12-2009

GIS Tools for Cartographic Representation of Spatial Data Uncertainty

Janelle P. Payne
University of Redlands

University of Redlands

GIS Tools for Cartographic Representation of Spatial Data Uncertainty

A Major Individual Project submitted in partial satisfaction of the requirements
for the degree of Master of Science in Geographic Information Systems

by

Janelle P. Payne

Mark P. Kumler, Ph.D., Committee Chair
R. Maxwell Baber, Ph.D.

December 2009

GIS Tools for Cartographic Representation of Spatial Data Uncertainty

by

Janelle P. Payne

The report of Janelle P. Payne is approved.

R. Maxwell Baber, Ph.D.

Mark P. Kumler, Ph.D., Committee Chair

December 2009

Acknowledgements

Numerous people were integral to my successful completion of this project. First and foremost, the project committee provided invaluable support. Dr. Max Baber helped steer the early phases of this project by providing valuable feedback and guiding project requirements as my advisor, and continued to provide useful support throughout the course of the project. Dr. Mark Kumler served as both advisor and committee chair; his methodical approach to project development, guidance throughout the project, and attention to detail also assisted me in successful completion of this project.

The project client, Mr. Charlie Frye from Environmental Systems Research Institute (ESRI), provided essential support defining requirements and scope of the project. Mr. Frye also shared his knowledge of the subject matter, software, and deployment in order to ensure successful completion of this project.

My wonderful family was unfailingly supportive while I was in the MS GIS program and my ultimate success in life and in this program is undoubtedly due to their support and encouragement.

Finally, I would like to thank artistically gifted Mr. Michael Franti and Spearhead, whose sweet lyrics motivated me during long hours spent in Lewis Hall 103; plus the good people of Chowchilla, California who provided me a much needed respite from project development; and also my colleagues at ESRI for their patient understanding and compassionate support during this particularly busy time in my life.

Abstract

GIS Tool for Representation of Spatial Data Uncertainty

by

Janelle P. Payne

Maps created in geographic information systems (GIS) are rendered with precisely defined features, but experienced GIS practitioners recognize that spatial data have relative error that is not always apparent to map readers. Limited awareness among many users regarding data error leads users to view and analyze data without regard for relative uncertainty. Tools and methods supporting map designer abilities to graphically communicate uncertainty associated with spatial data have not been readily available.

There exists a need for users to display quantifiable characteristics of relative uncertainty associated with spatial data affected via cartographic representation. Development for this project synthesized prominent research recommendations to provide map designers' with methods for conveying data uncertainty with scientifically tested symbolizations within the ArcGIS software. The ultimate goal of this development project is to increase map designers efficiency in illustrating data uncertainty, and stimulate conversation about GIS tools for representing this uncertainty to a wider audience.

Table of Contents

Table of Figures

List of Tables

List of Acronyms and Definitions

ESRI	Environmental Systems Research Institute
GIS	Geographic Information Systems
LPK	Layer Package
MXD	Map Document
NCGIA	National Center for Geographic Information and Analysis
NRCS	Natural Resources Conservation Service
SDTS	Spatial Data Transfer Standard
SSURGO	Soil Survey Geographic Database
USFWS	United States Fish and Wildlife Service
USGS	United States Geological Survey

Chapter 1 – Introduction

Maps created in geographic information systems (GIS) are typically rendered with precisely defined features, but experienced GIS practitioners recognize that spatial data have varying measures of relative error that are not always apparent to map readers. Limited awareness among many geographic information users regarding the associated error of the data leads users to view and analyze data without regard for relative uncertainty. Tools and methods supporting map designer abilities to graphically convey quantifiable uncertainty associated with spatial data have not been readily available.

Environmental Systems Research Institute (ESRI) recognizes a need for tools to display quantity and characteristics of relative uncertainty associated with spatial data in cartographic representation. There is considerable research on methods used for illustrating spatial data uncertainty. Application development for this project synthesized a number of prominent research recommendations to provide map designers with representation methods for conveying relative data uncertainty with scientifically tested symbolizations within ArcGIS. The ultimate goal of this development project was increased efficiency of map designers in illustrating data uncertainty, and likewise extending the conversation about GIS tools for representing spatial data uncertainty to a wider audience.

1.1 Client

The client for this project was Mr. Charlie Frye, Chief Cartographer at ESRI. Mr. Frye has a strong background in cartographic design and manages the ESRI Mapping Center, an online resource for ArcGIS users to obtain information and tools that assist and inform users on methods for better cartographic representation of their information in a GIS (ESRI, 2009b). As one of the early members of ArcGIS Online, the online forum which the representation methods were deployed on, he has provided valuable feedback to the development team for enhancing the services and sharing capabilities of the website. ArcGIS Online is a online resource that allows ESRI software users to find and share data and tools for map production (ESRI, 2009a).
Mr. Frye provided support for this project by:

- Providing guidance throughout the planning and development process
- Permitting access to relevant data
- Aiding in defining the scope in order to ensure that the project benefited the anticipated user community
- Supplying the requirements for providing the project components to users

1.2 Problem Statement

All spatial data have inherent error associated with their “positional accuracy, attribute accuracy, logical consistency, completeness and/or lineage” (United States Geological Survey, 1997, pp. 13-16). GIS users need a way to effectively communicate error and uncertainty in data. Visually representing spatial data uncertainty is particularly important for applications in which the data are highly susceptible to data collection or data entry errors. Additionally, different applications or data types will include different aspects of uncertainty and it is important to be able to display the type of uncertainty present in that data. For example, some natural resource data have areas where the boundary locations are not well-defined, but land ownership data may have entire areas that have some level of uncertainty.

Currently there is not a clearly defined process or symbology set to illustrate the uncertainty that may occur in data. Users that choose to represent spatial data uncertainty either spend a great deal of time creating their own custom symbology or, worse, users will choose to ignore the data uncertainty due to the lack of tools to streamline the representation process. This could potentially lead to misinterpretation and poor analysis of the data.

1.3 Proposed Solution

The proposed toolset is a comprehensive, industry-independent solution that allows users to illustrate a measure of uncertainty within data. Since the anticipated audience consists of users of the ESRI Mapping Center (ESRI, 2009b) who are generally ArcGIS software users, the tools will focus on the needs of these users. The symbology methods will streamline users' representation processes and allow them to view and analyze the data more accurately.

1.3.1 Goals and Objectives

The primary objective of this project was to increase the efficiency and effectiveness of users' cartographic representation process for spatial data uncertainty, while also bringing awareness of the topic of uncertainty to a broader audience. Many scholars have explored methods for illustrating uncertainty in data and it is not the purpose of this project to make yet another recommendation. The goal of this project was to develop techniques to assist users with representing the quality and degree of data uncertainty.

Currently users interested in depicting the quantity and/or quality of uncertainty in their data need to create their own processes and symbology. This project gives users procedures, including documentation, along with symbology for describing the uncertainty that is inherent in their data. The solution allows users to increase their productivity by taking advantage of step-by-step methodology and tools. This not only saves users time and money, but may also increase actual reliability of the interpretation of the data by providing a better understanding of the quality of its attributes.

The toolset includes two major components:

- Layer packages: a set of five specific representation methods for various data types
- Documentation: a combination of step-by-step guidance and resource website that allows users to determine better ways for describing the data uncertainty

1.3.2 Scope

Project development focused on creating the necessary tools to assist users in depicting the positional and attribute accuracy uncertainty in data. There are two major components of this project, namely symbology recommendation layers, and documentation.

Development for the symbology recommendations focused on the needs of existing ArcGIS users.

Data were collected from various sources and approved by the client. A total of five symbology methods were provided for different data quality and types. Specifically, development focused on symbology for:

1. Natural Resource Conservation Service (NRCS) Soil Survey Geographic (SSURGO) Database
    a. Data Type: Aggregated area
    b. Data Quality Concern: Positional accuracy and attribute accuracy
2. United State Geological Survey (USGS) earthquake data
    a. Data Type: Point
    b. Data Quality Concern: Attribute accuracy
3. Redlands Institute Salton Sea Sediment Composition
    a. Data Type: Continuous
    b. Data Quality Concern: Attribute accuracy

Access to final deliverables for this project was provided directly though the "Depicting Uncertainty" group on ArcGIS Online. The symbology layers and documentation were provided directly on ArcGIS Online.

The original scope of this project was modified to accommodate for technology changes during the development process. The original scope defined the project deliverables as a style set, an interactive toolbar for ArcGIS Desktop and documentation, with deployment to be made through the ESRI Mapping Center website. The release of ArcGIS 9.3.1 and ArcGIS Online provided a more appropriate option for project development and deployment.

1.3.3 **Methods**

A traditional development approach was taken for project development with the goal of developing a prototype early in the timeline and refining the products based on client and user feedback. Development took the following steps:

1. User needs assessment: Determine the needs of the client and potential users.
2. Literature review: Review pertinent literature to determine what leading scholars believe is an appropriate solution and synthesize the information for purposes of this project.
3. Design: Create a blueprint for all three project components.
4. Development: Create a prototype for the project components.
5. Testing: Test with sample data and provide to client for additional testing and approval.
6. Final product development: Modify project components based on testing and client feedback.
7. Testing: User acceptance testing for final deployment.
8. Product deployment: Load symbology, scripts, and documentation to "Depicting Uncertainty" group on the ArcGIS Online website for user consumption.

1.4 Audience

The intended audience for this paper are people who are familiar with GIS, spatial data, associated error and are, at least generally, familiar with the topic of uncertainty in spatial data. The users who will benefit from this paper are interested in cartographic representation methods for positional and attribute uncertainty in data. Furthermore, GIS users can apply some of the concepts of this paper to build and apply other types of cartographic representations to their data.

1.5 Overview of the Rest of this Report

The rest of this paper details the process that was used for research, development and implementation of this project. Chapter Two highlights some of the prominent research that has been done on the topic of depicting uncertainty in spatial data. The design and development of the project components is discussed in Chapter Three. The next chapter includes information regarding the data used for this project. Information regarding the implementation and methodology used for creating cartographic representation tools for depicting uncertainty is available in the fifth chapter and following chapters conclude with some final thoughts and recommendations for future work that could be done to expand on this project.

Chapter 2 – Background and Literature Review

There are many different types of spatial data and each data type has unique uncertainty characteristics leading to a plethora of representation options for users. A concisely organized set of examples of representation techniques is not currently available to users that are interested in the topic of spatial data uncertainty. It was necessary to review existing literature to identify the various types of spatial data uncertainty and to determine potential symbolization methods for this project.

2.1 Literature Review

When researching cartographic representation methods for spatial data uncertainty, many variables were considered. It became apparent that representation methods for spatial data are varied and largely conditional on the type of data and the type of uncertainty. All spatial data is ultimately transformed from a real world entity to the subject matter on a map. There are three main opportunities that can introduce uncertainty or error to data: transformations from 1) the real world entity to conception, 2) conception to measurement, and 3) from measurement to analysis (Longley, Goodchild, Maguire, & Rhind, 2005). Furthermore, it is commonly agreed among researchers that spatial data needs to presented in a minimum of two ways "(i) a map of variable of interest and (ii) some assessment of uncertainty in that map" (Foody & Atkinson, 2002, p. 1). Spatial data uncertainty is an important topic, yet many GIS users do not include a cartographic representation of uncertainty for their data. This is primarily because there are not many easily accessible options to assist users in representing data, even though there has been substantial research on the topic of uncertainty. The appropriate solution to this problem would synthesize the most common, scientifically tested recommendations made by scholars in order to minimize the deficiency that currently exists for users that need a way to more easily represent their data quality on a map.

Limited awareness among many geographic information users regarding the associated error of data leads them to view and analyze data without regard for relative uncertainty. Since uncertainty of mapped information has been a longstanding issue, there has been a considerable amount of research on this topic. Authors agree that the ability to "characterize and quantify the uncertainty" is important in order to illustrate the potential problems with using the data (Blais, 2002, p. 341). Although much of the existing research concentrates on statistical methods for describing uncertainty, there is also considerable research on qualitative methods for illustrating the same uncertainty.

Scientists often need to quantify their subject matter and uncertainty in geographic information is no different. Researchers have suggested many ways for determining and quantifying the amount of uncertainty of the data; there are recommendations for specific industries, for specific data types, and more general recommendations. The common approaches to quantifying data uncertainty include using spatial statistics, sampling methods, statistical models (Mowrer & Congalton, 2000), Boolean sets or fuzzy sets (Lowell & Jaton, 1999; Plewe, 2003) to quantify the uncertainty that the data holds. Specifically, statistical techniques, such as standard error can be used for "representing and communicating uncertainty" (Lowell & Jaton, 1999, p. 151). Once quantified, this information can be shared with the audience, commonly through metadata, marginalia, or associated attribute data. The amount of uncertainty for a data item can be difficult to display in a meaningful and easily interpretable way; this is where qualitative methods for displaying uncertainty of data are needed.

The potential consequences of the relative uncertainty of spatial information have been a concern for as long as there have been maps. Maps created in GIS are typically rendered with precisely defined features, but experienced GIS practitioners recognize that spatial data have varying measures of relative error that are not always apparent to map readers. The overall quality of a map can be attributed, in part, to the included data's "uncertainty, error, bias, precision, accuracy, scale, and quality," which are all terms that are used to describe the potential problems with spatial data and can be used interchangeably but have different meanings (Kimerling, Buckley, Muehrcke, & Muehrcke, 2009, p. 205). The term "uncertainty" can seem vague, but in the case of spatial data it refers to the difference between what exists in reality versus what is depicted on the map. Error refers to the qualitative measurements used to describe uncertainty, such as root mean square error (RMSE). Bias refers to a "systematic distortion" of the data that can occur when all points are skewed in the same direction and magnitude. Additionally, precision refers to the detail at which the data is collected and accounted for. Similar to uncertainty, accuracy refers to how well the data aligns with the reality. Scale is important to spatial data uncertainty in maps created at an inappropriate scale for the data, giving the impression that the data has a higher accuracy than it really does. Finally, quality is a general term that can refer to the data collection accuracy or precision, or can simply refer to the appropriateness of the data for the map. (Kimerling, Buckley, Muehrcke, & Muehrcke, 2009) Although this project has focused primarily on uncertainty, the quality, error, bias, precision, accuracy, and scale of spatial data are important to the development of tools to represent the data quality because each of these attributes can play a separate and significant role in the relative uncertainty of data.

Furthermore, there are different types of data uncertainty. Data quality assessment measures are commonly categorized into five areas for spatial data: lineage, positional accuracy, attribute accuracy, logical consistency and completeness and were originally published in the Spatial Data Transfer Standard, Logical Specifications report from the United States Geological Survey (United States Geological Survey, 1997). Although accuracy and completeness are commonly understood terms for describing geographic information, not all users are familiar with the term lineage to describe their data quality; a dataset's lineage describes how the data were collected and modified and this has implications to the uncertainty because it informs users of the source, scale, and date of the data and, in turn, its potential usefulness to a specific application. Through review of the scholarly research, it has become apparent that "most of the efforts to formalize an approach to uncertainty visualization with geovisualization (and GIScience more generally) derive from a long-term work on spatial data transfer standards (SDTS)" (MacEachren, et al., 2005, p. 143) and the five areas described for the United States Geological Survey SDTS Logical Specifications report.

