Summer 2010

Use of Multi-Criteria Decision Analysis with Fuzzy Measures in Historical GIS

Jeanne Marie Jones
San Jose State University

Follow this and additional works at: https://scholarworks.sjsu.edu/etd_theses

Recommended Citation

Jones, Jeanne Marie, "Use of Multi-Criteria Decision Analysis with Fuzzy Measures in Historical GIS" (2010). *Master's Theses*. 3840.
DOI: <https://doi.org/10.31979/etd.m54u-w8us>
https://scholarworks.sjsu.edu/etd_theses/3840

This Thesis is brought to you for free and open access by the Master's Theses and Graduate Research at SJSU ScholarWorks. It has been accepted for inclusion in Master's Theses by an authorized administrator of SJSU ScholarWorks. For more information, please contact scholarworks@sjsu.edu.

USE OF MULTI-CRITERIA DECISION ANALYSIS WITH FUZZY MEASURES
IN HISTORICAL GIS

A Thesis

Presented to

The Faculty of the Department of Geography

San José State University

In Partial Fulfillment

of the Requirements for the Degree

Master of Arts

by

Jeanne M. Jones

August 2010

© 2010

Jeanne M. Jones

ALL RIGHTS RESERVED

The Designated Thesis Committee Approves the Thesis Titled

USE OF MULTI-CRITERIA DECISION ANALYSIS WITH FUZZY MEASURES
IN HISTORICAL GIS

by

Jeanne M. Jones

APPROVED FOR THE DEPARTMENT OF GEOGRAPHY

SAN JOSÉ STATE UNIVERSITY

August 2010

Dr. M. Kathryn Davis	Department of Geography
----------------------	-------------------------

Dr. Gary Pereira	Department of Geography
------------------	-------------------------

Dr. Kathrine Richardson	Department of Geography
-------------------------	-------------------------

ABSTRACT

USE OF MULTI-CRITERIA DECISION ANALYSIS WITH FUZZY MEASURES IN HISTORICAL GIS

by Jeanne M. Jones

Geographic information systems (GIS) can enhance historical research by providing tools to explore the spatial relationships of locations in historical sources. However, no widespread methods currently exist for translating vaguely defined historical spatial information into GIS data formats and producing a location estimate. Other GIS techniques do exist that can model the necessary process. Multi-criteria decision analysis with fuzzy measures can be applied to vague historical records to approximate location. The Wieslander Vegetation Type Map dataset is used to demonstrate the model effectiveness. Results show that this technique successfully translated written descriptions of location into raster, or grid-based, surfaces within a GIS. Given the uncertainty of the qualitative descriptions, the technique resolved the text into a collection of locations instead of a single location, with a probability assigned to each location conveying the ambiguity associated with the results and the probabilistic nature of its interpretation.

TABLE OF CONTENTS

LIST OF FIGURES	vii
LIST OF TABLES	viii
INTRODUCTION	1
METHODS	6
Data	6
Uncertainty Ranges	12
Fuzzy Measure Generation	14
Suitability Surface Generation	15
Uncertainty and Sensitivity Analysis	16
Mexia Alaska Data	17
Mexia Ranges and Weights	18
RESULTS	24
Wieslander Transect Plot Location	24
Sensitivity Analysis	25
Mexia Plant Specimen Location	26
DISCUSSION	31
Uncertainty in Plot Location	31
Uncertainty in the Data	32
Patterns in the Data	37
Displaying Multiple Results	38
Advantage of Risk Weights and Fuzzy Measures	41
Mexia Data – Visualizing on a Larger Scale	42
CONCLUSION	45
REFERENCES	47
APPENDIX A – WIESLANDER UNCERTAINTY RANGES	49
APPENDIX B – SUITABILITY MAP RESULTS	55
Covelo	56
Mt. Diablo	59

San Simeon	63
Mt. Pinos	67
Preston Peak.....	71
Yosemite	76
Mexia – Alaska	80

LIST OF FIGURES

Figure 1: Example of historic quad map with plot locations	6
Figure 2: VTM 30-minute plot data locations	10
Figure 3: VTM 15-minute plot data locations	11
Figure 4: Compass rose.....	12
Figure 5: Sigmoidal and linear membership functions	15
Figure 6: Mt. McKinley (Denali) National Park in Alaska	18
Figure 7: Suitability surface results for three example plots	27
Figure 8: Sensitivity analysis – fuzzy function shape.....	28
Figure 9: Sensitivity analysis – width of fuzzy measure	29
Figure 10: Sensitivity analysis – order (risk) weights	30
Figure 11: Comparison of transect plot sizes with stamps	32
Figure 12: Historic quad map elevation discrepancy.....	33
Figure 13: Suitability surfaces with different interpretations of description	35
Figure 14: Comparison of features on historic and current quad.....	36
Figure 15: Comparison of topography on historic and current quad.....	36
Figure 16: Display of two suitability surfaces at Long Canyon	39
Figure 17: Display of three suitability surfaces at Fox Ridge	40
Figure 18: Use of less conservative risk weights.....	41
Figure 19: Extraction of “most likely” locations	43
Figure 20: Cumulative map of Mexia location estimates	44
Figure B-1: Covelo quad, Mendocino National Forest, California	56
Figure B-2: Mt. Diablo quad, Mt. Diablo State Park, California.....	59
Figure B-3: San Simeon quad, California.....	63
Figure B-4: Mt. Pinos quad, Los Padres National Forest, California	67
Figure B-5: Preston Peak quad, Redwood National Park, California.....	71
Figure B-6: Yosemite quad, Yosemite National Park, California	76
Figure B-7: Part of Mt. McKinley (Denali) National Park, Alaska.....	80

LIST OF TABLES

Table 1: Wieslander Study Plots with Field Note Entries	8
Table 2: Historic Basemaps and Associated Scale, Age, and Plot Error	12
Table 3: Example Locations of Study Plots with Uncertainty Ranges	14
Table 4: Mexia Plant Specimen Numbers with Location Description	19
Table 5: Mexia Specimen Numbers with Uncertainty Ranges	21
Table 6: Mexia / Alaska Factor Weights	23

INTRODUCTION

Geographic information systems (GIS) have the potential to enhance historical research by providing a powerful tool to explore the spatial relationships of locations found in historical records, journals, field notes, and other sources. Current GIS, however, require exact knowledge of location in order to display and analyze information. A wealth of historical spatial data remains inaccessible, vaguely defined in a format that cannot be translated into the precise data requirements of a GIS. A method for representing inexact locations within a GIS would provide access to this rich source of spatial information for visualization and analysis.

Presently, historical spatial data remain in the realm of qualitative research, with locations or distances semantically defined instead of precisely measured. Without formal standards or software procedures for representing vague information in a GIS, attempts to translate historical location data into a quantitative format have met with varying degrees of success. Gregory, Kemp, & Mostern (2003) describe the problem:

Unfortunately, in a historical context, location is often described as a place name, an ill-defined region such as “the southwest US,” or descriptively, as in “close to the banks of the river.” These all represent various forms of uncertainty. The simplistic approach to handling these in GIS is to attempt to infer the precise geographical coordinates and features required by GIS software. Thus, the place name is redefined to a point location taken from a modern map, the region is given precise boundaries, and “close to the river bank” is taken as being within an exact and arbitrary distance from a line representing the modern course of the river. These solutions are at best limited and at worst entirely inappropriate.
(p. 13)

The lack of tools or methods for representing historical ambiguity hinders the exploration of spatial relationships that a GIS can provide.

The quantitative nature of GIS has limited its usefulness in historical research. According to Gregory and Ell (2007), “GIS originated in disciplines that use quantitative and scientific approaches in a data-rich environment. Historical geography is rarely data-rich; in fact, data are frequently incomplete and error-prone” (p.1). Working with such data requires interpretation and decision-making as to its worth and reliability. To be useful, GIS need to model the data validity decision process of historical scholarship, to allow ambiguity, and to portray on a map the uncertainty inherent in the qualitative description (Gregory et al., 2003).

Malczewski (2006a) asserted that the growing field of GIS decision support models a similar decision-making process. Specifically, multi-criteria decision analysis (MCDA) provides procedures to examine multiple and possibly conflicting criteria within a decision support structure. It also allows prioritizing the importance of different criteria and evaluating alternative scenarios. MCDA is increasingly utilized to determine suitable locations based on a variety of information. Malczewski (2006b) noted that fuzzy logic techniques are an effective tool for representation within MCDA when the criteria are imprecisely defined.

