed script SpatialAutocorrelation...
Succeeded at Wed Apr 06 12:16:27 2011 (Elapsed Time: 1.00 seconds)
Table 12. OLS Results analysis window for one-kilometer-squared cell size determining model significance on the relationship between the dependent variable and the independent soils variable. Though the results show model significance OLS could not successfully calculate the relationship of the two variables causing the model to fail. Future research will have to account for this need for attribute reclassification before applying this bivariate relationship into the model. With the failure of OLS to execute this analysis there is neither an output feature class showing the relationship or a Spatial Autocorrelation Moran’s I tool output to analyze the residuals.
The following tables and figures represent bivariate OLS analysis with the dependent variable and each independent environmental variable for the 1 acre-squared Study Area:

Table 13. OLS Results window for one-acre-squared analysis determining model significance on the relationship between the dependent variable and the independent distance-to-water variable. The results show model significance under all mandatory classifications making this relationship a good relationship indicator.
Figure 14. Moran’s Tool Results text window for one-acre-squared analysis indicating a high z-score for residual values of the OLS analysis between the dependent site variable and independent distance-to-water variable. A high z-score value indicates the residuals of the data are clustered.
Table 14. OLS Results window for one-acre-squared analysis determining model significance on the relationship between the dependent variable and the independent aspect variable. The results show a lack of model significance under all mandatory classifications with the exception of the Jarque-Bera Statistic. This is indicative of a bad relationship indicator between the two variables.
Figure 15. Moran’s Tool Results text window for one-acre-squared analysis indicating a high z-score for residual values of the OLS analysis between the dependent site variable and independent aspect variable. A high z-score value indicates the residuals of the data are clustered.
Table 15. OLS Results window for one-acre-squared analysis determining model significance on the relationship between the dependent variable and the independent slope variable. The results show model significance under all mandatory classifications making this relationship a good relationship indicator.
Figure 16. Moran’s Tool Results text window for one-acre-squared analysis indicating a high z-score for residual values of the OLS analysis between the dependent site variable and independent slope variable. A high z-score value indicates the residuals of the data are clustered.
Table 16. OLS Results window for one-acre-squared analysis determining model significance on the relationship between the dependent variable and the independent elevation variable. The results show model significance under all mandatory classifications making this relationship a good relationship indicator.
Figure 17. Moran’s Tool Results text window for one-acre-squared analysis indicating a high z-score for residual values of the OLS analysis between the dependent site variable and independent elevation variable. A high z-score value indicates the residuals of the data are clustered.
Table 17. OLS Results window for one-acre-squared analysis determining model significance on the relationship between the dependent soil variable and the independent variable. The results show model significance under 5 of 6 of the mandatory classifications making this relationship a good relationship indicator. The Koenker (BP) Statistic is in fact above the .05 p-value at .110. Future research will still need to address the need for attribute reclassification before applying this bivariate relationship into the model.
Figure 18. Moran’s Tool Results text window for one-acre-squared analysis indicating a high z-score for residual values of the OLS analysis between the dependent site variable and independent soil variable. A high z-score value indicates the residuals of the data are clustered.
Table 18. OLS Results window for one-acre-squared analysis determining model significance on the relationship between the dependent variable and all independent variables. The results show model significance under all mandatory classifications making this relationship a good relationship indicator. The results have determined both the slope and distance-to-water variables are statistically significant for further analysis.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>StdError</th>
<th>t-Statistic</th>
<th>Probability</th>
<th>Robust SE</th>
<th>Robust t</th>
<th>Robust Pr VIF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-0.407727</td>
<td>0.569221</td>
<td>-0.712620</td>
<td>0.473588</td>
<td>0.56279</td>
<td>-0.76612</td>
<td>0.456536</td>
</tr>
<tr>
<td>SOIL_OBJECTID</td>
<td>0.001567</td>
<td>0.000594</td>
<td>1.841559</td>
<td>0.065657</td>
<td>0.00002</td>
<td>1.74503</td>
<td>1.772024</td>
</tr>
<tr>
<td>DISTANCE</td>
<td>-0.000667</td>
<td>0.000164</td>
<td>-4.557660</td>
<td>0.000002*</td>
<td>0.00045</td>
<td>-4.56657</td>
<td>0.000002*</td>
</tr>
<tr>
<td>ELEV_MRD</td>
<td>0.000039</td>
<td>0.000048</td>
<td>0.742876</td>
<td>0.457684</td>
<td>0.00094</td>
<td>0.75530</td>
<td>0.438111 15 1.56399</td>
</tr>
<tr>
<td>SLOPE</td>
<td>0.000870</td>
<td>0.000130</td>
<td>5.883834</td>
<td>0.000000*</td>
<td>0.002469</td>
<td>4.909100</td>
<td>0.000001*</td>
</tr>
<tr>
<td>ASPECTDEG</td>
<td>-0.000008</td>
<td>0.000049</td>
<td>-0.121362</td>
<td>0.903601</td>
<td>0.000077</td>
<td>-0.107697</td>
<td>0.914910 11 1.177973</td>
</tr>
</tbody>
</table>

OLS Diagnostics

- Number of Observations: 2020
- Number of Variables: 6
- Degrees of Freedom: 2002
- Akaike's Information Criterion (AIC): 326.692015
- Adjusted R-Squared: 0.041226
- Joint F-Statistic [3]: 21.431701
- Joint Wald Statistic [4]: 51.85198
- Breusch-Pagan Statistic [5]: 50.10598
- Lagrange-Dirichlet Statistic [6]: 2422.790159

Notes on Interpretation

* Statistically significant at the 0.05 level.
[1] Large VIF (＞7.5, for example) indicates explanatory variable redundancy.
[5] Significant p-values indicate biased standard errors; use robust estimates.

WARNING: Use the Spatial Autocorrelation (Moran's I) Tool to ensure residuals are not spatially autocorrelated.
Compiled script: OrdinaryLeastSquares...
Succeeded at Wed Apr 06 15:42:30 2011 (Elapsed Time: 12.00 seconds)
Figure 19. Moran’s Tool Results text window for one-acre-squared analysis indicating a high z-score for residual values of the OLS analysis between the dependent site variable and all independent variables. A high z-score value indicates the residuals of the data are clustered.
Figure 20. Scatter Plot Matrix representing each variable with one another in the one-acre-squared analysis. It is advisable to create a matrix for every attempt in using OLS to fully understand the relationship between variables. The Scatter Plot Matrix shows little or no relationship between the dependent site count variable and the independent environmental variables. This lack of relationship is believed to be a result from too few dependent sites within the Study Area. Future analysis could address this deficiency by selecting a Study Area with a higher number of variables which is believed would increase the model significance.