While the SDTS Logical Specifications report focuses primarily on discussing the different types of data quality assessment, the type of data can affect the recommendation for uncertainty symbolization in GIS, as well. It is commonly discussed that spatial data comes in two forms, raster or vector or, to put it another way, continuous or discrete. There has been much discussion regarding quantifying and depicting uncertainty for both of these data types. The approaches for representing the uncertainty in discrete and continuous data can be similar. For example, a recommended method for displaying uncertainty in raster data is called a “shadow map of uncertainty” (Berry, 1996, pp. 53-56) which recommends that the user indicate his or her level of uncertainty inherent in the data in order create a series of buffers or “shadows” (Figure 2-1). This leaves the user with a data set that includes a varied thickness layer around the original data that is supposed to illustrate the level of certainty in the attributes of this data.

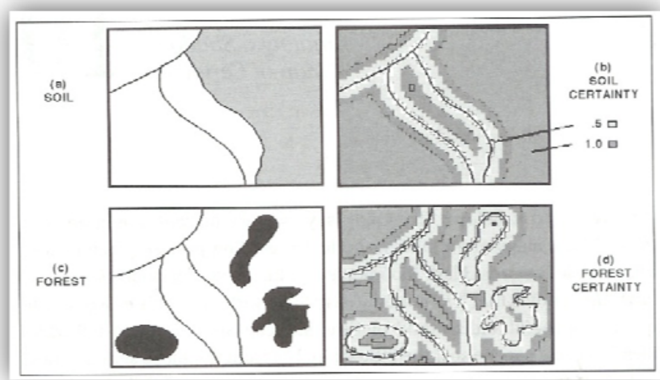


Figure 2-1: Shadow Map of Uncertainty (Berry, 1996)

Similar representation methods are recommended for vector data, namely the “error ellipse and epsilon error bands” (Zhang & Goodchild, 2002, pp. 78-87) which are widely used to describe positional errors for point and line data.

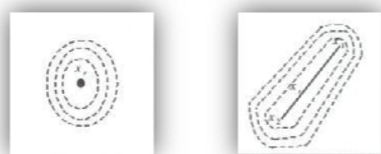


Figure 2-2: Error Ellipse & Epsilon Band (Zhang & Goodchild, 2002)

Qualitative methods often include representation methods such as dashed lines or fuzzy boundaries to indicate that the data has some level of vagueness or error and to indicate general uncertainty. Authors commonly recommend four methods for displaying uncertainty in data through cartographic representation: “contour crispness…fill clarity…fog…resolution” (MacEachren, 1992, pp. 15-16). These methods display the data in a fuzzy or unfocused way and each is slightly different to better describe the data characteristics and associated uncertainty of a particular type of data (Figure 2-3). An example of data that could benefit from this type of visualization is a choropleth map that “ignore[s] the uncertainty that results from finer-scale variation, generalization, misreporting, small numbers, and future unknowns” (De Cola, 2002, p. 364).

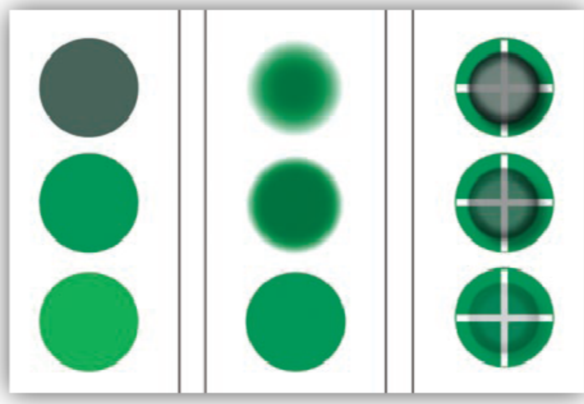
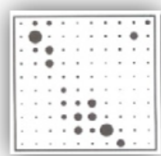
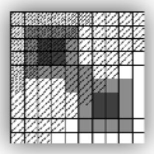
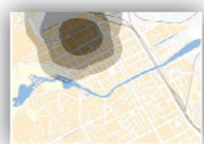
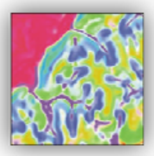
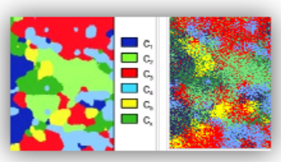
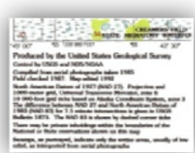


Figure 2-3: Use of saturation, crispness, and transparency to depict uncertainty (MacEachren, et al., 2005)

With that said, the most widely accepted recommendation for aligning the type of uncertainty (from the five STDS categories) with the type of data expands the categories of discrete and continuous data to include: “discrete data; categorical data for aggregation and overlay; partitioning and enumeration; and continuous interpolation” (MacEachren, et al., 2005, p. 144). This matrix assigns widely agreed upon representation recommendations to data types and the associated uncertainty type. Most of the cartographic representation recommendations from scholarly research focus on the data and uncertainty structure presented through the STDS Logical Specifications (United States Geological Survey, 1997) for the data quality descriptors and the National Center for Geographic Information and Analysis (NCGIA) Visualization of the Quality of Spatial Information report (Beard, Buttenfield, & Mackaness, 1994) for the data type descriptors. Additionally, recommended representation methods that align with STDS and NCGIA standards include the use of size, shape, texture, value, color saturation, color mixing, and additional marginalia, (Table 1) with specific recommendations such as error ellipses, epsilon bands, continuous tone vignettes, continuous tone isopleths, and blankets of error (MacEachren, et al., 2005), (Thomson, Hetzler, MacEachren, Gahegan, & Pavel, 2005).

Table 1. Examples of different methods for representing uncertainty in spatial data

Methods for Representing Uncertainty	
Size	 (MacEachren, How maps work: representation, visualization, and design, 1995)
Shape	 (Foody & Atkinson, 2002)
Texture	 (MacEachren, Vizualizing uncertain information, 1992)
Value	 (Yao & Jiang, 2005)
Saturation	 (Hengl, 2003)
Color Mixing	 (Hengl, Uncertainty visualization, 2009)
Marginalia	 (Kimerling, Buckley, Muehrcke, & Muehrcke, 2009)

Most research recommends a combination of both qualitative and quantitative measures. This indicates that a proper solution for this current project would also include methods to account for the quantitative measures while aiding users in illustrating the items in a qualitative manner. The majority of the recommendations made by scholars have focused on functional requirements for describing and visualizing uncertainty in spatial data, but there are other scholars who focused their research on uncertainty of geographic terminology and the problems resulting in attempting to represent the data in an accurate way. Additionally, very specific examples and applications have been created to help researchers account for uncertainty in spatial data for Spatial Decision Support Systems.

Spatial data uncertainty is often problematic in the area of historical GIS because of the imprecise definitions used in the historical descriptions of locations. Brandon Plewe that "incomplete, incoherence, ambiguous, vague, or conflicting definitions" as well as "vague or indefinite" measurements can be lead to substantial uncertainty in the final map output (Plewe, 2003). An example of vague geographic terminology would be "boundary 'in mountains'". It would be difficult, if not impossible for a user to accurately convey that historical boundary without having further information regarding what the terminology "in mountains" meant to the author of the historical document. This is a type of spatial data uncertainty; it is quite different from the definitions that are described by the SDTS and by NCGIA. It is also substantially different from the modern positional accuracy issues that are more common in today's GIS applications, such as GPS collection error.

Similar to historic geographic information, some SDSS applications will attempt to represent vague terminology in the output in order to support decision-making for an organization. For example, an application described by Ashley Morris and Piotr Jankowski attempts to classify features where the geographic terminology is vague. The application used different criteria to allow users to display the "nearness" of houses to a fire hydrant. The application allowed users to show how many houses were near a fire hydrant through a "Boolean representation (is near, or is not near), concentric circles where items within the core area are considered to be 100% near (value 1.0) and other circles have lesser degrees of nearness" (Morris & Jankowski, 2005). This application and associated symbolization attempts to produce data that can be used confidently even though the input terminology is unclear (Figure 2-4). Although a common GIS does not have a good method for managing the representation of lexical uncertainty of spatial data, representation recommendations offered by scholarly research for more common types to spatial data uncertainty can be appropriate. This project has not attempted to account for this type of uncertainty but some of the tools will still be applicable to users who want symbolization methods for representing vague spatial terminology in cartographic products.

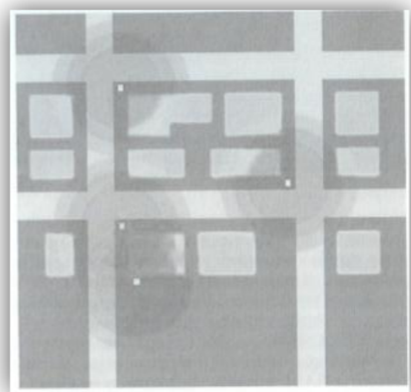


Figure 2-4: Use of multiple-ring buffers to illustrate certainty of "nearness" to a feature (Morris & Jankowski, 2005)

2.2 Background

Information fidelity is an extremely important topic to many industries because an organization's data is often the key to their successful operations. For example, a financial institution would be hard pressed to elicit any confidence from clients if it could not guarantee that its customer and financial transaction records were accurate and precise. In order to earn customers' confidence, financial institutions emphasize the quality achieved during data collection, management, and usage efforts and, as a result, this can often account for significant portions of time and financial budgets. Additionally, financial institutions will often use statistical measures to describe the confidence that they have in their data accuracy. Commonly these types of organizations will strive for "six sigma" a term used to describe the number of standard deviations that would account for an acceptable error rate (Six Sigma, 2009). Different types of organizations may spend greater or lesser efforts for their data collection and management. Typically, the more mission-critical that the information is to an organization, the more important it is to be certain. Similar to financial data, spatial data is collected, managed, and used in an effort to maintain its accuracy and precision, but all spatial data has inherent error and uncertainty associated with it. As the data becomes more mission-critical it becomes more important for the data to be as accurate as possible. Spatial data differs from financial data though, because maps are not often internal, organizational documents. Maps are often accessed by users who are not familiar with the quality of the data and assume that it is accurate.

Sharing data quality through visual representation methods is, arguably, even more important now than in the past. The public is becoming more aware of spatial data and has greater access to it than in the past, through readily available applications like Google Earth. It is possible that more people will be negatively affected or misled by data that is being offered for consumption without an explicit notice regarding the data quality and uncertainty. It can be argued that providing geographic information to the public along with a disclaimer, such as metadata, should be sufficient, but in many cases users don't read or care about metadata. Providing a visual cue to the data's quality forces the user to be more aware of the potential implications of using the data.

2.2.1 **Existing Recommendations**

Existing recommendations for representing spatial data uncertainty are extensive and varied, but the most widely accepted research offers specific representation methods for fourteen potential types of uncertainty, based on the type of data being used and the quality of the data (Figure 2-5). Even infrequent users of maps would probably be familiar with some of the recommendations that are suggested. Most likely, map users would be able to easily understand marginalia as a method for describing data quality, even if they were not very familiar with mapped information. Other representation methods, such as error ellipses, are less familiar to uninformed users but are a common recommendation for symbolization of discrete data, especially where statistical measures have been used to quantify the uncertainty of the data, and would be more easily understandable to GIS professionals.

Data Type / Data Quality	Positional Accuracy	Attribute Accuracy	Logical Consistency	Completeness	Lineage
Discrete Points and Lines	Size Shape (Error ellipses) (Epsilon bands)	Value Color Saturation (Feature code checks)	Color mixing Redundancy by overprinting Slivers by solid fills (Topological cleaning)	Mapping Technique Density traces Marginalia Generalization algorithm Mapping tolerance Buffer size	Mapping Technique Minimum Bounding Rectangles
Categorical Aggregation & Overlay (Tesselation, tiling, Areal coverages)	Texture Value (Certainty of boundary location)	Color mixing (Attribute code checks) (Topographic classifier)	lack error models	Mapping Technique Missing values Logical adjacency surface Marginalia Discrete model weights	
Partitioning & Enumeration (Metric class breaks)	not meaningful	Size = height (Blanket of error)	Size = height (Maximum likelihood prism maps)	Mapping Technique Missing values Missclassification matrix Classing scheme OAL/TA)	Marginalia Source of data Scale/Resolution Date Geometry
Continuous Interpolation (Surfaces and volumes)	no clear distinction b/w the two Value Color Saturation (Continuous tone vignettes) (Continuous tone isopleths)		Size = line wt Color Shape = compactness (TIN links)	not possible by definition Mapping Technique Surface of search attenuation Marginalia Interpolation algorithm	
	Graphical Syntax			**Graphical/Lexical Syntax**	

Figure 2-5: Matrix for Visualizing Cartographic Symbolization Recommendations (MacEachren, et al., 2005, p. 144)

There are countless symbolization recommendations for spatial data uncertainty but currently there is not a central repository where users can obtain examples and samples of the recommendations that are discussed in the literature. If spatial data uncertainty symbolization is going to become a more common practice among GIS users, there needs to be easily accessible options and a location where users can obtain information regarding uncertainty representation methods. It was the purpose of this project to dissect and synthesize tested symbolization methods (Figure 2-1) and make them available to a wider audience by presenting them outside of scholarly research articles.

2.2.2 **Existing Symbolization Process**

Current spatial data uncertainty symbolization is limited to academic and, to a lesser extent, government geographic information users. The primary reason for this is because these users are more likely to access scholarly articles regarding the different recommendations. They also understand that all spatial data has error associated with it, and highlighting the error does not detract from the value; in fact, it likely adds value to data. On the other hand, many other users would like to divert undue attention from the error associated with the data that they collected and are managing.

Currently users who are interested in symbolizing their spatial data uncertainty would need to research appropriate recommendations and compile the cartographic representation for use with their data. In reality, most users who are somewhat familiar with the process of symbolization for the relative uncertainty of geographic data may attempt to use a symbolization process that makes sense to them, but they will almost certainly not choose to research-tested methods of cartographic representation of their specific data uncertainty before adding it to the map. Additionally, users who do take the time to research tested methods of symbolization will need to create their own cartographic symbolization in GIS and this can be a time consuming process; creating the custom symbolization may comprise a substantial part of their map building process and the value of sharing the data quality information may become an afterthought in order to complete the project in a timely manner, especially if the symbolization will be used infrequently.

2.3 Summary

There has been a significant amount of research completed on the topic of spatial data uncertainty and a majority of the research agrees that appropriate methods for illustrating spatial data uncertainty cartographically take into account at least five different types of uncertainty sources: attribute accuracy, positional accuracy, completeness, logical consistency, and lineage, as well as different forms of data, including: point, line, polygon and field. The recommendations that existing research offer are important to users who are interested in implementing, improving, or streamlining their symbolization process for the uncertainty in their data. System design for this project has accounted for the most common suggestions of methods for cartographic representation of spatial data uncertainty, which were revealed through literature review.