Robinson (2003) also identified fuzzy sets as a technique for modeling ambiguous or uncertain geographic information. Fuzzy set theory is a departure from traditional Boolean logic, in which an element in a larger set (such as a location in a study area) either is or is not a member of a subset (such as the subset of locations in the study area “close to the river bank”). Fuzzy sets can represent the gradual transition from full membership (for locations at the river bank), through intermediate values (locations very

near, a little distance away, etc.), to no membership (locations very far away). This ability for partial membership provides not only an intuitive representation of uncertain boundaries, but also a mapping from natural language modifiers such as “near” and “far” into numeric values stored within a computer (Klir and Yuan, 1995). Robinson (2003) stated that fuzzy sets are well suited for capturing the vagueness of uncertain information, and that the sets allow that vagueness to be carried throughout the analysis.

Plewe (2002) investigated fuzzy sets as a technique to represent historical uncertainty in computational form. His explorations of ambiguity in historical geographic information revealed two aspects of uncertainty: the vague extent of boundaries, and the degree to which evidence about those boundaries is known to be true. Plewe favored representing both forms of uncertainty with fuzzy sets, while Zhang and Goodchild (2002) and Eastman (2001) argued that fuzzy sets are appropriate for vague boundaries, but probability theory is better suited for weighing the strength of evidence.

A broader form of fuzzy sets, called fuzzy measures, effectively combines the concepts of probability, uncertainty, and evidence. Yager (2002) characterized a fuzzy measure of a variable as a description of knowledge about that variable, and saw it as a framework for several forms of uncertainty representation, including probability theory and fuzzy sets. As its name implies, fuzzy measures are used to specify error or uncertainty in a measurement, where the exact value cannot be precisely defined, but only specified within some range. The values in a fuzzy measure represent the probability of the true measurement occurring at each position within the range.

Eastman (2001) applied the concept of fuzzy measures to MCDA suitability studies, where the degree of suitability of a location represented the likelihood, or the probability, that the location met the criteria. In this case, each of the criteria represented evidence to be weighed as part of the decision process. The real strength of fuzzy measures in decision making, according to Jiang and Eastman (2000), is the range of techniques for combining several measures into a single membership set representing the suitability of the location in light of all criteria. This range of techniques allows for different levels of decision-making risk to be factored into the outcome, based on the confidence in the evidence.

Yager (2002) noted that decision-making under uncertainty results in a collection of possible outcomes instead of a unique outcome. For qualitative historical spatial information, this provides the necessary portrayal of the inherent uncertainty in the underlying data. GIS with MCDA and fuzzy measures has potential as a useful tool for historical GIS by allowing the uncertainty in the data to be visualized and by providing a spatially distributed portrait of the results (Eastman, 2001).

Fuzzy measures provide the capability for quantitative representation of vague information in a GIS, and multi-criteria decision analysis with fuzzy measures shows promise as a technique for modeling imprecise historical information and producing a physical location estimate. This study explores the viability of the technique by first applying it to a set of locations from a historic dataset of forestry records containing both qualitative and quantitative locations. This provides a benchmark of the technique's effectiveness. Next, a more challenging set of historical locations, notes from a botanist's

1928 Alaska field journal, are modeled and mapped. The results demonstrate the ability of MCDA with fuzzy measures to translate imprecise written descriptions into useful location estimates within a GIS.

METHODS

Data

The dataset chosen for testing the model, the Wieslander Vegetation Type Mapping project at UC Berkeley (Wieslander VTM), contains not only written descriptions of location information, but also separate markers on quad maps for comparison. The Wieslander project began in the 1920s as an effort to document California forest resources. The U.S. Forest Service sent surveyors into the field to map the broad vegetation patterns over much of California's forested areas. In addition, they conducted sampling of vegetation in small plots, called transects, to gather detailed vegetation information. The surveyors described the locations of transect plots in field notes, and also stamped the locations on USGS quad maps (Figure 1).

Figure 1: Example of historic quad map with plot locations

This dataset has two very appealing qualities as a test case for uncertainty modeling of historic qualitative locations. First, surveyors working for the U.S. Forest Service created the original information under well-defined guidelines of data acquisition, and second, the original map documents with the plot location stamps have been rigorously analyzed for map error and the resulting error information made available (Kelly, Ueda, & Allen-Diaz, 2008). This dataset contains both text description of location variables and a separate stamp on a USGS quad map showing the plot location, with map error already determined. In effect, the VTM data provides both the imprecise description and a target on a map to measure the location estimation result against.

The VTM dataset contains thousands of plots on USGS quad maps that cover a large portion of California. Thirty-three plots were chosen from the set for estimation; the selection was nonrandom with the goal of providing a range of input values for elevation, slope, aspect, location type, and surveyor. Table 1 contains the recorded location descriptions from the surveyors' field notes (Wieslander VTM). Figures 2 and 3 show the locations of the quads within California, and Table 2 contains the scale, age, and plot error for the six quads in the study.

The original quad maps and plot data were downloaded from the Wieslander (Wieslander VTM) website, and provided the latitude and longitude of the study area for each plot. With this information, a ten-meter resolution digital elevation model (DEM) of the area was downloaded from the National Map dataset (USGS). Also, a digital raster graphic (DRG) of the most recent topographic map of the area was obtained from the same source. The DRG and the DEM provided the sources for generating raster layers of

elevation, slope, aspect, and location description, using the IDRISI Taiga and ESRI ArcGIS software applications.

Table 1: Wieslander Study Plots with Field Note Entries. The qualitative location description for each plot is shown, along with surveyor's name and date.

Plot key	Geographic Location	Date	Surveyor	Aspect	Slope (%)	Elev. (ft.)
COVELO						
43AC38	Billy Williams ridge	8/26/1932	Lewis	N	5	3000
43AB33	Hell Hole Canyon	9/8/1932	Yates	N	50	2750
43AB21	0.5 Mi - S of Leach Lake	9/25/1932	Lewis	W	5	6200
43AB13	0.75 Mi SE Ray Springs	10/3/1932	Lundh	E	45	2750
MT. DIABLO						
81BB220	0.125 NE Peak of Mt Diablo	7/22/1935	French	NE	45	3500
81BB217	Just North W of North Pk.	7/22/1935	French	NW	70	3500
81BA217	S. of Somersville	9/14/1934	Wieslander	N	30	1100
81BA22	1 Mile East Chaparral Springs	8/13/1932	Lundh	S	75	1500
81BB110	1 mi W. Deer Flat	3/29/1935	Wilson	W	50	2250
SAN SIMEON						
131DA29	0.25 mi N.W. from Mouth of Spanish	8/6/1930	Jensen	NE	30	600
131DA26	0.5 mi. E. from Junction of Van Gordon and Burnett Creeks.	8/6/1930	Jensen	SW	35	1200
131DA22	2 mi E. from Burnett Camp	8/6/1930	Jensen	SW	45	1500
131DA217	Blackoak Mountain	8/7/1930	Jensen	N	45	2100
131DA213	0.5 mi E. from Pine Mtn.	8/14/1930	Jensen	NE	10	3000
131DA313	1 mi. N and .333 mi W from Rocky Butte	8/14/1930	Jensen	W	35	2600

MT. PINOS						
155C37	Long Canyon (South Fork Apache Canyon).	4/11/1934	Sowder	N	60	4700
155C36	Long Canyon (South Fork Apache Canyon).	4/11/1934	Sowder	S	45	4400
155F65	Divide between Red Reef Canyon and Timber creek	7/27/1930	St John	SE	20	3700
155F51	0.5 miles Below mouth of Bear Canyon	7/26/1930	St John	N	5	2800
155E55	1 miles SW of Thorn Meadows	5/27/1930	Wilson	N	55	5600
155E54	W of Thorn Meadows	5/27/1930	Wilson	NE	60	5300
PRESTON PEAK						
8F29	0.5 mile SE of Red Mtn Meadow	8/25/1932	Johannsen	E	5	4100
8F12	0.5 mile W of Red Mtn.	8/26/1933	Johannsen	W	60	3400
8C38	Hurdy Gurdy Butte	8/1/1933	Lewis	S	75	4800
8C37	Table Mt.	8/1/1933	Lewis	NW	30	4500
8D23	Fox Ridge	8/9/1932	St John	S	10	1500
8D22	Fox Ridge	8/9/1932	St John	-	0	2100
8C29	Fox Ridge	8/9/1932	St John	S	5	3500
YOSEMITE						
77C27	E end of Crocker Ridge SW facing slope	8/19/1935	Belshaw	SW	62	6000
77B512	.5 Mi N Dark Hole	8/23/1936	Hawbecker	SE	20	8600
77B53	Junction of White Wolf & Pate Valley	7/9/1936	Hawbecker	NW	15	7300
77A65	.25 Mi E Table Lake	7/16/1936	Hawbecker	W	70	7500
77E44	1.75 Mi S of Chinquapin	9/17/1935	Peterson	SW	30	5500

Wieslander VTM 30-Minute Quads (1:125,000)

The grid on the map shows the areas considered part of the VTM project. Only the quad maps associated with grid locations filled with color contain vegetation plot data, and only the locations with a dot contain a complete set of files for analysis.