Chapter 3 – Systems Analysis and Design

A project and systems analysis was completed prior to the development of the components of this project. This was done to determine the users' needs for depicting uncertainty in spatial data and the major tasks that would be required to develop the tools. The systems analysis and design describes five major tasks that were completed in the development of this project and the three key deliverables that were provided to the client.

3.1 Problem Statement

Maps created with a GIS are typically filled with precisely defined points, lines, areas, and surfaces. However, some data are not nearly as accurate or certain as they are displayed in a GIS. Currently there is not a clearly defined process or toolset for users to display the accuracy or uncertainty of the data in a cartographic representation. In some cases, it would be desirable to display and analyze the areas of ambiguity with less visual prominence or in a different way than the data that has higher accuracy. Additionally, different data types have different aspects of uncertainty, and it is important to be able to display the type of uncertainty present in that data. Indicating the special spatial pattern of data uncertainty through a custom cartographic representation is crucial to different industries for the usefulness and accurate analysis of the data. ESRI has determined that there is a need to be able to depict the uncertainty of data through representation methods within the ArcGIS suite of products. This project focused on developing methods and symbology sets to illustrate the different types of spatial data uncertainty.

3.2 Major Tasks Summary

Development for this project utilized a traditional development lifecycle. Major tasks included requirements analysis, data collection, development, deployment and documentation. Each task was completed in its entirety before the next phase began. This approach permitted planning, and development to advance in support of the stated objectives of the project.

3.2.1 **Requirements Analysis**

A requirements analysis revealed the needs for both the client and anticipated users. Analysis included a thorough review of data and software requirements, as well as the anticipated schedule, risks, and assumptions. The requirements analysis for this project included the following steps:

1. Determining the client's needs
2. Determining the users' needs
3. Assessing users' current processes
4. Determining tested methods of representation

3.2.2 **Data Collection**

Data selection for this project focused on the anticipated user groups and the client's interests, determined from the requirements analysis. Ultimately, the data used for the development of this project primarily focused on the natural resource industry and were collected from a number of sources. Furthermore, it was important to use varied data types, such as point, line, area, and continuous surface data. Specifically, this project used three main data layers: earthquakes, soils, and sediment acetone levels. Each of the datasets had an element of uncertainty that was obtained directly from the data attributes, generated for purposes of this project, or from associated data layers.

3.2.3 **Development**

Project development focused entirely on utilizing tools that were available in ArcGIS Desktop. The software has preexisting tools for representing spatial data which were used for creating the sample symbolization methods for five types of spatial data uncertainty.

Development began with determining which existing tools could be used to create the appropriate symbology for illustrating the different types of data quality characteristics in different data types. Appropriate methods for representing various types of spatial data uncertainty were determined from a review of the literature and with agreement from the client.

3.2.4 **Deployment**

The finalized and client-approved symbology samples were deployed through ESRI's online sharing website, ArcGIS Online (www.arcgisonline.com) through a group that was created, and is managed by the author. The Depicting Uncertainty group on ArcGIS Online focuses on providing resources for users that are interested in the topic of spatial data uncertainty and different methods for representing it in a GIS. Samples of different representation methods are offered for download to users as Layer Packages that can be opened and evaluated in ArcGIS Desktop products. Additionally, the examples are provided as ArcGIS Desktop map document project files (.mxd) in order to give users a model of how the symbolization looks in a final product. Providing the project files to the users was particularly important in cases where data has been layered to create the illustration of data uncertainty because Layer Packages only allow for the sharing of one layer at a time.

The representation methodologies have been provided with each of the Layer Packages. The methodology has been provided as step-by-step instructions for creating the symbology that illustrates the quantity and quality of uncertainty in the datasets.

3.2.5 **Documentation**

Documentation and resources were provided through the ArcGIS Online group website. The primary resource provided through this website is a table that organizes representation recommendations based on the type of data being used and the type of uncertainty inherent in the data. Symbology samples were provided for some of the recommendations from this resource. Additionally, the table is a resource for users who are interested in obtaining information on other methods for representing spatial data uncertainty where sample representations have not been provided. Also, where appropriate, links have been provided for documents and websites that allow users to obtain more information regarding a particular symbolization method.

The complete details regarding the project development, processes, methodologies, and recommendations for future work are provided in this document. The completion of this document comprised a significant portion of the project and is one of the main deliverables. Along with information regarding the project design, development, implementation, and analysis, this document provides recommendations for future work that could be utilized by users who are interested in developing methods for other types of representation of spatial data uncertainty.

3.2.6 **Project Workflow**

The project workflow details the series of processes, along with decision points that required the client's approval, and were produced in order to direct a successful project plan and completion. The project plan culminated in the final documentation and all of the deliverables having been delivered to the client (Figure 3-1).

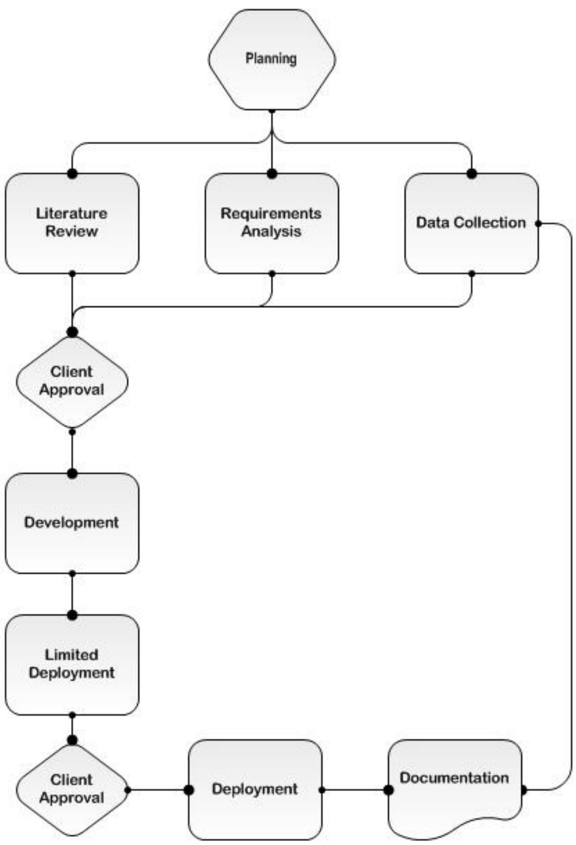


Figure 3-1: Project Workflow

3.3 System Requirements

The requirements analysis confirmed that the project development required three major components: a way for users to determine the appropriate representation method for their data; symbology recommendations and tools for various types of data; and documentation detailing the methodology for the symbology recommendations. The three components were provided through an online format that would allow users to freely access the data and examples.

One of the major requirements for this project was that it was developed using ArcGIS Desktop version 9.3.1 because the client required the symbology samples to be delivered in Layer Packages and this data format was not available in previous versions of the software. Furthermore, it was necessary to provide the various sample symbology layers for online download access. This was required by the client in order to allow for rapid deployment of the project deliverables, easy access for users, and an opportunity for user feedback. Additionally, the client preferred that the project development used tools that already existed in the software, and provide methods for making the representation process less complicated (Table 2).

Table 2. Project Requirements

Requirement	
Use ESRI ArcGIS Desktop software (version 9.3.1)	Mandatory
Deploy deliverables to allow for download access	Mandatory
Provide samples for multiple types of data	Mandatory
Provide ability for users to determine the type of uncertainty associated with their data type and quality	Mandatory
Incorporate data that are already available for general public access	Mandatory
Automate representation process	Desirable
Use existing tools to create representation samples	Desirable

3.4 Project Plan

The project plan describes the major tasks, schedule, primary deliverables, risks, and assumptions associated with this project. Most significantly, the final deliverables for this project were a little different from what was detailed in the project proposal but the major tasks remained similar throughout the project modifications. A reasonable adherence to the project schedule, and consideration of the potential project risks, and minor assumptions for this project allowed major milestones to be met and be completed in the planned timeframe.

3.4.1 **Initial Project Plan**

The project development changed slightly from the original proposal due to the client's request to utilize newer technology and tools that became available with the latest release of the development software. The originally proposed toolset was to be a comprehensive, industry-independent solution that would allow users to illustrate a measure of uncertainty within data. The toolset would guide users' symbology processes and allow them to view and analyze the data more accurately. The original toolset was anticipated to include three major components:

- Style: A general symbology set, including representation methods for many data types and user groups
- Toolbar: an interactive tool that will allow the user to further describe the quantitative and qualitative aspects of the data uncertainty
- Documentation: a combination of help and business process document that will allow novices and intermediate users to determine better ways for describing the data uncertainty within their organization.

Additionally, the schedule for this project was slated to run from mid-March through mid-October (Figures 3-2 & 3-3), with the user needs assessment and literature review ending in April and the design and development of the style, toolbar, and documentation; this process was estimated to take 95 days to complete. Additionally, two weeks of testing were planned, with anticipation of the toolset being deployed in approximately mid-October.

Task Name	Duration	Start	Finish
User Needs Assesment	30 days	Mon 3/16/09	Sun 4/26/09
Style	**30 days?**	**Tue 4/28/09**	**Mon 6/8/09**
Design	10 days	Tue 4/28/09	Mon 5/11/09
Development	20 days?	Tue 5/12/09	Mon 6/8/09
Toolbar	**40 days?**	**Tue 6/9/09**	**Mon 8/3/09**
Design	10 days?	Tue 6/9/09	Mon 6/22/09
Development	30 days?	Tue 6/23/09	Mon 8/3/09
Documentation	**25 days**	**Tue 8/4/09**	**Mon 9/7/09**
Design	3 days	Tue 8/4/09	Thu 8/6/09
Development	22 days	Fri 8/7/09	Mon 9/7/09
Testing	10 days?	Tue 9/8/09	Mon 9/21/09
Deployment	5 days?	Mon 10/5/09	Fri 10/9/09

Figure 3-2: Original Uncertainty Toolset Schedule

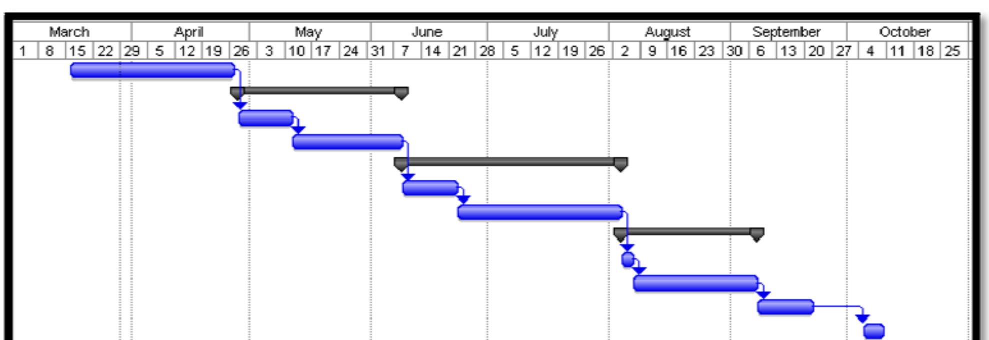


Figure 3-3: Original Uncertainty Toolset Schedule Chart

3.4.2 **Schedule**

Although the schedule for this project changed from the original project plan proposal, it provided a guideline for project development throughout the project design and development phases. The schedule change was primarily due to a change in the development and deployment requirements of the client, discussed previously. The total time allowed for project completion was one year and the schedule provided a general outline of the major tasks and their start and completion dates (Figures 3-4 & 3-5).

Notably, the user needs assessment ran to mid-July and took longer than originally expected. The bulk of the major tasks were focused on the later months of the project, with major design and development activities taking place from early September through late November. The significant tasks included the design and development of the representation samples, and the documentation. The project completion was expected to be no later than December 4, 2009. At this point the documentation and all deliverables were approved and received by the client and committee members.

Task Name	Duration	Start	Finish
User Needs Assesment	90 days	Mon 3/16/09	Fri 7/17/09
− Representation	**40 days**	**Tue 9/8/09**	**Mon 11/2/09**
Design	30 days	Tue 9/8/09	Mon 10/19/09
Development	10 days	Tue 10/20/09	Mon 11/2/09
− Documentation	**54 days**	**Tue 9/22/09**	**Fri 12/4/09**
Design	10 days	Tue 9/22/09	Mon 10/5/09
Development	40 days	Mon 10/12/09	Fri 12/4/09
Deployment	5 days	Mon 11/23/09	Fri 11/27/09

Figure 3-4: Depicting Uncertainty Schedule

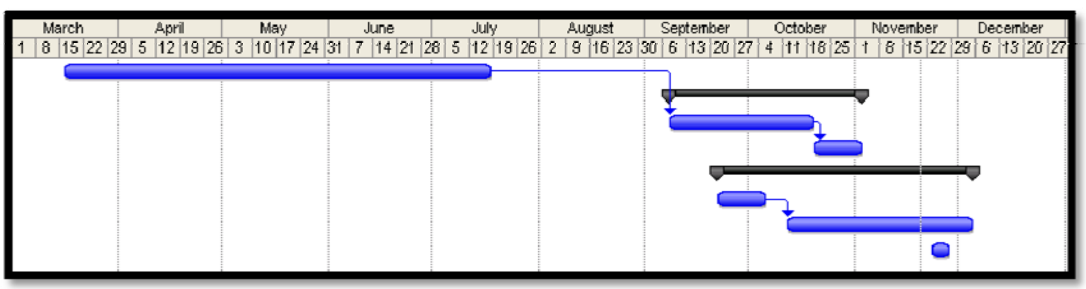


Figure 3-5: Depicting Uncertainty Schedule Chart

3.4.3 **Deliverables**

This project had two main deliverables: the representation samples, and documentation (Figure 3-1). In addition, the user needs assessment summary was provided to the client for review and approval.

1. User Needs Assessment Summary: a summary of findings from client meetings and literature review.

2. Representation Samples: a set of recommended representations that were to be delivered in Layer Package format for use with ArcGIS Desktop.
3. Documentation: information, both methodology and links to other resources, that both inform the user about the subject of uncertainty in spatial data and how to use the representation methods.

3.4.4 Risk Assessment

Several items posed a risk to the success and completion of this project. The potential for project scope creep and inexperience with the development framework were the largest, foreseeable risks for this project (Table 3). Mitigation of the stated risks was managed in order to ensure that their severity was minimized and did not exceed the expected levels. Project success was ensured primarily by maintaining contact with the client and periodically reviewing the project status against stated schedule and goals.

Table 3. Risk Assessment

Risk	Severity	Probability	Mitigation	Exposure
Project Creep	4	3	-Maintain frequent communication with client -Maintain frequent communication with advisors -Ensure that the client understands the MS GIS program requirements as it relates to the project	12
Complexity	3	3	-Develop prototype early in order to determine unexpected pitfalls -Determine alternative design options -Obtain client's feedback regarding project status	9
Inexperience with Development Tools	3	4	-Obtain additional resources for strengthening skills with applicable tools -Clarify product requirements with client	12

3.4.5 Assumptions

The success of this project relied on timely client response to key components of the project, such as those listed in the schedule section of this document. The client provided guidance throughout the development and deployment process on the ArcGIS Online website. The client also provided crucial feedback and key decision-making points, as described in the project workflow.