Figure 2: VTM 30-minute plot data locations (modified from Wieslander VTM)

Wieslander VTM 15-Minute Quads (1:62,500)

The grid on the map shows areas considered part of the VTM project. Only the quad maps associated with grid locations filled with color contain vegetation plot data, and only the locations with a dot contain a complete set of files for analysis.

Figure 3: VTM 15-minute plot data locations (modified from Wieslander VTM)

Table 2: Historic Basemaps and Associated Scale, Age, and Plot Error

Quad Name	Scale	Age	Plot Error (meters) (Wieslander VTM)
Mt. Diablo	1:62,500	1898 (reprint 1922)	134
San Simeon	1:62,500	1919	131
Covelo	1:62,500	1926	128
Mt. Pinos	1:125,000	1903 (reprint 1932)	332
Yosemite	1:125,000	1909 (reprint 1929)	254
Preston Peak	1:125,000	1922	263

Uncertainty Ranges

The digital data and the field note descriptions for the plots contain some amount of error and ambiguity. In order to quantify this, two ranges of values were estimated for the data format and each part of the description according to accuracy standards, field notes, and visual inspection. The first range represents the narrowest interval that could be derived with confidence for the value under consideration. For example, the Wieslander (Wieslander VTM) field guide instructed the surveyors to record aspect to the nearest of the eight compass points (Figure 4), so a measured aspect a little north of NE would be rounded to NE. The effect of this rounding ambiguity can be captured in the fuzzy measure for aspect by setting the inner range to include any aspect value from NNE to ENE.

Figure 4: Compass rose

The second range of values estimated for the data represents the interval within which there is at least some probability of the true value occurring. Considering again the surveyor's recorded aspect of NE, the field guide specified that aspect was to be measured with a compass. The second range incorporates the margin of error into the fuzzy measure of aspect to account for measurement technique. In essence, it attempts to quantify how wrong the surveyor's measurement could have been. In this example, a conservative estimate of ninety degrees, or from N to E, would define a range wide enough to account for any measurement error.

Table 3 shows the uncertainty ranges for three example plot descriptions taken from Table 1. As described above, the first, or inner, range for aspect extends to the nearest of the sixteen compass points on either side of the recorded value, and the outer range places the recorded value in the center of a swing of ninety degrees. Slope ranges measure 2.5% on either side of the recorded value for the inner range and 5% on either side for the outer. The elevation range takes into consideration the accuracy defined for the digital elevation model, the calculated historic map error, and the surveyors' rounding to the nearest contour line. A spread of one-to-two contour lines covers all these uncertainties, which on the 15-minute quads translates to 50 feet and 100 feet, respectively, and on the 30-minute quads becomes 100 feet and 200 feet. For the text description, if it simply referenced a nearby geographic feature, then a buffer around the feature defined the limit. If the description specified a distance from a feature, then a spread of typically one-quarter mile and one-half mile defined the inner and outer ranges. The complete table of ranges for all thirty-three points can be found in Appendix A.

Table 3: Example Locations of Study Plots with Uncertainty Ranges. For each entry, ‘a’ and ‘d’ represent the outer uncertainty range values, and ‘b’ and ‘c’ represent the inner range values.

Plot Number	Elevation	Slope	Aspect	Description
B220	3500 ft.	45%	NE	0.125 mi.
Mt. Diablo State Park (15-minute quad)	a. 3400 ft. b. 3450 ft. c. 3550 ft. d. 3600 ft.	a. 40% b. 42.5% c. 47.5% d. 50%	a. N b. NNE c. ENE d. E	a. 0 mi. b. 0.06 mi. c. 0.25 mi. d. 0.5 mi.
B21	6200 ft.	5%	W	0.5 mi.
Mendocino National Forest (15-minute quad)	a. 6100 ft. b. 6150 ft. c. 6250 ft. d. 6300 ft.	a. 0% b. 2.5% c. 7.5% d. 10%	a. SW b. WSW c. WNW d. NW	a. 0 mi. b. 0.25 mi. c. 0.75 mi. d. 1.0 mi.
C27	6000 ft. (5915 ft.*)	62%	SW	Buffer around Crocker Ridge
Yosemite National Park (30-minute quad)	a. 5715 ft. b. 5815 ft. c. 6015 ft. d. 6125 ft.	a. 57% b. 59.5% c. 64.5% d. 67%	a. S b. SSW c. WSW d. W	

*Note on Yosemite quad that elevations were later found to be 85 ft. too high.

Fuzzy Measure Generation

The ranges established during the uncertainty analysis contain not only the initial historical information or DEM value, but also knowledge about the uncertainties inherent in each value. These ranges define the fuzzy measure for each of the location criteria of elevation, slope, aspect, and text description. The IDRISI Taiga GIS provides a software module to create fuzzy measures from uncertainty ranges by defining four critical values, which are the upper and lower bounds of each range, and choosing a membership function to represent the shape of the measure. The two most common shapes chosen for

the functions are the sigmoidal and linear shapes, shown in Figure 5. These set membership functions allow for the gradual transition from outer to inner range, and provide a method for assigning a probability to each criterion value in the range. The IDRISI Fuzzy module uses the function and the critical values to translate each criterion raster image into its corresponding fuzzy raster representation.

Figure 5: Sigmoidal and linear membership functions, with critical values a,b,c,d

Suitability Surface Generation

The IDRISI GIS multi-criteria evaluation module (MCE) performed the aggregation of the fuzzy layers. This module allows a first set of weights to be assigned to each layer indicating the degree of importance of each factor in the final decision. The ordered weighted averaging option provides a second set of weights to assess decision making risk. These weights specify how many of the factors need to be present at a given pixel before the evidence is considered sufficient for consideration, and also allow a weighted combination of more than one choice.

The general aggregation process involves multiplying each layer by its factor weight, ranking the weighted values at each pixel from smallest to largest, and then calculating the final value for each pixel according to the rules defined by the second set of weights (Eastman, 2009). For all Wieslander data points, the factor weights were set

equally, and the risk weights were chosen first for the most conservative estimate (weights of 1, 0, 0, 0), and second from a less conservative estimate that combines the two lowest-valued factors (weights of 0.5, 0.5, 0, 0). This final suitability layer, displayed as a color variance on top of the current topographic map in ESRI ArcGIS 9.3, shows the spatial distribution of aggregation results and its relationship to the plot stamp area.

Uncertainty and Sensitivity Analysis

Uncertainty enters into the location estimation process in many ways. It may represent error that has a numerical value, such as the root mean square error determined for the VTM maps or the error in the contour, slope, and aspect layers generated from the digital imagery. Other sources of data error include rounding by the surveyors when recording their information, and vagueness in description that requires subjectivity to quantify, such as listing a location “near” a lake. This type of uncertainty can be accommodated in the fuzzy membership function by widening the critical values to include more locations when significant error is present or information is very vague (Figure 5).

A second source of uncertainty arises in the analysis model itself. One of the most frequent criticisms of fuzzy sets is the arbitrary selection of both the membership function shape and the critical values associated with it. To address this, the effect of the function shape and critical values is evaluated by varying the selection choices and comparing the final results. This same type of sensitivity analysis is used for the decision risk weights chosen during the aggregation process. The sensitivity of the model to risk

weight selection becomes apparent by varying the weights and comparing results. These sensitivity tests provide an indication of how critical the selection of these values is, and also provides an estimation of the confidence in the final historical location estimate.

Mexia Alaska Data

The model effectiveness is tested a second time on another set of historical spatial data. Ynes Mexia was a botanist and explorer who traveled through both North and South America collecting plant specimens for several herbaria collections. In the summer of 1928 she traveled through Mt. McKinley National Park (now Denali National Park) in Alaska (Figure 6) and made brief notes in her field journal on the locations of specimens collected there. The Mexia spatial data is more sparse and less precise than the Wieslander data, and no maps showing locations accompanied her field notes. This dataset provides an indication of how well fuzzy measures and multi-criteria decision analysis can represent more vaguely defined locations.

The University and Jepson Herbarium at UC Berkeley archives Mexia's field notebooks, and from her Mt. McKinley notebook (Mexia, 1928) twelve qualitative location descriptions were transcribed. These descriptions are shown in Table 4. The selection process was nonrandom, with the locations representing a range of features within the park. For the selected park areas, a sixty-meter resolution DEM was downloaded from the National Map dataset (USGS), along with a DRG of a topographic map. In addition, a spatial data file containing Alaska streams was obtained from the Alaska State Geo-Spatial Data Clearinghouse (Alaska Geo-Spatial). The DEM was the source for generating raster layers of elevation and slope with IDRISI Taiga, and the

DRG and streams provided the base layers for digitization of point locations, park road, rivers, and glaciers using ESRI ArcGIS.