3.5 Summary

The client's users need a more efficient way for representing the quality of the relative uncertainty that occurs in their spatial data. In order to effectively build tools that meet the needs of the users, a requirements analysis was conducted. The requirements analysis considered the tasks that would be essential to the successful completion of this project, including data collection, development, deployment, and documentation. The project requirements further detailed the functional and non-functional requirements for the major development tasks. The project schedule, list of deliverables, risk assessment, and assumptions assisted with keeping the project on task even though the project was modified since the original project proposal. The final deliverables, namely the representation samples and documentation, were developed with the goals and objectives of the client and user requirements in mind, and were created to meet the guidelines of the project and system requirements.

Chapter 4 – Data Considerations

Although this project did not require a substantial database design or have an extensive data collection requirement, data was still an important consideration. The project requirements analysis revealed various data types and associated error that could benefit from having their associated spatial data uncertainty represented cartographically. Data was selected for this project based on their potential to be incorporated into the types of data quality that were to be represented in the final deliverables.

4.1 Spatial Data Uncertainty

The literature review and requirements analysis for this project showed that there were numerous ways that a user could attempt to represent different types of spatial data uncertainty. It became apparent that it was important to have a clear and concise method for determining the type of representation that would be associated with different data types and different types of uncertainty. Having an organized method for determining the appropriate representation methods for various data types provided useful for the development of this project but also, later for a resource for users that were interested, more generally, in the topic of cartographic representation methods of spatial data uncertainty.

The most commonly agreed upon recommendations for depicting spatial data uncertainty in different data types were produced by Dr. Barbara Buttenfield. Table 4 organizes recommended methods for illustrating uncertainty in spatial data into twenty different categories based on the specific data type and data quality. Specifically, this table provides representation recommendations for point, line, polygon and continuous data, with quality concerns in the areas of “positional accuracy, attribute accuracy, logical consistency, completeness and lineage” (MacEachren, et al., 2005, pp. 143-144).

Table 4. Recommended Methods for Depicting Uncertainty in Spatial Data. Recreated from (MacEachren, et al., 2005), from Buttenfield & Weibel

<table>
<tr><th>Data Quality / Data Type</th><th>Positional Accuracy</th><th>Attribute Accuracy</th><th>Logical Consistency</th><th>Completeness</th><th>Lineage</th></tr>
<tr><td>Discrete Points & Lines</td><td>Size
Shape
(error ellipse, epsilon band)</td><td>Value
Color
Saturation
(feature code checks)</td><td>Color Mixing
Redundancy by overprinting.
Slivers by solid fills
(Topological cleaning)</td><td>Mapping Technique
Density Traces
Marginalia
Generalization algorithm
Mapping Tolerance
Buffer Size</td><td rowspan="4">Mapping Technique
Minimum Bounding Rectangles

Marginalia
Source of data
Scale/Resolution
Date
Geometry</td></tr>
<tr><td>Categorical Aggregation & Overlay</td><td>Texture
Value
(Certainty of boundary location)</td><td>Color Mixing
(Attribute code checks, topographic classifier)</td><td>Lack error models</td><td>Mapping Technique
Missing Values
Logical adjacency surface
Marginalia
Discrete model weights</td></tr>
<tr><td>Partitioning & Enumeration</td><td>Not meaningful</td><td>Size = height
(Blanket of error)</td><td>Size = height
(Maximum likelihood prism maps)</td><td>Mapping Technique
Missing Values
Logical adjacency surface
Classing scheme
OAL/TAI</td></tr>
<tr><td>Continuous Interpolation</td><td colspan="2">Value
Color Saturation
(continuous tone vignettes, continuous tone isopleths)</td><td>Size = line wt
Color
Shape = Compactness
(TIN links)</td><td>Not possible by definition</td></tr>
<tr><td></td><td colspan="2">← Graphical Syntax →</td><td colspan="3">← Graphical/Lexical Syntax →</td></tr>
</table>

The four data types that are accounted for in this table are discrete points & lines, categorical aggregation & overlay, partitioning & enumeration, and continuous interpolation (MacEachren, et al., 2005). Interestingly this research recommended similar representation methods for points and lines but offered separate recommendations for polygon data, depending on whether it is organized numerically or categorically. Furthermore, there are five categories of data quality described by this research and are widely accepted by many scholars. The five types of data quality were first introduced by the USGS in the SDTS Logical Specifications.

1. **Lineage**: Data quality is determined by the specific source of the data; this is often described by data collection date, publication date, data provider, and documentation of “mathematical transformations of coordinates” (United States Geological Survey, 1997, p. 13). This information is particularly significant to the associated quality for

data that is combined from different sources, collected on different dates, or merged from different coordinate systems. Typically information regarding the lineage of the data is provided to users through the associated metadata, in digital data, or through marginalia on printed map documents.

2. **Positional Accuracy**: Refers to the quality of data as it relates to the difference between the data location in digital space and its real world location or its "degree of compliance to the spatial registration standard" (United States Geological Survey, 1997, p. 14). The USGS recommends that the positional accuracy of data is obtained by using one of four methods: "1) deductive estimate, 2) internal evidence, 3) comparison to source, or 4) independent source of higher accuracy" (United States Geological Survey, 1997, p. 14). This information is often provided to users as a statement of the acceptable error tolerance or estimated error in the positional accuracy.
3. **Attribute Accuracy**: Quality in data attributes is explained simply as the discrepancy of spatial data attribution with that of the real world. The USGS recommends that accuracy reporting procedures should be the same for continuous data as it is for positional accuracy. Attribute accuracy for categorical data, on the other hand, should be obtained through "1) deductive estimate, 2) tests based on independent samples, or 3) tests based on polygon overlay" (United States Geological Survey, 1997, p. 14).
4. **Logical Consistency**: The relative quality of the association of the data, topological consistency and evaluation of the given attributes against possible attributes. Logical consistency errors can be detected and reported to users by using "tests of valid values, general tests for graphic data or specific topological tests" (United States Geological Survey, 1997, p. 15). For example, uncertainty can be introduced in data where the data have values attributed to them that are not reasonable for the data set, lines that do not connect or intersect where they should, or data has been entered more than once or not at all.
5. **Completeness**: Data quality is also affected by completeness, meaning what is and is not included in the dataset. As discussed in earlier chapters, all data is generalized by one or more of the following spatial operations: simplification, smoothing, aggregation, amalgamation, collapse, merging, refinement, exaggeration, enhancement, or displacement and any one of these operations will lead to an overall reduction of the information provided in the data or on the map (Kimerling, Buckley, Muehrcke, & Muehrcke, 2009). Generalization is necessary in order to convey the appropriate information for the subject-matter and intended scale but it also leads to a certain degree of data quality degradation. It is suggested that information regarding the specific generalization techniques that were used to create a dataset can be provided in terms of the smallest feature included, aggregation methods or selection choices, for example (United States Geological Survey, 1997).

4.2 Data Selection

Due to the length of time that was allowed for project development and the practical scope of the project, not all of the categories were accounted for in this project. Due to the nature of geographic information, anticipated user groups, and client feedback, it was clear that focusing the development on positional and attribute accuracy would be the most appropriate. Quality concerns due to the errors in the position or attributes of spatial data are commonly understood by GIS users, therefore representation methods for this type of data were considered to be useful to a larger audience than the other data types. Specifically, the data types and data quality that were included in this project were:

- Discrete point data with attribute accuracy concerns
- Categorically aggregated polygons with positional accuracy concerns
- Categorically aggregated polygons with attribute accuracy concerns
- Continuous field data with attribute accuracy concerns (Table 5)

Table 5. Data Types and Data Quality Selected for Project Development

Data Type \ Data Quality	Positional Accuracy	Attribute Accuracy	Logical Consistency	Completeness	Lineage
Discrete Points & Lines	Not Included	**Included**	Not Included	Not Included	Not Included
Categorical Aggregation & Overlay	**Included**	**Included**	Not Included	Not Included	Not Included
Partitioning & Enumeration	Not Included	Not Included	Not Included	Not Included	Not Included
Continuous Interpolation	Not Included	**Included**	Not Included	Not Included	Not Included

4.3 Data Sources

Data were collected that would provide suitable data quality attributes for representing the associated spatial data uncertainty and that would satisfy the project development requirement goals (Table 6). Specifically, the data that were selected was earthquake data, soils data and sediment composition. The particular types of data were selected, in part, because they could be provided to users without specific domain knowledge and the characteristics of the data and associated uncertainty could be understood.

Table 6. Depicting Uncertainty Project Data Sources

Data Type \ Data Quality	Positional Accuracy	Attribute Accuracy
Discrete Points & Lines		**Data:** California Earthquake History 1769 - Present **Provider:** USGS **Uncertainty Measure:** Year of event
Categorical Aggregation & Overlay	**Data:** SSURGO Soils for San Bernardino County **Provider:** NRCS **Uncertainty Measure:** Derived	**Data:** SSURGO Soils for San Bernardino County **Provider:** NRCS **Uncertainty Measure:** Derived
Continuous Interpolation		**Data:** Sediment Contamination (Acetone) for Salton Sea & Study Points **Provider:** Redlands Institute, USGS & USFWS **Uncertainty Measure:** Sample point density

The USGS earthquake data includes information from as far back as 1769 and the attribute of uncertainty with this dataset was the year of the event. The reason that this attribute was selected to represent the uncertainty of this data is because it is assumed that the more current events would have been measured with more accuracy than past events, due to technological improvements in the measuring devices. The NRCS soils data were selected due to their inherent error in both positional and attribute accuracy but the uncertainty was largely derived for purposes of this project. Finally, the Salton Sea sediment composition data includes attribute errors, primarily because the surface was interpolated from a series of sample locations. The uncertainty attributes, in this case, were derived from the density of sample locations, with the thought that the attributes of the data were likely to be more accurate in locations where there are a greater density of sample locations.

4.4 Summary

Spatial data uncertainty can be described in a variety of ways but this project development focused on research that effectively segmented uncertainty representation recommendations into 20 categories. In particular, it gave recommendations for data types: point, line, polygon, and continuous surface; and data quality categories: attribute accuracy, positional accuracy, logical consistency, lineage and completeness. Furthermore, it was determined that this project would focus exclusively on positional and attribute accuracy issues for point, polygon, and continuous data, due to the project scope and potential usefulness to anticipated user groups. As a result, this project focused on three key datasets: USGS earthquakes, NRCS soils, and USGS and USFWS lake sediment composition data.

Chapter 5 – Implementation

5.1 Cartographic Representation Methods Implementation

Representation methods for depicting the uncertainty in the attributes or positions of points, lines, polygons, or continuous surfaces vary depending on what would best illustrate the data's strengths and weaknesses. This project focused on implementing methods for representing spatial data uncertainty in ArcGIS.

5.1.1 **Attribute Accuracy in Discrete Points**

Point data can have error introduced to their attributes through incorrect data entry, assignment, or measurement. Project development focused on an earthquake dataset that included event data for earthquakes from 1769 through 2001 in order to illustrate the recommended representation method for discrete point data that has an inherent error associated with its attributes. In this example, the older events are assumed to have less certainty in the stated value of the magnitude. More recent earthquakes are assumed to have more accurate magnitudes because the measurement techniques for earthquake events have improved over time.

The recommended cartographic representation for discrete points with associated attribute errors is to use value, color, or saturation in order to distinguish the more certain data points from the less certain ones (Figure 5-1).

- Value refers to the use of one main color that changes to a darker color by adding a greater amount of black. An example of this would be a color ramp that starts with red and continues to black at the other end. When using value to represent the attribute uncertainty in spatial data it is normally most appropriate to use the brightest color to depict the points that have the most certainty in their attributes and the darker values to illustrate the points that have attribute values that are less certain.
- Color (also referred to as hue) can also be used to demonstrate variations in attribute uncertainty in a dataset. Varying colors should be used cautiously, though, as there is the possibility of confusing the audience even more. Appropriate use of color for this type of representation would entail using a series or continuum of colors that effectively communicates the high, medium, and low points of uncertainty in the data attributes. For example, users could symbolize their data using a color continuum from purple to pink; purple representing the data with the most accurate attributes because it is more prominent on light background colors.
- Saturation is similar to value except it uses varying levels of white instead of black to change the appearance of the main color. An example of this would be a color ramp that starts with red and continues to white at the other end. A symbol with more saturation—more of the main color and less white—is typically used to symbolize the data with the most accurate attributes and the lighter values would be used to illustrate the points where the data attribute values are less certain.

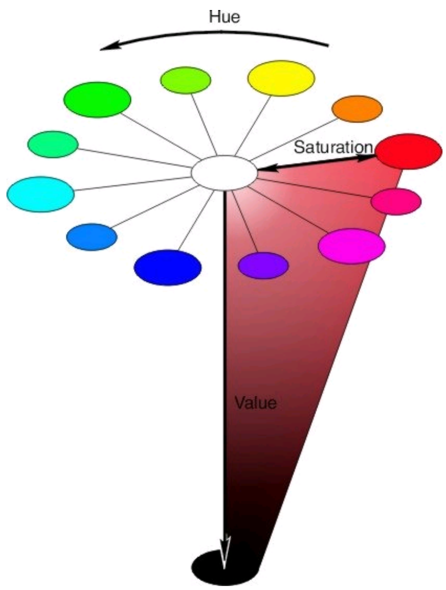
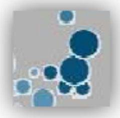
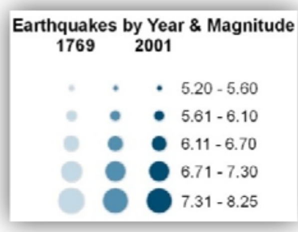


Figure 5-1: Hue, Saturation and Value (North Carolina State University, 2000)

The representation method for the earthquake data in this project used saturation to convey the relative amount of the attribute uncertainty in each of the points, and the size of the point symbol to convey the magnitude of the earthquake. Symbolization of the data accuracy and attributes required the use of Multiple Attribute symbolization in ArcGIS Desktop. Additionally, the legend for this type of representation requires special consideration because the multiple attribute symbolizations makes the legend unmanageable for practical use, therefore it is recommended that a second legend be created for use in the final map output (Table 7).

Table 7. Representation Method for Point Data with Attribute Uncertainty

Uncertainty Type: Discrete Points with Attribute Accuracy Quality Concerns	
Symbology 	**Symbolize Multiple Attributes** • Layer Properties > Symbology Tab > Show: Multiple Attributes > Quantity by category > Value Fields: 1) Uncertainty Field, 2) Data Category Field > Variation by: Symbol Size > Color Scheme
Layout 	**Add Layers to the Data Frame** • Copy & paste three layers in the Data Frame > Layer Properties > Symbology Tab > Show: Quantities > Graduated symbols > Change to match high, mid & low values of original layer **Create Legend with Added Layers** • Insert > Legend > Map Layers: Select the three new layers only > Set the number of columns in your legend: 3 > Double-click the Legend to open Legend Properties > Items Tab > Map Connection: Uncheck the Only display layers that are checked on in the Table of Contents **Modify the Legend Properties** • Right click the legend > Properties > Layer Properties > Style > Properties > General Tab > Uncheck Show Layer Name and Show Labels to achieve the desired appearance
License level required: ArcView	

The final map for this representation method was created on a medium grey background layer. It is important to consider the surrounding data's potential effect on user's visual interpretation of the data when selecting the use of value, saturation, or hue for the symbols. The symbols in Figure 5-2 range from light blue to dark blue, representing points of high uncertainty and low uncertainty, respectively. In this example the lighter symbols fade into the background visually which is appropriate because they should be analyzed with a lower weight due to the uncertainty in the attribute accuracy. Conversely, the darker points are those that have the most certainty in their attributes and should be analyzed with more weight than the lighter points.

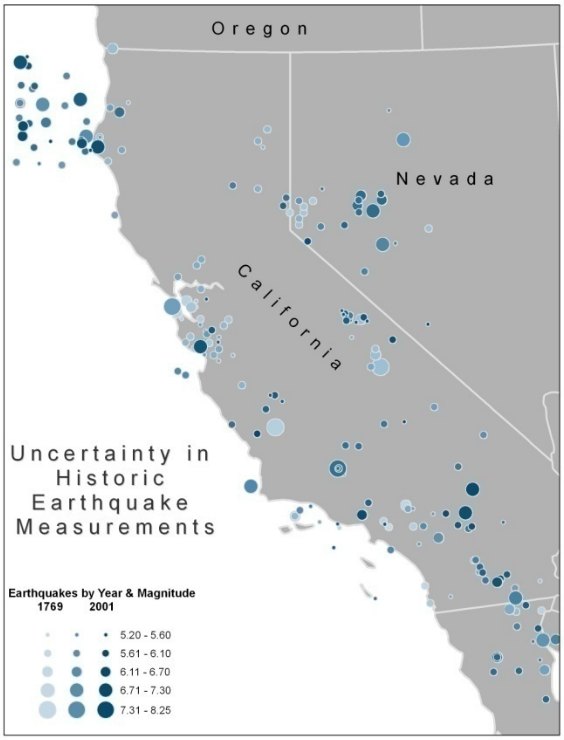


Figure 5-2: Depicting Points Data with Attribute Uncertainty

Using the concept of figure-ground, the data are symbolized in order to draw the user's attention to the most accurate data first and the least accurate data last. With this in mind, it is worth noting that if the background was darker the user would need to invert the color ramp used for the point data symbology. Doing this would make the less certain symbols dark and the more certain symbols lighter; on a dark background the light symbols would be more visible and the dark symbols would fade into the background. It is important when choosing a color scheme for representing this type of uncertainty to choose a sequential color scheme, either single hue or multi-hue, as opposed to a diverging color scheme. A diverging color scheme has the potential to confuse audiences about the relative certainty of data attributes, especially when the colors at the center of the color ramp are neutral.

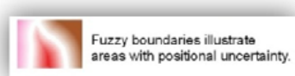
5.1.2 Positional Accuracy in Categorical Polygons

Categorical data that has been organized in polygons can have error associated with the accuracy of their positions. The error occurs when the positions of the spatial data are different than what occurs in the real world. With polygon data with categorical attributes the positional accuracy errors tend to occur along the borders and it is unlikely, but not impossible, that an entire polygon would have positional errors related with it. Examples of categorical data that might have error associated with their positions are soil and land cover data. The symbology methods for this example used NRCS SSURGO soil data with the positional accuracy concerns being derived in order to best represent the recommended symbology methods. In this example, individual areas were assigned arbitrarily to have less positional accuracy than the others. In soil data, some areas may have less positional accuracy based on the soil's characteristics or data collection techniques.

The use of texture and value is recommended for representing the positional accuracy of boundary locations in categorically aggregated data (MacEachren, et al., 2005). Value, as discussed in detail above, is the use of one key color that progresses to black in a color ramp. Texture in the form of image overlays, marker symbols, or buffered boundaries could all be used to illustrate the characteristics of the positional accuracy. It is important when deciding between the recommended types of representation to use one that is appropriate for the data type and data quality (Table 3) and that will be the most intuitive for users. For instance, when using texture as the method for depicting the positional uncertainty in data it is crucial to use a texture that assists users in visually analyzing the uncertainty characteristics but that is not confusing to audiences. Using inappropriate symbology can distract attention from the information being presented through the symbology in the map.

The representation development for this project used a vignette that fades from the main color near the body of the polygon to a neutral color toward the edge of the boundary (Table 8). The vignette uses a series of buffers to give the boundary locations a fuzzy look. The fuzzy appearance gives users a visual cue to the uncertainty of the boundary locations, whereas a crisp boundary would lead users to assume that the boundary location was very certain.

Table 8. Representation Method for Categorically Aggregated Data with Positional Uncertainty

Uncertainty Type: Categorical Polygons with Positional Accuracy Quality Concerns	
Symbology 	**Create Vignette** • "Fade to white background" effect (Appendix B) > Create buffers for vignette > Use negative distances in same proportions including -0.1 > Adding fields to a table > Calculating the transparency level > Modify expression to account for negative values (e.g. -((100 * [distance])/36)) > Symbolizing the buffers **Overlay Vignette on Symbolized Polygons** • Ensure that the vignette layer is on top of the polygon layer in the Data Frame
Layout 	**Create Symbol Thumbnail** • In the Data Frame View zoom into an area of the data that illustrates the symbology well > File > Export > Save • Insert > Picture **Add Text Explanation of Symbol Meaning** • Insert > Text > add text explaining the meaning of the symbology > align thumbnail and text to be adjacent in the layout
License level required: ArcView	

The vignette technique used for this representation (Figure 5-3) was modified from the one provided by the ESRI Mapping Center (ESRI, 2009b) for coastal boundary representation (Appendix B). The original method recommended by the ESRI Mapping

Center recommends creating multiple ring buffers using values of 5, 9, 13, 16, 19, 22, 24, 26, 28, 30, 31, 32, 33, 34, 35, & 36 or the same proportions. This project used values appropriate for the scale of the data but used the same proportions that were recommended by ESRI. Furthermore it is important to create a field in the attribute table to set the transparency levels, using the equation given by ESRI Mapping Center team as a guideline for calculating the transparency levels of the different buffers.

$$100-((100 * [distance])/36)$$

Completing these steps will ensure that the transparency levels can be set to move from the main color of the polygon to a neutral color for the buffers that are closer to the edge of the polygon boundary.

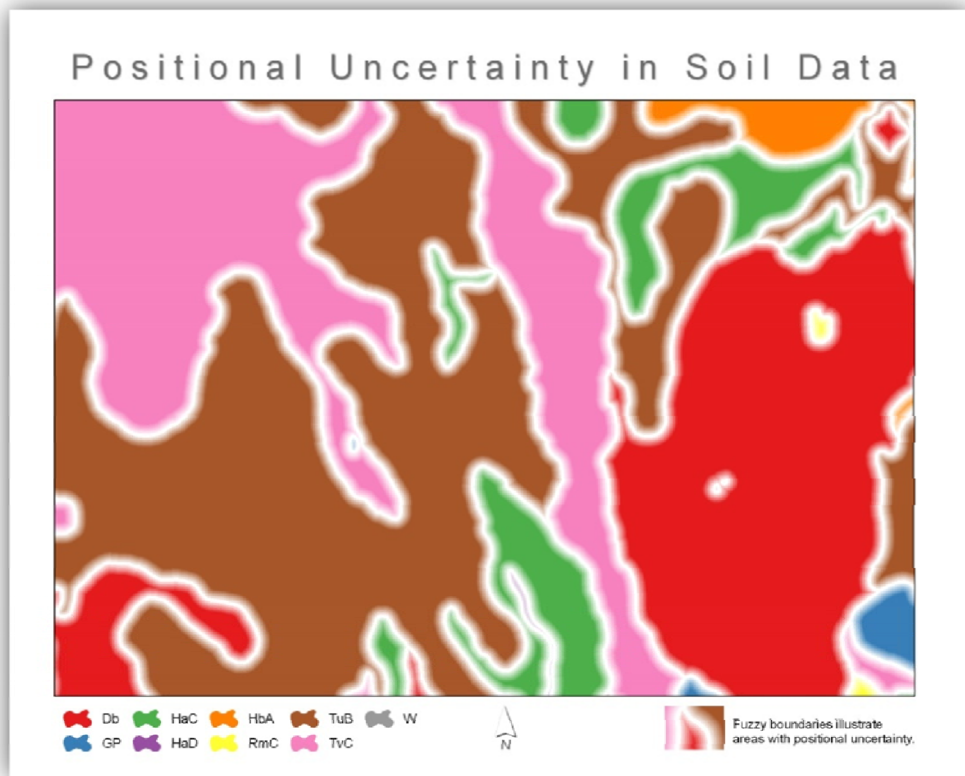


Figure 5-3: Depicting Polygon Data with Positional Uncertainty

The neutral color used in this case was white but other neutral colors could be used, such as grey, brown, or black if it made more sense for the color scheme used to illustrate the data categories. For example, a black and white rendering of the same data might use black as the neutral color. This would have the result of highlighting the areas of uncertainty because the black will stand out more than white as a neutral color; it is important to consider the potential effects of the neutral color on audience's visual analysis of the final map product. It is important to ensure that the neutral color that is chosen for the project does not give the appearance of another category of data. It should be clear to the audience that the neutral color depicts the spatial data uncertainty characteristics.

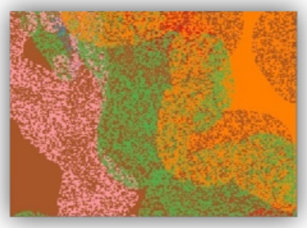
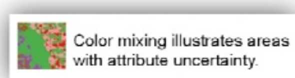
5.1.3 **Attribute Accuracy in Categorical Polygons (Method 1)**

Categorically aggregated data can have attribute errors associated with it through scale, data collection method, or incorrect attribute assignment. This project used the NRCS SSURGO soil data to demonstrate the recommended representation method for categorical areas that have error associated with their attributes. Attribute errors can be more prevalent near the boundaries of soil types; this is because soil data values are interpolated from sample collection sites and areas towards the center of the polygon are more certain than areas near the edges.

Color mixing is the recommended method for cartographically representing associated uncertainty in categorically aggregated polygons. Logically, this representation is recommended because one area's properties can actually blend into the others. For example, with land cover, a naturally forested area rarely stops abruptly where grassland begins. There is usually some blending between the two land covers even though the digital representation of their polygons may have clearly defined and abrupt boundaries. The same principle holds true for soils.

Color mixing can be completed in a number of ways but project development for this project focused on two separate methods that each made use of the same derived uncertainty value. The first method accomplished mixing colors along the boundaries by using a dot marker symbol for the buffers that were overlaid on the original polygons (Table 9). The buffer widths corresponded to varying degrees of uncertainty and were assigned random values ranging from 6 to 444 feet around the polygon. Polygons with wider buffers have more uncertainty than polygons with thinner buffers.

Table 9. Representation Method for Categorically Aggregated Data with Attribute Uncertainty

Uncertainty Type: Categorical Polygons with Attribute Accuracy Quality Concerns	
Symbology 	**Create Buffers** • Create buffers at varying widths to represent the different amount of uncertainty. This may require Add Field and Field Calculator in the attribute table. **Change Fill Symbol** • Layer Properties > Show: Categories > Unique values > Value Field > Field being represented >click Add all values > right click one of the symbols > change properties for all symbols • Symbol Selector > Properties > Symbol Property Editor > Type: Marker Fill Symbol > in the Marker Fill Tab select Marker > choose Circle 1 > Size: 7> Okay > select Outline > Width: 0 > Okay > select Random pattern > click the Fill Properties tab > Separation: X: 5 > Separation: Y: 5 • When back at Layer Properties dialog modify colors through Color Ramp **Overlay Symbolized Buffers on Polygons** • Ensure that the buffer layer is on top of the polygon layer in the Data Frame • Symbolize polygons using same color scheme as buffers
Layout   	**Create Inset Map** • Insert > Data Frame • Right click on the polygon layer > Copy > Right click on New Data Frame **Create Symbol Thumbnail** • In the Data Frame View zoom into an area of the data that illustrates the symbology well > File > Export > Save • Insert > Picture **Add Text Explanation of Symbol Meaning** • Insert > Text > add text explaining the meaning of the symbology > align thumbnail and text to be adjacent in the layout
License level required: ArcView	

One of the most difficult components of this symbolization method was trying to find a large number of colors that would be equally recognizable when the colors were mixed. It is important to ensure that the data are not mistakenly analyzed as being less important because of the color selected to represent the attribute; the use of many light and dark colors has the potential to confuse audiences. Ideally, the colors used to represent the data attributes would have similar visual prominence.

This cartographic representation of categorical data with attribute uncertainty used nine colors that were similar in saturation and hue. Accounting for the importance of color in portraying the different data categories allows each attribute to take nearly equal prominence visually.

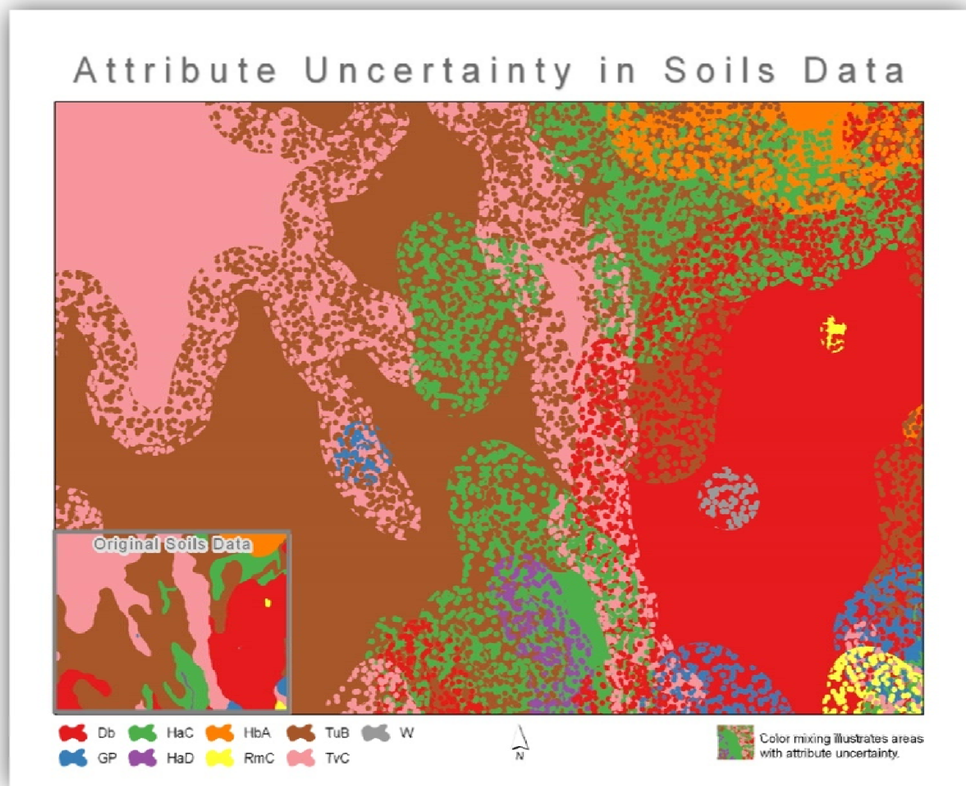


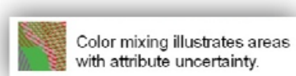
Figure 5-4: Depicting Polygon Data with Attribute Uncertainty

5.1.4 **Attribute Accuracy in Categorical Polygons (Method 2)**

Another method for illustrating mixed colors to denote attribute uncertainty along the boundaries of soil data is by using a line marker symbol for the buffers that were overlaid on the original polygons (Table 10).