Figure 6: Mt. McKinley (Denali) National Park in Alaska

Mexia Ranges and Weights

The Mexia dataset contains more uncertainty than the Wieslander data. No detailed USGS topographical maps of Alaska existed when Mexia gathered plant samples there in 1928. Sections of the park were still unexplored and therefore unmapped. Her elevation estimates were rough guesses at best, and in some cases were off by as much as 1500 feet. By 1928, the park road extended forty miles into the park from the entrance

(Brown, 1991), and her most accurate elevation values are in this area. However, most elevation estimates provided with her samples proved too inaccurate for use. Fortunately, little has changed in the park since its inception in 1917, and even the current topographical maps still date from the 1940s. Elevation values and digitized features extracted from the early USGS maps provided the bulk of the source information used to create the uncertainty ranges.

Table 4: Mexia Plant Specimen Numbers with Location Description. The qualitative location description for each number is shown, along with the date of collection.

Specimen Number	Date	Elevation	Locality	Description
MT. MCKINLEY NATIONAL PARK				
2009	6/17/1928	2600 ft.	Head Savage River	
2012	6/18/1928	2730 ft.	Inspiration Point	
2015	6/18/1928	2730 ft.	Inspiration Point	near stream
2019	6/18/1928	2730 ft.	Inspiration Point	on flat above Savage River near camp
2024	6/21/1928	2000 ?	Mt. Margaret, near summit	top of wind-swept peak
2029	6/22/1928		Savage River, near camp	river bed
2040	6/24/1928	1100 m.	Near Tetlanika river	small gulch. Off highway
2042	6/25/1928	1000 m.	Head Savage River	Narrow valley between high mtn., near stream
2066	6/25/1928	1100 m.	Head Savage River	damp stream valley
2076	7/5/1928	900 m.	Savage Camp	
(2092)	7/12/1928	1000 m.	Trail to Copper Mountain, Toclát River	grassy flat near river
2096	7/13/1928	950 m.	Foot of Copper Mountain	slopes near Muldrow Glacier

Unlike the Wieslander data, the Mexia data required interpretation of qualitative location information for each specimen number. As a botanist, she recorded location more as an indicator of plant growing environment than as actual location in physical space. She also frequently used the descriptive “near” which, depending on the immediate terrain, could take on a wide range of values. Table 5 shows the uncertainty ranges determined for the various locations, and Table 6 lists the factor weights assigned to the fuzzy measures and any necessary buffering of features. The most conservative risk weights of (1,0) or (1,0,0) were used, and in the case of only one fuzzy measure and one buffer, a simple overlay combined the two. Fuzzy measure ranges are very conservative to reflect the lack of detailed information and the expansive terrain of the park through which she traveled.

Table 5: Mexia Specimen Numbers with Uncertainty Ranges

Specimen Number	Elevation	Locality	Description
MT. MCKINLEY NATIONAL PARK			
2009	2600 ft. a. 1210 b. 2100 c. 3100 d. 4500	Head Savage River (River line 0 mi.) c. 0.25 mi. d. 1.0 mi.	Buffer upper end of Savage River
2012	2730 ft. a. 2530 b. 2630 c. 2830 d. 2930	Inspiration Point Buffer around Savage River / valley floor / 3000 ft. contour line	
2015	2730 ft. Same as #2012	Inspiration Point Buffer around Savage River / valley floor / 3000 ft. contour line	near stream (streamlines 0 mi.) c. 0.25 mi. d. 0.5 mi.
2019	2730 ft. Same as #2012	Inspiration Point Buffer around Savage River / valley floor / 3000 ft. contour line	on flat above Savage River near camp Slope: c. 5% d. 10% Savage Camp (0 mi.): c. 0.5 mi. d. 1.0 mi.
2024	2000 ? None, too inaccurate	Mt. Margaret, near summit (0 mi.) c. 0.5 d. 1.5	top of wind-swept peak Buffer 4000 ft. contour line around Primrose Ridge

2029	(900 m.) a. 700 b. 800 c. 1000 d. 1100	Savage River, near camp (0 mi.) c. 0.5 mi. d. 1.0 mi.	river bed (0 mi.) (width of riverbed: ~1000 ft.) c. 1000 ft. d. 2000 ft.
2040	1100 m.	Near Tetlanika river (river midline 0 mi.) c. 0.5 mi. d. 1.0 mi.	small gulch. Off highway (highway midline 0 mi.) c. 0.25 mi d. 0.5 mi
2042	1000 m. Same as #2009	Head Savage River	Narrow valley between high mtn., near stream
2066	1100 m. Same as #2009	Head Savage River	damp stream valley
2076	900 m. Same as #2029	Savage Camp (0 mi.) c. 0.25 mi. d. 0.5 mi.	
(2092) No number in text	1000 m.	Trail to Copper Mountain, Toclat River (0 mi.) c. 0.5 mi. d. 1.0 mi.	grassy flat near river (slope 0%) c. 5% d. 10%
2096	950 m.	Foot of Copper Mountain (use base elevation 4000 ft.) a. 3000 ft. b. 3500 ft. c. 4500 ft. d. 5000 ft.	slopes near Muldrow Glacier Buffer edge of glacier to higher elevations of Mt. Eielson (Copper Mtn.)

Table 6: Mexia / Alaska Factor Weights. The weights applied to each layer vary depending on confidence in the source information. The constraint narrowed the study area for each location.

Specimen Number	Factor Weights	Constraint
MT. MCKINLEY NATIONAL PARK		
2009	Elevation – 0.25 Distance to river – 0.75	Buffer around river and tributaries
2012	Elevation – 1.0	Buffer around river / valley floor / 3000 ft. contour line
2015	Elevation – 0.5 Distance to streams – 0.5	Buffer around river / valley floor / 3000 ft. contour line
2019	Elevation – 0.33 Distance to camp – 0.33 Slope – 0.34	Buffer around river / valley floor / 3000 ft. contour line
2024	Distance from summit – 1.0	Buffer around 4000 ft. contour line of Primrose Ridge
2029	Elevation – 0.2 Distance to river – 0.4 Distance to camp – 0.4	
2040	Distance to river – 0.5 Distance to road – 0.5	
2042	Elevation – 0.25 Distance to river – 0.75	Buffer around river and tributaries
2066	Elevation – 0.25 Distance to river – 0.75	Buffer around river and tributaries
2076	Elevation – 0.5 Distance to camp – 0.5	
(2092)	Slope – 0.5 Distance to river – 0.5	
2096	Elevation – 1.0	Buffer around Copper Mountain

RESULTS

Wieslander Transect Plot Location

The location estimate model returns a set of pixels representing geographic locations. Each pixel has a footprint of 10.5 meters by 8.5 meters (Universal Transverse Mercator projection) with an area of 89.25 square meters (960 square feet). Each pixel also has a probability assigned to it, ranging from 0 to 1, indicating the likelihood of that pixel containing the true location. The suitability results are displayed with a color band from light to dark, with dark representing a higher likelihood of the true historical location falling at a given point on the map. The results are overlaid on the current topographic map for visualization along with an estimate of the original stamp location and size. This estimate was drawn from a point marker digitized from the estimated center of the original stamp, and a circle added with radius set to the adjusted error radius determined in the Wieslander map analysis (Kelly et al., 2008).

The suitability results identified candidate locations within the plot stamps in twenty-four of the thirty-three stamp locations, and identified locations very near the plot stamps in seven more of the stamp locations. In all cases, part of the suitability surface appeared outside the plot stamp as well. The size and extent of the resulting surface was influenced by the local topography and the preciseness of the text description, with more precise descriptions around more unique features generating a smaller surface.

The most likely area for the transect plot, according to the written description, fell near the edge of the stamp on the map for twenty of the locations, while the other thirteen provided no additional information as to exact plot location. Figure 7 shows the locations

of the three example suitability surfaces along with the stamp area, and Appendix B contains the results for all locations in the dataset.

Sensitivity Analysis

The first sensitivity test of fuzzy function shape, using the base images for plot B220, showed only subtle differences between the resulting suitability maps. Results are displayed in Figure 8 for both the most conservative and less conservative risk weights. The intermediate values for the linear function are more pronounced than for the sigmoidal function, which transitions more sharply from low to high suitability. The sigmoidal function, by nature of its shape, groups the pixels more strongly into high, middle, and low suitability categories. Given the imprecise nature of the suitability process, this narrower grouping may be more helpful visually. However, the sensitivity to function shape appears to be low.