Table 10. Representation Method for Categorically Aggregated Data with Attribute Uncertainty

<table>
<tr><th colspan="2">Uncertainty Type: Categorical Polygons with Attribute Accuracy Quality Concerns</th></tr>
<tr><td>Symbology
</td><td>Create Buffers
Create buffers at varying widths to represent the different amount of uncertainty. This may require Add Field and Field Calculator in the attribute table.
Change Fill Symbol
Layer Properties > Show: Categories > Unique values > Value Field > Field being represented >click Add all values > right click one of the symbols > change properties for all symbols
Symbol Selector > Properties > Symbol Property Editor > Type: Line Fill Symbol > Units: Millimeters > in the Line Fill Tab select Line > Width: 2 > Okay > select Outline > Width: 0 > Okay > Separation: 1.5
When back at Layer Properties dialog modify colors through Color Ramp > right click one of the symbols > change properties for selected symbol(s) > Properties > Line Fill tab > Angle: 10, change angle for each feature based on the number of features being represented (e.g. nine features being represented, 180/9 = 20, 20 degree difference between each symbol angle, 10, 30, 50, 70, 90, 110, 130, 150 & 170)
Overlay Symbolized Buffers on Polygons
Ensure that the buffer layer is on top of the polygon layer in the Data Frame
Symbolize polygons using same color scheme as buffers</td></tr>
<tr><td>Layout



</td><td>Create Inset Map
Insert > Data Frame
Right click on the polygon layer > Copy > Right click on New Data Frame
Create Symbol Thumbnail
In the Data Frame View zoom into an area of the data that illustrates the symbology well > File > Export > Save
Insert > Picture
Add Text Explanation of Symbol Meaning
Insert > Text > add text explaining the meaning of the symbology > align thumbnail and text to be adjacent in the layout</td></tr>
<tr><td colspan="2">License level required: ArcView</td></tr>
</table>

This alternative method for representing attribute error in polygon data used the same varied buffers to show the areas of uncertainty. The line marker symbols were created at varied angles with an equal width and spacing between the lines (Figure 5-5). With either of the color mixing techniques, the user is led to understand that the areas that have a single color are areas where the attribute—soil type in this case—is fairly certain but the areas where the attribute assignment is questionable are symbolized with one of the color mixing techniques. In the example of the soil data, there are areas that may have up to four potential soil types that occupy the areas of uncertainty. The symbology is used to convey the fact that in certain areas the actual soil type is unknown but it is likely to be one of the two (or many) soil types that overlap that space on the map.

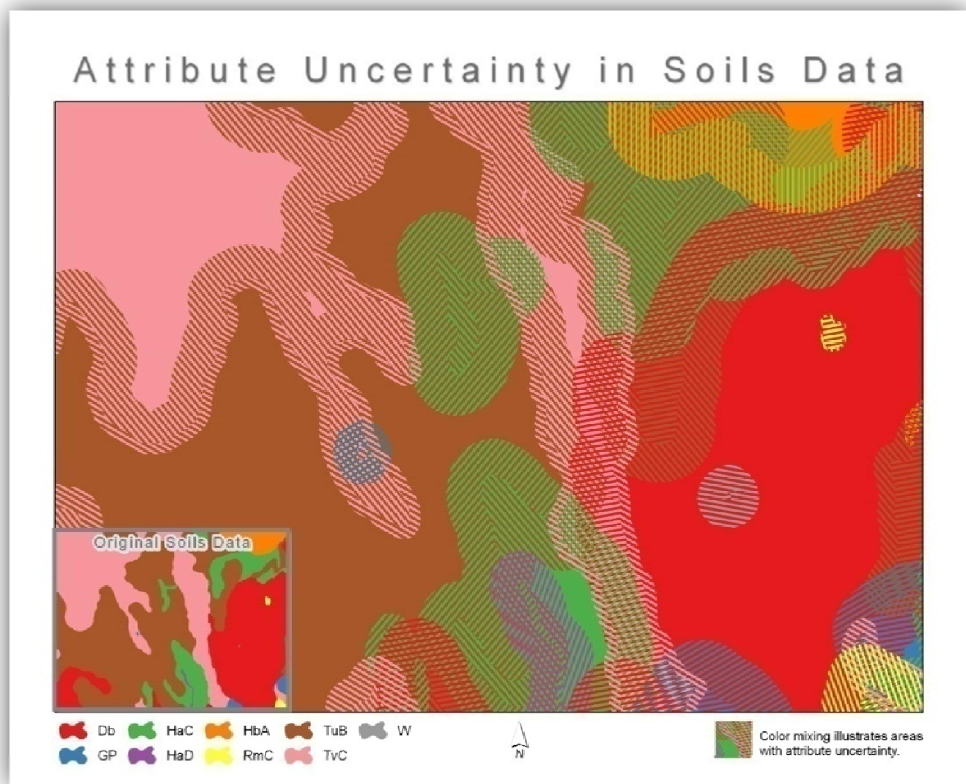


Figure 5-5: Depicting Polygon Data with Attribute Uncertainty

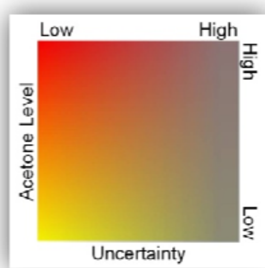
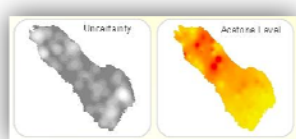
5.1.5 **Attribute Accuracy in Continuous Interpolation**

Surface data that is interpolated has error associated with its attributes because it is estimated from sample data. The only points on the surface that have a known attribute are the points where the sample was taken; all other points are mathematically generated. Interpolation methods use the sample locations to estimate the values of other locations on the surface; these estimated values are, by definition, uncertain. In general, the further that an estimated point is from a sample point, the less likely it is that the estimated value will be accurate. The dataset that was used to represent the attribute accuracy in a continuous surface consisted of a dataset of acetone levels in the sediment in the Salton Sea. Areas near many sample locations were considered to be more accurate than locations that were distant from sample locations.

The use of value or saturation in continuous tone vignettes or isopleths is recommended for representing attribute uncertainty in an interpolated surface. Isopleths or vignettes can be created using a color ramp that varies value or saturation in order to highlight user confidence in the attribute value in that area. This project used vignettes in order to illustrate the level of uncertainty in the attribute values of the interpolated surface. Creating the uncertainty surface required multiple steps, including: interpolating a density surface from the known sample locations layer; reclassifying the surface from floating to integer values; creating a polygon layer from the raster density layer;

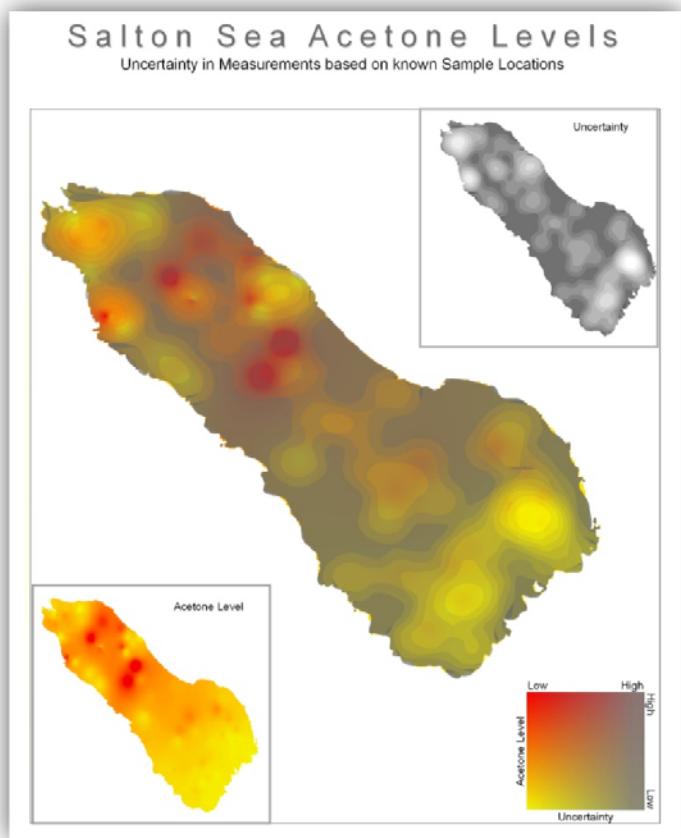
smoothing the polygon features, and; symbolizing the data with varying level of transparency (Table 11).

Table 11. Representation Method for Continuous Interpolation with Attribute Uncertainty

Uncertainty Type: Continuous Interpolation with Attribute Accuracy Quality Concerns	
Symbology 	**Create Sample Density** • Using sample point data, Spatial Analyst Toolbar > Spatial Analyst > Density > modify values, as needed > Okay **Reclassify Density Surface** • Spatial Analyst Toolbar > Reclassify > reclassify to Integer values > Okay **Convert Raster to Polygons** • Spatial Analyst Toolbar > Raster to Features > modify values, as needed > Okay **Smooth Polygon** • *(Optional): ArcToolbox > Data Management Tools > Generalization > Smooth Polygon **Set Layer Transparency** • Right-click the smoothed polygon layer > Open Attribute Table > Options > Add Field > Name "Xpar" for Transparency > Type: Long Integer> Okay > Right-click on the Xpar field > Field Calculator > Use expression (e.g. 100 - ((100 * [GRIDCODE])/9)) to set transparency from 0 to 100, 0 for the highest certainty polygons > Close Attribute Table • Right-click the smoothed polygon layer > Properties > Symbology Tab > Show: Features > Single Symbol > click the Symbol > Symbol Selector > Fill Color > Grey 60% > Okay > in the Symbology Tab select Advanced > Transparency > Choose a field: Xpar > Okay > click Okay to close Layer Properties **Overlay Symbolized Buffers on Polygons** • Ensure that the density polygon layer is on top of the raster layer in the Data Frame
Layout  	**Create a Bivariate Legend** • In Layout View draw a square using the New Rectangle tool on the Drawing Toolbar > Right click the square > Properties > Change Symbol > Properties > Type: Gradient Fill > Style: Linear > Style: Choose style that was used for the interpolated surface > Intervals: 50 > Percentage: 100 > Angle > Adjust if needed > Okay (three times to return to the Layout View • Copy and Paste the rectangle into a Microsoft Office document • Insert > Shapes > Rectangle > Draw a rectangle the same size as the gradient fill rectangle • Select the Shape to get the Drawing Tools, Format ribbon > Shape Fill > Gradient > Gradient Left • Shape Fill > Pattern > Gradient Tab > Colors > Color 1: Choose color that was used for density surface > Transparency > From: 0% > to: 100% > Okay • Arrange Transparency layer over gradient fill layer >Select both layers > Right click > Grouping > Group > Right click > Save as Picture • In ArcGIS Desktop, Insert > Picture • Add appropriate text around the new legend **Create Inset Map** • Insert > Data Frame • Right click on the polygon layer > Copy > Right click on New Data Frame • Repeat for each layer
License level required: ArcView & Spatial Analyst, *ArcInfo for optional step	

The final output for this representation example depicts the acetone levels using a color ramp that goes from red for high acetone levels to yellow for lower acetone levels. The uncertainty vignette is layered over the acetone surface as a series of grey polygons with varying levels of transparency to highlight or obscure the acetone levels as appropriate (Figure 5-6). This symbolization technique has the effect of clouding the areas of high uncertainty by using less transparency in the overlay and highlighting the areas of the greatest certainty with little or no transparency overlay.

Figure 5-6: Depicting Continuous Interpolation Data with Attribute Uncertainty



When creating continuous tone vignettes or isopleths it is important to be careful of the use of any color other than black or white. Any use of color in the uncertainty surface is likely to confuse audiences about the actual values of the interpolated surface. For example, overlaying a red surface on a blue surface is going to create areas that appear purple, which could appear to audiences as a separate attribute value. If it is crucial for users to include color as a variable in the vignettes or isopleths, it is recommended that the interpolated surface be rendered using a neutral color scheme such as black to white.

5.2 Spatial Data Uncertainty Representation Methods Deployment

ArcGIS Online is an online resource for ESRI software users who want to discover or share data or other resources, such as symbol layers. All of the representations that were created for this project were provided on this website, along with additional resources to assist users in representing the characteristics of the uncertainty in their spatial data. Each representation was provided as a Layer Package (.lkp) and a Map Document (.mxd). Provided along with access to the data were step-by-step instructions for creating the representation with ArcGIS and links to useful documents and tools to assist users in representing the uncertainty of their spatial data (Appendix B).

5.3 Summary

Five methods for representing different types of spatial data uncertainty were developed for this project. Those methods included the use of:

1. Saturation for discrete points with attribute errors
2. Texture for categorical aggregation polygons with positional errors
3. Color mixing using random dots for categorical aggregation polygons with attribute errors
4. Color mixing using line symbology for categorical aggregation polygons with attribute errors
5. Continuous tone vignettes for interpolated surface data with attribute errors

Each of the sample datasets was deployed through ESRI’s ArcGIS Online website, along with methodology documentation and resources for enhancing representation efforts. Providing the information through a website has given many users access to the tools provided, resulting in further awareness of the topic of spatial data uncertainty by a larger audience. This was one of the primary goals of the project.

Chapter 6 – Results and Analysis

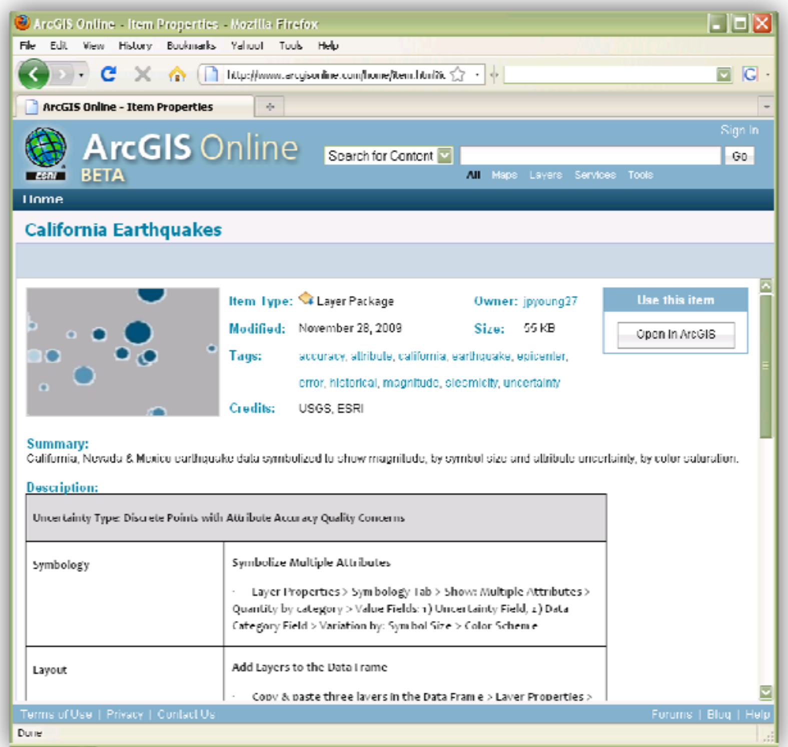
6.1 Integrating Representations into User Data

The format in which the representations and tools were going to be deployed changed through the term of this project. The original project plan included a custom style, toolbar, documentation, and static samples of the representation methods. The final project included instructions for creating the representations in ArcGIS Desktop and dynamic samples that display the representation methods. There were advantages and disadvantages for users interested in integrating the representation methods in their own data.

6.1.1 **Benefits of Representations**

The format that the data was provided in allowed users to explore the data and symbolization interactively. As discussed in previous chapters, the representation samples for this project were deployed to users in Layer Packages and Map Documents (Figure 6-1). The file formats that the data were provided in allow users to explore the data dynamically through either ArcGIS Desktop or ArcGIS Explorer, which is a free download. Users are able to explore the attributes and symbology settings of the individual data layers and can modify the layers if they so choose. The ability to modify features can be valuable for users who would like to evaluate how the different representation methods look with a different color scheme or transparency, for example. The original project plan would have provided the final outputs in a static format such as an Adobe Reader document, and the users would not have been able to explore the data at all in ArcGIS; the user’s ability to look at data in ArcGIS is an improvement over the original project plan.

Figure 6-1: User access to Layer Package through ArcGIS Online



6.1.2 Limitations of Representations

Even though the format in which the representations are provided allows users to explore the data and symbolization methods through ArcGIS software, the deployed formats do not allow the users to automatically apply the symbology to their own datasets. Users need to follow the instructions that are provided on the website if they are interested in applying a similar symbology to their own data. The representation Layer Packages and Map Documents are provided to users primarily for illustrative purposes. The original project plan included deployment of a style and toolbar that would have partially automated the representation process. However, the final version of the project educated users how to create the representations in ArcGIS, rather than automating the symbolization process for them. Detailed instructions on how to recreate the illustration of uncertainty in other datasets has been provided on the ArcGIS Online website. Comprehensive documentation on the website gives users instructions for how to create the individual symbolizations in their own data.

6.2 Deploying Representations to ArcGIS Online (Beta)

The representations were deployed to the ArcGIS Online website (Figure 6-2) for eventual consumption by users interested in representing the relative uncertainty of their spatial data quality. This website is new to ESRI and is currently in beta version, which has proven to have its advantages and disadvantages in the deployment of the final products for this project.

6.2.1 Benefits of Deployment Method

One of the primary purposes of the ArcGIS Online website is to allow users to “find maps, layer and tools” (ESRI, 2009a). Users can search for different types of data that are provided directly by ESRI or by other users and download and use the data for free. In addition to data, users are able to find and access layers that that include custom symbology. Providing the data layers from this project through this website had the benefit of reaching a larger audience than it would have if they had been deployed to the Mapping Center website alone, which was the original plan when the project commenced. The number of members in a group is a factor in the ultimate success of this project and tracking the number of users who have associated themselves with the group for this project is easily done through the administration tools provided on the website. Informing users about the topic of spatial data uncertainty is directly in line with one of the goals of this project and deploying the project deliverables to ArcGIS Online supports this goal and tracking the number of members is a factor in whether this goal was met or not.

Figure 6-2: Group for Depicting Spatial Data Uncertainty on ArcGIS Online

ArcGIS Online also provides benefits to the ongoing administration of all of the information that has been deployed for this project because the group owner can easily update the information that has been provided as the software changes or better methods for representing the type of spatial data uncertainty are developed. This ensures that users can have access to the latest information available, as it is deployed.

6.2.2 **Limitations of Deployment Method**

One of the primary limitations of deploying this project to the ArcGIS Online website is that it is new and not yet widely-known, and has the potential to limit the number of users accessing the site. This has been mitigated by posting information regarding the project tools on the more widely-accessed Mapping Center website. Administrative utilities for tracking the number of downloads, inviting users, and receiving feedback are also limited. It would be beneficial to be able to track the number of times that the symbology is downloaded in order to assess the success of this project. The client agrees that the ability to have a count of the number of times that a particular dataset is accessed would be useful and he has provided feedback to the ArcGIS Online development team; this functionality may be added in later revisions of the website. Furthermore, it would be ideal if users had the ability to provide feedback directly through the website, but user feedback tools are currently non-existent on the website. As an alternative, an email address was provided so that users could provide feedback on the representation methods and tools that were provided on the site.

6.3 Summary

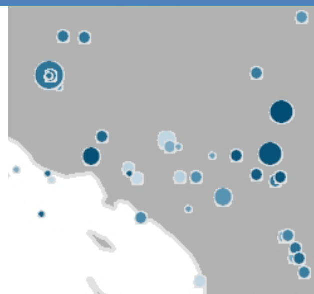
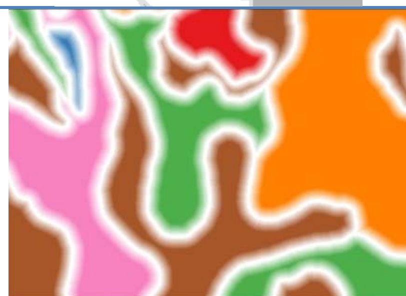
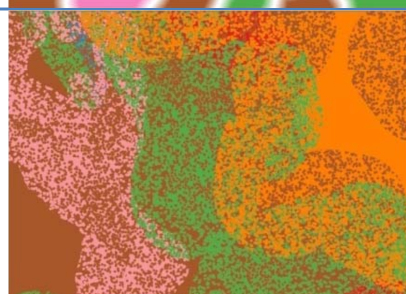
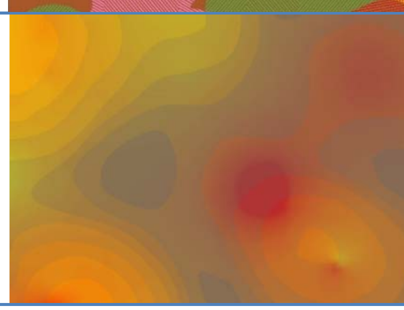
The representation methods and examples were deployed using ESRI Layer Packages and Map Documents on the ArcGIS Online website. Deploying the project deliverables using this method provided advantages and disadvantages of the project’s success. The main advantages were ease of user access, dynamic sample representations, and the ability to track the number of members of the group created for this project. The lack of appropriate administration tools for tracking the usage of data and a lack of online feedback tools are the main disadvantages to this project’s deployment. Although not included in the scope of the project, it would be ideal to provide users even more representation methods and data types, automate the representation process where possible, and test representation methods for comprehension.

Chapter 7 – Conclusions and Future Work

The primary goals of this project were to create tools and methods to allow map designers to be more efficient in representing the uncertainty that is inherent in their data, while presenting the topic of spatial data uncertainty to a larger audience. Development for this project relied on representation recommendations for spatial data uncertainty presented through research to produce methods for ArcGIS users to symbolize the quality of their data position and attributes. A majority of the research agrees that uncertainty can be present in different data types where the quality of attributes, position, completeness, logical consistency, or lineage is degraded. In addition to reviewing pertinent literature, a requirements analysis was completed prior to development to determine the client's needs, users' needs, and current symbolization processes where they exist. Although there are numerous categories of spatial data uncertainty, this project focused on the data types and uncertainty that would be the most useful to users; specifically, attribute and positional accuracy uncertainty was assessed in point, polygon, and surface data. Data collection efforts focused only on those categories. Final implementation provided five different methods for representing spatial data uncertainty and methods for illustrating areas where there are data quality concerns included color saturation, texture, color mixing, color value, and vignettes. Representation samples and a methodology were deployed to users through the ArcGIS Online website.

All of the mandatory requirements of this project were met. The representation samples can be downloaded for use in ArcGIS Desktop 9.3.1. The Layer Packages can only be used in version 9.3.1 or later, but the Map Documents can be used with earlier versions of the software. Providing a method for users to discover different types of representation methods was important to the client, particularly because the scope of this project did not allow for providing representation samples for every conceivable type of spatial data uncertainty. The five specific sample representation types that were the focus of this project development were provided for multiple data types, with different data quality concerns (Table 12). The example representation methods were created using data that is freely available, through ArcGIS Online, federal government websites, or other private websites. The data types were chosen because they had uncertainty characteristics that would be easily understood by audiences, even if users were not familiar with the particular dataset. Representation methods used preexisting tools and this project did not focus on automating the representation process, which was a desirable to the client. Automating the representation process would have been a much more crucial project requirement if modifying the symbology was a more complicated process, but the detailed methodology documentation suited the user and client requirements for this project.

Table 12. Summary of Project Representations

Final Representation Methods	
Discrete Points with Attribute Uncertainty	
Categorical Aggregation with Positional Uncertainty	
Categorical Aggregation with Attribute Uncertainty	
Categorical Aggregation with Attribute Uncertainty	
Continuous Interpolation with Attribute Uncertainty	

7.1 Future Work

Additional efforts on this topic should center on developing representation methods for the other categories of spatial data uncertainty, developing sample layers using other types of data, automation the symbology processes, and testing user comprehension of symbology methods

7.1.1 **Representation Methods**

This project focused on creating representation methods based on four of the recommendations made in the research but there are numerous other data type and quality combinations that could be developed into representation samples to assist users in illustrating their own data uncertainty. Other representation recommendations that could be developed into sample layers include:

- Positional accuracy uncertainty
    - Discrete points and lines
    - Continuous interpolation
- Attribute accuracy uncertainty
    - Partitioned and enumerated polygons
- Logical consistency uncertainty
    - Discrete points and lines
    - Partitioned and enumerated polygons
    - Continuous interpolation
- Completeness uncertainty
    - Discrete points and lines
    - Categorical aggregation and overlay polygons
    - Partitioned and enumerated polygons
    - Continuous interpolation
- Lineage uncertainty
    - Discrete points and lines
    - Categorical aggregation and overlay polygons
    - Partitioned and enumerated polygons
    - Continuous interpolation

Some of the categories, such as lineage, have the same representation recommendation for all data types and recommendations for other data types are excluded because the representation would not be useful to users, therefore future development could focus on developing ten more symbology methods.

7.1.2 **Data**

Even if future development efforts did not focus on additional representation methods, a user could focus developing additional examples for this project's representation samples. Providing users with even more examples of the symbology methods using different data sources could achieve the benefit of garnering even more interest from users, from different industries or those who are interested in other than from the natural resources data that was provided in this project. It would be particularly interesting to develop symbology samples using a set of industry-specific data. For example, this project’s samples could be used to develop additional examples that are focused on health data.

7.1.3 **Automation**

Additional development efforts could also focus on automating representation methods. This project documented the steps required to recreate the representation methods for users, but the representation methods would be much easier to generate if users could use one tool and input the necessary parameters. Automating the symbology methods would be particularly useful for representations that have multiple steps, such as the attribute accuracy representation for the interpolated surface. In this case a user could develop a script that would step through the process of creating the density surface, using the raster calculator to change from floating numbers to integers, transform the raster to a polygon, smooth the polygon edges, and use the field calculator to set the transparency levels. Each set is not especially difficult to complete, but automating the process would make symbolizing the data uncertainty easier.

7.1.4 **Testing**

Although this project focused on developing representations from methods that have been tested during research, none of the project representation methods were tested by users for comprehension. An appropriate testing process would include testing users’ comprehension of all of the recommended representation methods for one data type and quality. In addition, testing efforts should evaluate users’ comprehension of multiple representations for each recommendation in cases where there exist multiple symbology recommendations. For example, researchers recommend the use of value and color saturation in continuous tone isopleths or vignettes; a user should test the comprehension of representation methods that utilize: value in isopleths, saturation in isopleths, value in vignettes, and saturation in vignettes. Additionally, plans for testing should take in to consideration the amount of knowledge that audience has on the topic of uncertainty. It would be anticipated that audiences already familiar with the topic are going to better understand the representation methods than audiences who are not. Testing efforts should attempt to incorporate feedback from informed and uninformed users alike.

7.2 Conclusion

The client is now able to offer methods and resources to users who are interested in depicting spatial data uncertainty, something that was never before available to users. Enhancing users’ abilities to illustrate the quantity and quality of spatial data uncertainty through a GIS allows them to create map products in which the data accuracy can be visually inspected and accounted for. If not through cartographic representation, error would normally be described in the metadata, map marginalia, or textually on the map, but these types of explanations provide limited aids for somebody who needs to quickly assess the fidelity of the mapped data. Providing methods to ArcGIS users for representing the unique characteristics of data uncertainty serves to assist them in providing additional information regarding the data quality that was not easily done in the past. The provided representations of spatial data uncertainty also enhance the map users’ capability of making more informed decisions through visual inspection. By illustrating the data error on maps, users are able to effectively increase the quality and value of the map. The tools required to create the symbolization necessary to represent the different types of spatial data uncertainty were already available within ArcGIS Desktop; what was lacking were the methods for creating the representations recommended by scholarly research. This project focused primarily on providing users with access to the suggestions that have been made in the research, examples of some of the most common types of uncertainty in spatial data, and methods for implementing the representations in other kinds of data. By providing the methods and documentation online, the client is able to provide users with the information necessary to illustrate various types of spatial data uncertainty.

Works Cited

Beard, M. K., Buttenfield, B. P., & Mackaness, W. A. (1994). *Research initiative 7: visualization of the quality of spatial information.* Santa Barbara, CA: National Center for Geographic Information Analysis.

Berry, J. K. (1996). *Spatial reasoning for effective GIS.* Indianapolis: John Wiley & Sons.

Blais, J. (2002). Reliability considerations in geospatial information systems. *Geomatica*, 56(4), 341-350.

De Cola, L. (2002). Spatial forecasting of disease risk and uncertainty. *Cartography and Geographic Information Science*, 29(4), 363-380.

ESRI. (2009a). *ArcGIS Online*. Retrieved November 09, 2009, from http://www.arcgisonline.com/

ESRI. (2009b). *ESRI Mapping Center*. Retrieved January 28, 2009, from http://mappingcenter.esri.com/

Foody, G. M., & Atkinson, P. M. (2002). *Uncertainty in remote sensing and GIS.* West Sussex: John Wiley & Sons Ltd.