The second sensitivity test of fuzzy measure width, using the base images for plot C27, demonstrates the effect of different uncertainty ranges for the fuzzy measures. The pixels of the suitability map for the narrow function range are clustered between two contours on the side of Crocker Ridge (top image, Figure 9). The critical values of a, b, c, and d in the uncertainty ranges were doubled to create the suitability surface in the lower image of Figure 9. When the fuzzy function critical values were widened to expand the function width, three new clusters of pixels appeared. Also, a larger number of pixels in the original cluster have a high probability value. The wider function range in effect cast a wider net in the search for candidate locations, but did so discriminately. It did not simply enlarge the first area, but singled other groupings of pixels for

consideration. Therefore, there is some sensitivity to fuzzy function width, with the wider ranges pulling in more pixels for consideration.

The final test used plot B21 results to examine the sensitivity of the model to the risk weights. This test was a good visual indicator of the effect of the weight choices on the result. Figure 10 shows the suitability distribution for a variety of weight choices, ranging from fairly indiscriminate to very conservative. These weights are designed to provide a full range of combination possibilities, from a fuzzy logical AND, where only the smallest criterion factor at a given pixel is selected, to a fuzzy logical OR, where only the largest criterion factor is selected. The results show a shrinking of the suitability surface, but not a change in the underlying pattern. More selective weights ruled out the areas of weaker evidence, but left the areas of strongest evidence intact. From this analysis, the sensitivity of the model for areas of strongest evidence is low, and for areas of weakest evidence is high. For this study, the two most conservative sets of weights seemed sufficient to illustrate the underlying suitability pattern.

Mexia Plant Specimen Location

The suitability surfaces generated for the Mexia locations covered a much wider area than the Wieslander surfaces. In general, they were continuous spreads of pixels frequently covering more than two miles, instead of the individual clusters that often occurred with the Wieslander data. These spreads centered on or followed the physical features that Mexia used as reference points in describing location. The results for the twelve possible plant specimen locations are in Appendix B, after the Wieslander results.

Figure 7: Suitability surface results for three example plots. The maps show the location of the suitability surface in relation to the plot stamp for Mt. Diablo (top), Covelo (middle), and Yosemite (bottom) quads.

Sigmoidal fuzzy function, most conservative suitability estimate

Linear fuzzy function, most conservative suitability estimate

Sigmoidal fuzzy function, less conservative suitability estimate

Linear fuzzy function, less conservative suitability estimate

Figure 8: Sensitivity analysis – fuzzy function shape. The two upper images are the suitability results for plot 81BB220 (Mt. Diablo) when the most conservative combination was used, and the two lower images are for the same plot with a less conservative criteria combination. For the two images on the left, the sigmoidal fuzzy function shape was used to generate the fuzzy measures, and for the two images on the right, the linear fuzzy function shape was used for the fuzzy measures.

Figure 9: Sensitivity analysis – width of fuzzy measure. Both maps are the suitability results for 77C27 (Yosemite) in relation to the plot stamp. The fuzzy measures used to create the suitability surface in the lower map are twice as wide as the measures in the upper map.

Order weights: 0.4, 0.2, 0.2, 0.2

Order weights: 0.5, 0.25, 0.125, 0.125

Order weights: 0.5, 0.25, 0.25, 0

Order weights: 0.5, 0.5, 0, 0

Order weights: 1, 0, 0, 0

Each map shows the suitability results for 43AB21 (Covelo) with a different set of order, or risk, weights applied. Beginning with the upper left map and proceeding across and down, the weights were set more conservatively (less risk), until with the final map, at left, the most conservative weight values were applied.

Figure 10: Sensitivity analysis – order (risk) weights

DISCUSSION

Uncertainty in Plot Location

The plot stamps on the historical VTM maps do not reveal the true locations of the transect plots. According to the historical documents, the actual transect plots were 800 square meters in size in forested areas and just 400 square meters in areas of scrub and chaparral (Kelly & Allen-Diaz, 2005). When the 3.5 mm diameter stamps were placed on the maps to mark the locations, they delineated a geographic area that ranged in size (adjusted for error) from 42,000 square meters to over 670,000 square meters (Figure 11). Kelly et al. (2008) recommended a comparison of the historical map stamp location with its error-adjusted location on a current map, along with considerations of recorded values for slope and aspect. These factors taken together would help to more accurately locate the true plot area within the stamp neighborhood.

The process described by Kelly et al. is a multi-criteria decision analysis and represents the same process utilized in this study. A GIS can easily perform the task of locating areas with the target slope, aspect and elevation within the region delineated by the stamps. The maps in Appendix B highlight locations that meet the criteria, and the areas identified by the suitability results are smaller than the areas of the error-adjusted stamps. Using GIS as a tool for exploration can reduce time spent in the field searching for appropriate sites. The use of fuzzy measures with MCDA can provide a refined location estimation while at the same time compensating for estimation in the data.

Figure 11: Comparison of transect plot sizes with stamps. Sizes shown are relative to error-adjusted stamp markers. Also shown is the size of a pixel in the suitability surface relative to plot sizes and error-adjusted stamp sizes.

Uncertainty in the Data

Plewe (2002) asserted that historical data contain significant error and identified nine causes of error or uncertainty in historical records. The Wieslander data exhibited all these types of uncertainty, and this caused interpretation of the results to be much more challenging. While the fuzzy measures compensated for some types of error, others were related to the plot stamp location and made assessment of the results more difficult.

- **Observation limitations.** The surveyors' measuring devices had limited accuracy.

Elevation seemed the most fallible as it is a measurement that people are unable to

confirm with their senses. Figure 12 shows a common situation in the data where the recorded elevation does not match the elevation of the stamp.

Figure 12: Historic quad map elevation discrepancy. Stamp 17 extends from the 2850 ft. to the 3350 ft. contour lines, but marked elevation is 3500 ft.

- **Lack of evidence.** The actual location of the transect plot is an issue with the Wieslander data and one that leads to difficulty in utilizing the data set for vegetation change analysis (Kelly et al., 2008). Did the designers of the Wieslander study believe that a general geographic reference was sufficient, or was it obvious to them what the stamp placement meant? Current researchers can only infer correct location.
- **Lack of reference.** While many of the features used by surveyors for location reference persisted to the current quad maps, some features no longer appear on the

current maps. The surveyor Jensen frequently used small school buildings for reference in the San Simeon quad, and few if any of these structures remain today. Estimating the latitude and longitude of the structures from the old 15-minute quads would introduce a much larger amount of error in the fuzzy distance measure. These locations were not used in this study.

- **Questionable evidence.** No evidence was suspected to be biased or untrustworthy in this study, but there were clear differences in the quality of different surveyors' work. Jensen's six locations in San Simeon were all reliable, while the Yosemite data contained numerous errors. The Yosemite quad had over 650 transect plots recorded by a very large number of surveyors, and quality control was more of a problem with this data set.
- **Conflicting evidence.** While there were no problems with multiple data source conflicts, there were numerous conflicts between stamp location and listed elevation in the Wieslander data (Figure 12). They were not detected in this study until after all results were complete, since the author refrained from examining stamp placement on the historic quads while constructing the fuzzy measures so as not to bias the results.
- **Ambiguous evidence.** The language used in the text description could have different interpretations. In particular, the text description for plot 155F51 in the Mt. Pinos quad listed the plot as a half mile below the mouth of Bear Canyon. The author assumed, looking at a map, that "below" meant south of the canyon mouth, while evaluation of the results showed that the surveyor meant "below" to be lower in elevation, which is to the right on the map (Figure 13). In this situation, a description

which may be fairly obvious to an observer in the field is misinterpreted by someone working from a map.

Figure 13: Suitability surfaces with different interpretations of description. The red surface is the result of interpreting 'below' as south. The blue surface used 'lower in elevation' for south. The blue surface appears to be the correct interpretation.

- **Misinterpretation.** The process of interpreting descriptions, building fuzzy measures, and aggregating the data into a final suitability surface provides many opportunities for researcher error. In the case of 81BA22 in the Mt. Diablo quad, the surveyor described the location as one mile east of Chaparral Springs. The author chose the Chaparral Springs label on the current quad map for distance measurement instead of the streamline, and the result was a weak match (Figure 14).

Figure 14: Comparison of features on historic and current quad. A poor match of suitability surface to plot stamp 81BA22 resulted from using feature name as distance marker instead of the feature itself.

- **Transformation of phenomena.** While the digitizing of the plot centers allows the transformation of plot location from historic quad to current quad, the shift in scale alters the relationship of the stamp to the surrounding terrain (Figure 15). Also, the historic marker measures less than 500 feet from the trail, while the transformed marker sits nearly 1000 feet away.

Figure 15: Comparison of topography on historic and current quad. When evaluating results, the placement of stamp 8 on the historic quad (left) in relation to the ridge appears much different than the reconstructed stamp on the current quad (right). It also appears much closer to the trail.

- **Encoding error.** These errors are numerous in the Wieslander data. This is not surprising given the sheer volume of data digitized and transcribed into Excel records. Some errors seemed to originate from the transcriber, others from the digitizer, and still others from the original surveyor. Plots 155E55 and 155E54 in the Mt. Pinos results in Appendix B are examples of inconsistencies in stamp numbering, location, and database entries.

Plewe's (2002) categories of error and uncertainty help to highlight the challenge of working with historical records. A fuzzy approach to data representation, careful choice of factor weights with less reliable evidence, and conservative selection of risk weights can help mitigate the effects of some errors. A probabilistic approach to results visualization also conveys the concept that the results are, at best, good guesses.

Patterns in the Data

The first pattern to appear from the results was a tendency for the suitability surface to coincide with the edge of a plot stamp. While the majority of the results supported this idea, it was not conclusive. However, the tendency of several stamps on the historic quad maps to sit a small distance away from the stated elevation suggests that perhaps most surveyors chose to use the edge of the stamp to point to true plot location.

The thirty-three suitability results were then examined for any patterns relating to slope, aspect, elevation, surveyor, and type of features. No clear trend relating to slope appeared. Surfaces generated for areas of steep slope did not differ significantly from those of shallower slope. In some cases, areas of shallow slopes resulted in larger

patches and higher slopes produced narrower spreads, but this was not consistently true. Also, no trend relating to aspect, elevation, or surveyor appeared.

Topography, however, did factor significantly in the results. The pattern of the suitability surface, whether well-defined or widespread, narrow patches or large area, was highly dependent on the local topography. Canyons produced a larger area of small patches, while the sides of mountains produced a narrower, contiguous spread. Surprisingly, the results for some very long features like Fox Ridge in the Preston Peak quad generated small, well-defined surfaces even though the text descriptions were very general (see Appendix B, plots 8D22, 8D23, and 8C29). The continuously varying topography of the Wieslander study region allowed the suitability selection area to be narrowed down naturally. It remains to be seen if the technique would produce usable results in a broad, flat region like the San Joaquin Valley of California.

Displaying Multiple Results

The suitability results in Appendix B show only one surface and its corresponding plot stamp for each map. However, multiple plots were commonly grouped at an individual feature, and varied only in their recorded values of slope, aspect, and elevation. Two features from the Wieslander dataset were chosen as tests of both visualization and differentiation of results when more than one surface is displayed at a time. Long Canyon in the Mt. Pinos quad was the site for both plots 155C36 and 155C37. Figure 16 shows the result when both surfaces are displayed at the same time, using color to differentiate the between the two. Likewise, Figure 17 is a map of the surfaces for 8C29, 8D22, and 8D23 at Fox Ridge in the Preston Peak quad when all three

are displayed together. The fine-grained surfaces generated from the fuzzy measures allow intermixing with pixels from other results, and with the use of color, multiple surfaces can be mapped together and the results remain distinct. Care must be taken in interpretation though, since ArcGIS displays only the topmost layer at a pixel and does not blend layers. A mathematical combination of multiple surfaces would be required to determine and highlight areas of overlap.

Figure 16: Display of two suitability surfaces at Long Canyon. Suitability surfaces for two plots, both located in Long Canyon, are distinguishable on the same map through the use of color for differentiation.

Figure 17: Display of three suitability surfaces at Fox Ridge. Three separate results, 8D22, 8D23, and 8C29, are displayed together with color used to differentiate the different surfaces. The topographic map in the background has been faded to make the surfaces more visible.

Advantage of Risk Weights and Fuzzy Measures

The results in Appendix B show the suitability surfaces generated from the most conservative set of risk weights. For some locations, the evidence contained significant discrepancies and only a small surface was produced. Under these conditions, a less conservative set of weights allows more potential locations to be brought in for consideration and provides a better surface for visualizing the possible true location of the stamp plot. Figure 18 is an example of less conservative risk weights providing more information.

Figure 18: Use of less conservative risk weights. The most conservative risk weights produced weak results (left) while the less conservative risk weights provide more information (right).

The suitability surface results are also fuzzy measures themselves, created by combining the individual criteria fuzzy measures according to the factor weights and the risk weights. As fuzzy measures, they contain the range of true location probabilities from low to high. With this capability, the resulting surface can be filtered to show only the top 20% or 30% of locations, thereby narrowing the search range to the most likely potential locations. This utilizes one of the strengths of fuzzy measures, which is their ability to hold grades of membership in the set of “true location”. Figure 19 shows an example of the suitability results for plot 131DA313 in the San Simeon quad after the pixels with probabilities between 0.75 and 1.0, the ones with the highest likelihood of representing the true location, were extracted with the IDRISI Reclass module.

Mexia Data – Visualizing on a Larger Scale

The Mexia results are quite different visually from the Wieslander results. The clearly defined guidelines of Wieslander data collection, the four different parameters recorded at each location, and the mountainous topography all combine to produce a relatively compact and geographically small suitability surface. Mexia’s location descriptions are more general and contain fewer location “clues” to help narrow the search. As such, the resulting suitability surface is more generalized and covers a larger geographic area. While these results do not pinpoint precise location, they nevertheless give a visual indication of the area in the park where the specimen was collected and of the local topography.

In Mexia’s case, the individual results can be combined onto one map to display her movement and range as she explored the park. This type of visualization helps to

move the qualitative into the quantitative, while displaying the ambiguity of the original information. With the results in a GIS, other information can be added to gain a more complete picture of her explorations. Such additional information could consist of the digitized park road with dates of completion, camping facilities run by the park concession along the road, sites of ranger cabins and dates built, soil layers, and current vegetation information. As an example, Figure 20 shows all twelve Mexia results on the same map, along with dates to give a sense of her movement during her stay in the park.

Figure 20: Cumulative map of Mexia location estimates. The map contains all twelve suitability surfaces along with dates of collection to give a sense of Mexia's movement within the park.

CONCLUSION

Multi-criteria decision analysis with fuzzy measures successfully translated written descriptions of location into raster, or grid-based, surfaces within a GIS. Given the uncertainty of the qualitative descriptions, the technique resolved the text into a collection of locations instead of a single location. The probability assigned to each location in the suitability collection also conveyed the ambiguity associated with the results and the probabilistic nature of its interpretation.

Gregory et al. (2003) identified three characteristics that GIS needed to provide for historical scholarship: model the data validity process, allow ambiguity, and portray uncertainty. The multi-criteria decision analysis (MCDA) afforded the necessary flexibility to incorporate decisions about data validity into the modeling process. The fuzzy measures, by their nature, modeled the ambiguity of the evidence used for decision-making. And finally, the data aggregation of fuzzy measures within MCDA produced a collection of suitable locations representing the likelihood of a good match with the evidence. MCDA with fuzzy measures met the goals of data translation from qualitative to quantitative representation and of inherent uncertainty visualization.

The two datasets in the study gave different insight into the technique's usefulness. The Wieslander data demonstrated that the technique could provide a different interpretation of a vaguely defined location and possibly lead to new insight when combined with any existing quantitative information. This dataset also helped to clarify the difficulty of working with historical geographic data, a process that would not normally be visible when only qualitative descriptions are available. The Mexia data

showed the technique's effectiveness on more vaguely defined locations, where the resulting surface may be better suited for visualization than any analysis.

Further research is needed into the suitability of the MCDA raster results for data analysis. This will depend on the type of qualitative description translated into the GIS and on the goals of the researcher. Exploration of spatial relationships can be accomplished with both raster (surface grid) and vector (points, lines, or polygons) data. There are techniques for calculating a point approximation for a fuzzy surface, or for converting a raster collection of results into a polygon outline of the surface. These translations from raster to vector representation would open up a wider set of spatial analysis techniques for the data, but with the loss of information about the inherent probabilistic spread. Research into appropriate raster and vector analysis methods would provide a sophisticated set of GIS tools to enhance historical scholarship.