Hengl, T. (2009, June 22). *Uncertainty visualization*. Retrieved November 11, 2009, from spatial-analyst.net: http://spatial-analyst.net/wiki/index.php?title=Uncertainty_visualization

Hengl, T. (2003). Visualisation of uncertainty using the HSI colour model: computations with the colours. *Proceedings of the 7th International Conference on GeoComputation*, (pp. 8-10). Southampton.

Kimerling, A. J., Buckley, A. R., Muehrcke, P. C., & Muehrcke, J. O. (2009). *Map Use: Reading and Analysis.* Redlands, CA: ESRI Press.

Longley, P. A., Goodchild, M. F., Maguire, D. J., & Rhind, D. W. (2005). *Geographic information systems and science.* West Sussex: John Wiley & Sons Ltd.

Lowell, K., & Jaton, A. (1999). *Spatial accuracy assesment: land information uncertainty in natural resources.* Ann Arbor: Sleeping Bear Press.

MacEachren, A. M. (1995). *How maps work: representation, visualization, and design.* New York: The Guilford Press.

MacEachren, A. M. (1992). Vizualizing uncertain information. *Cartographic Perspective*, 13, 10-19.

MacEachren, A. M., Robinson, A., Hopper, S., Gardner, S., Murray, R., Gahegan, et al. (2005). Visualizing geospatial information uncertainty: what we know and what we need to know. *Cartography and Geographic Information Science*, 32, 139-160.

Morris, A., & Jankowski, P. (2005). Spatial decision making using fuzzy GIS. In F. E. Petry, V. B. Robinson, & M. A. Cobb, *Fuzzy modeling with spatial information for geographic problems* (pp. 275-298). Berlin: Springer.

Mowrer, H., & Congalton, R. G. (2000). *Quantifying spatial uncertainty in natural resources: theory and applications for GIS and remote sensing.* Ann Arbor: Sleeping Bear Press.

North Carolina State University. (2000, May 08). *Color principles - hue, saturation, and value*. Retrieved November 09, 2009, from Scientific visualization: http://www.ncsu.edu/scivis/lessons/colormodels/color_models2.html#secondary

Plewe, B. S. (2003). Representing datum-level uncertainty in historical GIS. *Cartography and Geographic Information Science*, 30, 319-334.
Six Sigma. (2009). Retrieved November 17, 2009, from Wikipedia: http://en.wikipedia.org/wiki/Six_sigma
Thomson, J., Hetzler, B., MacEachren, A., Gahegan, M., & Pavel, M. (2005). A typology for visualizing uncertainty. *Conference on Visualization and Data Analysis*, 16-20.
United States Geological Survey. (1997). *Spatial data transfer standard: logical specifications.* Reston, VA: United States Geological Survey.
Yao, X., & Jiang, B. (2005). Visualization of qualitative locations in geographic information systems. *Cartography and Geographic Information Science*, 32(4), 219 - 229.
Zhang, J., & Goodchild, M. (2002). *Uncertainty in geographical information.* London: Taylor & Francis.

Appendix A. ArcGIS Online – Depicting Spatial Data Uncertainty Group

Depicting Spatial Data Uncertainty

This is a group for ArcGIS users that would like information and resources for representing the relative uncertainty in spatial data.

Contact: janelle_p_payne@yahoo.com

Tags: uncertainty, error, bias, precision, accuracy, scale, quality, data, positional, attribute, symbology

Maps created in geographic information systems (GIS) are typically rendered with precisely defined features, but experienced GIS practitioners recognize that spatial data has varying measures of relative error that are not always apparent to map readers. Limited awareness among many geographic information users regarding the associated error of the data leads users to view and analyze data without regard for relative uncertainty. Tools and methods supporting map designer abilities to graphically convey quantifiable uncertainty associated with spatial data have not been readily available.

This is a group for ArcGIS users who would like to represent the relative uncertainty in their spatial data. This group was created to give users tools, resources and examples of symbolization methods that can be used to represent different types of spatial data uncertainty.

Data Quality & Data Type	Positional Accuracy	Attribute Accuracy	Logical Consistency	Completeness	Lineage
Discrete **Points & Lines**	Size Shape (error ellipse, epsilon band)	Value Color Saturation (feature code checks) Example: California Earthquakes	Color Mixing Redundancy by overprinting. Slivers by solid fills (Topological cleaning)	Mapping Technique Density Traces Marginalia Generalization algorithm Mapping Tolerance Buffer Size	Mapping Technique Minimum Bounding Rectangles Marginalia Source of data Scale/Resolution Date Geometry
Categorical **Aggregation & Overlay**	Texture Value (Certainty of boundary location) Example: Soils Vignette	Color Mixing (Attribute code checks, topographic classifier) Example: Soils Color Mixing with Dots or Lines	Lack error models	Mapping Technique Missing Values Logical adjacency surface Marginalia Discrete model weights	
Partitioning & Enumeration	Not meaningful	Size = height (Blanket of error)	Size = height (Maximum likelihood prism maps)	Mapping Technique Missing Values Logical adjacency surface Classing scheme OAL/TAI	
Continuous Interpolation	Value Color Saturation (continuous tone vignettes, continuous tone isopleths) Example: Salton Sea Sediment		Size = line wt Color Shape = Compactness (TIN links)	Not possible by definition	
	Graphical Syntax		Graphical/Lexical Syntax		

Adapted by B.P. Buttenfield; original version of the table first presented by B.P. Buttenfield and R. Weibel "Visualizing the Quality of Cartographic Data", Third International Geographical Information Systems Symposium (GIS/LIS 88), San Antonio, Texas, November, 1988.

Members are encouraged to provide feedback to the resources provided here and you can contact me via email with any questions or comments that you may have.

You must have ArcGIS 9.3.1 with the latest service pack to use the layer packages that are provided in this group. Additionally users can view the layer packages in latest release of ArcGIS Explorer.

10 items

Soils Vignette by jpyoung27 (November 22, 2009)
Use of vignette to illustrate positional accuracy characteristics in soils data.
Tags: accuracy, california, error, positional, precision, quality, uncertainty, soil
Type: Map Document
Open in ArcGIS

Soils Color Mixing with Lines by jpyoung27 (November 22, 2009)
Use of symbolized buffer to illustrate attribute accuracy characteristics in soils data.
Tags: accuracy, attribute, california, error, precision, quality, soil, symbology, uncertainty, buffer
Type: Map Document
Open in ArcGIS

Soils Color Mixing with Dots by jpyoung27 (November 22, 2009)
Use of symbolized buffer to illustrate attribute accuracy characteristics in soils data.
Tags: accuracy, attribute, california, error, precision, quality, soil, symbology, uncertainty, buffer
Type: Map Document
Open in ArcGIS

Continuous Tone Vignette for the Salton Sea by jpyoung27 (November 22, 2009)
Use of continuous tone vignette to illustrate attribute accuracy characteristics in sediment data.
Tags: accuracy, attribute, error, quality, symbology, uncertainty, vignette
Type: Map Document
Open in ArcGIS

California Earthquakes by jpyoung27 (November 27, 2009)
California, Nevada & Mexico earthquake data symbolized to show magnitude, by symbol size and attribute uncertainty, by color.
Tags: accuracy, attribute, california, earthquake, epicenter, error, historical, magnitude, seismicity, uncertainty

Type: Map Document

Open in ArcGIS

California Earthquakes by jpyoung27 (November 28, 2009)

California, Nevada & Mexico earthquake data symbolized to show magnitude, by symbol size and attribute uncertainty, by color.

Tags: accuracy, attribute, california, earthquake, epicenter, error, historical, magnitude, seismicity, uncertainty

Type: Layer Package

Open in ArcGIS

Continuous Tone Vignette for the Salton Sea-Acetone Levels by jpyoung27 (November 28, 2009)

Use of continuous tone vignette to illustrate attribute uncertainty in continuous data

Tags: accuracy, attribute, error, quality, symbology, uncertainty, vignette

Type: Layer Package

Open in ArcGIS

Soils Color Mixing with Dots by jpyoung27 (November 28, 2009)

Use of symbolized buffer to illustrate attribute accuracy characteristics in soils data.

Tags: accuracy, attribute, california, error, precision, quality, soil, symbology, uncertainty, buffer

Type: Layer Package

Open in ArcGIS

Soils Color Mixing with Lines by jpyoung27 (November 28, 2009)

Use of symbolized buffer to illustrate attribute accuracy characteristics in soils data.

Tags: accuracy, attribute, california, error, precision, quality, soil, symbology, uncertainty, buffer

Type: Layer Package

Open in ArcGIS

Soils Vignette by jpyoung27 (November 28, 2009)

Use of vignette to illustrate positional accuracy characteristics in soils data.

Tags: accuracy, california, error, positional, precision, quality, uncertainty, soil, vignette

Type: Layer Package

Open in ArcGIS

Appendix B. ESRI Mapping Center Buffer Vignette Technique

"Fade to white background" effect

Buffer vignettes symbolize the interface between two areas. They are often used to show the land-water interface by gradually fading blue at the coast into white or vice versa. You can also use them to fade the map out into a white background. These types of vignettes are created using buffers that are symbolized in a special way.

Note -- as described this only works on white backgrounds. See the two examples at the end to see how the data frame's background can be coordinated with the buffer vignette for different effects on along coastlines.

Create buffers for the vignette

1. In ArcToolbox, click Analysis Tools.
2. Click Proximity toolset.
3. Click Multiple Ring Buffer.
4. Double click the tool.
5. Set the input features that constitute the perimeter of your study area.
6. Call the output feature class bufferVignette.
7. To make 16 rings that decrease in size from 5 miles to one mile, enter these distances in ascending order: 5, 9, 13, 16, 19, 22, 24, 26, 28, 30, 31, 32, 33, 34, 35, 36. The last buffer is used to fill in the rest of the data frame as you will see below. Use the Plus sign to add each new number. (The numbers you use are scale dependent—it's really more a matter of measuring so use the Measure tool to determine the width of the buffer and then divide that distance using the proportions shown above as an example.)
8. Make sure you set the Buffer units to Miles.
9. Click OK.
10. Rename the new layer Buffer Vignette in the Table of Contents and move it below the State Outline and above the County Outline.

Adding fields to a table

1. Right-click the Buffer Vignette layer in the Table of Contents and click Open Attribute Table.
2. Click Options in the table to which you want to add a field.
3. Click Add Field.
4. Type the name of the field. Use Xpar. This is the abbreviation for "transparency".
5. Click the Type dropdown arrow and click the field type. Use long integer.
6. Click OK.

Calculating the transparency level

1. Right-click the layer or table you want to edit and click Open Attribute Table.
2. Right-click the Xpar field heading that you want to make a calculation for and click Calculate Values.
3. Click Yes when you see the dialog box. You can make calculations without being in an editing session; however, in that case, there is no way to undo the results.

4. Use the Fields list and Functions to build a calculation expression. Enter this expression: 100-((100 * [distance])/36). This will calculate transparency values that are a function of the distance so that the buffers farther away are less transparent.
5. Click Save.
6. Name the calculate statement Xpar_values.cal. Click Save.
7. Click OK.

Creating the universe polygon

Notice that you can see outside the extent of the last buffer. What you need to do to get rid of the rest of the image is to create one additional polygon in the buffers layer that will extend just outside the data frame.   There are MANY ways to skin this cat -- here is but one. If you have others that work for you, share them with us in the Comments at the bottom.

Adding the Editor toolbar

1. In ArcMap, click the View menu.
2. Point to Toolbars.
3. Click Editor. The toolbar is added.

Editing the buffer layer

1. Click the Editor menu and click Start Editing.
2. Choose the CA_buffer layer to edit it.
3. Use the Zoom In tool on the Layout toolbar to zoom in closer to the data frame with the buffer in it.
4. Use the Zoom In tool on the Standard toolbar to zoom in closer to the edges of the buffer layer.

Creating a polygon feature by digitizing

1. Click the Current Task dropdown arrow and click Create New Feature.
2. Click the Target layer dropdown arrow and click a line or polygon layer. This is the CA_buffer layer.
3. Click the Tool Palette dropdown arrow and click the Sketch tool.
4. Click on the map to digitize the feature's vertices. You want to create a polygon in the shape of a box with four vertices around the edge of the buffered state -- get close to the edge of the data frame, and if you get a warning that you are out of bounds, just make the box a little smaller.
5. When finished, right-click anywhere on the map and click Finish Sketch. The polygon is created on your map.
6. Now we will make the vertices to exact locations.

Moving a vertex by specifying x,y coordinates

1. Click the Current Task dropdown arrow and click Modify Feature.
2. Click the Edit tool and click the polygon you just digitized.
3. Click to select the vertex you want to move. Start with the one in the upper left.
4. Position the pointer over the vertex until the pointer changes.
5. Right-click and click Move To.
6. Type the x,y coordinates where you want to move the vertex. Use coordinates that are just outside the area of display for your map.
7. Right-click any part of the sketch and click Finish Sketch. The feature is reshaped.

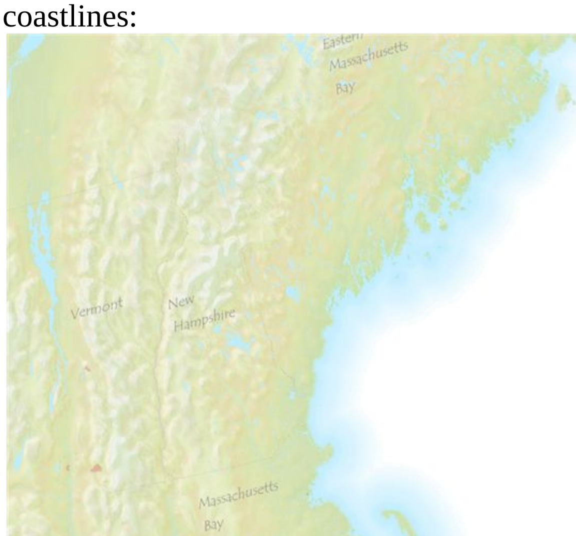
Calculating the transparency level of the universe polygon

1. In the Table of Contents, right-click the CA_Buffer layer that you want to edit and click Open Attribute Table.
2. Select the records for the polygon you just digitized.
3. Right-click the field heading for the Xpar field that you want to change the value for.
4. Type a value to set to the field. Use 0 (that's zero!)
5. Click Editor on the Editor toolbar and click Stop Editing.

Symbolizing the buffers

1. Right-click the CA_Buffers layer that you want to draw with a single symbol in the Table of Contents and click Properties.
2. Click the Symbology tab.
3. Click Features. Because Single Symbol is the only option, ArcMap automatically selects it.
4. Click the Symbol button to change the symbol. Select White.
5. Click Outline color and select No color.
6. Click OK on the Symbol Selector dialog box.
7. Click Advanced. Click Transparency.
8. Select Xpar as the field that you will use to vary the feature transparency based on field values in percent.
9. Click OK.

Here are a couple more examples of what can be done with buffer vignette layers on coastlines:



Above, the map background color is white and the buffers are blue.

Here the map background color is blue and the buffers are white.