REFERENCES

- Alaska State Geo-Spatial Data Clearinghouse. (Alaska Geo-Spatial). *Alaska Department of Natural Resources, Physical Features, Hydrography*, 63,360 [data file]. Retrieved Jan. 5, 2010 from <http://www.asgdc.state.ak.us/>
- Brown, W.E. (1991). *A History of the Denali - Mount McKinley Region, Alaska*. United States Department of the Interior, National Park Service, Southwest Regional Office, Santa Fe, New Mexico. Retrieved Jan. 2010 from http://www.nps.gov/history/history/online_books/dena/hrs.htm
- Eastman, J.R. (2001). Uncertainty management in GIS: Decision support tools for effective use of spatial data resources. In C. Hunsaker, M. Friedl, M. Goodchild, & T. Case (Eds.), *Spatial uncertainty in ecology: Implications for remote sensing and GIS applications* (pp. 379-390). New York, NY: Springer.
- Eastman, J.R. (2009). Decision support. In *IDRISI Taiga User's Guide*, Version 16.00. Worcester, MA: Clark University.
- Gregory, I. & Ell, P. (2007). *Historical GIS: technologies, methodologies and scholarship*. Cambridge: Cambridge University Press.
- Gregory, I., Kemp, K., & Mostern, R. (2003). Geographical information and historical research: Current progress and future directions. *Humanities and Computing*, 13, 7-22.
- Jiang, H. & Eastman, J.R. (2000). Application of fuzzy measures in multi-criteria evaluation in GIS. *International Journal of Geographical Information Science*, 14(2), 173-184.
- Kelly, M. & Allen-Diaz, B. (2005). Digitization of a historic dataset: the Wieslander California Vegetation Type mapping project. *Madroño*, 52(3), 191-201.
- Kelly, M., Ueda, K., & Allen-Diaz, B. (2008). Considerations for ecological reconstruction of historic vegetation: Analysis of the spatial uncertainties in the California Vegetation Type Map dataset. *Plant Ecology*, 194, 37-49.
- Klir, G. J. & Yuan, B. (1995). *Fuzzy sets and fuzzy logic: Theory and applications*. Upper Saddle River, NJ: Prentice-Hall.
- Malczewski, J. (2006a). GIS-based multicriteria decision analysis: a survey of the literature. *International Journal of Geographical Information Science*, 20(7), 703-726.

- Malczewski, J. (2006b). Ordered weighted averaging with fuzzy quantifiers: GIS-based multicriteria evaluation for land-use suitability analysis. *International Journal of Applied Earth Observations and Geoinformation*, 8, 270-277.
- Mexia, Y. (1928). Summer 1928 Alaska field notebook. Ynes Mexia Papers, Unprocessed Collection. University and Jepson Herbarium, University of California, Berkeley.
- Plewe, B. (2002). The nature of uncertainty in historical geographic information. *Transactions in GIS*, 6(4), 431-456.
- Robinson, V.B. (2003). A perspective on the fundamentals of fuzzy sets and their use in geographic information systems. *Transactions in GIS*, 7(1), 3-30.
- USGS National Map Dataset. (USGS). *Digital Elevation Model* [data file] and *Digital Raster Graphic* [data file]. Retrieved March - May 2009 and Nov 2009 - Jan. 2010, from <http://seamless.usgs.gov/index.php>
- Wieslander Vegetation Type Mapping Project. (Wieslander VTM). *Historic map digital images, plot point shapefiles, plot data Excel files* [data files] and *Field Guide* [PDF document]. College of Natural Resources, Geospatial Imaging and Informatics Facility (GIIF), UC Berkeley. Retrieved February – March 2009, from <http://vtm.berkeley.edu/>
- Zhang, J. & Goodchild, M. F. (2002). *Uncertainty in geographical information*. London: Taylor & Francis.
- Yager, R. (2002). Uncertainty representation using fuzzy measures. *IEEE Transactions on Systems, Man, and Cybernetics – Part B: Cybernetics*, 32(1), 13-20.

APPENDIX A – WIESLANDER UNCERTAINTY RANGES

Wieslander Study Plots with Uncertainty Ranges

Plot key	Elevation (ft.)	Slope (%)	Aspect	Description (mi.)
COVELO				
43AC38	3000 a. 2900 b. 2950 c. 3050 d. 3100	5 a. 0 b. 2.5 c. 7.5 d. 10	N a. NW b. NNW c. NNE d. NE	Ridgeline (0 mi.) c. 0.25 d. 0.5
43AB33	2750 a. 2650 b. 2700 c. 2800 d. 2850	50 a. 45 b. 47.5 c. 52.5 d. 55	N a. NW b. NNW c. NNE d. NE	Canyon base (0 mi.) c. 0.25 d. 0.5
43AB21	6200 a. 6100 b. 6150 c. 6250 d. 6300	5 a. 0 b. 2.5 c. 7.5 d. 10	W a. SW b. WSW c. WNW d. NW	0.5 mi. a. 0 b. 0.25 c. 0.75 d. 1.0
43AB13	2750 a. 2650 b. 2700 c. 2800 d. 2850	45 a. 40 b. 42.5 c. 47.5 d. 50	E a. NE b. ENE c. ESE d. SE	0.75 mi. a. 0.25 b. 0.5 c. 1.0 d. 1.25
MT. DIABLO				
81BB220	3500 a. 3400 b. 3450 c. 3550 d. 3600	45 a. 40 b. 42.5 c. 47.5 d. 50	NE a. N b. NNE c. ENE d. E	0.125 mi. a. 0 b. 0.06 c. 0.25 d. 0.5

81BB217	3500 a. 3400 b. 3450 c. 3550 d. 3600	70 a. 65 b. 67.5 c. 72.5 d. 75	NW a. W b. WNW c. NNW c. N	Just NW of peak c. 0.25 d. 0.5
81BA217	1100 a. 1000 b. 1050 c. 1150 d. 1200	30 a. 25 b. 27.5 c. 32.5 d. 35	N a. NW b. NNW c. NNE d. NE	Buffer south of Somersville
81BA22	1500 a. 1400 b. 1450 c. 1550 d. 1600	75 a. 70 b. 72.5 c. 77.5 d. 80	S a. SE b. SSE c. SSW d. SW	1 mi. a. 0 b. 0.5 c. 1.5 d. 2.0
81BB110	2250 a. 2150 b. 2200 c. 2300 d. 2350	50 a. 45 b. 47.5 c. 52.5 d. 55	W a. SW b. WSW c. WNW d. NW	1 mi. a. 0.5 b. 0.75 c. 1.25 d. 1.5
SAN SIMEON				
131DA29	600 a. 500 b. 550 c. 650 d. 700	30 a. 25 b. 27.5 c. 32.5 d. 35	NE a. N b. NNE c. ENE d. E	0.25 mi. a. 0 b. 0.125 c. 0.375 d. 0.5
131DA26	1200 a. 1100 b. 1150 c. 1250 d. 1300	35 a. 30 b. 32.5 c. 37.5 d. 40	SW a. S b. SSW c. WSW d. W	0.5 mi. a. 0 b. 0.25 c. 0.75 d. 1.0

131DA22	1500 a. 1400 b. 1450 c. 1550 d. 1600	45 a. 40 b. 42.5 c. 47.5 d. 50	SW a. S b. SSW c. WSW d. W	2 mi. a. 1.5 b. 1.75 c. 2.25 d. 2.5
131DA217	2100 a. 2000 b. 2050 c. 2150 d. 2200	45 a. 40 b. 42.5 c. 47.5 d. 50	N a. NW b. NNW c. NNE d. NE	Summit (0 mi.) c. 0.25 d. 0.5
131DA213	3000 a. 2900 b. 2950 c. 3050 d. 3100	10 a. 5 b. 7.5 c. 12.5 d. 15	NE a. N b. NNE c. ENE d. E	0.5 mi. a. 0 b. 0.25 c. 0.75 d. 1.0
131DA313	2600 a. 2500 b. 2550 c. 2650 d. 2700	35 a. 30 b. 32.5 c. 37.5 d. 40	W a. SW b. WSW c. WNW d. NW	1.05 mi. (Pythagorean T.) a. 0.55 b. 0.8 c. 1.3 d. 1.55
MT. PINOS				
155C37	4700 a. 4500 b. 4600 c. 4800 d. 4900	60 a. 55 b. 57.5 c. 62.5 d. 65	N a. NW b. NNW c. NNE d. NE	Buffer around Long Canyon
155C36	4400 a. 4200 b. 4300 c. 4500 d. 4600	45 a. 40 b. 42.5 c. 47.5 d. 50	S a. SE b. SSE c. SSW d. SW	Buffer around Long Canyon

155F65	3700 a. 3500 b. 3600 c. 3800 d. 3900	20 a. 15 b. 17.5 c. 22.5 d. 25	SE a. E b. ESE c. SSE d. S	Divide line (0 mi.) c. 0.18 d. 0.5
155F51	2800 a. 2600 b. 2700 c. 2900 d. 3000	5 a. 0 b. 2.5 c. 7.5 d. 1.0	N a. NW b. NNW c. NNE d. NE	0.5 mi a. 0 b. 0.25 c. 0.75 d. 1.0
155E55	5600 a. 5400 b. 5500 c. 5700 d. 5800	55 a. 50 b. 52.5 c. 57.5 d. 60	N a. NW b. NNW c. NNE d. NE	1 mi. a. 0.5 b. 0.75 c. 1.25 d. 1.5
155E54	5300 a. 5100 b. 5200 c. 5400 d. 5500	60 a. 55 b. 57.5 c. 62.5 d. 65	NE a. N b. NNE c. ENE d. E	Buffer west of Thorn Meadows
PRESTON PEAK				
8F29	4100 a. 3900 b. 4000 c. 4200 d. 4300	5 a. 0 b. 2.5 c. 7.5 d. 10	E a. NE b. ENE c. ESE d. SE	0.5 mi. a. 0 b. 0.25 c. 0.75 d. 1.0
8F12	3400 a. 3200 b. 3300 c. 3500 d. 3600	60 a. 55 b. 57.5 c. 62.5 d. 65	W a. SW b. WSW c. WNW d. NW	0.5 mi. a. 0 b. 0.25 c. 0.75 d. 1.0

8C38	4800 a. 4600 b. 4700 c. 4900 d. 5000	75 a. 70 b. 72.5 c. 77.5 d. 80	S a. SE b. SSE c. SSW d. SW	Buffer around Hurdygurdy Butte
8C37	4500 a. 4300 b. 4400 c. 4600 d. 4700	30 a. 25 b. 27.5 c. 32.5 d. 35	NW a. W b. WNW c. NNW d. N	Buffer around Table Mountain
8D23	1500 a. 1300 b. 1400 c. 1600 d. 1700	10 a. 5 b. 7.5 c. 12.5 d. 15	S a. SE b. SSE c. SSW d. SW	Buffer around Fox Ridge
8D22	2100 a. 1900 b. 2000 c. 2200 d. 2300	0 c. 2.5 d. 5	- none	Buffer around Fox Ridge
8C29	3500 a. 3300 b. 3400 c. 3600 d. 3700	5 a. 0 b. 2.5 c. 7.5 d. 10	S a. SE b. SSE c. SSW d. SW	Buffer around Fox Ridge
YOSEMITE*				
77C27	6000 (5915) a. 5715 b. 5815 c. 6015 d. 6115	62 a. 57 b. 59.5 c. 64.5 d. 67	SW a. S b. SSW c. WSW d. W	Buffer around east end of Crocker Ridge

77B512	8600 (8515) a. 8315 b. 8415 c. 8615 d. 8715	20 a. 15 b. 17.5 c. 22.5 d. 25	SE a. E b. ESE c. SSE d. S	0.5 mi. a. 0 b. 0.25 c. 0.75 d. 1.0
77B53	7300 (7215) a. 7015 b. 7115 c. 7315 d. 7415	15 a. 10 b. 12.5 c. 17.5 d. 20	NW a. W b. WNW c. NNW d. N	Buffer around trail junction
77A65	7500 (7415) a. 7215 b. 7315 c. 7515 d. 7615	70 a. 65 b. 67.5 c. 72.5 d. 75	W a. SW b. WSW c. WNW d. NW	0.25 mi. a. 0 b. 0.125 c. 0.375 d. 0.5
77E44	5500 (5415) a. 5215 b. 5315 c. 5515 d. 5615	30 a. 25 b. 27.5 c. 32.5 d. 35	SW a. S b. SSW c. WSW d. W	1.75 mi. a. 1.25 b. 1.5 c. 2.0 d. 2.25

*Note on Yosemite quad that elevations were later found to be 85 ft. too high.

APPENDIX B – SUITABILITY MAP RESULTS

The suitability surfaces generated for each location in the study are displayed on a topographic map of the area. For the Wieslander data, an adjusted estimate of the original plot stamp is displayed as well. The stamp is centered on a point digitized from the center of the original stamps on the historic maps (Wieslander VTM), and its radius is set to the adjusted error radius determined by error analysis (Kelley et al., 2008).

This section contains the results for each of the thirty-three Wieslander locations in the California historic quads of Covelo, Mt. Diablo, San Simeon, Mt. Pinos, Preston Peak, and Yosemite, followed by the twelve Mexia locations in Denali National Park in Alaska. The suitability surface is displayed as a color variance from light yellow to dark red, with the darker color indicating locations with a higher probability of matching the true location, according to the evidence in the written descriptions.

These written descriptions are shown in the text box for each map along with the date for the transect plot study or plant specimen collection and the name of the individual who wrote the description. For the Wieslander data, the second line contains the text description, followed by the “aspect / slope / elevation” recorded for the plot. Aspect is to the eight compass points, slope is in percent, and elevation is in feet. The final line is the historic quad name and the size of the quad, either 15-minute or 30-minute. For the Mexia data, the first number is the plant specimen number, and after her name and collection date is the general area of collection. This is followed by elevation in feet or meters and a more specific description of the immediate area where the sample was collected.

Covelo

Figure B-1: Covelo quad, Mendocino National Forest, California. The Covelo quad is an historic 15-minute USGS topographic quad. It covers part of the Coastal Mountain Range in northwestern California in an area that is now Mendocino National Forest. Elevations range from 1200 feet to almost 7000 feet in this largely unpopulated area. The forest vegetation consists of Douglas fir and Ponderosa pine.

Covelo - plots 43AC38, 43AB33

Covelo - plots 43AB21, 43AB13

Mt. Diablo

Figure B-2: Mt. Diablo quad, Mt. Diablo State Park, California. The Mt. Diablo quad is an historic 15-minute USGS topographic quad. It covers part of the Coast Range on the eastern edge of the San Francisco Bay region, in an area that is now Mt. Diablo State Park. Elevations range from nearly sea level to 3849 feet (the peak of Mt. Diablo). The landscape in this urban park is covered with oaks and pines, grassland, and chaparral.

Mt. Diablo - plots 81BB220, 81BB217

Mt. Diablo - plots 81BA217, 81BA22

Mt. Diablo – plot 81BB110

San Simeon

Figure B-3: San Simeon quad, California. The San Simeon quad is an historic 15-minute USGS topographic quad. It covers part of the Santa Lucia Range along the coast in central California. Elevations range from sea level to 3590 feet. The natural vegetation consists of pine forest, oak woodland, chaparral and grasslands.

San Simeon - plots 131DA29, 131DA26

San Simeon - plots 131DA22, 131DA217

San Simeon - plots 131DA213, 131DA313

Mt. Pinos

Figure B-4: Mt. Pinos quad, Los Padres National Forest, California. The Mt. Pinos quad is an historic 30-minute USGS topographic quad. It covers part of the Coast Mountains in central / southern California. Elevations range from 1100 feet to 8800 feet. The natural vegetation consists of pine forest and chaparral.

Mt. Pinos - plots 155C37, 155C36

Mt. Pinos - plots 155F65, 155F51

Mt. Pinos - plots 155E55, 155E54

Preston Peak

Figure B-5: Preston Peak quad, Redwood National Park, California. The Preston Peak quad is an historic 30-minute USGS topographic quad. It covers part of the north coast of California. Elevations range from sea level to 7240 feet. The natural vegetation consists of redwood and Douglas fir forests, spruce, oak and grasslands.

Preston Peak - plots 8F29, 8F12

Preston Peak - plots 8C38, 8C37

Preston Peak - plots 8D23, 8D22

Preston Peak – plot 8C29

Yosemite

Figure B-6: Yosemite quad, Yosemite National Park, California. The Yosemite quad is an historic 30-minute USGS topographic quad. It covers part of the Sierra Nevada Range in California. Elevations range from 1066 feet to 10,846 feet. The natural vegetation varies from scrub and chaparral to pine forest to alpine meadows.

Yosemite - plots 77C27, 77B512

Yosemite - plots 77B53, 77A65

Yosemite – plot 77E44

Figure B-7: Part of Mt. McKinley (Denali) National Park, Alaska. The area shown is in the northern section of the park, with the high peaks of the Alaska Range located to the south. Elevations in the area shown range from 1092 feet to 12,096 feet. Both taiga (spruce, aspen, birch, willow) and tundra (shrubs and wildflowers) make up the natural vegetation in this subarctic climate.

Alaska / Mt. McKinley – specimen numbers 2012, 2015

Alaska / Mt. McKinley – specimen numbers 1919, 2024

Alaska / Mt. McKinley – specimen numbers 2029, 2040

Alaska / Mt. McKinley – specimen numbers 2009, 2042, 2066, 2076

Alaska / Mt. McKinley – specimen numbers 2092, 2096